

Can learning a new language make you better at maths? A meta-analysis of foreign language learning and numeracy skills during early adolescence

Research Article

Cite this article: Nucette, A., Hamamura, T., Leita, S. and Biedermann, B. (2025). Can learning a new language make you better at maths? A meta-analysis of foreign language learning and numeracy skills during early adolescence. *Bilingualism: Language and Cognition* 28, 923–953. <https://doi.org/10.1017/S1366728924000701>

Received: 30 January 2024
Revised: 15 August 2024
Accepted: 17 August 2024
First published online: 16 December 2024

Keywords:

systematic review; meta-analysis; foreign language learning; second language learning; mathematical skills

Corresponding author:

Alejandra Nucette;
Email: alejandra.nucette@postgrad.curtin.edu.au

  This research article was awarded Open Data and Open Materials badges for transparent practices. See the Data Availability Statement for details.

We would like to thank the two anonymous reviewers for their contributions to earlier versions of the manuscript.

Alejandra Nucette¹ , Takeshi Hamamura², Suze Leita¹ and Britta Biedermann¹ 

¹School of Allied Health, Faculty of Health Sciences, Curtin University, Bentley, WA, Australia and ²School of Population Health, Faculty of Health Sciences, Curtin University, Bentley, WA, Australia

Abstract

This systematic review and meta-analysis explore the correlation between foreign language instruction and mathematical skills in young adolescents, highlighting the significance of high school mathematical education and the adaptability of the adolescent brain. Focused on students starting second language programs between ages 8 and 13, following PRISMA guidelines, this review included 25 studies (1978–2020) with 785,552 participants. Using a random-effects model, the overall effect size revealed a statistically significant relationship between our variables, indicating a threefold higher likelihood of passing or achieving higher grades in mathematical tests for language-learning students. Moderating variables analyses identified socioeconomic status (SES) and intervention length as influencers of observed heterogeneity, with SES being the primary factor. Sensitivity analyses, including adding potentially missing studies and removing outliers, confirmed the robustness of the overall effect. Nonetheless, additional research is needed to enhance global diversity and comprehensively understand the interplay between language learning and cognitive function.

Highlights

- Systematic review and meta-analysis, 25 studies (1978–2020), 785,552 children.
- Random-effects model: moderate, significant effect ($OR = 2.99, p < 0.0001$).
- Mixed-effects model explained over 65% of between-study heterogeneity.
- Sensitivity and subgroup analyses resulted in positive, significant correlations.
- Further research is needed to identify confounding factors and enhance global diversity.

1. Introduction

Mathematics has long been considered a challenge within the educational system, often eliciting apprehension among students, administrators and government officials. This complex discipline is both feared and actively pursued in the quest for improved educational outcomes (Freeman et al., 2019). In contrast to subjects such as mathematics and science, other subjects, including foreign language learning, are seen as less critical, particularly in English-speaking countries and among schools with low socio-economic profiles across the OECD countries (OECD, 2023a). Consequently, we raise the question: What if the cognitive exercise involved in language acquisition could enhance numerical aptitude?

The significance of mathematical education, particularly at the secondary school level, is widely acknowledged as a crucial determinant of students' academic and vocational success. Research has demonstrated a positive correlation between fundamental mathematical skills and academic achievement across various high school subjects (Anderton et al., 2017; Badru & Owodunni, 2021; Kanwal et al., 2022) and in university programs (Ballard & Johnson, 2004; Johnson & Kuennen, 2006; Yunker et al., 2009). Furthermore, proficiency in mathematics has been linked to improvements in many facets of life, including the transition between secondary school and university (Anderton et al., 2017), performance in university studies (Delaney & Devereux, 2020) and outcomes in the labour market, such as increased salaries and higher-level jobs (Joensen & Nielsen, 2009).

Over the past 15 years, the importance of Science, Technology, Engineering and Mathematics (STEM) studies has gained global recognition for its pivotal role in technological advancement, economic competitiveness and social welfare (Freeman et al., 2019). Consequently, there has been a significant increase in STEM education research in both developed and developing

© The Author(s), 2024. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

countries (Zhan et al., 2022). Nonetheless, challenges persist, particularly in enhancing students' attitudes towards mathematics and facilitating mathematical learning for those who find it daunting (Tytler, 2020).

Similarly, education experts have emphasised that learning a second language offers the added social benefits of acquiring linguistic and cultural skills, which have become increasingly crucial in today's globalised and multicultural society (Davies, 2018; Midgley, 2017). Research has also demonstrated that acquiring foreign language proficiency can open up better job opportunities (Araújo et al., 2015). However, students do not always perceive foreign language learning as a valuable asset for their future careers and studies, especially when considering local or domestic employment opportunities (OECD, 2020).

As such, if there is a positive correlation between language acquisition during high school years and mathematical skills, this may help boost teenagers' perception of learning a second language as a complementary method to improve their mathematical achievement.

1.1. Second language learning

The acquisition of a second language (L2) and the factors influencing one's success in this endeavour have been extensively studied. Historically, the literature was dominated by the critical or sensitive period hypothesis, which suggested that achieving native-like proficiency in an L2, particularly in pronunciation and grammar, becomes nearly impossible after early childhood (Johnson & Newport, 1989; Scovel, 1988).

However, this theory has encountered challenges on three main fronts. Firstly, recent studies indicate that if there is indeed a sensitive period for L2 learning, it extends beyond the previously assumed boundaries, continuing until approximately 17 to 18 years of age (Hartshorne et al., 2018; Master et al., 2020). Secondly, researchers have demonstrated that differences in cognitive ability, the context in which language is acquired and even native language, rather than one's age, play the most significant role in achieving native-like proficiency (Bialystok & Hakuta, 1999; Birdsong, 2006; Figueiredo et al., 2016; Flege et al., 1999; Hakuta, 2001; Hakuta et al., 2003), even when it comes to phonology (Abu-rabia et al., 2004; Birdsong, 2006; Darcy et al., 2015). Finally, the advent of neuroimaging techniques has provided evidence indicating that the brain's plasticity during the early stages of adolescence bears notable similarities to that observed during the first three years of life (Fuhrmann et al., 2015; Selemon, 2013; Steinberg, 2014), possibly rendering it well-suited for the acquisition of new languages. This heightened plasticity is further enhanced by the advantages conferred by white matter growth and pruning, ultimately leading to heightened cognitive efficiency (Selemon, 2013; Simpson-Kent et al., 2020; Squeglia et al., 2013).

1.1.1. Second language learning and cognitive ability

The effect of L2 acquisition on cognition has been a topic of intense debate for over a century. Initially, bilingualism was seen as a cognitive burden (Darcy, 1953; Saer, 1923). Later studies, however, highlighted the cognitive advantages of bilingualism (for a timeline see Barac & Bialystok (2011)). Nonetheless, some researchers argue that the evidence supporting bilingualism's impact on executive control is inconsistent (Andreou et al., 2020; Dick et al., 2019; Paap et al., 2015; Paap & Sawi, 2014).

A current theory proposes that the inconsistency in results regarding bilingualism arises from its dynamic nature (Beatty-

Martínez & Titone, 2021; Bialystok & Craik, 2022; Di Pisa et al., 2021). Bilingualism evolves based on factors like duration and intensity of language learning and usage, which lead to variations in brain changes (Pliatsikas et al., 2020) and subsequent non-linguistic cognitive benefits (DeLuca et al., 2019, 2020; Gullifer et al., 2018).

Although most of the research has studied the advantages that bilingualism may confer, some researchers have sought to determine the level of language mastery needed to activate these benefits. Neuroimaging studies have shown that even early stages of L2 learning can induce brain changes, such as grey matter density and white matter integrity, observable across all ages and sensitive to variables like age of acquisition and performance level (Li et al., 2014). These changes could be seen after just three months of language training, and since brain changes have been associated with executive control enhancement (Luna et al., 2015), the author speculated that these brain changes might result in enhancing executive functions (EFs).

Consistent with this hypothesis, studies have found that language learners exhibited EF scores lower than early or simultaneous bilingual but higher than any monolingual speaker groups (White & Greenfield, 2017). However, others found that no significant gains were achieved when the L2 input was too limited (Poarch & van Hell, 2012). These differences may be explained by Bialystok and Barac's (2012) and Sorge et al. (2017) work on language learners, which proposes that the interaction between language learning and executive functioning is largely dependent on the learners' language use and exposure.

1.2. Mathematical ability and learning

Mathematical ability is defined as the capacity to learn, apply and retain mathematical theories and ideas, enabling effective problem-solving in any numerical context (Karsenty, 2014). Despite changes in teaching techniques, the core mathematical content remains consistent. Davis (1978) identified four fundamental areas of mathematical knowledge, which continue to be the main focus of math education, aiming to facilitate students' understanding:

Mathematical Concepts: Basic knowledge, like counting or arithmetic (e.g., addition, or subtraction).

Mathematical Generalisations: Patterns or relationships between numbers or basic knowledge, such as the distributive law or recognising that adding two odd numbers will always result in an even number.

Mathematical Procedures: Sequence of mathematical operations carried out in order, like BIMDAS (Brackets, Indices, Multiplication, Division, Addition and Subtraction) for solving algebraic equations.

Number Facts: Simple calculations between two numbers, usually committed to memory, such as $5 + 5 = 10$, or $4 \times 5 = 20$.

At high-school level, the context of this review, these areas come together to help students acquire fundamental numerical knowledge (e.g., counting, magnitude, etc.), understand arithmetic, geometry and algebra and develop proficiency in mathematical word-problem solving and reasoning (Peng et al., 2015). To assess these mathematical skills, most tools include common measures such as basic operations (i.e., arithmetic), algebra, geometry, mathematical word problems, logic and reasoning and practical applications (Breux & Lichtenberger, 2016; Brown et al., 2012; Hresko et al., 2003).

With advancements in brain imaging and neuroscience, the focus has shifted from teaching methods to understanding how

students learn. Geary (2004), for example, observed that learning disabilities often stem from deficits in knowledge competencies, linked to disruptions in neural structures related to the Central Executive (EF – inhibitory control, cognitive flexibility, working memory), the Language System (phonetic–articulatory system) or the Visuospatial System. This understanding has led to a framework suggesting that EF plays a critical role in mathematical learning and may predict numerical ability.

1.2.1. Mathematical learning and cognitive ability

Mathematics is the academic domain with the most robust and consistent correlation with cognitive ability across all age groups. A longitudinal study of 562 four-year-old children found that central executive skills (inhibitory control, attention flexibility and working memory) were more strongly associated with mathematical achievement than with literacy (Fuhs et al., 2014). Similarly, cognitive flexibility has been shown to predict mathematical performance in elementary school students aged seven to ten (Cantin et al., 2016) and in adolescents (Gathercole et al., 2004).

Diverse methods have been used to study the relationship between cognitive skills and mathematics (for reviews see Bull and Lee (2014) and Cragg and Gilmore (2014)). Some studies combine EF processes into one latent variable, finding a significant positive association with math performance, particularly during early childhood (Fuhs et al., 2014; Mazzocco & Kover, 2007; Mercader et al., 2018). However, this correlation weakens with age, potentially because different cognitive processes relate to different mathematical skills at different ages, reducing the overall effect of the latent variable (Bull & Lee, 2014).

Other researchers measure EF processes separately to match them independently to mathematical performance. For instance, it has been suggested that base levels of inhibitory control skills and their growth positively predicted children's early math skills (Choi et al., 2018). Conversely, lack of inhibition was highly correlated with poor mathematical ability in children aged six and eight years (Bull & Scerif, 2001). A meta-analysis found a substantial and significant relationship between cognitive flexibility and math (Yeniad et al., 2013). Additionally, working memory, both visuospatial and verbal, has also been closely related to mathematical ability, often showing a stronger association than with other central executive processes (Lin, 2018; St Clair-Thompson & Gathercole, 2006).

1.3. Language learning and mathematical ability: The role of executive function

The effect of bilingualism on specific EF skills, namely inhibition (ability to control rehearsed or automatic behaviours), cognitive flexibility (ability to switch tasks or strategies when problem-solving) and working memory (mental manipulation of recently acquired information) (Miyake et al., 2000), has been extensively studied. Similarly, the link between cognitive processes and mathematical ability has been the aim of numerous studies. However, the specific cognitive skills that enhance individual math skills remain unclear, and only some general interactions between these variables have been identified.

1.3.1. Inhibitory control

Studies have shown that bilinguals exhibited faster response times and less interference in trials with and without conflicts when compared to monolinguals, even among very low SES individuals without literacy skills (Barac et al., 2014; Nair et al., 2017; Poarch &

van Hell, 2012). Additionally, Planckaert et al. (2023), in their systematic review, found a higher prevalence of the bilingual advantage in inhibition and switching tasks among children under six years of age, suggesting this advantage may diminish as children advance in age.

Inhibitory control is crucial for solving mathematical word problems and, to a lesser extent, for basic number knowledge and arithmetic. Mathematical word problems often include extraneous information, and the ability to disregard irrelevant information is essential for finding the correct answer (Bull & Lee, 2014). Inhibitory control aids in managing order of magnitude, particularly in fractions, by helping to suppress incorrect automatic responses (e.g., recognising that $1/2$ is larger than $1/10$, despite 10 being larger than 2) (Bull & Lee, 2014; Nguyen et al., 2019). It also helps with the correct application of order of operations in arithmetic, preventing students from solving equations in the sequence they appear (e.g. $1 + 2 \times 3 - 2 = 7$ rather than 5, which is the correct response) (Nguyen et al., 2019).

1.3.2. Cognitive flexibility

Barac et al. (2014) concluded in their review that cognitive flexibility, along with inhibition, were the EF processes with the most robust reported advantages for bilinguals. This conclusion is supported by previous studies, such as Barac and Bialystok (2012), which showed that bilinguals outperformed their monolingual peers in task switching, regardless of language similarity, cultural background or language of schooling. Additionally, Adi-Japha et al. (2010) found that the enhancement of shifting skills due to bilingualism extends to cognitive flexibility in non-linguistic domains.

Cognitive flexibility or shifting, is hypothesized to be linked to high performance in algebra, word-problem solving and mathematical reasoning. Complex algebraic exercises and mathematical word problems require multiple steps to be solved, and each step may require a different approach. Thus, a heightened ability to switch between solving strategies is associated with achievement in these areas (Bull & Lee, 2014; Cragg & Gilmore, 2014; Nguyen et al., 2019).

1.3.3. Working memory

Blom et al. (2014) and Morales et al. (2013) found that bilingual children outperformed their monolingual counterparts in working memory, especially when the tasks were more demanding (Morales et al., 2013) and even when the children were socioeconomically disadvantaged (Blom et al., 2014). More recently, meta-analyses performed by Grundy and Timmer (2016) and Monnier et al. (2022) showed a significant small to medium effect size favouring an advantage due to bilingualism, which appeared more frequently in children, and was largely moderated by the language in which the tasks were performed.

Some authors have found conflicting results. For instance, Antón et al. (2019) observed no disparity between monolingual and bilingual individuals in simple working memory tasks (e.g., remembering visuospatial patterns). However, when more complex tasks were introduced (e.g., remembering and reversing a string of numbers), bilinguals exhibited a distinct advantage over their monolingual counterparts, even after controlling for known confounding variables.

Working memory has been extensively correlated with mathematical performance, particularly with arithmetic and mathematical word-problem solving. Bellon et al. (2019) found that updating accuracy predicted arithmetic accuracy, as working memory is necessary for recalling arithmetic facts and storing interim

solutions. Moreover, single and multi-digit calculations were found to rely on working memory, regardless of the operation required to arrive at the solution (Peng *et al.*, 2015). Interestingly, the reliance on working memory for solving basic operations decreases with age; children depend more on working memory for these tasks than adults do (Cragg & Gilmore, 2014).

Additionally, working memory significantly influences mathematical word-problem solving, indicating that this skill does not solely depend on language-based working memory resources as previously thought (Peng *et al.*, 2015).

1.4. *Language learning and mathematical ability in adolescents: Is there a link?*

As evident from the substantial body of research linking both language learning and mathematical ability to the central EF, the idea of exploring the potential of learning new languages to enhance mathematical achievement is indeed grounded in empirical evidence.

While there is research to support this hypothesis, it is important to note that only a relatively small number of researchers have explored this topic, and the majority of these investigations have primarily centred on preschool-aged children, such as those conducted by Choi *et al.* (2018) and Hartanto *et al.* (2018), as well as primary school children, as evidenced by the studies of Iqbal (2022) and Stewart (2005). This tendency may be attributed to the heightened neuroplasticity observed during these developmental stages. Interestingly, the findings from these studies have consistently demonstrated affirmative outcomes.

For instance, Hartanto *et al.* (2018) examined the correlation between bilingualism and mathematical achievement, employing various metrics, including teacher evaluations, emergent numeracy skills and standardised test scores. Their research revealed that bilingualism positively contributed to children's mathematical aptitude and reasoning abilities, after controlling for potential confounding variables like age, gender and socio-economic status. Similarly, Choi *et al.* (2018), in their investigation involving low-income preschoolers, discovered that bilingual children exhibited superior performance in mathematics assessments when compared to their monolingual counterparts, even in cases where bilingual children started with lower base skills. Furthermore, both literature reviews conducted by Stewart (2005) and Iqbal (2022), which focused on primary school children, highlighted a positive association between bilingualism and foreign language acquisition and heightened mathematical ability, as indicated by higher achievement test scores.

