Research article
UDC 372.851
https://doi.org/10.21702/rpj.2026.3.11

Instructional Strategies for Teaching Mathematics to Preschoolers: A Meta-Analysis of Empirical Evidence

Eugene Borokhovski[1*], Éric Dion[2], David Pickup[1], Isabelle Plante[2], Patric Charland[2], Stéphane Cyr[2]

[1]Centre for the Study of Learning and Performance (CSLP), Concordia University, Montreal, Canada

[2]Université du Québec à Montréal (QCAM), Montréal, Canada

[*]Corresponding author: eugene.borokhovski@concordia.ca

Abstract

Introduction. This meta-analysis aimed at summarising research evidence regarding the effectiveness of innovative instructional practices of teaching Mathematics to preschoolers.

Method. Analyses and aggregation of quantitative data from 145 primary empirical studies, identified and selected through comprehensive searches and systematic review process, allowed for producing weighted average effect sizes in the following categories of learning outcomes: (1) Mathematics Precursors, (2) Number Sense, (3) Geometry Knowledge and Skills, and (4) Mathematics Composite Scores.

Results. Derived by applying the random effects analytical model, these effect sizes were respectively: (1) $g^+ = 0.429$ ($k = 40$, $p < .01$); (2) $g^+ = 0.455$ ($k = 168$, $p < .01$); (3) $g^+ = 0.372$ ($k = 30$, $p < .01$); and (4) $g^+ = 0.374$ ($k = 40$, $p < .01$).

Discussion. Significant heterogeneity of findings within each outcome category warranted further exploration of the source(s) of systematic variation by means of moderator variable analyses, which identified as the most influential (in favor of the experimental conditions) factors: content-specific (as opposed to developmental) instructional practices, as well as explicit instruction rather than guided discovery learning; non-reliance on technological tools (lower effects of technology use in comparison with technology-free learning environments); and use of physical manipulatives for teaching Mathematics. These patterns of results, however, are not without caveats, which is discussed using specific examples from the studies included in the meta-analysis. In brief, innovative researcher-developed experimental instructional

intervention in early Mathematics demonstrated a strong potential to be advantageous in comparison with traditional ("business-as-usual") instruction. In the context of moderator variable analyses, different instructional approaches, tailored for specific learning goals and student populations in various contexts should be considered.

Keywords

meta-analysis, effect size, Mathematics, instructional practices, learning outcomes, preschoolers, research evidence

Funding

Quebec Ministry of Education, contract **T25526.**

For citation

Borokhovski, E., Dion, É., Pickup, D., Plante, I., Charland, P., & Cyr, S. (2026). Instructional strategies for teaching mathematics to preschoolers: A meta-analysis of empirical evidence. *Russian Psychological Journal*, 23(3), 227–258. https://doi.org/10.21702/rpj.2026.3.11

Introduction

Statement of Study Objectives

The major rationale for undertaking this systematic review was the timely need to collect, analyze, and aggregate research evidence regarding the most effective instructional practices in teaching Mathematics to preschoolers. Its outcomes are expected to inform the readership (including the research community, educators, and policy makers) about 'What works' effectively in early Mathematics education in order to equip the stakeholders (primarily, school teachers and administrators) with the evidence-based guidelines to implement in their respective classrooms the most promising instructional strategies and use the most effective learning materials.

To achieve these goals, we employed the methodology of quantitative synthesis, also known as meta-analysis (MA), guided by the following research questions:

How effective could the innovative instructional strategies (teacher/researched designed experimental condition) be in advancing preschoolers' knowledge and skills in Mathematics when compared to the current teaching practices ('business as usual' control conditions)?

What specific activities/qualities of the instructional interventions in question (i.e., substantive moderator variables) for what categories of learners (i.e., demographic moderator variables) and under what particular educational circumstances (i.e., contextual moderator variables) enhance or inhibit the effect of innovative teaching strategies in early Mathematics?

Summary of the Previous Research

Previous research in the field of Mathematics education has largely focused on elementary and secondary formal compulsory schooling. In contrast, the earliest stages of teaching Mathematics, i.e., at preschool levels (pre-kindergarten and kindergarten instruction) has not yet received sufficient attention.

Two relatively recent MAs have attempted to summarize relevant research that focused on this category of very young students (Nelson & McMaster, 2018; Wang et al., 2016). Both concluded that the introduction of well-designed and implemented researcher-developed activities led to substantial gains (i.e., significant effect sizes) in early numeracy skills. The effective researcher-developed activities typically relied on carefully sequenced and clear explanations of key concepts.

However, there are issues with these previous quantitative summaries of primary research on early Mathematics interventions that deserve additional attention. For example, MA by Nelson and McMaster (2018) encompasses studies of young learners with the list of moderator variables indiscriminately applied across grade levels instead of targeting study characteristics that are unique for preschoolers. Also, both named prior MAs are somewhat limited in terms of details about some important moderator variables, including psychometric quality of the instruments used to assess learning outcomes, their alignment with the content of the instruction and learning activities, verification of fidelity of implementation, parental involvement, experience of and training for teachers, among others.

Perhaps, the most consequential limitation of all in both MAs is the lack of consistency and clarity of dealing with the control conditions to which the experimental ones (innovative researcher-developed) are being compared. In far too many studies included in the previous MAs control conditions are simply referred to as ‘regular instruction’ / ‘business as usual’ etc., without providing sufficient description of teaching / learning activities implemented in them. In contrast, the current MA makes serious efforts to locate in the reviewed studies all available information about control conditions and categorize them. Without clear understanding of the type of control condition in any given study, it is impossible, for example to test the hypothesis of especially beneficial early interventions – whether, indeed, they are effective as a foundation for further learning of Mathematics or they are just being compared to a less informative (or obviously inferior) instruction in the corresponding control conditions.

Similarly, detailed knowledge of both experimental and control condition is needed to meaningfully inform one of the most influential and persistent discussions in education – for example, about relative effectiveness of direct instruction versus guided discovery learning – alongside with a multitude of other issues that are of applied value for educational practitioners, particularly with respect to teaching and learning Mathematics.

It is typically assumed that formal education begins in the first year of the elementary school, but recently this practice has been seriously questioned, at least across North American jurisdictions. Why should formal schooling wait until grade one and not to be

picked-up earlier – for preschoolers (kindergarteners), especially with respect to training in numeracy? One of the most obvious reasons for initiating Mathematics instruction earlier is that children from low-income families on average start the first year of formal schooling with less preparation than their peers from more advantageous backgrounds. There is a strong concern among researchers and policy makers alike that the initial disadvantage of children from low-income families may not be overcome easily but instead will worsen accumulating problems and challenges for this category of learning over the school years and gradually lead to demotivation and failure. From this standpoint, it is important to ensure that children from low-income families arrive in first grade with as good preparation as the rest of their peers. In this project we intend to explore, based on a systematic accumulation of the relevant research evidence if pre-school (kindergarten) classroom interventions that reach large numbers of children of diverse backgrounds is a viable way to accomplish this challenging task.

Whereas a consensus has been built around the idea of general relevance of interventions focused on preparation for more effective learning and, thus, prevention of academic failure, several specific points still need to be clarified. Empirical research has tested different forms of early interventions with varying effects on children's readiness for compulsory school Mathematics education. Since prevention takes time, effort and resources, it is important to ensure that a reasonable 'return on investment' is achieved.

It has been shown repeatedly that not only some innovative instructional practices can be particularly effective but also that they differ in their effects depending on the population of learners and category of learning outcomes. For example, Fuchs et al. (2001), while demonstrating the overall positive promise of peer-assisted learning strategies, reported a striking difference between general population of learners (typical achievers) and low-performing students (g = 0.44 and g = 0.38, respectively), on the one hand, and high-achieving students (g = –0.16), on the other, meaning that the latter category of kindergarteners does not benefit from or even may be susceptible to negative effect of employing this instructional strategy. According to Jordan et al. (2012) and later Dyson et al. (2013), small-group interventions in Number Sense (based on carefully selected vocabulary, use of manipulatives, number recognition and number sequencing games, partly adapted from *Math Expressions* curriculum—Fuson, 2009) were quite successful (g = 1.10 & g = 0.64) for underperforming kindergarten students from low-income families. Targeting the same learning outcomes of early Number Sense development in a similar (high poverty) sample could also be achieved (g = 1.38) through employing the intervention based on spatial, 'one or two more/less' and part-part-whole relations (Sood & Jitendra, 2011).

