Annenberg Institute
BROWN UNIVERSITY

EdWorkingPaper No. 25-1198

School-based language, math, and reading interventions for executive functions in children and adolescents: A systematic review

Jens Dietrichson
The Danish Center for Social Science Research

Julie Kaas Seerup
The Danish Center for Social Science Research

Sofie Elgaard Iisager Jensen
The Danish Center for Social Science Research

Johan Klejs
The Danish Center for Social Science Research

Elizabeth Bengtsen
The Danish Center for Social Science Research

Martin Williams Strandby
The Danish Center for Social Science Research

Morten Kjær Thomsen
University of Oxford

Executive functions are a set of cognitive skills and processes used when directing behaviour towards the attainment of a certain goal. A large literature has documented positive associations between executive functions and a variety of desirable outcomes throughout life, including academic achievement. However, training executive functions appears to have limited effects on academic achievement, and the nature of this association remains unclear. We use a systematic review and meta-analysis to examine if preschool and school-based interventions training language, literacy, and/or mathematical skills improve children's and adolescents' executive functions. We include 51 studies in the data synthesis (47 are randomised controlled trials). Using inverse-variance weighted random-effects models, we find a statistically significant weighted average effect size on pre-validated measures of executive functions (0.14, 95% CI = [0.07, 0.22]). The effect is robustly positive in all sensitivity analyses, including tests of publication bias. We also find substantial heterogeneity, which persists in moderator analyses. This means we cannot identify specific types of interventions that are more effective than others in improving executive functions. Our results support theories that emphasise the unidirectional effects from academic skills to executive functions or a bidirectional relation. Further research is needed to identify the mechanisms through which academic skills training affect executive functions.

VERSION: February 2026

Suggested citation: Dietrichson, Jens, Julie Kaas Seerup, Sofie Elgaard Iisager Jensen, Johan Klejs, Elizabeth Bengtsen, Martin Williams Strandby, and Morten Kjær Thomsen. (2026). School-based language, math, and reading interventions for executive functions in children and adolescents: A systematic review. (EdWorkingPaper: 25-1198). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/x2kr-yh69

School-based language, math, and reading interventions for executive functions in children and adolescents: A systematic review*

Jens Dietrichson[1], Julie Kaas Seerup[1], Sofie Elgaard Iisager Jensen[1], Johan Klejs[1], Elizabeth Bengtsen[1], Martin Williams Strandby[1], Morten Kjær Thomsen[2]

1. VIVE – The Danish Center for Social Science Research, Copenhagen, Denmark
2. Department of Sociology, Leverhulme Centre for Demographic Science, and Nuffield College, University of Oxford, Oxford, UK

2026-01-30

Abstract

Executive functions are a set of cognitive skills and processes used when directing behaviour towards the attainment of a certain goal. A large literature has documented positive associations between executive functions and a variety of desirable outcomes throughout life, including academic achievement. However, training executive functions appears to have limited effects on academic achievement, and the nature of this association remains unclear. We use a systematic review and meta-analysis to examine if preschool and school-based interventions training language, literacy, and/or mathematical skills improve children's and adolescents' executive functions. We include 51 studies in the data synthesis (47 are randomised controlled trials). Using inverse-variance weighted random-effects models, we find a statistically significant weighted average effect size on pre-validated measures of executive functions (0.14, 95% CI = [0.07, 0.22]). The effect is robustly positive in all sensitivity analyses, including tests of publication bias. We also find substantial heterogeneity, which persists in moderator analyses. This means we cannot identify specific types of interventions that are more effective than others in improving executive functions. Our results support theories that emphasise the unidirectional effects from academic skills to executive functions or a bidirectional relation. Further research is needed to identify the mechanisms through which academic skills training affect executive functions.