Given the recognition that early adolescence may represent a period of heightened neuroplasticity and central EF changes (Luna *et al.*, 2015; Steinberg, 2014), it is reasonable to build up on existing research and investigate the possibility of a positive association between second language learning and mathematical ability among adolescents.

1.5. *Research question*

Recognising the significance of mathematical education during high school and the potential of the adaptable adolescent brain, we chose to focus our study on students who initiated their second language programs at around 8–13 years of age, which includes the onset of early puberty as defined by Dorn *et al.* (2006), and were tested during secondary school years.

Thus, we carried out a systematic review and meta-analysis to answer our research question: Do young adolescents who received formal instruction in a foreign language exhibit improved numerical skills compared to those who did not?

2. Method

This systematic review was conducted and reported using the guidelines set out in the Johanna Briggs Institute (Aromataris & Munn, 2020) and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement (Page *et al.*, 2021). Our protocol was registered with PROSPERO, under the ID: CRD42020172859.

2.1. *Search strategy*

For this review, studies were identified by searching electronic databases (PsycINFO, PubMed, ERIC, Wiley Online Library, EBS-COhost, Taylor and Francis online and Scopus) and grey literature (EMBASE, Web of Science, MedNar) for published peer-reviewed journal articles, reports, and books, as well as unpublished masters' theses, doctoral dissertations, conference proceedings and presentations. Article alerts from relevant databases were also monitored to identify peer-reviewed papers.

The search terms were derived from the PICO protocol, and depending on the database, a combination of subject headings (MeSH), keywords and wildcards was used to ensure the search criteria were robust. Each search consisted of all the terms linked using Boolean operators, limited by language (English only, due to the difficulty of the authors to extract data from papers in other languages) and type of publication (only journal articles, reports, dissertations, books, speeches and conference proceedings were included) with no restrictions of date of publication.

The searches were performed between August and November 2022, while the search alerts were monitored until the end of December in the same year (2022). Additional information is provided in the [Supplementary Materials](#). The search terms and an example can be found in [Table S1 \(Appendix S1\)](#). The complete list of retrieved articles is available in [Appendix S2](#).

2.2. *Inclusion and exclusion criteria*

All studies that met the following criteria were considered to answer our research question and were included.

2.2.1. *Types of study*

All types of studies reported in English that explore the quantitative relationship between language learning and academic achievement (i.e., in numeracy, mathematics or science).

2.2.2. *Participants/population*

Participants were typically-developing early adolescents (including children from 9 to 17 years of age) enrolled in a formal school program.

2.2.3. *Intervention(s), exposure(s)*

Participants in the study were required to engage in a formal foreign language learning program, either within a school setting or through alternative means, where the second language was not the exclusive medium of instruction. It is important to highlight that, for inclusion in the study, students must have received explicit

instruction on the structural aspects of the second language (such as grammar and sentence formation), as opposed to mere exposure to the language. Consequently, the research focused on examining the impact of bilingual education (BE), dual language (DL) and two-way bilingual (TWB) programs initiated during elementary education, where fundamental language skills are taught.

A brief explanation of the interventions within the included studies is provided below.

- a. **Transitional Bilingual Education (TBE):** A US-based program where students are taught English as a Second Language (ESL) while learning subjects like math or science in their native language until their English proficiency is sufficient for them to function successfully in the mainstream¹ with additional support classes (Medina et al., 1985).
- b. **Two-Way Bilingual Education Program Model (TWB):** This US bilingual education model uses both Spanish and English as the medium of instruction, maintaining a 50/50 to 60/40 ratio between Spanish and English-speaking students since its inception in 1993. Instruction in literacy, math, science and social studies alternates weekly between Spanish and English (Cobb et al., 2006).
- c. **Bilingual Education (BE) Program:** Designed for native speakers of languages other than English with limited English proficiency, this program aims to prepare students for success in mainstream classrooms while maintaining access to their native language. Instruction is provided in two or more languages, with the amount of instruction given in each language varying by school. Prior to 1995, it operated as a traditional transitional bilingual education (TBE) program, focusing on exiting within 3 years (de Jong, 2004).
- d. **Dual Language Bilingual Program (DL):** A program, implemented in the US, where roughly equal numbers of English learners and fluent/native English-speaking students are taught together, offering a higher percentage of instruction in Spanish. The goal is full bilingualism and biliteracy for both groups (HISD, 2020).
- e. **English as a Second Language Program (ESL):** A self-contained program focusing on English mastery to prepare students for success in an English-speaking environment. Students typically divide their day between regular English classes and specialised ESL instruction (de Jong, 2004).
- f. **Foreign Language Learning (FL):** This program enables students to learn a non-native language outside of the geographical area where it is commonly spoken (e.g., learning French in the UK) (Karaoglan Yilmaz et al., 2020).

Notably, conventional high school programs lacking second language education (or support for immigrant students) and bilingual programs commencing in high school were excluded from our review.

2.2.4. Outcome(s)

Any studies reporting short or long-term academic achievement, regarding numerical skills, ability, reasoning, competence, achievement, or performance measured quantitatively (e.g., results from standardised mathematics tests, such as SAT) were included. Qualitative studies, or those which did not report on numerical achievement, were not considered.

Therefore, studies were excluded if they

- a. did not provide quantitative results (e.g., editorials, qualitative studies);
- b. were not fully available in English, even if the abstract was;
- c. reported on performance on skills/subjects that are not numerical (e.g., English literature, music, sports);
- d. reported on the current academic performance of students who were being taught in a second language (e.g., foreign students in conventional schools) or
- e. included a sample specifically selecting students diagnosed with any form of atypical development that could affect their learning (e.g., language delay).

Once all titles and abstracts were collated, each team member reviewed a set of ten. To ensure a high level of inclusion/exclusion agreement, the results of this set were discussed, and following this discussion, all the authors independently reviewed a second set of ten titles and abstracts. As the agreement among the team reached 85%, the selection continued.

All studies (title and abstract) were uploaded to Rayyan (Ouzzani et al., 2016), where they were screened by at least two independent reviewers, working in blind mode, unable to see other reviewers' decisions. A 95% agreement (Cohen's kappa 0.57) was reached in screening. Following this, reference lists of related articles were checked, and a Google Scholar search was conducted to ensure the search was exhaustive. This process identified 71 new studies (a complete list is provided in Appendix S3), which were also reviewed independently by two team members (96% agreement, Cohen's kappa 0.85). Both times, areas of disagreement were resolved by a third team member. Common areas of disagreement involved the availability of data to analyse (Olsen & Brown, 1992) and misinterpretations concerning the start of interventions (Padilla et al., 2022).

A total of 58 studies were retrieved to be read in full and assessed for eligibility. Ultimately, 25 studies were selected to be included in this review. The number of results obtained and reasons for exclusion can be found in the PRISMA flowchart (Haddaway et al., 2022) (Figure 1).

2.3. Data extraction

A data extraction sheet was developed to record relevant data from the selected studies. This included the study description (author, title, year, type and location), population characteristics (description, participant number, gender, SES, age at the time of testing and languages), methods (aim, intervention type and duration), outcomes (assessment tools, their scale and confounding variables) and results (primary and other significant findings, statistical method and effect size). Conclusion and limitations for each study were also documented.

Eligibility and Risk of Bias (ROB) columns were included: eligibility ensured adherence to the selection criteria (intervention, population and outcomes), while ROB shows the results of the ROBINS-I tool results, as well as conflict of interest and sample size bias.

Where necessary, data reported within each paper was segregated into separate results; when a study presented multiple results, the information extracted for each result was recorded in a separate row, labelled with a letter after the article ID (e.g., 4a and 4b represent two different outcomes included in Cobb et al. (2006)). As such, a total of 49 results were identified and recorded.

Two team members independently extracted data, followed by a collaborative risk of bias assessment. This assessment aimed to evaluate study quality but did not exclude any studies. Consensus on critical studies was achieved through thorough discussions. The

¹In education, "mainstream" refers to the regular or general school classes designed for children without special needs. In this setting, students do not receive additional support services, whether language-related or otherwise.

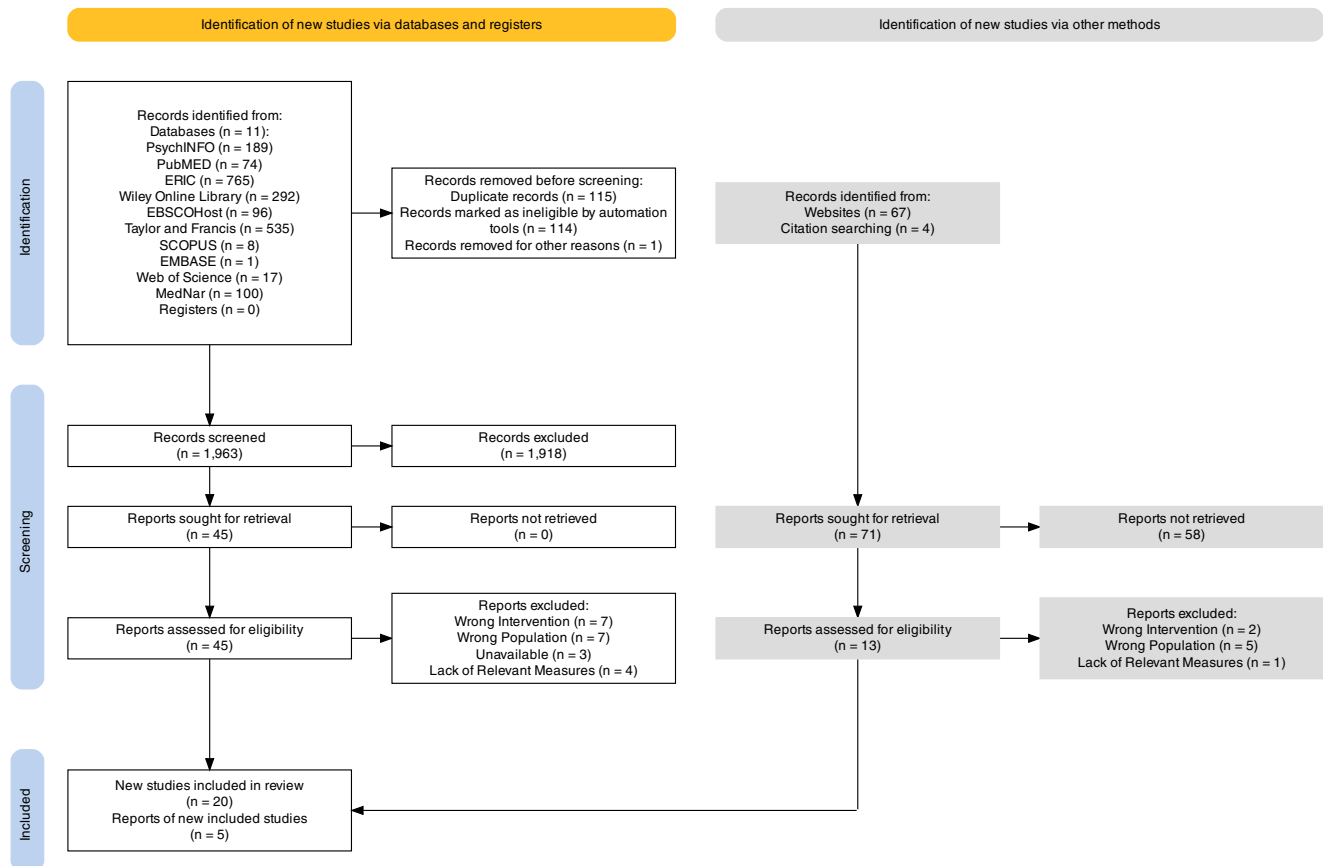


Figure 1. PRISMA Flowchart.

detailed results of the ROBINS-I tool and the data extraction sheet are available in [Appendices S4](#) and [S5](#).

2.4. Publication bias and small study effect

Although great care was taken to include all the available literature on our topic, there remains a concern about publication bias, where studies with significant results tend to be published more frequently than those which do not (Cooper, 2017). To evaluate the prospect of publication bias, a contoured enhanced funnel plot was generated following the methodology outlined by Peters et al. (2008). This approach aimed to discern any bias against statistically non-significant results.

Moreover, for a quantitative assessment of potential bias, both Trim and Fill analysis and random/mixed effect Egger's tests were employed (Rothstein et al., 2005). These analyses were conducted to gauge and address any potential impact of bias on the overall findings.

2.5. Statistical analyses

2.5.1. Effect size calculations

The selected primary studies documented the impact of language learning on numerical skills using diverse methods, encompassing various formats of statistical analyses such as *t*-values, *F*-values exclusively, *F*-values in conjunction with *p*-values and Cohen's *d*, or by providing descriptive data. We transformed all these statistics into odds ratios (OR), a measure of effect size that shows the strength of the association between exposure and target event due

to its ease of calculation and interpretation (especially true in its logarithmic form) (Norton et al., 2018).

To ensure the fulfilment of the independence assumption for all results recorded during the data extraction stage, we verified they were obtained using different participants, outcome tests or age groups. This led to removing outcomes from studies where either the sample had been used in previous analyses or had been pooled to produce an overall effect. Additionally, some studies provided outcome data on several samples, but the available information did not suffice to calculate individual effect sizes; in these cases, the data were combined across the samples. Such is the case of the work by de Jong (2004), which did not provide precise information on their control group, making it impossible to calculate individual effect sizes for the samples studied (i.e., year 4 and 8).

If there was sufficient data to directly calculate ORs (and log ORs), we recorded the number of students who had the treatment (i.e., foreign language learning) and those who had not (control group), as well as those who had achieved a passing mark on the outcome instrument (i.e., mathematical skills test), and those who failed to create 2×2 tables (see [Table S2](#) in [Appendix S1](#) for an example). Then, the R package 'metafor' (Viechtbauer, 2010) was used to compute ORs (logORs) and variances.

When the study data was in a continuous format, and there was insufficient information to directly calculate ORs, Cohen's *d* and variances were computed instead (Lipsey & Wilson, 2000; Wilson, 2023). The resulting effect sizes were converted into logORs,

following the mechanism, and associated statistical formulae, described by Borenstein et al. (2009) (for additional information, see Figure S1 in Appendix S1). Calculations and results are available in Appendix S6.

2.5.2. Moderating variables

During the process of data extraction, certain variables were identified as potential moderators due to their possible influence on the reported results and their availability in the data. These variables included the type and duration of the intervention (measured in years), the socio-economic status of the treatment group (as reported), the age of the participants at the time of testing, the type of publication, the type of outcome tests used, the gender distribution (as percentage of females) and whether the first language spoken by the control group was the dominant language or not (e.g., for studies in the USA, control groups comprised of native English speakers, native Spanish speakers or both native Spanish and English speakers).

While all variables were directly extracted from the study content, simplifications were necessary for modelling. For instance, intervention types were grouped into BE programs (balanced instruction in both first and second languages) and Second/Foreign language programs (mainly second language instruction). Similarly, 16 types of mathematical skills tests were categorised into standardised and non-standardised tests, depending on their reach and application. A summary of the coding strategy is provided in Table S3 of Appendix S1.

2.5.3. Meta-analytical models

All meta-analytical analyses were conducted using two R packages: 'metafor' (Viechtbauer, 2010) and 'meta' (Balduzzi et al., 2019). Because of the substantial variation in population characteristics (such as age, native and second language), intervention designs (e.g., length and type of language programs) and outcome measures (different types of mathematical skills tests) across studies, and given the assumption of the absence of a 'true' effect size (Borenstein et al., 2009), we opted for a random-effects model to aggregate the calculated effect sizes and estimate an overall effect size.

We utilised the restricted maximum likelihood (REML) estimation to estimate the heterogeneity variance. This choice was based on the literature's suggestion that it produces almost unbiased results (Langan et al., 2019; Tanriver-Ayder et al., 2021; Viechtbauer, 2010). To test individual coefficients and their confidence intervals, we applied the Hartung, Knapp, Sidik and Jonkman (HKSJ) method, which has received widespread recommendations, especially when the number of studies is small (Balduzzi et al., 2019; Int'Hout et al., 2014; Viechtbauer, 2010).

To assess between-study heterogeneity, we used the I^2 statistic, reported as a percentage of the variability in effect sizes (Higgins et al., 2023). A high level of variability in effect sizes, unrelated to sampling error, was expected. To explore this heterogeneity, we constructed a mixed-effects model. We individually tested moderating variables to examine their impact on between-study variability. Significant variables were then incorporated into the model, which we subsequently tested for their relative contribution (Borenstein et al., 2009). The R code and the data input are available in the Supplementary Materials, in Appendices S7 and S8, respectively.

3. Results

3.1. Systematic review

Out of the selected 25 studies, published between 1978 and 2020, 44 individual results were extracted (see Table 1 for details). A total of 785,552 children were included in the analysis; 91,997 were language learners, while 690,555 were in the control groups. Sample sizes ranged from 6 to 47,844 for the language learning group ($Me = 58$) and 6 to 281,642 for the control group ($Me = 9,877.5$). The age range of the participants at the time of the study spanned from 10 to 17 years ($M = 13.97$, $Me = 13.5$, $Std\ Dev = 1.63$), while the duration of the interventions ranged from 1 to 7 years ($M = 3.65$, $Me = 4$).