We have undertaken this MA to identify and describe classroom interventions that effectively promote preschool numeracy. In order to do so, we looked as objectively and systematically as possible at the results of all available primary empirical studies implemented with sufficient rigor under realistic conditions, so that their findings could be reliably generalized and used for developing meaningful recommendations for educational practice. Also, as stated earlier, it is imperative to account for possibly differential effects of the experimental interventions on various outcomes in a broad range of samples—as early

competency in Mathematics cannot be subsumed by just one generic construct. In fact, it is composed by a diverse set of skills and includes understanding (rather a developing one) of a number of concepts that would lay a foundation for acquiring more complex Mathematic constructs and mastering advanced operations in the upcoming compulsory schooling. It is important to note that this review does not address students with special needs (e.g., formally diagnosed with learning disabilities), those are quite thoroughly summarized elsewhere (e.g., Gersten et al., 2009).

Refined Research Questions

Considering findings and limitation of the previous research on the topic, the current MA aimed to address the following, more specific, questions:

- Do innovative researcher-developed (experimental) instructional intervention in early Mathematics have advantage over more traditional teaching practices as measured by a number of age-appropriate performance assessment tools?
- What is the magnitude of the average weighted effects size specifically for the following outcome types: (1) Precursors (or logical foundation) of further competencies in Mathematics (e.g., classification, seriation, and oddity skills); (2) Number Sense understanding and mastery (both generic and as manifested in performing specific tasks); (3) Geometry knowledge and skills; and (4) Mathematics composite outcomes (total scores of tests that address various knowledge and skills)?
- Does the magnitude of these effects depend on the category of experimental (e.g., 'direct instruction', 'guided discovery') and the corresponding control (e.g., 'business as usual', active control with or without coverage of the matched content in Mathematics) influence the direction and magnitude of these effects?
- What substantive (e.g., whole-class versus small-group activities, use of manipulatives, individualized feedback), contextual (e.g., country of the study, treatment duration and intensity) and demographic (e.g., gender composition and socio-economic status of the participating samples) moderate these effects and to what extent?

Methods

In conducting this MA we followed the standards recommended by the Campbell Collaboration (2019), the major international body for supporting, coordinating, and promoting conduct of systematic reviews of research evidence in social sciences. Typically, design and implementation of a systematic review consist of a series of steps, as outlined by Cooper (2017) – from study selection, through effect size extraction, aggregation, and various analyses toward interpretation and presentation of the findings (see also, Bernard et al., 2014; Borenstein et al., 2009; Borokhovski & Bernard, 2013a, 2013b). These steps and some associated statistical information are presented in greater detail in Appendix A.

Search Strategy, Implementation, and Outcomes

Based on the primary research question about relative effectiveness of educational practices on Mathematics learning outcomes of young students we developed and implemented the following search strategy. To build a collection of relevant publications for the review, searches were conducted in an iterative manner across an array of bibliographic databases that included both subject-specific and multi-disciplinary resources. No date limits were placed on the results, though they were restricted to publications in English. The following databases have been searched: ERIC, Education Source, Academic Search Complete, Communication Abstracts, ProQuest Central, ProQuest Dissertations & Theses Global, PsycINFO, and Web of Science.

Search results were required to contain at least one term from each of five sets of keywords for the following main concepts: 1) Mathematics, 2) Classroom-based, 3) General Outcomes, 4) Early Childhood population, and 5) Quantitative Methodology. For example:

("Number Conservation" OR "Numerical Conservation" OR Math* OR Numeracy OR "Number Sense" OR Counting OR Manipulative*)

AND

(Train* OR Teach* OR Classroom OR "in-class" OR Instruct* OR "whole class" OR classwide OR "small group*" OR "Emotional Training" OR "Emotional Regulation" OR "Cognitive Training" OR "Theory of the Mind" OR "Zone of Proximal Development" OR scaffold*)

AND

(Outcomes OR Prepared* OR Cogniti* OR Development OR Achieve* OR Aptitude OR Failure OR Attain* OR Grades OR Grading OR Effective* OR Succe* OR "Math* Competenc*" OR Performance OR "Emotional Competenc*" OR "Executive Function*" OR "Concept Formation" OR Reasoning)

AND

(Preschool* OR Pre-Kindergarten OR Kindergarten OR "early childhood education")

AND

("Comparison Group" OR Quasi-Experiment* OR Quasiexperiment* OR "control group" OR "post-test" OR "pre-test" OR posttest OR pretest OR "random* assign* OR "randomiz*" OR RCT OR "experimental comparison" OR "pre-post design")

All searches were conducted by a library science expert and followed standards first articulated by Cooper (1998) and later developed into a set of best practices by the Campbell Collaboration (Kugley et al., 2017; see also Pickup et al., 2018). Wherever possible, the strategy was tailored to the available features of a particular database, for instance by using official controlled vocabulary subject headings or publication type filters (e.g., to remove magazine and newspapers articles). Table 1 below summarizes the outcomes of all searches across sources used before and after removal of duplicates.

Table 1

Search outcomes by database

Source	Results	Results After Duplicates Removed
ERIC	361	352
ERIC 2	53	46
ERIC 3	4	4
Education Source	103	67
Education Source 2	49	29
Education Source 3	1	0
Academic Search Complete	67	9
Academic Search Complete 2	40	8
Academic Search Complete 3	1	0
Communication Abstracts	3	2
Communication Abstracts 2	0	0
Communication Abstracts 3	0	0
ProQuest Central	78	48
ProQuest Central 2	34	17
ProQuest Central 3	2	2
ProQuest Dissertations	123	96
ProQuest Dissertations 2	15	10
ProQuest Dissertations 3	3	1
PsycINFO	169	104
PsycINFO 2	255	49
PsycINFO 3	1	0
Web of Science	149	73
Web of Science 2	202	33
Web of Science 3	5	0
Branching	68	42
Erudit	221	221
Google	124	74
What Works Clearinghouse	8	8
TOTAL:	2131	1295

Inclusion / Exclusion Criteria

Primary empirical studies (not reviews, surveys, or conceptual articles) included in this meta-analysis satisfied the following inclusion criteria. They had to:

- Be conducted in formal pre-school educational settings (pre-kindergarten and kindergarten) whose findings could be generalized to guide teaching practices that would not require additional resources, i.e., studies featuring one-on-one or outside of the class-time tutoring were excluded;
- Deal with samples of students from general population (of variable levels of knowledge and skills), but not special needs students (e.g., diagnosed with autism spectrum disorders, various forms of learning disabilities);
- Address at least two different instructional strategies to enable comparison between innovative (experimental) and more traditional (control) conditions;
- Implement adequate research design – either randomized or high-quality quasi-experimental trials with more than a single class / teacher per condition to compensate for a potential teacher / personality effect;
- Quantitatively assess students' learning outcomes in at least one of the above-mentioned categories: (1) Mathematics Precursors, (2) Number Sense, (3) Geometry, and (4) Composite Scores.
- Provide sufficient description of instructional strategies carried out in both experimental and control conditions to enable coding for moderator variables.

Failure to meet any of these criteria resulted in study exclusion, either at the abstract screening or full-text review level, carried out independently by pairs of trained research assistants (graduate students in UQAM and Concordia in Montreal, Canada). The reason for rejection (after discussing and resolving occasional disagreements) was documented for subsequent reporting (please, see the next sub-section).