* The full coding and risk of bias assessment of every study as well as the data used in the analysis and the analytic code (in R) will be made available at osf.io upon publication. Two of the review authors (Dietrichson, Seerup) are co-authors of an included study. None of these two authors were involved in the risk of bias assessment of this study. There are no other conflicts of interest. VIVE - The Danish Center for Social Science Research, Denmark provided funding. Morten Thomsen acknowledges funding from the Leverhulme Trust (Grant RC-2018-003) for the Leverhulme Centre for Demographic Science. We are grateful to the editor Sarah Miller and several anonymous peer reviewers for very helpful comments and suggestions on how to improve our protocol. We thank Bjørn Arleth Viinholt for helping us develop the search strategy and Mikkel H. Vembye for helping out with the risk of bias assessment of one study. We are grateful to Feiyan Chen, Douglas H. Clements, Eder Ricardo Da Silva, Douglas Fuchs, Christopher Lonigan, Irina Kovalcikova, Flavia H. Santos, and H. Lee Swanson for taking the time to answer our questions. The following review team members provided excellent research assistance: Sarah Abigail Bay, Flora Berthelsen, Magnus Spåbæk Bundgaard, Maluhs Haulund Christensen, Klara Steen Henriksen, Lars Julius Norman Holme Jørgensen, Emilie Erreboe Hougaard, Lau Madsen, and Thomas Olsen. Correspondence concerning this article should be addressed to Jens Dietrichson, Herluf Trolles Gade 11, 1052 Copenhagen K, Denmark. E-mail: jsd@vive.dk.

Introduction

Executive functions (EF) are a set of cognitive skills and processes used when directing behaviour towards the attainment of a certain goal (Miyake et al., 2000; Zelazo, 2015). EF is a term referring to multiple advanced cognitive skills and processes, but three core executive functions generally include inhibitory control, working memory, and cognitive flexibility (e.g., Diamond, 2013). Inhibitory control is the control of one's attention, behaviour, impulses, thoughts, and emotions (ibid.). Working memory is the processes involved in the control, regulation, and active maintenance of task-relevant information in the service of complex cognition (Castles et al., 2018). Cognitive flexibility is the ability to change perspective and adjust in the face of changed circumstances (Diamond, 2013; McClelland & Cameron, 2012). It builds on inhibitory control and working memory, and develops later in life than the other two skills (Diamond, 2013; Silva et al., 2022). Thus, there is overlap between EF skills as well as between measures used in their assessment (Miyake et al., 2000). Furthermore, closely related concepts like self-regulation, self-control, emotional control, problem-solving and planning involve the application of all three core EF skills (Allan et al., 2014; Diamond & Ling, 2016; McClelland & Cameron, 2012, 2019; Hernández et al., 2018).

Numerous studies have found positive associations between executive functions and a variety of desirable outcomes throughout life (see Diamond, 2013, for an overview). A large number of reviews have found a substantial association between children's EF and their academic achievements (Blair & Razza, 2007; Emslander & Scherer, 2022; Follmer, 2018; Jacob & Parkinson, 2015; Peng et al., 2016; Peng et al., 2018; Peng & Kievit, 2020). In addition to academic achievement, there are positive associations with, for example, career success (Prince et al., 2007), marriage satisfaction (Eakin et al., 2004), and physical and mental health (Moffitt et al., 2011). Mirroring the associations with desirable outcomes are the negative associations with neurodevelopmental disorders, such as attention-deficit/hyperactivity disorder (ADHD; Nigg et al., 2020), autism spectrum disorder (Zelazo & Carlson, 2020), and learning disabilities (Peng & Swanson, 2022). Therefore, researchers and practitioners across fields have taken an interest in the nature of these associations, and ultimately, how EF might be trained and what the effect of such training might be.

A large number of research interventions as well as commercial products build on the premise that when measures of EF predict desirable outcomes, then training EF skills should improve such outcomes (Simons et al., 2016). However, the assumption of a causal effect from training EF skills to other skills lacks broad-based empirical support. EF training typically improves performance on the trained tasks but the evidence is not compelling that such training substantially improves performance on more distantly related tasks like academic achievement or general cognitive performance (for reviews supporting this statement, see e.g., Cortese et al., 2015; Jacob & Parkinson, 2015; Katz et al., 2018; Melby-Lervåg & Hulme, 2013; Melby-Lervåg et al., 2016; Rapport et al., 2013; Redick et al., 2015; Sala & Gobet, 2017; Schwaighofer et al., 2015; Shipstead et al., 2012; Simons et al., 2016). In a recent meta-analysis, Kassai et al. (2019) failed to find robust, statistically significant evidence that training one component of EF influences other EF skills. This evidence does not necessarily mean that there are never any transfer effects of EF training. Most meta-analyses indicate small, positive and statistically insignificant effects and there are examples of successful EF training interventions (see e.g., Berger et al., 2025; Schunk et al., 2022 for recent large-scale interventions; and Au et al., 2015, and Peng & Miller, 2016, for meta-analyses indicating positive effects for certain types of EF training).