Regarding study design, 54% of the effect sizes were from secondary analyses of existing data. Quasi-experimental studies contributed 18% of the effect sizes, while cross-sectional surveys accounted for 11%. Experimental studies represented 9% of the effect sizes, and longitudinal quantitative studies contributed 7% to the overall dataset.

Concerning the nature of interventions implemented, 14% of the studies used Second/Foreign Language programs (English as Second Language and Foreign Language), whereas the remaining 86% of interventions were characterised as Bilingual programs (Transitional Bilingual Education, Bilingual Education, TWB Education and DL).

The participants' linguistic backgrounds were somewhat homogeneous, as 34% of the interventions reported including only native Spanish speakers as the language learners, with English as the second language, while 48% of the studies involved both Spanish and English native speakers in a DL setting. Additionally, 7% of the studies included language learners with multilingual background, with nearly 50 different home languages recorded including Spanish, Bosnian, Vietnamese, Portuguese, Chinese, Hindi and Japanese, all learning English as a second language. Furthermore, 9% of the studies reported results on English native speakers, learning a variety of modern languages, including Spanish, French and Italian. The remaining 2% of the studies involved native Korean speakers learning English as a second language.

All studies reported results based on assessments conducted as part of the students' school activities. The tools, such as Renaissance Star 360° Maths (RenaissanceLearning, 2024) and STAAR (TEA, 2024) tests, used by the selected studies covered fundamental mathematical domains, including number properties and operations, measurement, geometry, data analysis and probability, as well as algebraic thinking and problem-solving. While a substantial array of tests was used, with at least 16 assessments identified, most of these assessments (77%) were standardised tests, such as the Stanford Achievement Test (SAT), RS 360 Maths and STAAR tests, mandated for use across all schools within the country or state where the evaluation occurred. The remaining results were derived from assessments conducted at district level (5%) or within individual schools (18%).

3.2. Overall effect and between-study heterogeneity

As previously explained, a random-effects model was used to aggregate the computed effect sizes and estimate the overall effect of language learning on mathematical skills. This model yielded a moderate and statistically significant effect size of 1.09, 95% $CI = [0.75, 1.44]$, $p < 0.0001$ (see Figure 2). When converted back to ORs, this result suggests that, following the treatment, students engaged in language learning exhibited an approximately threefold increased likelihood of attaining a passing grade (or achieving higher results in standardised tests) than their peers, who were

Table 1. Systematic review results summary

Author (year)	Publication type	Type of study	Country (state/city)	Population description	Number of participants (N = Total, LL = Language Learners, NLL = Non-LL)	SES	Gender mix	Age in years (mean) at the time of testing	Languages spoken	Intervention/ exposure	Length of intervention	Assessment tool	Statistical methods used	a. Type of effect size (reported) b. Effect size value (reported) c. Author's interpretation d. LogOR (calculated)
Alvarado (1985)	Dissertation (ProQuest)	Experimental	USA (Eastern Connecticut)	Years 7 and 8 students of Puerto Rican descent, who attended a bilingual program during their elementary school years.	N = 41 LL = 18 NLL = 23	Low	50% female	13 (as per year level)	Spanish English	Transitional Bilingual Education program	>= 2 years	Stanford Achievement Test (SAT) - Math section	Repeated measures of analysis of variance	a. ANOVA <i>F</i> -value (Study) b. <i>F</i> = 0.512; <i>p</i> < 0.500 c. No significant difference between groups d. LogOR = −0.3824
Ardasheva (2012)	Journal article	Descriptive and quantitative	USA (Kentucky)	Years 6, 7, and 8 students who had exited an English as a second language program (Former ELLs)	N = 17,970 LL = 500 NLL = 17,470	Low	51% female	13.49	48 home languages of which the top four were Spanish (47.2%), Bosnian (10.1%), Vietnamese (5.3%), and Maymay (4.7%) English	English as a Second Language (content-based)	>= 4 years	Kentucky Core Content Tests (KCCT) - Total Math section	Two-level hierarchical linear model (HLM)	a. Cohen's <i>d</i> and 95% confidence intervals (CIs) (Study) b. <i>d</i> = 0.42, 95% CI [0.33, 0.51] c. Significant difference favouring intervention group d. LogOR = 0.7618
Chan (1982)	Report	Secondary analysis of existing data (longitudinal)	USA	Years 10 and 12 students, Spanish native speakers, who only attended school in the US and required some form of ESL program, dual language	N = 442 LL = 274 NLL = 168	Medium	Not given	16 (as per year level)	Spanish English	Dual-Language Bilingual Program	Not given	High school and beyond (HS&B) database maths test	Multiple regression analysis	a. Test of significance (<i>p</i> -value) at α = 0.05 (Study) b. <i>p</i> = 2.67 (>2.99 significant) c. Approaches significant difference between groups d. LogOR = 0.9146 (Calculated based on Medium SES groups)
Cobb (2006)	Journal article	Ex post facto, non-randomized, matched-pairs alternative group design	USA (Northern Colorado)	Years 6 to 7, native English speaker students enrolled in a two-way bilingual elementary school	N = 22 LL = 11 NLL = 11	Low	Not given	12 (as per year level)	Spanish English	Two-way Bilingual Immersion Program	>= 4 years	Colorado Student Assessment Program (CSAP) mathematics subtest	ANOVA with repeated measures	a. Cohen's <i>d</i> (Study) b. <i>d</i> = 0.15 c. Small difference favouring the intervention group d. LogOR = 0.2721

(Continued)

Table 1. (Continued)

Author (year)	Publication type	Type of study	Country (state/city)	Population description	Number of participants (N = Total, LL = Language Learners, NLL = Non-LL)	SES	Gender mix	Age in years (mean) at the time of testing	Languages spoken	Intervention/ exposure	Length of intervention	Assessment tool	Statistical methods used	a. Type of effect size (reported) b. Effect size value (reported) c. Author's interpretation d. LogOR (calculated)
														*) Small to medium effect size was found when both NES and NSS were considered
Cobb (2006)	Journal article	Ex post facto, non-randomised, matched-pairs alternative group design	USA (Northern Colorado)	Years 6 to 7, native Spanish speaker students enrolled in a two-way bilingual elementary school	N = 12 LL = 6 NLL = 6	Low	Not given	12 (as per year level)	Spanish English	Two-way Bilingual Immersion Program	>= 4 years	Colorado Student Assessment Program (CSAP) mathematics subtest	ANOVA with repeated measures	a. Cohen's d (Study) b. d =1.42 c. Very large difference favouring intervention group d. LogOR = 2.5756 *) Small to medium effect size was found when both NES and NSS were considered.
de Jong (2004)	Journal article	Report (Secondary analysis of existing data)	USA (Northeastern)	Years 4 and 8 students, who attended an English as a Second Language or Bilingual program and were deemed ready for inclusion in a mainstream program. Only year 8 students were considered	N = 56 (LL)	Low	Not given	13.5 (as per year level)	Various (Spanish, Portuguese, Russian, Chinese, Hindi, Japanese, etc.) English	English as a Second Language and Bilingual Education Language programs	3 years	State-level standardised test - Mandate state test (Specifics were not given)	T-test analysis	a. t-test (Study) b. p = 0.04 c. Relatively significant difference favouring intervention group* d. LogOR = 1.0375 *) Significant difference favouring the intervention group in grade 4. NOTE: Individual effect sizes could not be calculated due to the lack of information on the control group.

(Continued)

Table 1. (Continued)

Author (year)	Publication type	Type of study	Country (state/city)	Population description	Number of participants (N = Total, LL = Language Learners, NLL = Non-LL)	SES	Gender mix	Age in years (mean) at the time of testing	Languages spoken	Intervention/ exposure	Length of intervention	Assessment tool	Statistical methods used	a. Type of effect size (reported) b. Effect size value (reported) c. Author's interpretation d. LogOR (calculated)
Falter (1991)	Journal article	Experimental	Canada (Ontario)	Year 7 students enrolled in French immersion classes.	N = 275 LL = 131 NLL = 144	Not given	Not given	12.5 (as per year level)	English French	Foreign language immersion	>= 6 years	State-level standardised test - Maths test based upon the Ministry of Education guidelines (Specifics were not given)	Correlation and Analysis of variance	a. ANOVA F-value, p -value and eta square (Study) b. $F = 8.72$, $p < 0.01$, $\eta^2 = 0.13$ c. Very significant difference between groups d. LogOR = 0.7458 *) Significant difference favouring the intervention group for grade 7 only
Flores (1981)	Dissertation (ProQuest)	Experimental	USA (Florida)	Year 12 Spanish-speaking students who attended a bilingual program	N = 95 LL = 48 NLL = 47	Low	Not given	17.5 (as per year level)	Spanish English	Bilingual Education program	6 years	Standford Achievement Test (SAT) - Math section	t -test	a. t -value (Study) b. $t = -1.75$ (Y11) c. The result approaches statistical significance d. LogOR = -0.6513
Houston Independent School District (HISD) (2020)	Report	Report (Secondary analysis of existing data)	USA (Texas)	Students both Spanish and English native speakers, attending years 3 to 12 who exited the Dual Language program in 2020 (Exited DL) Only year 7 students were considered	N = 10,463 LL = 37 NLL = 10,426	Low/medium*	Not given	12.5 (as per year level)	Spanish English	Dual-Language Bilingual Program	>= 4 years	District-level assessments (DLAs) Math section - based on State of Texas Assessments of Academic Readiness (STAAR)	Descriptive statistics only	a. N/A b. N/A c. N/A d. LogOR = 1.5334
Houston Independent School District (HISD) (2020)	Report	Report (Secondary analysis of existing data)	USA (Texas)	Students both Spanish and English native speakers, attending years 3 to 12 who exited the Dual Language	N = 8,443 LL = 21 NLL = 8,422	Low/medium*	Not given	13.5 (as per year level)	Spanish English	Dual-Language Bilingual Program	>= 4 years	District-level assessments (DLAs) Math section - based on State of Texas Assessments of Academic	Descriptive statistics only	a. N/A b. N/A c. N/A d. LogOR = 1.5555

(Continued)

Table 1. (Continued)

Author (year)	Publication type	Type of study	Country (state/city)	Population description	Number of participants (<i>N</i> = Total, LL = Language Learners, NLL = Non-LL)	SES	Gender mix	Age in years (mean) at the time of testing	Languages spoken	Intervention/exposure	Length of intervention	Assessment tool	Statistical methods used	a. Type of effect size (reported) b. Effect size value (reported) c. Author's interpretation d. LogOR (calculated)
				program in 2020 (Exited DL) Only year 8 students were considered								Readiness (STAAR)		
Houston Independent School District (HISD) (2020)	Report	Report (Secondary analysis of existing data)	USA (Texas)	Students both Spanish and English native speakers, attending years 3 to 12 who exited the Dual Language program in 2020 (Exited DL) Only years 6 to 12 students were considered	<i>N</i> = 52,838 LL = 221 NLL = 52,617	Low/medium*	Not given	14.5 (as per year level)	Spanish English	Dual-Language Bilingual Program	>= 4 years	Renaissance Star 360 * (R360) English Maths section	Descriptive statistics only	a. N/A b. N/A c. N/A d. LogOR = 1.9924
Houston Independent School District (HISD) (2020)	Report	Report (Secondary analysis of existing data)	USA (Texas)	Students both Spanish and English native speakers, attending years 3 to 12 who exited the Dual Language program in 2020 (Exited DL) Only years 9 to 11 students were considered	<i>N</i> = 10,197 LL = 34 NLL = 10,163	Low/medium*	Not given	16 (as per year level)	Spanish English	Dual-Language Bilingual Program	>= 4 years	School achievement test - End-of-Course Subjects: Algebra I	Descriptive statistics only	a. N/A b. N/A c. N/A d. LogOR = 1.4171
Houston Independent School District (HISD) (2019)	Report	Report (Secondary analysis of existing data)	USA (Texas)	Students both Spanish and English native speakers, attending years 3 to 12 who exited the Dual Language program in 2019 (Exited DL) Only year 7 students were considered	<i>N</i> = 12,475 LL = 58 NLL = 12,417	Low/medium*	Not given	12.5 (as per year level)	Spanish English	Dual-Language Bilingual Program	>= 4 years	State of Texas Assessments of Academic Readiness (STAAR)	Descriptive statistics only	a. N/A b. N/A c. N/A d. LogOR = 3.9621

(Continued)

Table 1. (Continued)

Author (year)	Publication type	Type of study	Country (state/city)	Population description	Number of participants (N = Total, LL = Language Learners, NLL = Non-LL)	SES	Gender mix	Age in years (mean) at the time of testing	Languages spoken	Intervention/exposure	Length of intervention	Assessment tool	Statistical methods used	a. Type of effect size (reported) b. Effect size value (reported) c. Author's interpretation d. LogOR (calculated)
Houston Independent School District (HISD) (2019)	Report	Report (Secondary analysis of existing data)	USA (Texas)	Students both Spanish and English native speakers, attending years 3 to 12 who exited the Dual Language program in 2019 (Exited DL) Only year 8 students were considered	N = 10,615 LL = 23 NLL = 10,592	Low/medium*	Not given	13.5 (as per year level)	Spanish English	Dual-Language Bilingual Program	>= 4 years	State of Texas Assessments of Academic Readiness (STAAR)	Descriptive statistics only	a. N/A b. N/A c. N/A d. LogOR = 2.2336
Houston Independent School District (HISD) (2019)	Report	Report (Secondary analysis of existing data)	USA (Texas)	Students both Spanish and English native speakers, attending years 3 to 12 who exited the Dual Language program in 2019 (Exited DL) Only years 9 to 11 students were considered	N = 14,806 LL = 67 NLL = 14,739	Low/medium*	Not given	16 (as per year level)	Spanish English	Dual-Language Bilingual Program	>= 4 years	School achievement test - End-of-Course Subjects: Algebra I	Descriptive statistics only	a. N/A b. N/A c. N/A d. LogOR = 2.1321
Houston Independent School District (HISD) (2018)	Report	Report (Secondary analysis of existing data)	USA (Texas)	Students both Spanish and English native speakers, attending years 3 to 12 who exited the Dual Language program in 2018 (Exited DL) Only year 7 students were considered	N = 12,921 LL = 58 NLL = 12,863	Low/medium*	Not given	12.5 (as per year level)	Spanish English	Dual-Language Bilingual Program	>= 4 years	State of Texas Assessments of Academic Readiness (STAAR)	Descriptive statistics only	a. N/A b. N/A c. N/A d. LogOR = 2.9007
Houston Independent School District (HISD) (2018)	Report	Report (Secondary analysis of existing data)	USA (Texas)	Students both Spanish and English native speakers, attending years 3 to 12 who exited the Dual Language	N = 10,468 LL = 36 NLL = 10,432	Low/medium*	Not given	13.5 (as per year level)	Spanish English	Dual-Language Bilingual Program	>= 4 years	State of Texas Assessments of Academic Readiness (STAAR)	Descriptive statistics only	a. N/A b. N/A c. N/A d. LogOR = 0.9680

(Continued)

Table 1. (Continued)

Author (year)	Publication type	Type of study	Country (state/city)	Population description	Number of participants (N = Total, LL = Language Learners, NLL = Non-LL)	SES	Gender mix	Age in years (mean) at the time of testing	Languages spoken	Intervention/exposure	Length of intervention	Assessment tool	Statistical methods used	a. Type of effect size (reported) b. Effect size value (reported) c. Author's interpretation d. LogOR (calculated)
				program in 2018 (Exited DL) Only year 8 students were considered										
Houston Independent School District (HISD) (2018)	Report	Report (Secondary analysis of existing data)	USA (Texas)	Students both Spanish and English native speakers, attending years 3 to 12 who exited the Dual Language program in 2018 (Exited DL) Only years 9 to 11 students were considered	N = 16,127 LL = 98 NLL = 16,029	Low/medium*	Not given	16 (as per year level)	Spanish English	Dual-Language Bilingual Program	>= 4 years	School achievement test - End-of-Course Subjects: Algebra I	Descriptive statistics only	a. N/A b. N/A c. N/A d. LogOR = 1.4477
Houston Independent School District (HISD) (2017)	Report	Report (Secondary analysis of existing data)	USA (Texas)	Students both Spanish and English native speakers, attending years 3 to 12 who exited the Dual Language program in 2017 (Exited DL) Only year 7 students were considered	N = 12,607 LL = 77 NLL = 12,530	Low/medium*	Not given	12.5 (as per year level)	Spanish English	Dual-Language Bilingual Program	>= 4 years	State of Texas Assessments of Academic Readiness (STAAR)	Descriptive statistics only	a. N/A b. N/A c. N/A d. LogOR = 1.6219
Houston Independent School District (HISD) (2017)	Report	Report (Secondary analysis of existing data)	USA (Texas)	Students both Spanish and English native speakers, attending years 3 to 12 who exited the Dual Language program in 2017 (Exited DL) Only year 8 students were considered	N = 10,818 LL = 58 NLL = 10,760	Low/medium*	Not given	13.5 (as per year level)	Spanish English	Dual-Language Bilingual Program	>= 4 years	State of Texas Assessments of Academic Readiness (STAAR)	Descriptive statistics only	a. N/A b. N/A c. N/A d. LogOR = 0.9666