In other words, we considered experimental and quasi-experimental (those that relied on proper methodological or statistical controls) quantitative empirical studies to compare two or more forms of Mathematics instruction in regular pre-school classrooms, while excluding studies of low-methodological quality (e.g., studies without statistical control for initial group differences), studies examining the effectiveness of instructional approaches that could not be implemented in a regular classroom context (e.g., one-on-one instruction, home schooling) and studies that did not provide sufficient information on either the researcher-developed instruction in the experimental or business-as-usual instruction in the control classrooms (e.g., references to commercial textbooks used in either condition).

Screening / Selecting Studies

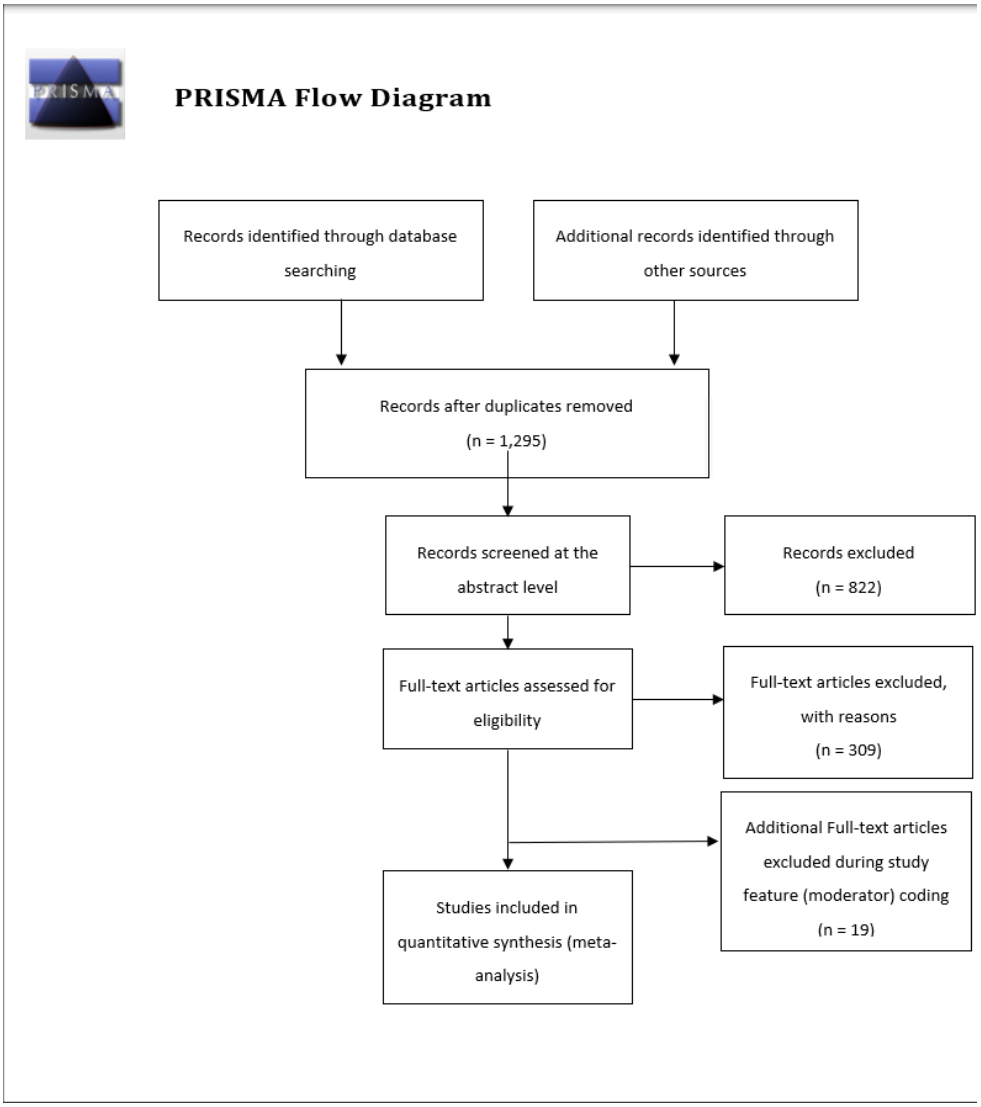
The flow-chart PRISMA diagram below illustrates decisions made by implementing independent pair-wise abstract screening and full-text review procedures.

Major reasons for discarding studies from the MA were:

- At the abstract screening stage (822 studies or 63.5% of total number identified through systematic searches) – (1) relevance of intervention and outcomes to the review area of interest, i.e., teaching and learning Mathematics, (2) non-empirical nature of papers under review, and (3) absence of a comparison group;
- At the full-text review stage (309 studies or 65.3% of reviewed at full-text level) – (1) inadequate research design, (2) insufficient (for effect size extraction) statistical information, and (3) qualitative nature of reported data;
- At the study features coding stage (additional 17 studies removed from the dataset) – (1) lack of detailed description of either experimental or control conditions needed to meaningfully categorize them and (2) specific aspects of intervention implementation that would disqualify the study from analyses (i.e., make them incompatible with other studies and/or impossible to generalize effects to suggest using the same instructional interventions in regular educational practice – namely, 'one teacher per condition' or 'one-on-one instruction' that would require additional resources). Also, two 'data duplicates' (studies using the same samples and the same outcome measures) were identified and removed from the collection at this point.

Moderator Variables

As stated earlier, in order to explain variability in findings four categories of moderator variables were coded (also through independent double-coding procedure) and analysed according to the mixed effects analytical model. Methodological moderator variables (e.g., study research design, psychometric quality of the assessment tools employed, precision of the effect size extraction procedure) were used to check the consistency of data distribution to make sure that fluctuations due to methodological quality did not seriously affect the outcomes. The other three classes of moderator variables, intended to systematically explore and explain variability surrounding weighted average effect size in each outcome category, were: (1) substantive (e.g., direct instruction versus guided discovery, whole-class versus small-group activities, use of manipulatives, use of technology, individualized feedback); (2) contextual (e.g., country of the study, treatment duration and intensity, parental involvement), and (3) demographic (e.g., gender composition and socio-economic status of the participating samples). The complete codebook outlining specific levels of all moderator variables is available upon request.



Results

As depicted by the flow chart in the Method section above, overall one hundred and forty-five primary empirical studies were selected for inclusion in the current MA. Inter-rater agreement rates at the abstract screening and the full-text review stages were: 85.3 % (Cohen's kappa = 0.71) and 86.7 % (Cohen's kappa = 0.73), respectively. Together these studies produced two hundred and seventy-eight individual effect sizes (k = 278) in the four major outcome categories. Effect size extraction reliability was estimated at 88.5 % (Cohen's kappa = 0.77). Finally, all effect sizes were coded for the complete set of moderator variables with the average inter-rater agreement rate of 83.2 % (Cohen's kappa = 0.66). These rates represent the initial (independent) pair-wise comparisons of the corresponding decision-making steps and could

be judged as sufficiently high to confidently rely upon and to use data that resulted from these decisions in the subsequent analytical procedures. At each stage of the review discrepancies were documented, discussed and resolved, by inviting the third opinion from one of the senior researchers on the team, when necessary.

The Overall Effect Sizes

Table 2 below presents the overall findings for all four major outcome categories: (1) Mathematics Precursors (k = 40); (2) Number Sense Indices (k = 168); (3) Geometry Knowledge and Skills (k = 30); and (4) Composite Measures of Mathematics (k = 40). The corresponding weighted average effect sizes are reported there according to both Random Effects and Fixed Effect analytical models. Given the apparent non-uniformity (sample- / settings- and instruction-wise) of studies in our collection, the former provides more accurate estimates of the effects in the studied population, while the latter serves primarily to estimate the degree of heterogeneity of the respective effect size distributions – these statistics are reported in the last line of each subsection of the table below.