In conclusion, we know that measures of EF and academic achievement are associated, but we lack robust evidence that strengthening EF also improves academic skills. Another hypothesis is that training academic skills improves EF. That is, the association would be explained by a causal effect of academic skills on EF skills. Theoretically, such effects may be expected for at least three reasons: First, neurocognitive skills like EF develop when they are used (Zelazo, 2015), and, for example, learning to read, speak, and understand a language, and performing maths operations involve the use of EF skills (e.g., Castles et al., 2018; Clements et al., 2016; Peng & Kievit, 2020). Second, training academic skills may create new cognitive routines that are useful also for solving EF tasks (Gathercole et al., 2019). Third, learning academic skills increases domain-specific knowledge, which in turn may improve performance on EF tasks (Oberauer et al., 2018).

In this systematic review, we examined the effects of preschool and school-based interventions aimed at enhancing language, literacy, and mathematical skills on the development of children's and students' EF. That is, does training language, literacy, and maths skills improve EF?

The interventions

Interventions of interest to this review aimed, at least in part, at enhancing language, literacy or maths skills of children and adolescents attending pre, primary, secondary, or high-school. Furthermore, interventions were applied, at least in part, in a preschool or school setting and were administered by teachers, teaching assistants, preschool teachers, or the like. For example, solely increasing the amount of homework or the amount of time parents are encouraged to train their children's skills at home did not constitute an eligible intervention. Interventions varied in duration and methodology and did not have to be pre-validated in terms of having an effect on language, literacy, or maths skills.

While we restricted the review in terms of the content of interventions (i.e., to language, literacy, and/or maths), the instructional methods used in the interventions were not restricted. We included for example tutoring, whole-class interventions, and computer-assisted (CAI) interventions that used software programs or apps in the instruction.

Interventions that changed the content rather than the format of instruction, for example by emphasising phonological awareness in early literacy training, were also included as long as the aim was to improve language, literacy, or maths skills. To avoid confounding components, we excluded interventions including components that directly trained EF skills.

Interventions had to be evaluated by at least one measure of EF. As numerous others have noted before us, studies differ substantially in the specific dimensions they include and identify when conducting research on EF (e.g., Garon et al., 2008; Jacob & Parkinson, 2015; Jurado & Rosselli, 2007). We included only measures that were pre-validated on a different sample than the intervention sample but otherwise used a broad definition of EF. This meant that we included measures of inhibitory control, working memory, and cognitive flexibility (and synonyms to these skills), as well as more composite skills, such as self-regulation.

In summary, we included a range of different preschool and school-based interventions aimed at increasing language, literacy, and/or maths skills, but only if the potential effects of such interventions were assessed using pre-validated measures of EF.

How the interventions might work

If training EF skills has limited effects on academic skills, how are the robust associations between measures of EF and academic achievement to be explained?

One hypothesis is that academic skills and EF are both caused by one or more other variables (Brunner et al., 2021; Jacob & Parkinson, 2015; Lawson et al., 2018; Sirin, 2005). One example of such a variable may be parental socioeconomic status (SES). There is abundant evidence that growing up in high-SES families, on average, provides environments more conducive to academic achievement than growing up in low-SES families. These environmental advantages include richer language and literacy environments (Bus et al., 1995; Golinkoff et al., 2019; Hart & Risley, 2003), parents having higher expectations of their children's academic achievement (Bradley & Corwyn, 2002; Slates et al., 2012), and having better access to resources such as high-quality early childhood education, health care, nutrition, and enriching spare-time activities (e.g., Esping-Andersen et al., 2012; Morgan et al., 2012). However, as there are also many low-SES children and youth who thrive in school (Dietrichson et al., 2017), it may be the activities more often carried out in high-SES families, rather than the SES per se, that influence child development (Lawson et al., 2018; Munakata & Michaelson, 2021). Interestingly, activities that high-SES parents typically do more of, such as child-directed speech, reading together with their children, and helping children with their homework are typically aimed at improving language, literacy, and maths. That is, (pre)academic skills–not EF. Nevertheless, these activities might actually work as EF training sessions (Romeo et al., 2022; Valcan et al., 2018).