(Continued)

Table 1. (Continued)

Author (year)	Publication type	Type of study	Country (state/city)	Population description	Number of participants (N = Total, LL = Language Learners, NLL = Non-LL)	SES	Gender mix	Age in years (mean) at the time of testing	Languages spoken	Intervention/exposure	Length of intervention	Assessment tool	Statistical methods used	a. Type of effect size (reported) b. Effect size value (reported) c. Author's interpretation d. LogOR (calculated)
Houston Independent School District (HISD) (2017)	Report	Report (Secondary analysis of existing data)	USA (Texas)	Students both Spanish and English native speakers, attending years 3 to 12 who exited the Dual Language program in 2017 (Exited DL) Only years 9 to 11 students were considered	N = 16,378 LL = 115 NLL = 16,263	Low/medium*	Not given	16 (as per year level)	Spanish English	Dual-Language Bilingual Program	>= 4 years	School achievement test - End-of-Course Subjects: Algebra I	Descriptive statistics only	a. N/A b. N/A c. N/A d. LogOR = 2.6288
Houston Independent School District (HISD) (2016)	Report	Report (Secondary analysis of existing data)	USA (Texas)	Students both Spanish and English native speakers, attending years 3 to 12 who exited the Dual Language program in 2016 (Exited DL) Only year 7 students were considered	N = 11,793 LL = 108 NLL = 11,685	Low/medium*	Not given	12.5 (as per year level)	Spanish English	Dual-Language Bilingual Program	>= 4 years	State of Texas Assessments of Academic Readiness (STAAR)	Descriptive statistics only	a. N/A b. N/A c. N/A d. LogOR = 1.5339
Houston Independent School District (HISD) (2016)	Report	Report (Secondary analysis of existing data)	USA (Texas)	Students both Spanish and English native speakers, attending years 3 to 12 who exited the Dual Language program in 2016 (Exited DL) Only year 8 students were considered	N = 9,658 LL = 66 NLL = 9,592	Low/medium*	Not given	13.5 (as per year level)	Spanish English	Dual-Language Bilingual Program	>= 4 years	State of Texas Assessments of Academic Readiness (STAAR)	Descriptive statistics only	a. N/A b. N/A c. N/A d. LogOR = 1.5154
Houston Independent School District (HISD) (2016)	Report	Report (Secondary analysis of existing data)	USA (Texas)	Students both Spanish and English native speakers, attending years 3 to 12 who exited the Dual Language	N = 13,896 LL = 100 NLL = 13,796	Low/medium*	Not given	16 (as per year level)	Spanish English	Dual-Language Bilingual Program	>= 4 years	School achievement test - End-of-Course Subjects: Algebra I	Descriptive statistics only	a. N/A b. N/A c. N/A d. LogOR = 2.0000

(Continued)

Table 1. (Continued)

Author (year)	Publication type	Type of study	Country (state/city)	Population description	Number of participants (N = Total, LL = Language Learners, NLL = Non-LL)	SES	Gender mix	Age in years (mean) at the time of testing	Languages spoken	Intervention/exposure	Length of intervention	Assessment tool	Statistical methods used	a. Type of effect size (reported) b. Effect size value (reported) c. Author's interpretation d. LogOR (calculated)
				program in 2016 (Exited DL) Only years 9 to 11 students were considered										
Houston Independent School District (HISD) (2014)	Report	Report (Secondary analysis of existing data)	USA (Texas)	Students both Spanish and English native speakers, attending years 3 to 12 who exited the Dual Language program in 2014 (Exited DL) Only year 8 students were considered	N = 9,482 LL = 18 NLL = 9,464	Low/medium*	Not given	13.5 (as per year level)	Spanish English	Dual-Language Bilingual Program	>= 3 years	State of Texas Assessments of Academic Readiness (STAAR)	Descriptive statistics only	a. N/A b. N/A c. N/A d. LogOR = 0.6412
Houston Independent School District (HISD) (2014)	Report	Report (Secondary analysis of existing data)	USA (Texas)	Students both Spanish and English native speakers, attending years 3 to 12 who exited the Dual Language program in 2014 (Exited DL) Only years 9 to 11 students were considered	N = 13,422 LL = 67 NLL = 13,355	Low/medium*	Not given	16 (as per year level)	Spanish English	Dual-Language Bilingual Program	>= 3 years	School achievement test - End-of-Course Subjects: Algebra I	Descriptive statistics only	a. N/A b. N/A c. N/A d. LogOR = 2.3775
Houston Independent School District (HISD) (2013)	Report	Report (Secondary analysis of existing data)	USA (Texas)	Students both Spanish and English native speakers, attending years 3 to 12 who exited the Dual Language program in 2013 (Former Two-way) Only year 8 students were considered	N = 12,420 LL = 19 NLL = 12,401	Low/medium*	Not given	13.5 (as per year level)	Spanish English	Two-Way Bilingual Immersion Program	>= 3 years	State of Texas Assessments of Academic Readiness (STAAR)	Descriptive statistics only	a. N/A b. N/A c. N/A d. LogOR = 2.5110

(Continued)

Table 1. (Continued)

Author (year)	Publication type	Type of study	Country (state/city)	Population description	Number of participants (N = Total, LL = Language Learners, NLL = Non-LL)	SES	Gender mix	Age in years (mean) at the time of testing	Languages spoken	Intervention/exposure	Length of intervention	Assessment tool	Statistical methods used	a. Type of effect size (reported) b. Effect size value (reported) c. Author's interpretation d. LogOR (calculated)
Houston Independent School District (HISD) (2013)	Report	Report (Secondary analysis of existing data)	USA (Texas)	Students both Spanish and English native speakers, attending years 3 to 12 who exited the Dual Language program in 2013 (Exited TWBIP) Only years 9 and 10 students were considered	N = 11,927 LL = 82 NLL = 11,845	Low/medium*	Not given	15.5 (as per year level)	Spanish English	Two-Way Bilingual Immersion Program	>= 3 years	School achievement test - End-of-Course Subjects: Algebra I	Descriptive statistics only	a. N/A b. N/A c. N/A d. LogOR = 0.4971
Houston Independent School District (HISD) (2013)	Report	Report (Secondary analysis of existing data)	USA (Texas)	Students both Spanish and English native speakers, attending years 3 to 12 who exited the Dual Language program in 2013 (Exited TWBIP) Only year 11 students were considered	N = 9,329 LL = 59 NLL = 9,270	Low/medium*	Not given	16.5 (as per year level)	Spanish English	Two-Way Bilingual Immersion Program	>= 3 years	Texas Assessment of Knowledge and Skills (TAKS)	Descriptive statistics only	a. N/A b. N/A c. N/A d. LogOR = 1.9909
Lamia (2015)	Dissertation (ProQuest)	Blended methodology study	USA (New York)	Year 9 students from a single school, who had not had French, Italian or Spanish instruction prior to year 8.	N = 108 (LL)	Medium	49% female	13.5 (as per year level)	English French, Italian, Spanish	Foreign language courses	1 year	State-level standardised test - New York Grade 8 foreign language and Grade 8 Math scores.	Correlation and Analysis of variance (Pearson's r)	a. Pearson's r and R ² (study) b. r = 0.554, r ² = 30.7% c. There's a Moderate positive correlation between foreign language proficiency scores and attitudes and Maths scores. Also, 30.7% (Moderate effect) of that variation can be explained by the adherence to the treatment. d. LogOR = 2.4142

(Continued)

Table 1. (Continued)

Author (year)	Publication type	Type of study	Country (state/city)	Population description	Number of participants (N = Total, LL = Language Learners, NLL = Non-LL)	SES	Gender mix	Age in years (mean) at the time of testing	Languages spoken	Intervention/exposure	Length of intervention	Assessment tool	Statistical methods used	a. Type of effect size (reported) b. Effect size value (reported) c. Author's interpretation d. LogOR (calculated)
Lane (1995)	Report	Longitudinal	USA	Latino Year 8 students from a QUASAR project school attending a bilingual program Only Spring 94 data was considered	N = 57 LL = 6 NLL = 51	Not given	50% female	13.5 (as per year level)	Spanish English	Bilingual Education program	>= 4 years	QUASAR Cognitive Assessment Instrument (QCAI) - Grade 8 version	Descriptive statistics only	a. N/A b. N/A c. N/A d. LogOR = -0.2018
Lee (2010)	Dissertation (ProQuest)	Descriptive and Quantitative (second-hand use of existing data)	South Korea	Female South Korean grade 11 students from a single school	N = 391 (LL)	Medium	100% female	16	Korean English Chinese, Japanese	English as a Second Language program	7 years	School achievement test -Grade 11 achievement	Pearson product-moment correlation analysis	a. Pearson's r (study) and R ² (calculated) b. Adv Math r = 0.771, r ² = 59.44% Gen Math r = 0.767, r ² = 58.83% c. There's a High positive correlation between foreign language proficiency scores and Maths scores. Also, ~ 59% (Strong effect) of that variation can be explained by the adherence to the treatment. d. LogOR = 4.3368
Medina et al. (1985)	Journal article	Experimental	USA (Arizona)	Mexican American Spanish-speaking children in grades 6, 8, and 12 who attended a bilingual program	N = 43 LL = 19 NLL = 24	Medium	Not given	14.5 (as per year levels)	Spanish English	Transitional Bilingual Education program	5 years	Metropolitan Achievement Test (Computation, Concepts and Problem-Solving subtests) in grades 6 and 8 and the same subtests of the California Achievement Test in grade 12	t-test for independent samples	a. t-test (study) b. Y12 = -0.6, Y8 = -1.28, Y6 = -.56 c. No Significant differences were found in the performance of the two groups in any of the year levels and MAT subtests. d. LogOR = -0.3341

(Continued)

Table 1. (Continued)

Author (year)	Publication type	Type of study	Country (state/city)	Population description	Number of participants (<i>N</i> = Total, LL = Language Learners, NLL = Non-LL)	SES	Gender mix	Age in years (mean) at the time of testing	Languages spoken	Intervention/exposure	Length of intervention	Assessment tool	Statistical methods used	a. Type of effect size (reported) b. Effect size value (reported) c. Author's interpretation d. LogOR (calculated)
Medina (1993)	Journal article	Descriptive and Quantitative	USA (Arizona)	Mexican American Spanish-speaking children in grades 4 to 8, of low SES, who participated in a bilingual program	<i>N</i> = 676 (LL)	Low	Not given	10 (as per year levels)	Spanish English	Transitional Bilingual Education program	Unclear	La Prueba Riverside de Realizacion en Español (La Prueba) - Mathematics	Analysis of Variation (F-ratios)	a. F-ratio (study) b. Grade 1–8 Math, <i>F</i> = 8.93, <i>p</i> < 0.01 c. There is a small but significant effect of grade levels on math results in Spanish, probably showing the long-term benefits of the program. d. LogOR = 0.4229
Olson (1989)	Report	Secondary analysis of existing data (Longitudinal)	USA	Mexican American sophomores (grade 9 students), who spoke Spanish as L1 and were part of an English language learning program.	<i>N</i> = 306 LL = 168 NLL = 138	Low	Not given	14.5 (as per year level)	Spanish English	Bilingual Education program (various)	Unclear	High school and beyond (HS&B) database maths test	Descriptive statistics only	a. N/A b. N/A c. N/A d. LogOR = 0.0040
Powers (1978)	Dissertation (ProQuest)	Ex-post facto study	USA (Arizona)	Mexican American students attending year 7 who attended a bilingual education program.	<i>N</i> = 87 LL = 44 NLL = 43	Low	57% female	12.5 (as per year level)	Spanish English	Bilingual Education program	3–5 years	Stanford Achievement Test (SAT) - Math section	t-test and analysis of variance	a. t-test (study) b. SAT <i>t</i> = 0.81, <i>p</i> = 0.056 c. No Significant differences were found in the performance of the two groups in any of the maths tests. d. LogOR = 0.3151
Powers (1978)	Dissertation (ProQuest)	Ex-post facto study	USA (Arizona)	Mexican American students attending year 7 who attended a bilingual education program.	<i>N</i> = 87 LL = 44 NLL = 44	Low	57% female	12.5 (as per year level)	Spanish English	Bilingual Education program	3–5 years	Wide Range Achievement Test - Arithmetic subtest	t-test and analysis of variance	a. t-test (study) b. WRAT <i>t</i> = −1.92, <i>P</i> = 0.058 c. No Significant differences were found in the performance of the two

(Continued)

Table 1. (Continued)

Author (year)	Publication type	Type of study	Country (state/city)	Population description	Number of participants (N = Total, LL = Language Learners, NLL = Non-LL)	SES	Gender mix	Age in years (mean) at the time of testing	Languages spoken	Intervention/exposure	Length of intervention	Assessment tool	Statistical methods used	a. Type of effect size (reported) b. Effect size value (reported) c. Author's interpretation d. LogOR (calculated)
														groups in any of the maths tests. d. LogOR = -0.7467
Slama et al. (2017)	Report	Longitudinal	USA (Texas)	Hispanic English learner students who were in grade 1 in traditional public schools in 2005/06 (Grade 8 at the end of the study)	N = 320,878 LL = 39,236 NLL = 281,642	Low	48% female	11.5 (as per year level)	Spanish English	Bilingual Education program (various)	>= 2 years	Texas Assessment of Knowledge and Skills (TAKS)	Descriptive statistics only	a. N/A b. N/A c. N/A d. LogOR = 0.4950
Slama et al. (2017)	Report	Longitudinal	USA (Texas)	Hispanic English learner students who were in grade 1 in traditional public schools in 2005/06 (Grade 8 at the end of the study)	N = 147,457 LL = 47,844 NLL = 99,613	Low	48% female	13.5 (as per year level)	Spanish English	Bilingual Education program (various)	>= 2 years	State of Texas Assessments of Academic Readiness (STAAR)	Descriptive statistics only	a. N/A b. N/A c. N/A d. LogOR = -0.4689
Smith (2016)	Dissertation (ProQuest)	Quantitative research, comparative design	USA	Bilingual students ages 14–18, in grades 9 through 12, who attended a DL program	N = 78 LL = 52 NLL = 26	Not given	Not given	16	Spanish English	Dual-Language Bilingual Program	1 – 12 years	Smarter Balanced Assessment Consortium (SBAC) - Maths	Factorial ANOVA	a. ANOVA <i>F</i> -value, <i>p</i> -value and eta square (study) b. $F = 6.739$ $p = 0.14$, $\eta^2 = 0.174$ c. Significant effect of enrolment status (all English levels) favouring Dual Language Education d. LogOR = 1.1309 NOTE: Group-specific effect sizes couldn't be calculated as for some subgroups $n = 1$

(Continued)

Table 1. (Continued)

Author (year)	Publication type	Type of study	Country (state/city)	Population description	Number of participants (N = Total, LL = Language Learners, NLL = Non-LL)	SES	Gender mix	Age in years (mean) at the time of testing	Languages spoken	Intervention/exposure	Length of intervention	Assessment tool	Statistical methods used	a. Type of effect size (reported) b. Effect size value (reported) c. Author's interpretation d. LogOR (calculated)
Tobias (2012)	Dissertation (ProQuest)	Quasi-experimental (second-hand use of existing data)	USA (Tennessee)	Foreign language learners who were in grade 6 in 2009 (grade 8 in 2011)	N = 100,519 LL = 906 NLL = 99,613	Low	49% female	13.5 (as per year level)	Spanish English Other	Foreign language courses	>= 1 year	Mathematics Tennessee Comprehensive Assessment Program (TCAP)	t-test	a. t-test (study) b. $t = 5.85$, $\alpha < 0.5$ c. A significant difference was found favouring the treatment group d. LogOR = 0.5552
Tobias (2012)	Dissertation (ProQuest)	Quasi-experimental (second-hand use of existing data)	USA (Tennessee)	English learners who were in grade 6 in 2009 (grade 8 in 2011)	N = 99,737 LL = 124 NLL = 99,613	Low	44% female	13.5 (as per year level)	Spanish English	English as a Second Language program	>= 1 year	Mathematics Tennessee Comprehensive Assessment Program (TCAP)	t-test	a. t-test (study) b. $t = 2.87$, $\alpha < 0.5$ c. A significant difference was found between the two groups on the TCAP results. d. LogOR = -0.2216
Vega (2015)	Dissertation (ProQuest)	Longitudinal quasi-experimental	USA	Native English speakers attending a Two-way immersion program from year 6 to 10. Only year 10 students were included	N = 90 LL = 45 NLL = 45	Medium	53% female	15.5 (as per year level)	Spanish English	Two-Way Bilingual Immersion Program	>= 4 years	Colorado Student Assessment Program (CSAP) mathematics subtest	2 x 2 mixed ANOVA	a. Cohen's d (study)* b. Y6 d = 0.26, Y7 d = 0.03, Y8 d = 0.06, Y9 d = 0.07, Y10 d = 0.07 c. Very small effect sizes were seen at all grade levels for English speakers d. LogOR = 0.1183
Vega (2015)	Dissertation (ProQuest)	Longitudinal quasi-experimental	USA (Colorado)	Native Spanish speakers attending a Two-way immersion program from year 6 to 10 Only year 10 students were included	N = 88 LL = 44 NLL = 44	Low	53% female	15.5 (as per year level)	Spanish English	Two-Way Bilingual Immersion Program	>= 4 years	Colorado Student Assessment Program (CSAP) mathematics subtest	2 x 2 mixed ANOVA	a. Cohen's d (study)* b. Y6 d = 0.42, Y7 d = 0.47, Y8 d = 0.53, Y9 d = 0.37, Y10 d = 0.50 c. Moderate effect sizes were seen at all grade levels for Spanish speakers d. LogOR = 0.8966

*Houston Independent School District (HISD) census information show that around 58-60% of the population in those years were economically disadvantaged, while the rest were predominantly middle class.

not involved in language lessons, as observed in their performance on mathematical skills tests.