Table 2

Overall Weighted Average Effect Size for All Four Outcome Types and Associated Heterogeneity Statistics

Population Estimates	*k*	*g*+	*SE*	Lower 95th	Upper 95th
Precursors of Mathematics:					
Fixed Effect Model	40	0.279**	0.02	0.23	0.33
Random Effects Model	40	0.429**	0.06	0.31	0.55
Heterogeneity Analysis	Q_T = 195.22 (df = 39), p < .001, I^2 = 80.02				
Number Sense Outcomes:					
Fixed Effect Model	168	0.394**	0.01	0.37	0.42
Random Effects Model	168	0.455**	0.03	0.39	0.52
Heterogeneity Analysis	Q_T = 1003.57 (df = 167), p = .001, I^2 = 83.36				

Population Estimates	*k*	*g+*	*SE*	Lower 95th	Upper 95th
Geometry Knowledge and Skills:					
Fixed Effect Model	30	0.247**	0.03	0.20	0.30
Random Effects Model	30	0.372**	0.07	0.24	0.50
Heterogeneity Analysis	Q_T = 138.37 (*df* = 29), *p* < .001, I^2 = 79.04				
Composite (Mixture of Several Types) Mathematics Outcomes:					
Fixed Effect Model	40	0.238**	0.02	0.21	0.27
Random Effects Model	40	0.374**	0.07	0.23	0.51
Heterogeneity Analysis	Q_T = 636.29 (*df* = 39), *p* < .001, I^2 = 93.87				

* *p* < .05; ** *p* < 0.01

Publication Bias and Sensitivity Analyses

After aggregating effect sizes within each outcome category into the corresponding weighted average point estimate and prior to engaging into further exploring the surrounding variability, we conducted necessary publication bias and sensitivity analyses. The latter discovered no obvious outliers in ether of the collections, whereas the former found potentially "missing" null-effects in all but one collection of outcomes, to a different extent though. Below we report these results with the suggested adjustments. It is worth noting that adjustments to the magnitude of the overall weighted average effects are based on statistical assumptions about potentially "missing" studies that were not located in actual searches. With this in mind, we also performed checks for the robustness of the calculated effect sizes in each outcome collection.

Collection of Precursors: Duval and Tweedie's *Trim and Fill* analytical routine implied eleven studies potentially missing from right side of the distribution needed to balance it. Imputing effects from those "missing" studies would result in bringing the overall effect size in this outcome category from the observed 0.279 to the adjusted 0.187 – under the fixed analytical model and from 0.429 to 0.207 – under the random effects model. Both adjusted effects, however, would remain statistically significant. Moreover, *Classic Fail-safe* analysis revealed that 1,747 potentially "missing" null-effect studies would be needed to render the

observed overall weighted average effects size non-significant (i.e., to bring the calculated *p*-value above the pre-set alpha of 0.05), i.e., indicating that these findings were quite robust. The results of these same analyses for the other three collections are briefly summarized below.

Collection of Number Sense outcomes: Duval and Tweedie's *Trim and Fill* adjusted effects were 0.346 (down from the observed 0.394) – under the fixed effect model and 0.354 (down from the observed 0.455) – under the random effects model. Both adjusted effects are statistically significant. *Classic Fail-safe N* = 45,566.

Collection of Geometry Knowledge and Skills outcomes: Duval and Tweedie's *Trim and Fill* adjusted effects were 0.171 (down from the observed 0.247) – under the fixed effect model and 0.198 (down from the observed 0.372) – under the random effects model. Both adjusted effects are statistically significant. *Classic Fail-safe N* = 828.

Collection of Composite Mathematic outcomes: Duval and Tweedie's *Trim and Fill* found no need for adjusting the overall average effect size under either fixed or random analytical model, both statistically significant. *Classic Fail-safe N* = 2,746.

Given the robustness of the findings and that no actual source of publication bias could be identified, we proceeded with the rest of the analyses for unadjusted values.

Analyses of Methodological Moderator Variables

Methodological moderators were first analyzed for each collection of effect sizes to ensure their consistency – i.e., to confirm that fluctuations in various aspects of methodological quality did not distort the patterns of outcomes. With just a couple of minor exceptions, it was the case across all four data sets. Specifically: (1) within the Geometry outcome category, six quasi-experimental studies produced the average effect significantly larger than the rest twenty-four randomized control trials did (p = .002); (2) within the Precursors outcome category, precisely calculated effect sizes were on average larger than the approximated ones (p = .018) – it is not surprizing given that our team used more conservative estimates that would rather underestimate than overinflate the approximated effects; and (3) within the Number Sense outcome category, statistically significant difference (p = .032) was detected between effect sizes derived from standardized and researcher-developed assessment tools in favour of the latter. In all these cases, before engaging into analyses of substantive, contextual and demographic moderators, decisions to either adjust for or explain the detected differences were made by the research team.

Analyses of Substantive, Contextual and Demographic Moderator Variables

This section is organized by outcome type, where within each category, only significant findings for substantive, contextual, and demographic moderator variables are reported in the table format. In some analyses number of cases per level of the corresponding moderator variable may not add up to the total number of cases in the collection, as the N/A (information

necessary for coding was not available in the original report) option was removed from the analyses.

Collection of Mathematics Precursors

Within this outcome category, significant differences were observed for the following moderator variables: *Intervention Type* – in favor of content-specific interventions*; Alignment of the Outcome Measures with the Content of the Instruction* – in favor of higher overlap between what was taught and what was measured*; Use of Physical Manipulatives* – in favor of the usage*; Use of Computer-Based Technology* – in favor of technology-free interventions*; Individualized Feedback* – in favor of providing such feedback (Table 3); *Intensity of Experimental Intervention* – in favor of moderate intensity (3 times per week)*;* and *Geographic Region* – in favor of studies conducted in USA (Table 4).

Table 3
***Substantive** Moderator Variables and Associated Heterogeneity Statistics*

Levels of Moderator Variables	Number of Effects (*k*)	Category Average *ES* (*g*+)	Lower 95th *CI*	Upper 95th *CI*	$Q_{Between}$
Intervention Type: Content-Specific vs. Developmental					
Content-Specific Interventions	36	0.445**	0.31	0.58	
Developmental Interventions	3	0.089	-0.08	0.25	
Between Groups: *df* = 1					10.91, *p* = .001
Were Precursors Taught in the Intervention?					
Yes	28	0.596**	0.39	0.81	
No	10	0.313**	0.16	0.47	
Between Groups: *df* = 1					4.59, *p* = .032

Levels of Moderator Variables	Number of Effects (k)	Category Average ES (g+)	Lower 95th CI	Upper 95th CI	$Q_{Between}$
Instructional Use of Physical Manipulatives					
Yes	28	0.522**	0.35	0.69	
No	10	0.276**	0.11	0.44	
Between Groups: df = 1					4.17, p = .041
Instructional Use of Technology (Computer-Based)					
Yes	9	0.175	-0.07	0.42	
No	26	0.592**	0.41	0.78	
Between Groups: df = 1					7.08, p = .008
Provision of Individualized Feedback					
Yes	14	0.717**	0.38	1.06	
No	26	0.319**	0.20	0.43	
Between Groups: df = 1					4.73, p = .030

* p < .05; ** p < .01

Table 4
***Contextual** Moderator Variables and Associated Heterogeneity Statistics*

Levels of Moderator Variables	Number of Effects (*k*)	Category Average *ES* (*g*+)	Lower 95th *CI*	Upper 95th *CI*	$Q_{Between}$
Intensity of Experimental Intervention: Classes per Week					
High (4 or 5 times a week)	5	0.404	-0.04	0.85	
Low (1 or 2 times a week)	8	0.129	-0.18	0.44	
Most common (3 times a week)	15	0.658**	0.43	0.89	
Between Groups: *df* = 2					7.33, *p* = .026
Geographic Region: USA vs. Other Countries					
USA	21	0.596**	0.39	0.81	
Other	19	0.313**	0.16	0.47	
Between Groups: *df* = 1					4.59, *p* = .032

* $p < .05$; ** $p < .01$

None of the considered demographic moderator variables came out to show significant differences across their levels in this outcome category.