In line with this notion, a second hypothesis–the one we examined in this review–is that training academic skills improves EF. One reason to expect such effects is that neurocognitive skills like EF develop when they are used; with repeated use, the neural circuits involved in the mental operation become more efficient (Zelazo, 2015). As learning to read, speak, and understand a language, and performing maths operations involve the use of EF skills (e.g., Castles et al., 2018; Clements et al., 2016; Peng & Kievit, 2020), academic interventions may improve EF skills. Furthermore, if more academic training involves more use of EF skills, then longer academic interventions may improve EF skills more than shorter interventions. However, the duration of the intervention may have countervailing effects. A long intervention may for example increase stigma or make children demotivated (see e.g., Dietrichson et al., 2017; Dietrichson et al., 2021; Roberts et al., 2022; Wanzek et al., 2006; and Wanzek et al., 2013, for reviews finding negative, nonlinear, or null associations between duration and effects on academic skills).

Training one set of skills may not be enough to improve another set (James, 1890; Woodworth & Thorndike, 1901). In other words, improving skills on one task may not transfer to other tasks. Taatgen (2013) theorized that transfer between tasks will occur when training one task develops a set of operators, which are also useful for a new task. Similarly, Gathercole et al. (2019) argued that training one set of cognitive skills (in their case working memory) provides benefits to other skills when the training involves learning new cognitive routines that can be applied to novel or not-yet-learned tasks involving the other skill. "Near" transfer to similar tasks is therefore more likely than "far" transfer to dissimilar tasks (see Barnett & Ceci, 2002, for a typology of what constitutes near and far transfer). More training would only be an improvement up to the point where the new routine has been learnt, and would not transfer at all if the routine cannot be applied to the novel task.

Although the mechanisms that produce transfer between cognitive skills are not well-understood (Katz et al., 2018; Simons et al., 2016), there are several candidates for how training language, literacy, and maths skills may transfer to EF skills. Self-directed speech, the "outer speech" used by young children to guide themselves while performing tasks and the "inner speech" used by older children and adults, seems to be important for cognitive functions and the regulation of behaviour (e.g., Luria, 1959; Vygotsky, 1980). Better language skills could improve EF skills by enhancing self-directed speech (Bishop et al., 2014; Weiland et al., 2014). For example, language skills may help children formulate more complex verbal rules that enable the remembering of task sequences and the activation of the relevant task set before operations (Cragg & Nation, 2010; Zelazo & Frye, 1998; Zelazo, 2015), or help children override overlearned responses in favor of a novel response, that is, to self-regulate (Doebel & Zelazo, 2016; Luria, 1959). Doebel and Zelazio (2016) found that 3-year-olds who were exposed to contrasting negations (of the form "not X, Y") scored higher on measures of EF skills than children who were either only exposed to contrasting stimuli, or read storybooks with an adult. Melby-Lervåg and Hulme (2010) found that second graders improved serial and free recall after training phoneme awareness and vocabulary, but not after training rhymes. They argued that improved phonological and semantic memory representations of the words, which rhyme training did not provide, may explain the pattern of results. Doebel (2020) provided a more general example: that training may give children experiences that make them value exercising a skill more. For example, if children note that paying close attention during instruction helped them learn to read, then they might value attention skills higher afterward.

In sum, language, literacy, and maths interventions may affect EF skills by creating new cognitive routines that transfer to EF tasks. If verbal processes play an important part in the development and exercise of self-regulation and other EF skills, then interventions improving language skills may improve EF skills. As literacy training may improve phonological, vocabulary, and comprehension skills (Morrison et al., 2019; Stanovich, 1986), literacy interventions may therefore affect EF skills through similar channels. Learning mathematics involves training in logical and statistical reasoning, which have been shown to transfer to the solution of novel problems (e.g., Simons et al., 2016) and may in general involve the development of new cognitive routines that also can be applied to EF tasks (e.g., Clements et al., 2016; Demetriou et al., 2014).