Figure 2 shows the distribution of effect sizes from the selected studies, accompanied by their respective 95% confidence intervals.

The dotted reference line, positioned at a log OR of 0, serves as a benchmark for no discernible difference in the outcomes of the mathematical skill tests. Consequently, all effect sizes situated to the right of this reference line indicate outcomes that favour the

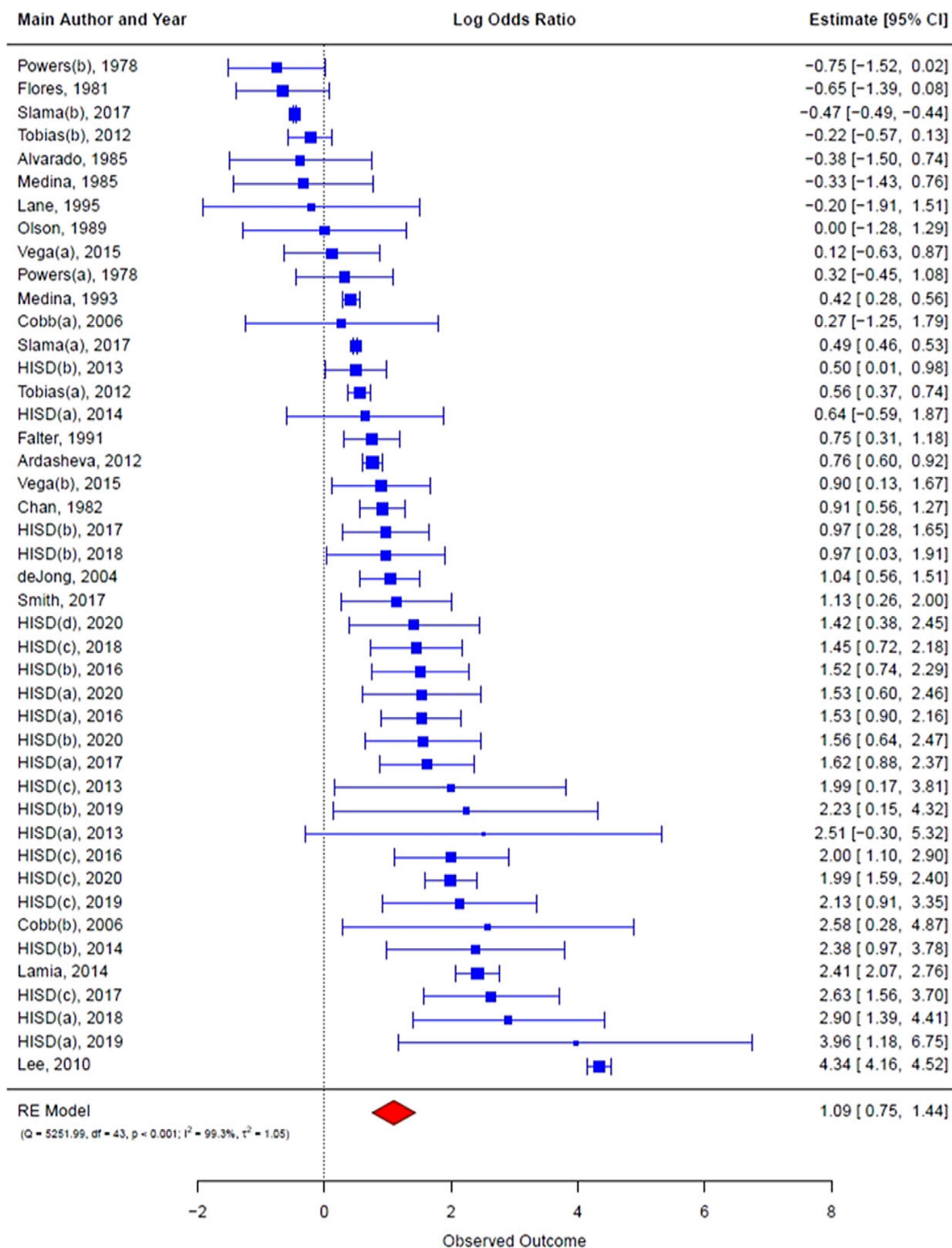


Figure 2. Forest plot for the effect of foreign language learning on mathematical skills: Studies are denoted by square symbols and horizontal lines, indicating their effect size and confidence intervals respectively. The size of the squares is proportional to the weight of each study within the model, reflecting their sample size. The central position of the diamond shape conveys the overall effect, while the 95% confidence intervals are represented by the horizontal extents of the diamond's corners.

treatment, while those on the left depict results that favour the control group.

Adopting a random-effects model was deemed appropriate due to the considerable variability in effect sizes observed across the selected studies (Borenstein et al., 2009). This choice was further justified by the substantial heterogeneity detected ($Q_{43} = 5251.99$, $p < 0.0001$, $\tau^2 = 1.05$ SE = 0.27), as well as the high degree of total variability in the true effects, $I^2 = 99.31\%$.

3.3. Analyses of moderating variables

Given the high level of heterogeneity exhibited in the random-effect model, a mixed-effect model was used to investigate the roles of the selected moderating variables. Their effects were assessed individually, as summarised in Table 2.

Four of the nine variables analysed demonstrated an explanatory impact on the heterogeneity. Interestingly, gender emerged as the most influential factor, contributing to 58.33% of the observed differences in effect sizes. This indicates that studies with a higher proportion of female students reported larger effect size (Figure 3D). However, it is essential to highlight that this metric was derived from approximately a quarter of the studies ($n = 13$), underscoring the need for a cautious interpretation of this finding.

Among the remaining statistically significant results, the impact of socioeconomic status (SES) (45.56%) and the length of the intervention (9.69%) were observed. SES is commonly acknowledged as a crucial influencing factor, given its association with adverse effects on educational outcomes, often observed in schools or children with lower socioeconomic levels (OECD, 2019). This pattern was observed in the current study, with a larger effect size reported among studies that examined students from higher SES background (Figure 3A). Similarly, the results indicate there is a relationship between the length of the intervention and improved outcomes (Figure 3B), aligning with the notion that longer interventions yield more favourable results.

The observed effect for the year of publication (Figure 3C) raises the possibility that recent years may exhibit a bias towards publishing research with positive and significant correlations (see Section 3.4 for further details). However, it is also possible that a cluster of studies with similar results is influencing the data. While this inference is not conclusive, it prompts a discussion on the potential effect of subsets on the overall combined effect. A further analysis on this point is presented below.

Next, a mixed-effect model incorporating all relevant variables was constructed to gain a comprehensive understanding of the data. Nevertheless, the presence of collinearity needed to be reviewed; hence, Pearson's correlations were computed among the relevant variables (Olivoto, 2020). No significant collinearity concerns emerged with the exception of Gender mix, which exhibited a strong correlation with both outcome test type ($r = 0.98$, $p < 0.001$) and length of intervention ($r = 0.78$, $p < 0.01$). Consequently, gender mix was omitted from the mixed-effect model (see Figure S2 in Appendix S1 for details).

The constructed mixed-effect model successfully explained a considerable proportion of the observed heterogeneity ($R^2 = 66.16\%$) using most of the effect sizes in the meta-analysis ($k = 38$).

Nevertheless, a significant residual between-study variance persisted, as indicated by the results of the residual heterogeneity test ($QE_{29} = 153.08$, $p < 0.0001$), signifying unaccounted sources of variability.

In this model, both the SES and the publication year maintained their statistical significance, as indicated in Table 3. Furthermore, the length of the intervention, which exhibited a near-significant association in isolation, achieved a p -value of 0.0397 in the multivariate model, thereby solidifying its role as a significant factor in the correlation between language learning and mathematics.

The observed result concerning the year of publication may stem from a concentration of studies from the same source. As shown in Figure 1, a series of studies was produced by the Houston Independent School District (HISD) between 2013 and 2020. A follow-up analysis was conducted to examine whether this subset of relatively more recent studies differs significantly from other studies included in the review, utilising the same model characteristics employed for the total sample.

In the case of the HISD studies ($k = 21$), the applied model yielded an estimate of 1.59 ($p < 0.0001$) with low heterogeneity ($I^2 = 49\%$). Relatively small heterogeneity makes sense as these studies adhered to a consistent methodology, including factors such as the type and duration of intervention, as well as the nature and timing of outcome tests. In contrast, for the non-HISD studies ($k = 23$), the pooled effect size was smaller but still positive and statistically significant (estimate = 0.61, $p = 0.0206$). However, it is important to note that the heterogeneity within this subgroup was considerably higher ($I^2 = 99\%$), as anticipated due to variations in intervention approaches, participant age and types of outcome tests. As such, the findings of this analysis provide a potential

Table 2. Mixed-effect model: Individual variables assessment

	R^2	b	SE	t	p	95% CI		k
Length of intervention	9.69%	0.2658	0.1367	1.9449	0.0592*	-0.0109	0.5425	40
Type of intervention	5.07%	0.6676	0.4525	1.4755	0.1475	-0.2455	1.5807	44
Control group language/s	0.0%	0.2840	0.1921	1.4788	0.1467	-0.1036	0.6716	44
Socio-economic status	45.56%	1.0721	0.2086	5.1398	<0.0001*	0.6502	1.4941	41
Age	3.06%	0.1303	0.1036	1.2577	0.2154	-0.0788	0.3393	44
Publication type	0.00%	-0.0603	0.2647	-0.2280	0.8208	-0.5944	0.4738	44
Outcome test type	0.00%	0.4130	0.4491	0.9196	0.3630	-0.4933	1.3192	44
Gender mix	58.33%	0.0748	0.0181	4.1347	0.0017*	0.0350	0.1147	13
Year of publication	22.16%	0.0430	0.0110	3.8899	0.0004*	0.0207	0.0653	44

Moderating variables analyses. R^2 refers to the amount of heterogeneity accounted for per moderator, and b is the variation relative to the combined effect.

*Statistically significant result ($p < .05$).

*Approaches statistical significance.

explanation for the moderating role of the year of publication. It suggests that studies from HISD in recent years tended to report larger effect sizes, signifying that this trend was not solely a result of the year of publication.

3.4. Publication bias analyses

Our study employed multiple methods to assess the potential impact of publication bias on the estimated overall effect. One such approach involved the use of a contour-enhanced funnel plot. The analysis of this plot revealed asymmetry, with a lack of studies on the left side of the funnel in both statistically non-significant and significant areas.

The absence of non-significant results in the bottom left region suggests the possibility of publication bias. Another contributing

factor to the asymmetry is implied, given the scarcity of studies in the top-left section of the chart (Figure 4A). However, this asymmetry is less likely attributable to publication bias towards non-significant results, as the missing studies would fall into the area of high statistical significance.

Similar to the previously mentioned publication year effect, it is important to note that these results seem to support the notion that there might be a tendency to publish positive results over negative ones.

To complement the funnel plot analysis, a trim and fill test was employed to identify the location of missing studies on the chart and determine the quantity needed for symmetry. As expected, the analysis indicated the addition of six studies to the left side of the chart to achieve funnel plot symmetry (see Figure 4B). However, most of the filled studies were situated in the region of higher

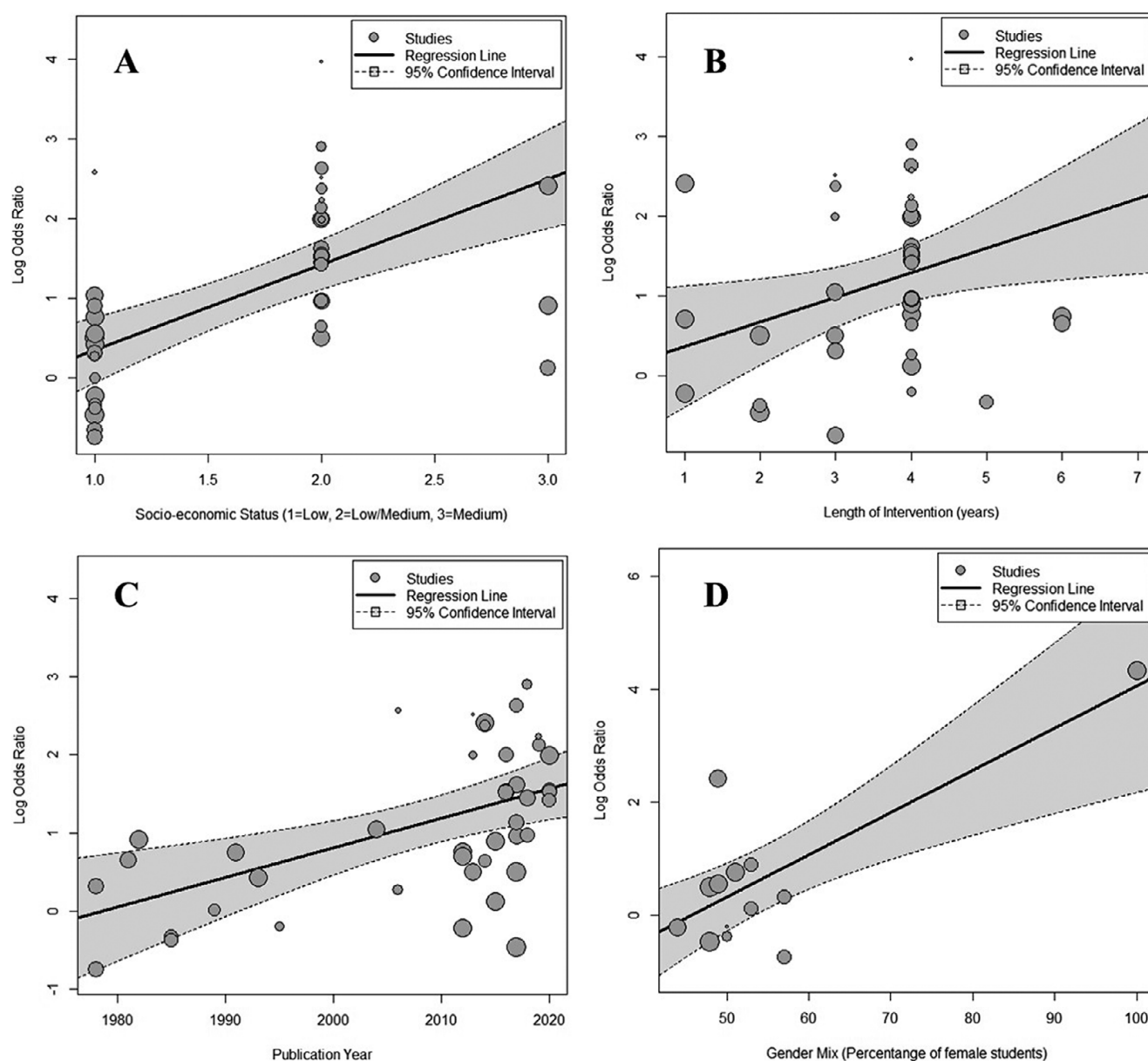


Figure 3. Moderating variables bubble plots, (A) SES, (B) Length of intervention, (C) Publication year, and (D) Gender mix. Each bubble represents a study, with its size proportional to its weight in the model, indicating the sample size. The regression line depicts the predicted Log OR as a function of each moderator, accompanied by 95% confidence intervals represented by the surrounding grey area.

Table 3. Mixed-effect model: All non-correlated variables

	<i>b</i>	SE	<i>t</i>	<i>p</i>	95% CI	
Length of intervention	0.2783	0.1292	2.1541	0.0397*	0.0141	0.5425
Type of intervention	0.4764	0.5115	0.9313	0.3594	−0.5698	1.5225
Control group language/s	−0.4091	0.4292	−0.9533	0.3483	−1.2868	0.4686
Socio-economic status	0.7576	0.2893	2.6184	0.0139*	0.1659	1.3494
Age	−0.0529	0.1052	−0.5029	0.6188	−0.2680	0.1622
Publication type	−0.2381	0.3632	−0.6554	0.5174	−0.9809	0.5048
Outcome test type	−0.0112	0.3718	−0.0302	0.9762	−0.7717	0.7493
Year of publication	0.0444	0.0214	2.0718	0.0473*	0.0006	0.0882

Moderating variables analyses. R^2 refers to the amount of heterogeneity accounted for per moderator, and *b* is the variation relative to the combined effect.

*Statistically significant result ($p < .05$).

statistical significance, suggesting that the asymmetry may not only be attributed to publication bias (Peters et al., 2008). Alternative factors, such as location biases (especially language bias, as only studies in English were included in this analysis) or true heterogeneity (possibly due to differences in the intensity of the interventions), are plausible (Borenstein et al., 2009; Rothstein et al., 2005).