Collection of Number Sense Outcomes

Within this outcome category, significant differences were observed for the following moderator variables: *Instructional Approach* – in favor of explicit instruction; *Type of Control Condition* – favoring various forms of active control; *Story-based Problems* – in favor of this format of presenting problems to students (Table 5); *Teacher's Experience* – in favor of more experienced instructors (Table 6); and *Learners' Abilities* – in favor of severely underperforming students (Table 7).

Table 5
***Substantive** Moderator Variables and Associated Heterogeneity Statistics*

Levels of Moderator Variables	Number of Effects (*k*)	Category Average *ES* (*g*+)	Lower 95th *CI*	Upper 95th *CI*	$Q_{Between}$
Instructional Approach: Explicit Instruction vs. Guided Discovery					
Explicit Instruction	98	0.471**	0.40	0.54	
Guided Discovery	23	0.285**	0.17	0.40	
Between Groups: *df* = 1					7.01, *p* = .008
Control Conditions for Content-Specific Instruction					
Active Control without Math	19	0.442**	0.24	0.65	
Active Control with Math	7	0.479*	0.06	0.90	
'Business-as-Usual' without Math	11	0.278**	0.18	0.38	
'Business-as-Usual' with Math	88	0.485**	0.41	0.56	
Between Groups: *df* = 3					10.67, *p* = .014
Were Story-Based Problems an Explicit Content Component?					
Yes	41	0.597**	0.47	0.73	
No	127	0.409**	0.34	0.48	
Between Groups: *df* = 1					6.18, *p* = .013

* $p < .05$;
** $p < .01$

Table 6
***Contextual** Moderator Variables and Associated Heterogeneity Statistics*

Levels of Moderator Variables	Number of Effects (*k*)	Category Average *ES* (*g*+)	Lower 95th *CI*	Upper 95th *CI*	$Q_{Between}$
Experimental Group Teacher's Experience					
Extensive: Over 5 Years	48	0.462**	0.36	0.56	
Limited: Under 5 Years	7	0.235**	0.16	0.37	
Between Groups: *df* = 1					12.79, $p < .001$

* $p < .05$; ** $p < .01$

Table 7
***Demographic** Moderator Variables and Associated Heterogeneity Statistics*

Levels of Moderator Variables	Number of Effects (*k*)	Category Average *ES* (*g*+)	Lower 95th *CI*	Upper 95th *CI*	$Q_{Between}$
Learners' Abilities					
High Achievers	3	0.046	-0.30	0.39	
Low Achievers	4	0.399**	0.14	0.66	
At-Risk of Failure in Math Students	49	0.543**	0.44	0.65	
Unselected Sample	110	0.428**	0.35	0.51	
Between Groups: *df* = 3					8.62, $p = .035$

* $p < .05$; ** $p < .01$

Collection of Geometry Outcomes

Within this outcome category, significant differences were observed for the following moderator variables: *Intervention Type* – in favor of content-specific instruction (Table 8); Grade Level – in favor of older kids; and *Ethnicity of participating samples* – in favor of visible minorities and diverse samples (Table 9).

Table 8
***Substantive** Moderator Variables and Associated Heterogeneity Statistics*

Levels of Moderator Variables	Number of Effects (k)	Category Average ES (g+)	Lower 95th CI	Upper 95th CI	$Q_{Between}$
Intervention Type: Content-Specific vs. Developmental					
Content-Specific Interventions	26	0.423**	0.28	0.56	
Developmental (or Mixed) Interventions	4	0.034	-0.19	0.26	
Between Groups: df = 1					8.31, p = .004

* p < .05; ** p < .01

Table 9
***Demographic** Moderator Variables and Associated Heterogeneity Statistics*

Levels of Moderator Variables	Number of Effects (k)	Category Average ES (g+)	Lower 95th CI	Upper 95th CI	$Q_{Between}$
Grade Level					
Pre-Kindergarteners	13	0.239*	0.10	0.38	
Kindergarteners	17	0.527**	0.31	0.74	
Between Groups: df = 1					4.85, p = .028

Sample Demographics: Ethnicity					
Predominantly Minorities	10	0.489**	0.18	0.79	
Predominantly White	3	-0.009	-0.30	0.28	
Diverse Samples	9	0.483**	0.29	0.67	
Between Groups: *df* = 2					8.50, *p* = .014

* *p* < .05; ** *p* < .01

There were no contextual moderator variables with statistically significant differences across their levels capable of explaining systematic variation in effect sizes for this category of learning outcomes.

Collection of Mathematics Composite Outcomes

Within this outcome category, significant differences were observed for the following moderator variables: *Use of Physical Manipulatives* – in favor of the usage*; Use of Computer-Based Technology* – in favor of technology-free interventions (Table 10)*; Implementation Fidelity Checks* – in favor of studies, in which implementation fidelity was ensured by independent observers; *Intensity-by-Duration of Intervention Interaction* – in favor of longer interventions with relatively low (once or twice a week) intensity (Table 11)*; Learners' Abilities* – in favor of severely underperforming students; and *Geographic Region* – in favor of studies conducted in USA (Table 12).

Table 10
***Substantive** Moderator Variables and Associated Heterogeneity Statistics*

Levels of Moderator Variables	Number of Effects (*k*)	Category Average *ES* (*g*+)	Lower 95th *CI*	Upper 95th *CI*	$Q_{Between}$
Instructional Use of Physical Manipulatives					
Yes	24	0.530**	0.26	0.80	
No	15	0.174**	0.11	0.23	
Between Groups: *df* = 1					6.42, *p* = .011

Levels of Moderator Variables	Number of Effects (k)	Category Average *ES* (g+)	Lower 95th *CI*	Upper 95th *CI*	$Q_{Between}$
Instructional Use of Technology (Computer-Based)					
Yes	9	0.168	-0.08	0.41	
No	31	0.499**	0.36	0.64	
Between Groups: df = 1					5.31, p = .021

* p < .05; ** p < .01

Table 11

Contextual *Moderator Variables and Associated Heterogeneity Statistics*

Levels of Moderator Variables	Number of Effects (k)	Category Average *ES* (g+)	Lower 95th *CI*	Upper 95th *CI*	$Q_{Between}$
Implementation Fidelity Checks					
Teachers' Self-Reports	4	0.075	-0.01	0.16	
Third-Party Formalized Observations	17	0.472**	0.20	0.75	
Combination of Self-Reports and Third-Party Observations	12	0.219**	0.08	0.36	
No Check	7	0.362	-0.01	0.73	
Between Groups: df = 3					10.41, p = .015
Intensity-by-Duration of Intervention Interaction					
Long (over Semester) x High Intensity	8	0.289*	0.10	0.48	
Long (over Semester) x Low Intensity	10	0.654**	0.37	0.93	

Levels of Moderator Variables	Number of Effects (k)	Category Average *ES* (g+)	Lower 95th *CI*	Upper 95th *CI*	$Q_{Between}$
Medium (Semester) x High Intensity	8	0.143*	0.07	0.22	
Medium (Semester) x Most Common (3 times a week) I	3	0.328	-0.03	0.68	
Between Groups: df = 3					13.50, p = .004

* p < .05; ** p < .01

Table 12

Demographic *Moderator Variables and Associated Heterogeneity Statistics*

Levels of Moderator Variables	Number of Effects (k)	Category Average *ES* (g+)	Lower 95th *CI*	Upper 95th *CI*	$Q_{Between}$
Learners' Abilities					
At-Risk of Failure in Math Students	7	0.499**	0.31	0.69	
Unselected Sample	32	0.246**	0.16	0.33	
Between Groups: df = 1			5.59, p = .018		
Geographic Region: USA vs. Other Countries					
USA	34	0.419**	0.26	0.58	
Other	6	0.147*	0.04	0.25	
Between Groups: df = 1					7.69, p = .006

* p < .05; ** p < .01

Discussion

Just to remind, the research questions of this systematic review addressed: (1) The overall comparative effectiveness of innovative instructional interventions in early teaching of Mathematics; (2) Their effectiveness assessed for different categories of learning outcomes; and (3) The systematic influence (if any) of methodological, substantive, contextual, and demographic moderator variables on the magnitude of the corresponding effect sizes. In brief summary, the result of the undertaken MA with respect to its major research questions demonstrated the following.