Successful language, literacy, and maths interventions furthermore increase domain-specific knowledge. Oberauer et al. (2018) reviewed evidence that knowledge from past experience has substantial effects on the performance of working memory tests. Manifestations of this effect included that prior learning improved “chunking” (i.e., combining items into larger sets), that known words were easier to remember than unknown, and that repetition improved performance. Some theories of working memory also emphasize that working memory capacity is jointly determined by two components, the domain-general central executive, and the retrieval of domain-specific knowledge from long-term memory (see e.g., discussion in Peng & Swanson, 2022). If academic interventions are successful, they improve learning and knowledge, which may therefore be a channel through which such interventions may improve working memory, and potentially, other EF skills. As cognitive flexibility builds on working memory skills (Diamond, 2013), improved domain-specific knowledge may improve cognitive flexibility through the same channels.

The mechanisms explaining why academic interventions may affect EF skills do not rule out that training EF skills also improves academic skills. On the contrary, several authors have hypothesized a bidirectional relationship between academic skills and EF skills (e.g., Castles et al., 2018; Clements et al., 2016; Connor, 2016; Peng & Kievit, 2020). A bidirectional, or reciprocal, relationship between academic skills and EF would seem to predict that training EF skills should also improve academic skills but, as mentioned, the causal evidence for this direction is not robust. However, the effects found in previous meta-analyses of EF training programs (primarily working memory training) on academic achievement in children and youth are typically positive and small, not precisely estimated null effects (Cortese et al., 2015; Melby-Lervåg & Hulme, 2013; Melby-Lervåg et al., 2016; Rapport et al., 2013; Sala & Gobet, 2017; Schwaighofer et al., 2015). Thus, these reviews do not rule out small positive effects of training working memory on academic achievement.

Some theories emphasize the unidirectional relationship from academic skills to EF skills, or downplay the possibilities for transfer from training EF skills to academic skills (Demetriou et al., 2014; Gathercole et al., 2019). Gathercole et al. (2019) argued that the new cognitive routines learnt through working memory training programs are unlikely to apply to language, literacy, and maths tasks. As academic skills rely on an extensive array of cognitive routines, they are unlikely "to be trained with anything other than real-life experience" (Gathercole et al., 2019, p. 38). Van der Maas et al. (2006) argued that changing one single variable (i.e., a single EF skill) in a complex system (such as academic skills) may be ineffectual. If the transfer between skills depends on the content of learning, the similarity of the contexts in which that learning is applied, and the interaction between the content and context, then specific content would transfer less often (Simons et al., 2016). If language, literacy, and mathematical learning involve a more complex array of skills and the learning of more cognitive routines than EF training interventions typically have provided, then transfer from academic skills training to EF skills may be more likely than the other way around.

Longitudinal studies examining whether the relationship between EF and academic skills is uni- or bidirectional have for example found unidirectional associations from expressive vocabulary to EF skills (e.g., Jones et al., 2020), from EF skills to receptive vocabulary (Weiland et al., 2014), and from self-regulation to reading achievement (Hernández et al., 2018). Bidirectional associations have been found between reading comprehension and self-regulation (Connor et al., 2016), maths achievement and self-regulation (Hernández et al., 2018), and language skills and EF skills (Romeo et al., 2022). Some studies have only found insignificant associations between the development of EF skills and language skills (Gooch et al., 2016) and others found only reciprocal associations among high-performing students (Zhang & Peng, 2023). However, these longitudinal studies lack the exogenous variation in both EF and academic skills needed to identify a causal relation (Hernández et al., 2018; Jones et al., 2020). The interventions we examined provided exogenous variation in one direction, that is, evidence of the effects of training language, literacy, and maths skills on EF skills.