Lastly, we used a regression test designed to assess funnel plot asymmetry, commonly known as Egger's test. Specifically, the random/mixed effects version, as recommended by Rothstein et al. (2005) and Viechtbauer (2010), was employed due to the model used for estimating the overall effect. This test aimed to quantify the potential association between effect sizes and standard error, which often serves as an indicator of funnel plot asymmetry. The obtained results were consistent with the trim-and-fill method, revealing a regression intercept with a *p*-value of 0.09 ($t = 1.76$). This finding suggests limited evidence supporting the presence of bias towards studies with statistically non-significant results. Hence, confirmation of publication bias could not be established.

3.5. Sensitivity analysis

Sensitivity analyses were performed in two stages: firstly, by including studies to achieve symmetry based on the results of the fill and trim, and secondly, by excluding effect size outliers. A summary of the sensitivity analyses is provided in Table S4 of Appendix S1.

As previously mentioned, the fill and trim test indicated the potential omission of six studies from the left side of the funnel plot. The overall effect decreased upon their inclusion, yet it remained positive and statistically significant ($k = 50$, $\log OR = 0.86$, $p < 0.0001$).

In the second sensitivity analysis, the *dmeter* package was used to remove outliers, defined as studies whose 95% confidence interval extended beyond the 95% confidence interval of the combined effect (Harrer et al., 2019). This process resulted in the exclusion of 12 studies, considerably lowering the between-study heterogeneity ($I^2 = 61.35\%$). Despite this, the pooled effect continued to be moderately positive and statistically significant ($k = 32$, $\log OR = 1.10$, $p < 0.0001$).

It is important to emphasise that this outcome highlights the robustness of the association between language learning and mathematical skills in young adolescents, despite the data suggesting the presence of publication bias.

4. Discussion

The objective of this study was to synthesise existing research on the impact of formal second language learning during early

adolescence on mathematical skills. The 25 included studies in the systematic review exhibited considerable methodological diversity, especially evident in areas such as sample selection, number of participants, assessment tools used and the statistical methods employed.

This diversity resulted in a highly heterogeneous mix of effect sizes and confidence intervals. When meta-analysed, these yielded a medium-sized effect, according to the field-specific benchmark established by Plonsky et al. (2021). Specifically, students who underwent formal language learning, either participating in a BE program or a second/foreign language program, were three times more likely to achieve higher grades on a mathematical test compared to those without such exposure.

It is well known that meta-analytical results are contingent upon the reliability of the synthesised data. If the area of study has been subjected to any type of selection bias, such as publication bias, the results may be affected (Borenstein et al., 2009). In this study, both the moderating variables (specifically, year of publication) and funnel plot analyses suggest a tendency to favour the publication of positive results over negative ones and significant findings over non-significant ones. However, it is crucial to note that the results of the trim and fill analysis as well as Egger's test indicate publication bias could not be confirmed.

Given the pronounced heterogeneity observed in the model, an in-depth exploration of potential factors influencing the overall result was conducted. Analyses addressing study clusters, filling in potentially missing studies (as a result of the trim and fill analysis) and excluding outliers consistently yielded positive and statistically significant effect sizes. These results revealed that students learning a second language were at least 1.84 times more likely to outperform their counterparts in control groups when assessed in a mathematical skills test. Although we recognise it is possible that future meta-analyses in this area may yield a different result, this data suggests that most studies analysing the impact of language learning on mathematical skills should be able to demonstrate a correlation of comparable magnitude to ours.

As numerous authors have recently stated, language learning and bilingualism are dynamic phenomena, existing along a continuum influenced by a myriad of factors (Beatty-Martínez & Titone, 2021; de Bruin, 2019; Gullifer et al., 2018; Titone & Tiv, 2023). In line with the ongoing debate regarding the cognitive benefits attributed to learning a second language, it is understood that such effects may be contingent upon said factors. Within the constraints of the data collected, several covariates were analysed, revealing SES and intervention duration as the most influential and statistically significant, contributing to 46% and 10% of the

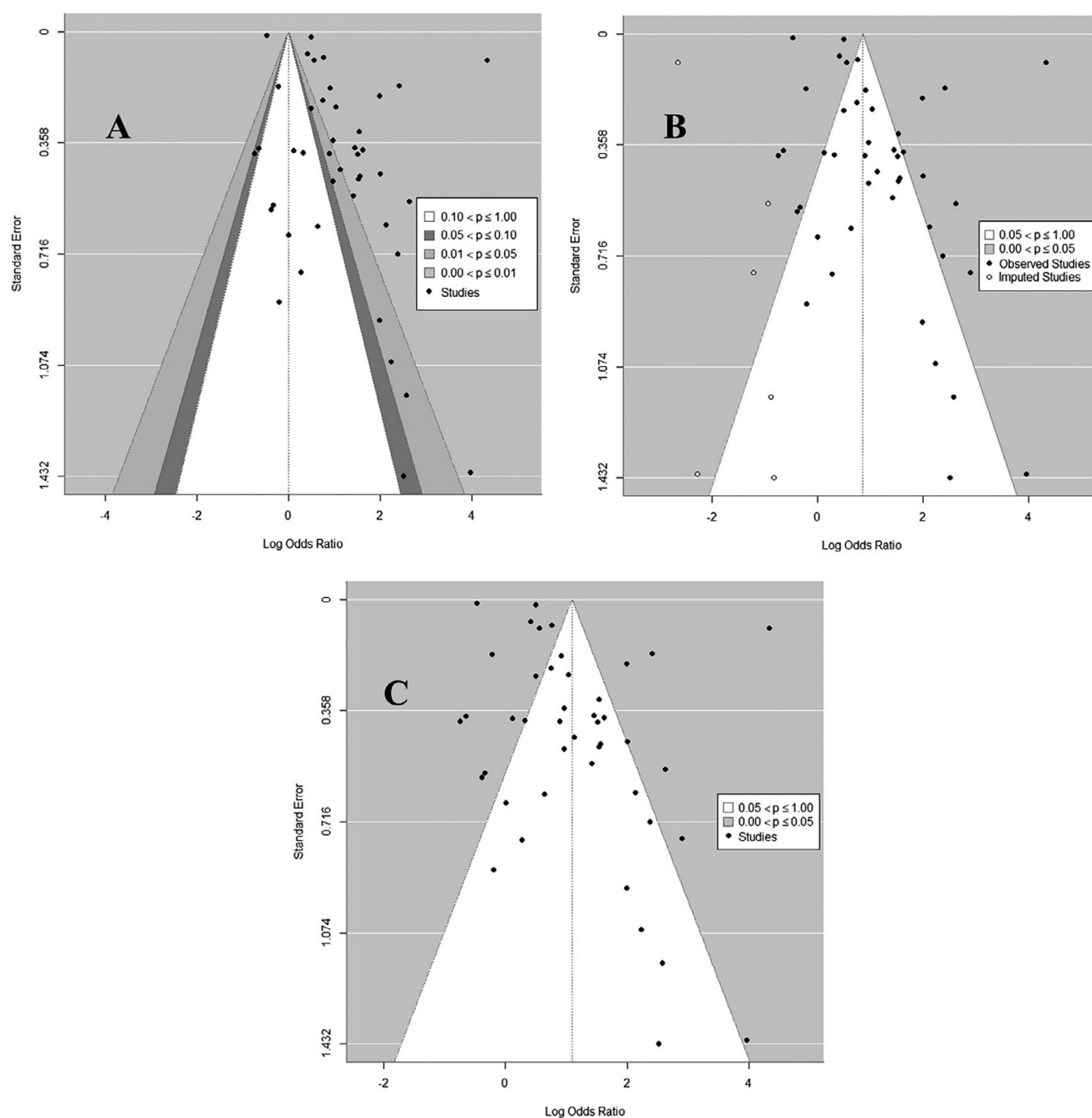


Figure 4. Contour-enhanced funnel plot (A), centred at 0 (null hypothesis value) to aid interpretation, and (B) Trim and fill analysis funnel plot. The white and shaded sections represent areas of statistical non-significance and significance, respectively. For comparison, a standard funnel plot (C) is included.

between-study heterogeneity, respectively. Specifically, higher SES levels and longer formal language training programs both correlated with higher test results.

The assertion that SES and the duration of the intervention affect student performance, and thus effect sizes, is substantiated by existing literature. Firstly, empirical evidence from international standardised tests reveals that, even among high-performing nations, SES significantly impacts test results. Specifically, lower socio-economic status, as measured by a composite of social, financial, cultural and human capital factors, correlates with diminished student performance (Mullis et al., 2020; OECD, 2019, 2023b). Moreover, research indicates that increased exposure and practice

in a second language enhance the likelihood of attaining cognitive benefits derived from linguistic experiences (Bialystok, 2015). Consequently, our observation that longer interventions yield higher effect sizes aligns logically with this established framework.

Other factors documented in the literature as influential in shaping the bilingual experience, such as the type of language intervention/exposure – particularly full language immersion (Farabolini et al., 2023; Figueiredo et al., 2016) – and age, which generally favours learners under six years of age (Planckaert et al., 2023), were also analysed. However, these factors did not exhibit a statistically significant influence on our model and resultant effect size.

Regarding the lack of differential effect based on the type of language program, we believe this result is attributable to the nature of the data. That is, the majority of the effect sizes were based on samples that included students with an immigrant background participating in various types of bilingual and second language programs aimed at effectively introducing them to mainstream programs. An interpretation of this finding might be that when the intervention is a second language program designed as immigration aid (i.e., when the dominant language is the one being learnt), it provides the same level of exposure, in terms of understanding how the language works and how it can be used as an everyday communication tool, as bilingual programs. This similarity in exposure likely results in similar reported effect sizes.

Interestingly, the remaining programs, whose populations were learning languages that were not necessarily widely spoken in their geographical area, reported effect sizes that were mostly within the confidence intervals of the overall effect. This may suggest that under specific circumstances, a foreign language learning setting might provide benefits similar to those of full immersion. This could be particularly true when foreign language programs incorporate cultural and practical language components, ensuring comprehensive engagement and communication skills designed for immediate real-world application. Additionally, as foreign language learning is often optional, the learners' high motivation may play an important role in their success.

The absence of a clear pattern between participants' age at the time of testing and effect sizes – meaning that neither older nor younger students performed significantly better on mathematical tests – is somewhat unexpected. This, however, could be explained by the presence of other factors that influence cognitive ability at specific ages, such as gender, or by factors that may be more influential but were not recorded in the study, such as home language (dominancy, proficiency, etc.), cultural values and attitudes towards language learning or heritage language. Another interpretation of this finding is that within the age range studied, the age of testing is not a determining variable in the association between language learning and mathematical ability.

This meta-analysis is the first, to our knowledge, to aggregate studies examining the association between language learning and mathematical skills during early adolescence. While our findings are promising, it is imperative to acknowledge that causation cannot be definitively inferred. Our study does not establish whether students excel in mathematics due to participating in language courses, or if those proficient in mathematics are more inclined to opt for language courses. Notably, more than half of the effect sizes in our dataset were derived from populations where students were learning a second language as a migration aid. These individuals engaged in language learning not solely out of intrinsic motivation but due to a change in circumstances, which highlights the potential cognitive benefits of second language acquisition.

Additionally, existing research offers some support for the benefits of language learning on mathematical ability. Certain authors have identified a positive correlation between language learning and mathematical skills, albeit in younger age groups (Iqbal, 2022; Lin, 2018; Stewart, 2005; Tobias, 2012). Furthermore, the literature extensively documents the positive impact of language learning on non-linguistic cognitive skills (Poarch & van Hell, 2012; White & Greenfield, 2017; Woumans *et al.*, 2018; Yurtsever *et al.*, 2023), with mathematical achievement accurately predicted by these skills (Bellon *et al.*, 2019; Cantin *et al.*, 2016; Cragg *et al.*, 2017).

This correlation is substantiated by international standardised tests data, as evidenced by consistently superior performance in the SAT mathematics section among students who engaged in language learning (CollegeBoard, 2019, 2020, 2021, 2022, 2023). Likewise, the ten highest-performing countries in mathematics in PISA 2022 are nations where more than 85% of students are at least learning one foreign language in their school settings (OECD, 2020, 2023b).

5. Limitations and future directions

Although this review adhered to rigorous guidelines and a robust searching strategy, it is important to acknowledge its limitations. Most included studies were conducted in North America, with only one exception in Asia, leaving other regions unrepresented. Greater geographical diversity would have been beneficial, given the significant global variation in attitudes towards foreign language learning (OECD, 2020, 2023a). This limitation can be partially attributed to the deliberate selection of English as the language criterion for the sourced studies. Future meta-analyses incorporating research in various language could enhance the geographical diversity and robustness of findings.

Furthermore, this review also identified substantial heterogeneity between studies, with mixed-effects modelling attributing some differences to variations in SES, the length of intervention and publication year. However, a significant residual heterogeneity persisted, suggesting the existence of unaccounted-for variables. Consequently, additional research is needed to systematically identify additional confounding factors, such as gender, thereby elucidating the nature of the association between language learning and mathematical skills.

Finally, it is important to note that due to the dynamic nature of cognitive skills and language learning, their interaction is inherently complex. Therefore, experimental research conducted under controlled settings is essential to establish causation. Rigorous methods such as matched-paired sample selection, homogenous interventions and time-lagged correlations, along with a thorough evaluation of the role of EF, may enhance the robustness and decisiveness of findings. Nevertheless, challenges remain in comparing bilingual populations, as their language learning experiences and cognitive realities may vary significantly across and within studies (Titone *et al.*, 2024). This variability may complicate the ability to draw definitive conclusions about this fundamentally complex area of study.

6. Conclusion

This study synthesised research on the impact of formal second language learning during early adolescence on mathematical skills. The systematic review of 25 studies revealed significant methodological diversity, leading to heterogeneous effect sizes. The subsequent meta-analysis showed a medium-sized effect, indicating that students in bilingual or second language programs performed three times better on math tests than those without such exposure. Methodological diversity among studies, particularly in sample selection and statistical methods, was observed. However, sensitivity analyses confirmed the overall effect remained robust, with language learners outperforming their monolingual counterparts at least 1.84 times. SES and intervention duration were identified as influential factors affecting the outcomes.

Nevertheless, limitations included a geographic bias and residual heterogeneity in the data. Future research should address

these limitations and investigate causation through experimental studies. Our findings emphasise the significance of language learning in educational policy and practice, suggesting a positive correlation between language learning and mathematical skills that warrants further exploration.

Supplementary material. To view supplementary material for this article, please visit <http://doi.org/10.1017/S1366728924000701>.

Data availability statement. The data that support the findings of this study [Appendix S8], along with the R code used [Appendix S7], are openly available in the Supplementary Materials.