Innovative researcher-developed experimental instructional intervention in early Mathematics demonstrated a strong potential to be advantageous in comparison with traditional ("business-as-usual") instruction – i.e., significant positive effect sizes (standardized mean difference in learning outcomes between students in experimental and control conditions) were observed in all investigated outcome categories.

Specifically, the following effects – Hedges' g^+ values calculated according to the random effects analytical model (as the most appropriate given the objectives of this MA and the nature of primary empirical studies included in it) – were discovered: (1) $g^+ = 0.429$, $k = 40$, $p < .01$ – for the collection of Precursors outcomes; (2) $g^+ = 0.455$, $k = 168$, $p < .01$ – for the collection of Number Sense outcomes; (3) $g^+ = 0.372$, $k = 30$, $p < .01$ – for the collection of Geometry Knowledge and Skills outcomes; and (4) $g^+ = 0.374$, $k = 40$, $p < .01$ – for the collection of Composite Mathematics outcomes. All these findings were significantly heterogeneous, thus warranting further exploration of the systematic sources of variability through the subsequent moderator variable analyses.

With respect to the latter, here we take a look primarily at those moderators, whose significant influence was detected in more than one outcome category. First, it is important to acknowledge that so-called *content-specific* interventions – i.e., those with the focus on Mathematic notions taught and competencies to be mastered – were more effective than what was considered *developmental* interventions – i.e., those pursuing the purpose of more general development, such as cognitive (non-Mathematical skills / knowledge, e.g., common erudition, self-regulation); social (e.g., communicative attitudes and skills); sport or music-related. It was evident for two out of the four outcome categories, namely: for Mathematics Precursors and for Geometry. Actually, content-specific instruction came on top in all four outcome categories, just that difference has not reached the level of statistical significance for Number Sense and Composite Mathematics scores. On the surface, this pattern of results seems rather obvious – one gains expertise in what she/he is directly exposed to (what is practiced). Partly, it also explains higher effect sizes in the studies where instructional content was closely aligned with the assessment tools of learning outcomes ($g^+ = 0.60$, $p < .01$, $k = 28$ – in the Precursors outcome category).

However, the reality may be more complicated, and various instructional approaches stemming from well-established theories in developmental psychology (including those of Jean Piaget and Lev Vygotsky) are not to be easily ignored. It is most likely that the applied value of

specific instructional approaches, especially for learners of that young age, should be tested in a context – what is to be learned, by whom, and under what particular circumstances. Indirectly it is supported by the fact that, despite many developmental interventions resulted in low or even negative effect sizes, there are also examples of quite successful studies that employed this type of teaching strategies. Among those are: small-group communication and exchange over the shared computer screen (Gomez et al., 2013; g = 0.58); music-enriched Montessori instruction (Harris, 2007; g = 2.67); domain-general working memory training (Kroesbergen et al., 2014; g = 0.41) – all in the Number Sense outcome category. It is just that content-specific interventions in early Mathematics produced positive results (higher effect sizes) more consistently than the developmental ones across outcome categories. The following could be named as the most illustrative examples. The early numeracy training program that included different paper-and-pencil tasks and games designed to enhance early numerical abilities such as counting, number-line representation, one-to-one correspondence between quantities and numerals, and quantity comparison (Passolunghi & Costa, 2016) resulted in g = 1.14. Liang et al. (2019) studied a game-based instruction in Mathematics, designed for number sense acceleration by means of employing self-made number cards, small bamboo baskets, beans, pills, etc., and documented an effect size of g = 0.85. Another content specific intervention (ROOTS_50) consisted of 50 lessons, designed to build students' whole number proficiency by emphasizing concepts from the Counting & Cardinality and Operations & Algebraic Thinking Domains, was significantly more effective in comparison with traditional teaching Mathematics to young kids: g = 1.09 (Doabler et al., 2016). All these effect sizes belong to the Number Sense outcome category. Also, there were: Head Start program composed of small-group activities with direct participation of teachers who presented the learning materials and activity templates, provided stories, mathematical information, and guiding elaboration questions, further expanded into whole-class work (Brendefur et al., 2013; g = 0.51 – in the Geometry category and g = 0.59 – in the Precursors category) or Project M2 (Casa et al., 2017) that involved intensive class discussions on identifying attributes of shapes (including 3D) and distances, estimating, comparing and ordering lengths, etc. (g = 2.77 – in the Mathematics Composite category).

Somewhat related to the above, are the findings in favor of Explicit Instruction (g^+ = 0.47, k = 98) in comparison with more constructivism-driven Guided Discovery (g^+ = 0.29, k = 23) in the Number Sense outcome category – $Q_{between}$ = 7.01 (though both were statistically significant at p < .01) – despite some effective interventions in the latter category, explicit explanations of the terms, concepts , and rules appear to be more effective for pre-schoolers. However, further more nuanced, attention is needed to specific pedagogical approaches whether they are aligned more with the constructivist or the direct instruction paradigm (e.g., Gresalfi & Lester, 2009 – for a longstanding discussion on the issue).

Also, of undeniable practical value is quite consistent negative influence on learning outcomes of introducing computer-based technology in teaching Mathematics to young kids. In two outcome categories the average effects in the studies that employed technological tools were significantly lower than in the studies that did not (g^+ = 0.18, *ns*, k = 9 vs. g^+ = 0.59, p < .01,

k = 26; $Q_{between}$ = 7.08 – in the Precursors outcome category and g^+ = 0.17, k = 9, *ns* vs. g^+ = 0.50, p < .01, k = 31; $Q_{between}$ = 5.31 – in the Mathematics Composite outcome category). So, despite relative popularity of software applications for teaching and learning and the 'novelty effect', often used to explain better outcomes of employing new fashionable devises, use of technology in instruction of Mathematics to young learners seems to be rather counterproductive. For one, the content of the instruction is also new at this age, and combining it with a new delivery tool could cancel the effect of either (if not even be overwhelming), also given how widespread various technological tools (including mobile ones) and associated applications are in everyday life outside of classrooms even among the youngest learners, novelty may simply not apply at all.

On contrary, usage of physical manipulatives (material object that could illustrate / explain the content taught), seems to be effective in early Mathematics instruction, quite in line with various theoretical explanations – either facilitating progression through preoperational developmental (in Piagetian terms) or helping to internalize new knowledge and skills with needed help from teachers or in active exchange with peers (as in logic of the 'zone of proximal development' by Vygotsky). Indeed, the difference between employing and not employing physical manipulatives, in favor of the former was quite evident in the Precursors and Mathematics Composite outcome categories: g^+ = 0.52, p < .01, k = 28 vs. g^+ = 0.28, p < .01, k = 10; $Q_{between}$ = 4.17 and g^+ = 0.53, p < .01, k = 24 vs. g^+ = 0.17, p < .01, k = 15; $Q_{between}$ = 6.42, respectively (though interventions without manipulatives were also effective, just to a lesser degree).

Furthermore, some data on the effectiveness of provision of elaborate individualized feedback (e.g., g^+ = 0.72, p < .01, k = 14 – in the Precursors outcome category) could be suggestive for considering broader usage of this instructional component in preparing and implementing early Mathematics instruction. Similarly, it seems that presenting tasks / problems in a form of stories might be helpful for better understanding of the related concepts and more successful problem-solving – as in the Number Sense outcome category (g^+ = 0.60, p < .01, k = 41).

There are findings that show the effectiveness of innovative Mathematics teaching strategies on learning outcomes of various groups of underperforming / underprivileged students, for instance: g^+ = 0.54, p < .01, k = 49 and g^+ = 0.50, p < .01, k = 7 – for at-risk groups in the Number Sense and Mathematics Composite outcome categories, respectively; or g^+ = 0.49 , p < .01, k = 10 for visible minorities in the Geometry outcome category. Those did not come as a surprise either, as many of experimental interventions are designed with the specific goal to elevate Mathematics knowledge and skills of these particular categories of young learners.

Finally, we found some evidence in support of a commonsense observation that 'all is good in moderation' – either moderately intensive (3 times a week) interventions (e.g., g^+ = 0.66, p < .01, k = 15 – in the Precursor outcome category) or longer, but less intensive interventions (e.g., g^+ = 0.65, p < .01, k = 10 – in the Mathematics Composite outcome category) appeared to be more effective than their counterparts of much higher / lower length / intensity.

Thus, we found that different moderator variables (substantive, contextual, and demographic) varied in their respective effects on the studied outcome categories. Some came up to significantly influence the effect of experimental (researcher-designed innovative)

interventions on learning early Mathematics, whereas some other manifested themselves at the level of a tendency (statistically non-significant, but indicative). The latter deserve more careful consideration, possibly supplemented by in-depth qualitative analyses.

Conclusion

There are a couple of concluding remarks that, we believe, is important to present in summarizing finding of the current systematic review. Complex and challenging pedagogical tasks / objectives (early Mathematics teaching is, undoubtedly, one of them – as it has serious implication for many future academic developments and social / professional adaptation) require proportionately complex solutions – instructional practices that are carefully designed and implemented (preferably with various checks for fidelity of implementation). A single factor (as powerful as some instructional strategies could be) hardly is capable of delivering unambiguously desirable results. Different instructional approaches, tailored for specific learning goals (knowledge and skills to acquire) and student populations in various contexts should be considered. No less importantly, effective practical (instructional / pedagogical / administrative) decisions should be evidence-based – i.e., grounded in most rigorous empirical research. Large-scale comprehensive MAs, supported by focused qualitative inquiries is one of the most reliable sources of such evidence.

References

Bernard, R. M., Borokhovski, E., Schmid, R. F., Tamim, R. M. & Abrami, P. C. (2014). A meta-analysis of blended learning and technology use in higher education: From the general to the applied. *Journal of Computing in Higher Education, 26*(1), 87–122. https://doi.org/10.1007/s12528-013-9077-3

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

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods, 1*(2), 97–111. https://doi.org/10.1002/jrsm.12

Borokhovski, E. F., & Bernard, R. M. (2013a). Quantitative syntheses in social sciences: Methodology and practice of meta-analysis. Part 1. *Psychology-Economics-Law*, *2*, 6-15. (In Russ.).

Borokhovski, E. F., & Bernard, R. M. (2013b). Quantitative syntheses in social sciences: Methodology and practice of meta-analysis. Part 2. *Psychology-Economics-Law, 3*, 41-56. (In Russ.).

*Brendefur, J., Strother, S., Thiede, K., Lane, C., & Surges-Prokop, M. J. (2013). A professional development program to improve math skills among preschool children in Head Start. *Early Childhood Education Journal, 41*, 187–195. https://doi.org/10.1007/s10643-012-0543-8

Campbell Collaboration. (2019). *Campbell systematic reviews: Policies and guidelines* (Version 1.4).

Casa, T. M., Firmender, J. M., Gavin, M. K., & Carroll, S. R. (2017). Kindergarteners' achievement on geometry and measurement units that incorporate a gifted education approach. *Gifted Child Quarterly, 61*(1), 52–72. https://doi.org/10.1177/0016986216671806

Cooper, H. (1998). *Research synthesis and meta-analysis: A step-by-step approach*. SAGE Publications.

Cooper, H. (2017). *Research synthesis and meta-analysis: A step-by-step approach* (5th ed.). SAGE Publications.

Doabler, C. T., Clarke, B., Kosty, D. B., Kurtz-Nelson, E., Fien, H., Smolkowski, K. & Baker, S. K. (2016).

Testing the efficacy of a tier 2 mathematics intervention: A conceptual replication study. *Exceptional Children*, *83*(1), 92–110. https://doi.org/10.1177/0014402916660084

Dyson, N. I., Jordan, N. C., & Glutting, J. (2013). A number sense intervention for low-income kindergartners at risk for mathematics difficulties. *Journal of Learning Disabilities, 46,* 166 –181. https://doi.org/10.1177/0022219411410233

Fuchs, L. S., Fuchs, D., & Karns, K. (2001). Enhancing kindergartners' mathematical development: Effects of peer-assisted learning strategies. *The Elementary School Journal, 101,* 495–510. https://doi.org/10.1086/499684

Fuson, K. C. (2009). *Math expressions*. Houghton Mifflin Harcourt.

Gersten, R., Chard, D., Jayanthi, M., Baker, S., Morphy, P., & Flojo, J. (2009). Mathematics instruction for students with learning disabilities or difficulty learning mathematics: A synthesis of the intervention research. *Review of Educational Research, 79*(3), 1202–1242. https://doi.org/10.3102/0034654309334431

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

Gomez, F., Nussbaum, M., Weitz, J. F., Lopez, X., Mena, J., & Torres, A. (2013). Co-located single display collaborative learning for early childhood education. *Computer Supported Learning 8*, 225–244. https://doi.org/10.1007/s11412-013-9168-1

Gresalfi, M. S., & Lester, F. (2009). What's worth knowing in mathematics? In S. Tobias & T. M. Duffy (Eds.), *Constructivist Theory Applied to Instruction: Success or Failure?* (pp. 264-290). Routledge.

Harris, M. A. (2007). Differences in mathematics scores between students who receive traditional Montessori instruction and students who receive music enriched Montessori instruction. *Journal for Learning through the Arts, 3*(1). https://doi.org/10.21977/D93110059

Hedges, L. V., Shymansky, J. A. & Woodworth, G. (1989). *A practical guide to modern methods of metaanalysis* (Stock Number PB-52). National Science Teachers Association. (ERIC Document Reproduction Service No. ED309952).

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

Jordan, N. C., Glutting, J., Dyson, N., Hassinger-Das, B., & Irwin, C. (2012). Building kindergartners' number sense: A randomized controlled study. *Journal of Educational Psychology, 104,* 647–660. https://doi.org/10.1037/a0029018

Kroesbergen, E. H., van 't Noordende, J. E., & Kolkman, M. E. (2014). Training working memory in kindergarten children: Effects on working memory and early numeracy. *Child Neuropsychology, 20*(1), 23–37. https://doi.org/10.1080/09297049.2012.736483

Kugley, S., Wade, A., Thomas, J., Mahood, Q., Jørgensen, A.-M. K., Hammerstrøm, K., & Sathe, N. (2017). Searching for studies: A guide to information retrieval for Campbell systematic reviews. *Campbell Systematic Reviews, 13*(1), 1–73. https://doi.org/10.4073/cmg.2016.1

Liang, Y., Zhang, L. J., Long, Y., Deng, Q., & Liu, Y. J. (2019). Promoting effects of RtI-based mathematical play training on number sense growth among low-SES preschool children. *Early Education and Development*, *31*(3), 335–353. https://doi.org/10.1080/10409289.2019.1664261

Nelson, G., & McMaster, K. L. (2018). The effects of early numeracy interventions for students in preschool and early elementary: A meta-analysis. *Journal of Educational Psychology, 111*(6), 1001–1022. https://doi.org/10.1037/edu0000334

Passolunghi, M. C. & Costa, M. C. (2016). Working memory and early numeracy training in preschool children. *Child Neuropsychology*, *22*(1), 81–98. https://doi.org/10.1080/09297049.2014.971726

Pickup, D. I., Bernard, R. M., Borokhovski, E., Wade, A. C., & Tamim, R. M. (2018). Systematically searching empirical literature in the social sciences: Results from two meta-analyses within the domain of education. *Russian Psychological Journal, 15*(4), 245–265. https://doi.org/10.21702/rpj.2018.4.10

Pigott, T. (2012). *Advances in meta-analysis*. Springer. https://doi.org/10.1007/978-1-4614-2278-5

Sood, S., & Jitendra, A. K. (2011). An exploratory study of a number sense program to develop kindergarten students' number proficiency. *Journal of Learning Disabilities, 46,* 328–346. https://doi.org/10.1177/0022219411422380

Wang, A. H., Firmender, J. M., Power, J. R., & Byrnes, J. P. (2016). Understanding the program effectiveness of early mathematics interventions for prekindergarten and kindergarten environments: A meta-analytic review. *Early Education and Development, 27*(5), 692–713. https://doi.org/10.1080/10409289.2016.1116343

Appendix A

Methodology of Meta-Analysis

A meta-analysis is a special class of systematic review that relies on quantifiable data from a multitude of primary studies addressing a common core research question or a set of closely related research questions. Meta-analysis aggregates systematically collected effect sizes from individual studies to estimate the magnitude of a difference between groups of interest in the entire studied population, and explains the variability that surrounds the overall effect size by thoroughly coding and analyzing methodological, substantive, demographic, and contextual moderator variables. The main research question(s) should be stated and substantiated a-priori to inform search strategies, to specify inclusion criteria, and to meaningfully guide the review process through all its steps (e.g., Cooper, 2017).