Competing interest. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- Abu-rabia, S., & Kehat, S. (2004). The critical period for second language pronunciation: is there such a thing? *Educational Psychology*, 24(1), 77–97. <https://doi.org/10.1080/0144341032000146467>
- Adi-Japha, E., Berberich-Artzi, J., & Libnawi, A. (2010). Cognitive flexibility in drawings of bilingual children. *Child Development*, 81(5), 1356–1366. <http://www.jstor.org/stable/40800677>
- Alvarado, R. B. (1985). *A comparison of educational progress between Hispanics in non-bilingual and bilingual school programs*. University of Connecticut: United States -- Connecticut. <https://www.proquest.com/dissertations-theses/comparison-educational-progress-between-hispanics/docview/303385526/se-2>
- Anderton, R., Hine, G., & Joyce, C. (2017). Secondary school mathematics and science matters: predicting academic success for secondary students transitioning into university allied health and science courses. *International Journal of Innovation in Science and Mathematics Education (IJISME)*, 25(1). <https://openjournals.library.sydney.edu.au/CAL/article/view/11317/11058>
- Andreou, M., Tsimpli, I. M., Masoura, E., & Agathopoulou, E. (2020). Cognitive mechanisms of monolingual and bilingual children in monoliterate educational settings: evidence from sentence repetition. *Frontiers in Psychology*, 11, 613992. <https://doi.org/10.3389/fpsyg.2020.613992>
- Antón, E., Carreiras, M., & Duñabeitia, J. A. (2019). The impact of bilingualism on executive functions and working memory in young adults. *PLoS One*, 14(2), e0206770. <https://doi.org/10.1371/journal.pone.0206770>
- Araújo, L., Flisi, S., Dinis da Costa, P., & Soto-Calvo, E. (2015). *Languages and employability*. Publications Office. <https://doi.org/10.2788/504155>
- Ardasheva, Y., Tretter, T. R., & Kinny, M. (2012). English language learners and academic achievement: revisiting the threshold hypothesis. *Language Learning*, 62(3), 769–812. <https://doi.org/10.1111/j.1467-9922.2011.00652.x>
- Aromataris, E., & Munn, Z. (2020). *JBI Manual for Evidence Synthesis*. <https://doi.org/10.46658/JBIMES-20-01>
- Badru, A. K., & Owodunni, S. A. (2021). Influence of Mathematical language ability and Parental supports on students' academic achievement in Secondary School Sciences (Physics, Chemistry and Biology) in Ogun State, Nigeria. *Education and Self Development*, 16(1), 10–20. <https://doi.org/10.26907/esd16.1.03>
- Balduzzi, S., Rücker, G., & Schwarzer, G. (2019). How to perform a meta-analysis with [R]: a practical tutorial. *Evidence-Based Mental Health*, 22(4), 153–160. <https://doi.org/10.1136/ebmental-2019-300117>
- Ballard, C. L., & Johnson, M. F. (2004). Basic math skills and performance in an introductory economics class. *The Journal of Economic Education*, 35(1), 3–23. <https://doi.org/10.3200/JECE.35.1.3-23>
- Barac, R., & Bialystok, E. (2011). Cognitive development of bilingual children. *Language Teaching*, 44(1), 36–54. <https://doi.org/10.1017/S0261444810000339>
- Barac, R., & Bialystok, E. (2012). Bilingual effects on cognitive and linguistic development: role of language, cultural background, and education. *Child Development*, 83(2), 413–422. <http://www.jstor.org/stable/41416093>
- Barac, R., Bialystok, E., Castro, D. C., & Sanchez, M. (2014). The cognitive development of young dual language learners: a critical review. *Early Childhood Research Quarterly*, 29(4), 699–714. <https://doi.org/10.1016/j.ecresq.2014.02.003>
- Beatty-Martínez, A. L., & Titone, D. A. (2021). The quest for signals in noise: leveraging experiential variation to identify bilingual phenotypes. *Languages (Basel)*, 6(4). <https://doi.org/10.3390/languages6040168>
- Bellon, E., Fias, W., & De Smedt, B. (2019). More than number sense: The additional role of executive functions and metacognition in arithmetic. *Journal of Experimental Child Psychology*, 182, 38–60. <https://doi.org/10.1016/j.jecp.2019.01.012>
- Bialystok, E. (2015). Bilingualism and the development of executive function: the role of attention. *Child Development Perspectives*, 9(2), 117–121. <https://doi.org/10.1111/cdep.12116>
- Bialystok, E., & Craik, F. I. M. (2022). How does bilingualism modify cognitive function? Attention to the mechanism. *Psychonomic Bulletin & Review*, 29(4), 1246–1269. <https://doi.org/10.3758/s13423-022-02057-5>
- Bialystok, E., & Hakuta, K. (1999). Confounded age: linguistic and cognitive factors in age differences for second language acquisition. In D. Birdsong (Ed.), *Second Language Acquisition and the Critical Period Hypothesis*. Lawrence Erlbaum Associates, Publishers.
- Birdsong, D. (2006). Age and second language acquisition and processing: a selective overview. *Language Learning*, 56(s1), 9–49. <https://doi.org/10.1111/j.1467-9922.2006.00353.x>
- Blom, E., Küntay, A. C., Messer, M., Verhagen, J., & Leseman, P. (2014). The benefits of being bilingual: Working memory in bilingual Turkish–Dutch children. *Journal of Experimental Child Psychology*, 128, 105–109. <https://doi.org/10.1016/j.jecp.2014.06.007>
- Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley & Sons. <https://doi.org/10.1002/9780470743386>
- Breaux, K. C., & Lichtenberger, E. O. (2016). *Essentials of KTEA-3 and WIAT-III Assessment*. John Wiley & Sons.
- Brown, V. L., Cronin, M. E., & Bryant, D. P. (2012). *TOMA-3: Test of Mathematical Abilities (3rd ed.)*. PRO-ED.
- Bull, R., & Lee, K. (2014). Executive functioning and mathematics achievement. *Child Development Perspectives*, 8(1), 36–41. <https://doi.org/10.1111/cdep.12059>
- Bull, R., & Scerif, G. (2001). Executive functioning as a predictor of children's mathematics ability: inhibition, switching, and working memory. *Developmental Neuropsychology*, 19(3), 273–293. https://doi.org/10.1207/s15326942dn1903_3
- Cantin, R. H., Gnaedinger, E. K., Gallaway, K. C., Hesson-McInnis, M. S., & Hund, A. M. (2016). Executive functioning predicts reading, mathematics, and theory of mind during the elementary years. *Journal of Experimental Child Psychology*, 146, 66–78. <https://doi.org/10.1016/j.jecp.2016.01.014>
- Chan, K. S., & So, A. Y. (1982). *The Impact of Language of Instruction on the Educational Achievement of Hispanic Students*. National Center for Bilingual Research. <https://www.proquest.com/docview/63404369>
- Choi, J. Y., Jeon, S., & Lippard, C. (2018). Dual language learning, inhibitory control, and math achievement in Head Start and kindergarten. *Early Childhood Research Quarterly*, 42, 66–78. <https://doi.org/10.1016/j.ecresq.2017.09.001>
- St Clair-Thompson, H. L., & Gathercole, S. E. (2006). Executive functions and achievements in school: Shifting, updating, inhibition, and working memory. *Quarterly Journal of Experimental Psychology*, 59(4), 745–759. <https://doi.org/10.1080/17470210500162854>
- Cobb, B., Vega, D., & Kronauge, C. (2006). Effects of an elementary dual language immersion school program on junior high achievement. *Middle Grades Research Journal*, 1(1), 27–47. <https://www.proquest.com/docview/61928144>
- CollegeBoard. (2019). *SAT Suite of Assessments Annual Report (Total Group)*. <https://reports.collegeboard.org/media/pdf/2019-total-group-sat-suite-assessments-annual-report.pdf>
- CollegeBoard. (2020). *SAT Suite of Assessments Annual Report (Total Group)*. <https://reports.collegeboard.org/media/pdf/2020-total-group-sat-suite-assessments-annual-report.pdf>
- CollegeBoard. (2021). *SAT Suite of Assessments Annual Report (Total Group)*. <https://reports.collegeboard.org/media/2022-04/2021-total-group-sat-suite-of-assessments-annual-report%20%281%29.pdf>
- CollegeBoard. (2022). *SAT Suite of Assessments Annual Report (Total Group)*. <https://reports.collegeboard.org/media/pdf/2022-total-group-sat-suite-of-assessments-annual-report.pdf>

- CollegeBoard. (2023). *SAT Suite of Assessments Annual Report (Total Group)*. <https://reports.collegeboard.org/media/pdf/2023-total-group-sat-suite-of-assessments-annual-report%20ADA.pdf>
- Cooper, H. (2017). *Research Synthesis and Meta-Analysis: A Step-by-Step Approach* (5th ed.). SAGE Publications, Inc. <https://doi.org/10.4135/9781071878644>
- Cragg, L., & Gilmore, C. (2014). Skills underlying mathematics: The role of executive function in the development of mathematics proficiency. *Trends in Neuroscience and Education*, 3(2), 63–68. <https://doi.org/10.1016/j.tine.2013.12.001>
- Cragg, L., Keeble, S., Richardson, S., Roome, H. E., & Gilmore, C. (2017). Direct and indirect influences of executive functions on mathematics achievement. *Cognition*, 162, 12–26. <https://doi.org/10.1016/j.cognition.2017.01.014>
- Darcy, I., Park, H., & Yang, C.-L. (2015). Individual differences in L2 acquisition of English phonology: The relation between cognitive abilities and phonological processing. *Learning and Individual Differences*, 40, 63–72. <https://doi.org/10.1016/j.lindif.2015.04.005>
- Darcy, N. (1953). A review of the literature on the effects of bilingualism upon the measurement of intelligence. *The Pedagogical Seminary and Journal of Genetic Psychology*, 82(1), 21–57. <https://doi.org/10.1080/08856559.1953.10533654>
- Davies, E. (2018). *Bilingual Benefits. The Australian Education Reporter*. <http://www.edureporter.com.au/bilingual-benefits/>
- Davis, E. (1978). A model for understanding understanding in mathematics. *The Arithmetic Teacher*, 26(1), 13–17. <https://doi.org/10.5951/AT.26.1.0013>
- de Bruin, A. (2019). Not all bilinguals are the same: a call for more detailed assessments and descriptions of bilingual experiences. *Behavioral Sciences (Basel)*, 9(3). <https://doi.org/10.3390/bs9030033>
- de Jong, E. J. (2004). After exit: Academic achievement patterns of former English language learners. *Education Policy Analysis Archives*, 12, 50. <https://doi.org/10.14507/epaa.v12n50.2004>
- Delaney, J. M., & Devereux, P. J. (2020). Math matters! The importance of mathematical and verbal skills for degree performance. *Economics Letters*, 186, 108850. <https://doi.org/10.1016/j.econlet.2019.108850>
- DeLuca, V., Rothman, J., Bialystok, E., & Platsikas, C. (2019). Redefining bilingualism as a spectrum of experiences that differentially affects brain structure and function. *Proceedings of the National Academy of Sciences*, 116(15), 7565–7574. <https://doi.org/10.1073/pnas.1811513116>
- DeLuca, V., Rothman, J., Bialystok, E., & Platsikas, C. (2020). Duration and extent of bilingual experience modulate neurocognitive outcomes. *NeuroImage*, 204, 116222. <https://doi.org/10.1016/j.neuroimage.2019.116222>
- Di Pisa, G., Pereira Soares, S. M., & Rothman, J. (2021). Brain, mind and linguistic processing insights into the dynamic nature of bilingualism and its outcome effects. *Journal of Neurolinguistics*, 58. <https://doi.org/10.1016/j.jneuroling.2020.100965>
- Dick, A., Garcia, N., Pruden, S., Thompson, W., Hawes, S., Sutherland, M., Riedel, M., Laird, A., & Gonzalez, R. (2019). No evidence for a bilingual executive function advantage in the nationally representative ABCD study. *Nature Human Behaviour*. <https://doi.org/10.1038/s41562-019-0609-3>
- Dorn, L. D., Dahl, R. E., Woodward, H. R., & Biro, F. (2006). Defining the boundaries of early adolescence: a user's guide to assessing pubertal status and pubertal timing in research with adolescents. *Applied Developmental Science*, 10(1), 30–56. https://doi.org/10.1207/s1532480xads1001_3
- Falter, P., & Persinger, M. A. (1991). Improved arithmetic skills of french-immersion grade-school students are confounded by higher intelligence. *Perceptual and Motor Skills*, 72(3), 772–774. <https://doi.org/10.2466/pms.1991.72.3.772>
- Farabolini, G., Taboh, A. R., Ceravolo, M. G., & Guerra, F. (2023). The association between language exposure and nonword repetition performance in bilingual children: A systematic review and meta-analysis. *Bilingualism: Language and Cognition*, 26(3), 621–635. <https://doi.org/10.1017/S1366728922000906>
- Figueiredo, S., Alves Martins, M., & Silva, C. F. d. (2016). Second language education context and home language effect: language dissimilarities and variation in immigrant students' outcomes. *International Journal of Multilingualism*, 13(2), 184–212. <https://doi.org/10.1080/14790718.2015.1079204>
- Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language*, 41(1), 78–104. <https://doi.org/10.1006/jmla.1999.2638>
- Flores, D. J. (1981). *An investigation of the long-term effects of bilingual education upon achievement, language maintenance and attitudes*. University of Florida: United States -- Florida. <https://www.proquest.com/dissertations-theses/investigation-long-term-effects-bilingual/docview/303119231/se-2>
- Freeman, B., Marginson, S., & Tytler, R. (2019). *An international view of STEM education STEM Education 2.0* (pp. 350–363). Brill. https://doi.org/10.1163/9789004405400_019
- Fuhrmann, D., Knoll, L. J., & Blakemore, S.-J. (2015). Adolescence as a sensitive period of brain development. *Trends in Cognitive Sciences*, 19(10), 558–566. <https://doi.org/10.1016/j.tics.2015.07.008>
- Fuhs, M. W., Nesbitt, K. T., Farran, D. C., & Dong, N. (2014). Longitudinal associations between executive functioning and academic skills across content areas. *Developmental Psychology*, 50(6), 1698–1709. <https://doi.org/10.1037/a0036633>
- Gathercole, S. E., Pickering, S. J., Knight, C., & Stegmann, Z. (2004). Working memory skills and educational attainment: evidence from national curriculum assessments at 7 and 14 years of age. *Applied Cognitive Psychology*, 18(1), 1–16. <https://doi.org/10.1002/acp.934>
- Geary, D. C. (2004). Mathematics and learning disabilities. *Journal of Learning Disabilities*, 37(1), 4–15. <https://doi.org/10.1177/00222194040370010201>
- Grundy, J. G., & Timmer, K. (2016). Bilingualism and working memory capacity: A comprehensive meta-analysis. *Second Language Research*, 33(3), 325–340. <https://doi.org/10.1177/0267658316678286>
- Gullifer, J. W., Chai, X. J., Whitford, V., Pivneva, I., Baum, S., Klein, D., & Titone, D. (2018). Bilingual experience and resting-state brain connectivity: Impacts of L2 age of acquisition and social diversity of language use on control networks. *Neuropsychologia*, 117, 123–134. <https://doi.org/10.1016/j.neuropsychologia.2018.04.037>
- Haddaway, N. R., Page, M. J., Pritchard, C. C., & McGuinness, L. A. (2022). PRISMA2020: An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and Open Synthesis. *Campbell Systematic Reviews*, 18(2), e1230. <https://doi.org/10.1002/cl2.1230>
- Hakuta, K. (2001). Chapter 10. A critical period for second language acquisition? In D. B. Bailey Jr., J. T. Bruer, F. J. Symons & J. W. Lichtman (Eds.), *Critical thinking about critical periods* (pp. xiv+299). Paul H Brookes Publishing.
- Hakuta, K., Bialystok, E., & Wiley, E. (2003). Critical evidence: a test of the critical-period hypothesis for second-language acquisition. *Psychological Science*, 14(1), 31–38. <https://doi.org/10.1111/1467-9280.01415>
- Harrer, M., Cuijpers, P., Furukawa, T., & Ebert, D. D. (2019). *dmeter: Companion R Package For The Guide 'Doing Meta-Analysis in R'*. <http://dmeter.protectlab.org/>
- Hartanto, A., Yang, H., & Yang, S. (2018). Bilingualism positively predicts mathematical competence: Evidence from two large-scale studies. *Learning and Individual Differences*, 61, 216–227. <https://doi.org/10.1016/j.lindif.2017.12.007>
- Hartshorne, J. K., Tenenbaum, J. B., & Pinker, S. (2018). A critical period for second language acquisition: Evidence from 2/3 million English speakers. *Cognition*, 177, 263–277. <https://doi.org/10.1016/j.cognition.2018.04.007>
- Higgins, J., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M., & Welch, V. (2023). *Cochrane Handbook for Systematic Reviews of Interventions version (Vol. 6) (updated August 2023)*. Cochrane. www.training.cochrane.org/handbook
- HISD. (2013). *Dual Language Program Evaluation: Developmental Bilingual & Two-Way Bilingual Immersion, 2012-2013. Research Educational Program Report*. Houston Independent School District. <https://www.proquest.com/reports/dual-language-program-evaluation-developmental/docview/2461142070/se-2>
- HISD. (2014). *Dual Language Program Evaluation, 2013-2014. Research Educational Program Report*. Houston Independent School District. <https://www.proquest.com/reports/dual-language-program-evaluation-2013-2014/docview/2461137802/se-2>
- HISD. (2016). *Dual Language Program Evaluation, 2015-2016. Research Educational Program Report*. Houston Independent School District. <https://www.proquest.com/reports/dual-language-program-evaluation-2015-2016/docview/2461142000/se-2>
- HISD. (2017). *Dual Language Program Evaluation, 2016-2017. Research Educational Program Report*. Houston Independent School District.