Step 1—Formulating the problem:

In addition to the overview of the state of the research in the area of interest, it includes the critical task of stating specific research questions, developing operational definitions of the main concepts and outcomes, determining key words that would guide subsequent literature searches, and formulating comprehensive inclusion criteria for admitting relevant studies to the review.

Step 2—Searching the literature:

Such search efforts necessarily should be systematic and as exhaustive as possible including not only targeted (adaptive, but consistent) searches of relevant electronic databases and manual searches of tables of content of major journals in the respective field of research, but also attention to so-called “grey literature” by means of consulting conference proceedings, unpublished dissertations, “branching” from previous major reviews, etc.

Step 3—Gathering information from studies:

At this step the actual review of retained documents proceeds, including effect size extraction and coding of moderator variables (all these require some form of consistency check among coders to ensure sufficient reliability of the review process and outcomes).

Step 4—Evaluating the quality of studies:

Conducted in parallel with Step 3, review activities include assessing methodological quality of studies under review and addressing the issue of potential dependencies and confounds among samples, treatment conditions and outcome measures.

Step 5—Analyzing and integrating the outcomes of research:

Specifically for meta-analyses this step presents challenges of selecting and applying a proper analytical model for aggregating effect sizes from individual independent samples and of implementing necessary procedures of sensitivity and publication bias analyses, whereas qualitative syntheses require meaningful and consistent categorization and cross-referencing of the data.

Step 6—Interpreting the evidence:

The key component of this step is relating findings to the main research questions and to the context outlined prior to undertaking the review to emphasize its unique contribution to the field and implications for practice.

Step 7—Presenting the results:

This step, though it may be perceived as pretty straightforward, requires clarity and transparency in describing not only findings, but the review procedures as well.

The basic metric and the unit of analysis in meta-analytical research is an effect size. Most frequently used in comparative studies in social sciences in general, and especially in Education, is the *d*-type effect size (also known as Cohen's *d*), which is the standardized difference between the means of two groups (usually the treatment and the control – $\bar{X}_t$ and $\bar{X}_c$ respectively) that are compared to each other on a consistent set of relevant outcomes, i.e., as depicted by *Equation 1*:

$$d = \frac{\bar{X}_E - \bar{X}_C}{SD_{Pooled}} \qquad \textit{Equation 1}$$

The denominator in this formula is the pooled standard deviation of both groups (in other words, the standardization term), calculated according to *Equation 2*:

$$SD_{Pooled} = \sqrt{\frac{(n_E - 1)SD_E^2 + (n_C - 1)SD_C^2}{(n_E - 1) + (n_C - 1)}} \qquad \textit{Equation 2}$$

An effect size of *d*-type can also be calculated or estimated from various inferential statistics, such as *t*-tests, *F*-ratios and/or associated *p*-values (exact or in relation with the preset alpha-levels) using an array of formulas (e.g., Glass et al., 1981; Hedges et al., 1989; Bernard et al., 2014). There is also a correction to Cohen's *d* for potential bias that could result from studies based on small samples. It is known as Hedges' *g* and calculated using *Equation 3*:

$$g \cong d\left(1 - \frac{3}{4N - 9}\right)$$

Equation 3

At the synthesis phase of a meta-analysis, to produce a truly representative overall average, effect sizes are weighted and analyzed according to one of the two analytical models: (1) The fixed-effect model or (2) The random-effects model. Comprehensive description of the conceptual foundations and procedural aspects of these models can be found in the related literature (e.g., Bernard et al., 2014; Borenstein et al., 2010; Borenstein et al., 2009; Pigott, 2012). For the purposes of the current review, we use the random-effects model as more suitable for the situation of great variety of samples, settings and contexts in the reviewed research. The overall analysis is preceded by necessary methodological checks, including sensitivity and publication bias analyses and then is followed by the analyses of moderator variables implemented within the logic of the so-called mixed effect analytical model. One of the key features of the systematic reviews in general and of a meta-analysis, in particular, is that all its procedures are carried out by at least two researchers, acting independently, comparing their respective decisions and, in case of disagreements, documenting, discussing, and resolving them in the most rigorous and consistent fashion possible, inviting, when necessary, a third qualified opinion from another member of the research team.

Received: March 18, 2025
Revision received: November 11, 2025
Accepted: March 30, 2026

Author Contributions

Eugene F. Borokhovski – Conceptualization, Methodology, Investigation, Formal analysis, Writing – original draft, Writing – review & editing, Supervision, Funding acquisition

Éric Dion – Conceptualization, Methodology, Investigation, Formal analysis, Writing – review & editing, Funding acquisition

David I. Pickup – Conceptualization, Methodology, Investigation

Isabelle Plante – Conceptualization, Writing – review & editing, Funding acquisition

Patrick Charland – Conceptualization, Writing – review & editing, Funding acquisition

Stéphane Cyr – Conceptualization, Writing – review & editing, Funding acquisition

Author Details

Eugene F. Borokhovski – PhD in Experimental Psychology, Affiliate Associate Professor, Department of Education, Systematic Reviews Project Manager, Centre for the Study of Learning and Performance (CSLP), Concordia University, Montreal, Canada; Web of Science ResearcherID: GQY-9237-2022; Scopus Author ID: 36860439400; ORCID ID: https://orcid.org/0000-0002-5615-0417; e-mail: eugene.borokhovski@concordia.ca

Éric Dion – PhD in Psychology, Full Professor, Department of Special Education and Vocational Training, Université du Québec à Montréal (UQAM), Montreal, Canada; ORCID ID: https://orcid.org/0000-0001-5912-7100; e-mail: dion.e@uqam.ca

David I. Pickup – Master of Library and Information Studies, Information Specialist, Centre for the Study of Learning and Performance (CSLP), Concordia University, Montreal, Canada; ORCID ID: https://orcid.org/0000-0001-5213-1238; e-mail: david.pickup@concordia.ca

Isabelle Plante – PhD in Education, Professor, Département de didactique, Faculté des sciences de l'éducation, Université du Québec à Montréal (UQAM), Montreal, Canada; ORCID ID: https://orcid.org/0000-0003-4612-7018; e-mail: plante.isabelle@uqam.ca

Patrick Charland – MSc, PhD, Full Professor, Département de didactique, Co-holder, UNESCO Chair in Curriculum Development, Université du Québec à Montréal (UQAM), Montreal, Canada; ORCID ID: https://orcid.org/0000-0001-5893-7443; e-mail: charland.patrick@uqam.ca

Stéphane Cyr – PhD in Education, Professor, Department of Mathematics, Co-holder, UNESCO Chair in Curriculum Development, Université du Québec à Montréal (UQAM), Montreal, Canada; ORCID ID: https://orcid.org/0000-0003-1794-5978; e-mail: cyr.stephane@uqam.ca

Conflict of Interest Information

The authors have no conflicts of interest to declare.