- <https://www.proquest.com/reports/dual-language-program-evaluation-2016-2017/docview/2461142192/se-2>
- HISD. (2018). *Dual Language Program Evaluation Report, 2017-2018. Research Educational Program Report*. Houston Independent School District. <https://www.proquest.com/reports/dual-language-program-evaluation-report-2017-2018/docview/2461134894/se-2>
- HISD. (2019). *Dual Language Program Evaluation Report, 2018-2019. Research Educational Program Report*. Houston Independent School District. <https://www.proquest.com/reports/dual-language-program-evaluation-report-2018-2019/docview/2459002613/se-2>
- HISD. (2020). *Dual Language Bilingual Program Evaluation Report, 2019-2020. Research Educational Program Report*. Houston Independent School District (HISD). <https://www.proquest.com/reports/dual-language-bilingual-program-evaluation-report/docview/2535437113/se-2>
- Hresko, W., Schlieve, P., Herron, S., Swain, C., & Sherbenau, R. (2003). Comprehensive mathematical abilities test (CMAT). *Assessment for Effective Intervention*. <https://doi.org/10.1177/0737247706031004>
- Int'Hout, J., Ioannidis, J. P. A., & Borm, G. F. (2014). The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. *BMC Medical Research Methodology*, 14(1), 25. <https://doi.org/10.1186/1471-2288-14-25>
- Iqbal, A. (2022). *Bilingualism and mathematical ability in typically developing primary school children: A systematic review*. M.Sc. Education Oxford University: United Kingdom. <https://ora.ox.ac.uk/objects/uuid:d780d7be-265c-4a55-8133-dbf213d10942/files/dbn9997307>
- Joensen, J. S., & Nielsen, H. S. (2009). Is there a causal effect of high school math on labor market outcomes? *Journal of Human Resources*, 44(1), 171–198. <https://doi.org/10.3368/jhr.44.1.171>
- Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, 21(1), 60–99. [https://doi.org/10.1016/0010-0285\(89\)90003-0](https://doi.org/10.1016/0010-0285(89)90003-0)
- Johnson, M., & Kuennen, E. (2006). Basic math skills and performance in an introductory statistics course. *Journal of Statistics Education*, 14(2). <https://doi.org/10.1080/10691898.2006.11910581>
- Kanwal, W., Qamar, A. M., Nadeem, H. A., Khan, S. A., & Siddique, M. (2022). Effect of conceptual understanding of mathematical principles on academic achievement of secondary level chemistry students. *Multicultural Education*, 8(3), 242–254. <https://doi.org/10.5281/zenodo.6370449>
- Karaoglan Yilmaz, F. G., Koseoglu, B., Ayvali, C., & Ozturk, T. (2020). The effect of using digital stories in teaching English as a second language: Digital stories in teaching English. In S. M. Yilan & K. Koruyan (Eds.), *ICT-Based Assessment, Methods, and Programs in Tertiary Education* (pp. 191–208). IGI Global. <https://doi.org/10.4018/978-1-7998-3062-7.ch010>
- Karsenty, R. (2014). Mathematical ability. In S. Lerman (Ed.), *Encyclopedia of Mathematics Education* (pp. 372–375). Springer Netherlands. https://doi.org/10.1007/978-94-007-4978-8_94
- Lamia, P. (2015). *The influence of foreign language instruction on student achievement in eighth grade English language arts and math*. <https://www.proquest.com/dissertations-theses/influence-foreign-language-instruction-on-student/docview/1684415809/se-2>
- Lane, S. et al. (1995). *An Examination of the Performance Gains of Culturally and Linguistically Diverse Students on a Mathematics Performance Assessment within the QUASAR Project*. <https://www.proquest.com/reports/examination-performance-gains-culturally/docview/62642389/se-2>
- Langan, D., Higgins, J. P. T., Jackson, D., Bowden, J., Veroniki, A. A., Kontopantelis, E., Viechtbauer, W., & Simmonds, M. (2019). A comparison of heterogeneity variance estimators in simulated random-effects meta-analyses. *Research Synthesis Methods*, 10(1), 83–98. <https://doi.org/10.1002/jrsm.1316>
- Lee, H. (2010). *The relationship between Korean EFL high school students' English achievement and other subject achievement*. Alliant International University Shirley-M. Hufstetler School of Education. <https://www.proquest.com/dissertations-theses/relationship-between-korean-efl-high-school/docview/863582967/se-2>
- Li, P., Legault, J., & Litcofsky, K. A. (2014). Neuroplasticity as a function of second language learning: Anatomical changes in the human brain. *Cortex*, 58, 301–324. <https://doi.org/10.1016/j.cortex.2014.05.001>
- Lin, G. C. (2018). *The Roles of Language and Executive Function on Early Mathematics Among Emergent Bilinguals* [Doctor of Philosophy]. University of California: Irvine, CA, USA. <https://www.proquest.com/dissertations-theses/roles-language-executive-function-on-early/docview/2099611668/se-2>
- Lipsey, M. W., & Wilson, D. B. (2000). *Practical Meta-Analysis* (1st ed., Vol. 49). SAGE Publications, Inc. <https://www.campbellcollaboration.org/escalc/html/EffectSizeCalculator-Home.php>
- Luna, B., Marek, S., Larsen, B., Tervo-Clemmens, B., & Chahal, R. (2015). An integrative model of the maturation of cognitive control. *Annual Review of Neuroscience*, 38, 151–170. <https://doi.org/10.1146/annurev-neuro-071714-034054>
- Master, S. L., Eckstein, M. K., Gotlieb, N., Dahl, R., Wilbrecht, L., & Collins, A. G. E. (2020). Distangling the systems contributing to changes in learning during adolescence. *Developmental Cognitive Neuroscience*, 41, 100732. <https://doi.org/10.1016/j.dcn.2019.100732>
- Mazzocco, M. M. M., & Kover, S. T. (2007). A longitudinal assessment of executive function skills and their association with math performance. *Child Neuropsychology*, 13(1), 18–45. <https://doi.org/10.1080/09297040600611346>
- Medina Jr., M. (1993). Spanish achievement in a maintenance bilingual education program: Language proficiency, grade and gender comparisons. *Bilingual Research Journal*, 17(1-2), 57. <https://doi.org/10.1080/15235882.1993.10162648>
- Medina, M., Saldate, M., & Mishra, S. (1985). The sustaining effects of bilingual instruction: A follow-up study. *Journal of Instructional Psychology*, 12(3), 132.
- Mercader, J., Miranda, A., Presentación, M. J., Siegenthaler, R., & Rosel, J. F. (2018). Contributions of motivation, early numeracy skills, and executive functioning to mathematical performance. A longitudinal study. *Frontiers in Psychology*, 8, 2375. <https://doi.org/10.3389/fpsyg.2017.02375>
- Midgley, W. (2017, 24 Mar). Which languages should Australian children be learning to get ahead? *ABC News*. <https://www.abc.net.au/news/2017-03-24/which-languages-should-australian-children-be-learning/8383146>
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: a latent variable analysis. *Cognitive Psychology*, 41(1), 49–100. <https://doi.org/10.1006/cogp.1999.0734>
- Monnier, C., Boiché, J., Armandon, P., Baudoin, S., & Bellocchi, S. (2022). Is bilingualism associated with better working memory capacity? A meta-analysis. *International Journal of Bilingual Education and Bilingualism*, 25(6), 2229–2255. <https://doi.org/10.1080/13670050.2021.1908220>
- Morales, J., Calvo, A., & Bialystok, E. (2013). Working memory development in monolingual and bilingual children. *Journal of Experimental Child Psychology*, 114(2), 182–202. <https://doi.org/10.1016/j.jecp.2012.09.002>
- Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2020). *TIMSS 2019 International Results in Mathematics and Science*. <https://timssandpirls.bc.edu/timss2019/international-results/>
- Nair, V. K., Biedermann, B., & Nickels, L. (2017). Effect of socio-economic status on cognitive control in non-literate bilingual speakers. *Bilingualism: Language and Cognition*, 20(5), 999–1009. <https://doi.org/10.1017/S1366728916000778>
- Nguyen, T., Duncan, R. J., & Bailey, D. H. (2019). Theoretical and methodological implications of associations between executive function and mathematics in early childhood. *Contemporary Educational Psychology*, 58, 276–287. <https://doi.org/10.1016/j.cedpsych.2019.04.002>
- Norton, E. C., Dowd, B. E., & Maciejewski, M. L. (2018). Odds ratios—Current best practice and use. *JAMA*, 320(1), 84–85. <https://doi.org/10.1001/jama.2018.6971>
- OECD. (2019). *PISA 2018 Results (Volume II)*. <https://doi.org/10.1787/b5fd1b8f-en>
- OECD. (2020). *How language learning opens doors*. <https://www.oecd.org/pisa/foreign-language/opens-doors.pdf>
- OECD. (2023a). *Education GPS: Learning time & disciplinary climate*. Retrieved 08/11/2023, 2023, from <https://gpseducation.oecd.org/revieweducationpolicies/#!node=41718&filter=all>
- OECD. (2023b). *PISA 2022 Results. The State of Learning and Equity in Education (Volume I)*. https://read.oecd-ilibrary.org/view/?ref=1235_1235421-gumq51fbgo&title=PISA-2022-Results-Volume-I
- Olivoto T, L. A. (2020). metan: An R package for multi-environment trial analysis. *Methods in Ecology and Evolution*, 11(6), 783–789. <https://doi.org/10.1111/2041-210X.13384>

- Olsen, S., & Brown, L. (1992). The relation between high school study of foreign languages and ACT English and mathematics performance. *ADFL Bulletin*, 23(3), 47–50.
- Olson, S. L. (1989). *Long-Term Effects of Bilingual Education on a National Sample of Mexican-American Sophomores: A Component Analysis*. <https://www.proquest.com/reports/long-term-effects-bilingual-education-on-national/docview/63134848/se-2>
- Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan—a web and mobile app for systematic reviews. *Systematic Reviews*, 5(1), 210. <https://doi.org/10.1186/s13643-016-0384-4>
- Paap, K. R., Johnson, H. A., & Sawi, O. (2015). Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. *Cortex*, 69, 265–278. <https://doi.org/10.1016/j.cortex.2015.04.014>
- Paap, K. R., & Sawi, O. (2014). Bilingual advantages in executive functioning: problems in convergent validity, discriminant validity, and the identification of the theoretical constructs. *Frontiers in Psychology*, 5, 962. <https://doi.org/10.3389/fpsyg.2014.00962>
- Padilla, A. M., Chen, X., Swanson, E., Peterson, M., & Peruzzaro, A. (2022). Longitudinal study of Spanish Dual Language Immersion graduates: Secondary school academic and language achievement. *Foreign Language Annals*, 55(2), 408–434. <https://doi.org/10.1111/flan.12615>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, 372(71). <https://doi.org/10.1136/bmj.n71>
- Peng, P., Namkung, J., Barnes, M., & Sun, C. (2015). A meta-analysis of mathematics and working memory: moderating effects of working memory domain, type of mathematics skill, and sample characteristics. *Journal of Educational Psychology*, 108, 455. <https://doi.org/10.1037/edu0000079>
- Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2008). Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. *Journal of Clinical Epidemiology*, 61(10), 991–996. <https://doi.org/10.1016/j.jclinepi.2007.11.010>
- Planckaert, N., Duyck, W., & Woumans, E. (2023). Is there a cognitive advantage in inhibition and switching for bilingual children? A systematic review. *Frontiers in Psychology*, 14. <https://doi.org/10.3389/fpsyg.2023.1191816>
- Pliatsikas, C., DeLuca, V., & Voits, T. (2020). The many shades of bilingualism: language experiences modulate adaptations in brain structure. *Language Learning*, 70(S2), 133–149. <https://doi.org/10.1111/lang.12386>
- Plonsky, L., Sudina, E., & Hu, Y. (2021). Applying meta-analysis to research on bilingualism: An introduction. *Bilingualism: Language and Cognition*, 24(5), 819–824. <https://doi.org/10.1017/S1366728920000760>
- Poarch, G. J., & van Hell, J. G. (2012). Executive functions and inhibitory control in multilingual children: evidence from second-language learners, bilinguals, and trilinguals. *Journal of Experimental Child Psychology*, 113(4), 535–551. <https://doi.org/10.1016/j.jecp.2012.06.013>
- Powers, S. (1978). *The influence of bilingual instruction on academic achievement and self-esteem of selected Mexican-American junior high school students*. The University of Arizona: United States -- Arizona. <https://www.proquest.com/dissertations-theses/influence-bilingual-instruction-on-academic/docview/287995743/se-2>
- Renaissance Learning. (2024). *Star Math: Evidence-based K-12 Math Assessments*. Retrieved 05/06, 2024, from <https://www.renaissance.com/products/star-math/>
- Rothstein, H., Sutton, A. J., & Borenstein, M. (2005). *Publication bias in meta-analysis prevention, assessment and adjustments*. Wiley. <https://doi.org/10.1002/0470870168>
- Saer, D. J. (1923). The effect of bilingualism on intelligence. *British Journal of Psychology: General Section*, 14, 25–38. https://pure.mpg.de/rest/items/item_2376853/component/file_2376852/content
- Scovel, T. (1988). *A time to speak: A psycholinguistic inquiry into the critical period for human speech*. Harper Collins Publishers Ltd
- Selemon, L. D. (2013). A role for synaptic plasticity in the adolescent development of executive function. *Translational Psychiatry*, 3(3), e238. <https://doi.org/10.1038/tp.2013.7>
- Simpson-Kent, I. L., Fuhrmann, D., Bathelt, J., Achterberg, J., Borgeest, G. S., & Kievit, R. A. (2020). Neurocognitive reorganization between crystallized intelligence, fluid intelligence and white matter microstructure in two age-heterogeneous developmental cohorts. *Developmental Cognitive Neuroscience*, 41, 100743. <https://doi.org/10.1016/j.dcn.2019.100743>
- Slama, R., Molefe, A., Gerdeman, D., Herrera, A., Brodziak de los Reyes, I., August, D., & Cavazos, L. (2017). *Time to proficiency for hispanic English learner students in Texas*. REL 2018-280. Regional Educational Laboratory Southwest. <https://www.proquest.com/reports/time-proficiency-hispanic-english-learner/docview/2009554575/se-2>
- Smith, J. E. (2016). *The role of dual language programs in the development of ethnic identity of bilingual students – A quantitative study*. Northcentral University: United States -- California. <https://www.proquest.com/dissertations-theses/role-dual-language-programs-development-ethnic/docview/1858814811/se-2>
- Sorge, G. B., Toplak, M. E., & Bialystok, E. (2017). Interactions between levels of attention ability and levels of bilingualism in children's executive functioning. *Developmental Science*, 20(1), e12408. <https://doi.org/10.1111/desc.12408>
- Squeglia, L. M., Jacobus, J., Sorg, S. F., Jernigan, T. L., & Tapert, S. F. (2013). Early adolescent cortical thinning is related to better neuropsychological performance. *Journal of the International Neuropsychological Society*, 19(9), 962–970. <https://doi.org/10.1017/S1355617713000878>
- Steinberg, L. (2014). *Age of opportunity: Lessons from the new science of adolescence*. Houghton Mifflin Harcourt.
- Stewart, J. H. (2005). Foreign language study in elementary schools: Benefits and implications for achievement in reading and math. *Early Childhood Education Journal*, 33(1), 11–16. <https://doi.org/10.1007/s10643-005-0015-5>
- Tanriver-Ayder, E., Faes, C., van de Castele, T., McCann, S. K., & Macleod, M. R. (2021). Comparison of commonly used methods in random effects meta-analysis: application to preclinical data in drug discovery research. *BMJ Open Science*, 5(1), e100074. <https://doi.org/10.1136/bmjopen-2020-100074>
- TEA. (2024). *STAAR Mathematics Resources*. Retrieved 05/06, 2024, from <https://tea.texas.gov/student-assessment/testing/taaar/taaar-mathematics-resources>
- Titone, D., Hernández-Rivera, E., Iniesta, A., Beatty-Martínez, A. L., & Gullifer, J. W. (2024). Are language–cognition interactions bigger than a breadbox? Integrative modeling and design space thinking temper simplistic questions about causally dense phenomena. *Behavioral and Brain Sciences*, 47, e60. <https://doi.org/10.1017/S0140525X23002145>
- Titone, D. A., & Tiv, M. (2023). Rethinking multilingual experience through a systems framework of bilingualism. *Bilingualism: Language and Cognition*, 26(1), 1–16. <https://doi.org/10.1017/S1366728921001127>
- Tobias, K. S. (2012). *The Effect of Foreign Language Study in Tennessee Middle Schools on Mathematics Achievement*. <https://www.proquest.com/dissertations-theses/effect-foreign-language-study-tennessee-middle/docview/1433293545/se-2>
- Tytler, R. (2020). STEM education for the twenty-first century. In J. Anderson & Y. Li (Eds.), *Integrated approaches to STEM education: An international perspective* (pp. 21–43). Springer International Publishing. https://doi.org/10.1007/978-3-030-52229-2_3
- Vega, L. D. (2015). *Effects of an elementary two way bilingual Spanish-English immersion school program on junior high and high school student achievement*. <https://www.proquest.com/dissertations-theses/effects-elementary-two-way-bilingual-spanish/docview/1667939951/se-2>
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package (version 4.2-0). *Journal of Statistical Software*, 36(3). <https://doi.org/10.18637/jss.v036.i03>
- White, L. J., & Greenfield, D. B. (2017). Executive functioning in Spanish- and English-speaking Head Start preschoolers. *Developmental Science*, 20(1). <https://doi.org/10.1111/desc.12502>
- Wilson, D. B. (2023). *Practical meta-analysis effect size calculator (Version 2023.11.27)*. Campbell Collaboration. <https://www.campbellcollaboration.org/escalc/html/EffectSizeCalculator-Home.php>
- Woumans, E., Ameloot, S., Keuleers, E., & Van Assche, E. (2018). The relationship between second language acquisition and nonverbal cognitive abilities. *Journal of Experimental Psychology: General*. <https://doi.org/10.1037/xge0000536>

- Yeniad, N., Malda, M., Mesman, J., Van Ijzendoorn, M. H., & Pieper, S. (2013). Shifting ability predicts math and reading performance in children: A meta-analytical study. *Learning and Individual Differences*, **23**, 1–9. <https://doi.org/10.1016/j.lindif.2012.10.004>
- Yunker, J. A., Yunker, P. J., & Krull, G. (2009). The Influence of Mathematical Ability on Performance in Principles of Accounting. *The Accounting Educators' Journal*, **19**. <https://www.aejournal.com/ojs/index.php/aej/article/view/101>
- Yurtsever, A., Anderson, J. A. E., & Grundy, J. G. (2023). Bilingual children outperform monolingual children on executive function tasks far more often than chance: An updated quantitative analysis. *Developmental Review*, **69**, 101084. <https://doi.org/10.1016/j.dr.2023.101084>
- Zhan, Z., Shen, W., Xu, Z., Niu, S., & You, G. (2022). A bibliometric analysis of the global landscape on STEM education (2004–2021): towards global distribution, subject integration, and research trends. *Asia Pacific Journal of Innovation and Entrepreneurship*, **16**(2), 171–203. <https://doi.org/10.1108/APJIE-08-2022-0090>