As mentioned, the content and context of training may be important for the transfer between skills (Barnett & Ceci, 2002). Clements et al. (2016) argued that the relationship between mathematics and EF is stronger than between literacy and EF. In a similar vein, Hernández et al. (2018) hypothesized that EF skills may be most useful when applied to novel situations, and maths, at least in primary and secondary school, may be less automatised than language and literacy processes. As language is not automatised when children are very young, and many literacy skills are not automatised until children are fluent readers, this hypothesis would also suggest age-dependent effects.

The neural basis of EF skills is another reason to expect heterogeneity of effects across age. Although EF skills develop throughout adolescence (Diamond, 2013), the plasticity of the nervous system declines with age (e.g., Zelazo, 2015). Short-term memory processes are essential for early skill acquisition but less important once the cognitive processes behind a skill have been automated (Van Der Maas et al., 2006). Early interventions may therefore have larger effects on EF skills (Ahmed et al., 2021). This discussion also suggests interaction effects between the type of intervention and age. For example, language interventions may have the largest effects on the EF skills of very young children and literacy interventions around the start of primary school when most children acquire basic decoding skills.

In their review, Peng and Kievit (2020) found that reading and mathematics skills predict cognitive skills and vice versa, but that this bidirectional relationship was weaker for disadvantaged students. They hypothesized that the differences in learning experiences and opportunities between advantaged and disadvantaged children explain the weaker bidirectionality. That is, advantaged children or not-at-risk children, including high-SES children, who start out with stronger cognitive and academic skills may be more likely to trigger and benefit from cognitive-academic bidirectionality. A similar mechanism may be in play regarding the association between EF skills and language, literacy, and mathematics skills.

Study characteristics may also influence the effects (e.g., Cheung & Slavin, 2016). A potential moderator is the type of control group. Using an active control group that performs similar activities as the treatment group in all aspects but the “working ingredient” (i.e., a placebo condition) may be advantageous when the aim of a study is to pinpoint the mechanism through which an intervention affects a skill. Because a placebo condition may shut down other possible mechanisms, such as changed expectations and motivations, it increases the chances of isolating the hypothesized mechanism (Simons et al., 2016). However, in field experiments in preschools and schools, using a placebo condition is not necessarily an advantage. Education interventions, whether intentionally or not, may improve academic and EF skills precisely through the changes in motivation and expectations that using a placebo condition aims to preclude (Diamond, 2014; Bøg et al., 2021). Closing down these mechanisms by using a placebo condition may thus change what is being estimated.

Treatment as usual (TAU) control groups may be more at risk of Hawthorne and John Henry-effects: that is, the treatment and control groups behave differently because they know that they are participating in a study (Glennerster & Takavarasha, 2013). While such effects are difficult to completely avoid in education interventions, placebo control groups also know that they are participating in a study and, if the placebo treatment works well, believe that they participate on equal terms with the treatment group. In such instances, Hawthorne and John Henry-effects may therefore be mitigated.

Further examples of study characteristics that may moderate effect sizes include the study design – in our case, whether the study was a randomised controlled trial (RCT), quasi-randomised controlled trial (QRCT), or a quasi-experimental study (QES) – the type of measure (whether the children are measured directly, or assessed by someone who knows them well, like a teacher or a parent), and measurement timing. The direction of the influence on effect sizes of these moderators is theoretically ambiguous with the exception of measurement timing: longer follow-ups are typically associated with smaller effect sizes (e.g., Bailey et al., 2020; Dietrichson et al., 2025; Hart et al., 2024).

In sum, language, literacy, and maths interventions may improve EF skills for at least three reasons: first, because learning to read, speak, and understand a language, and performing maths operations involve the use of EF skills; second, because training language, literacy, and maths create new cognitive routines that are useful also for solving EF tasks; and third, because the interventions improve domain-specific knowledge. As discussed in this section, there were reasons to expect effect sizes to be moderated by the subject of the intervention, the type of EF skill measured, child age or grade, the at-risk status of participants, as well as study characteristics. We examined if these variables explained effect size heterogeneity in a confirmatory moderator analysis, and examined additional variables in exploratory analyses.

Why it is important to do this review

Education spending constitute a large proportion of total government spending in many countries around the globe (OECD, 2020). EF skills are fundamental cognitive skills underlying all forms of goal-directed behaviour (Miyake et al., 2000; Zelazo, 2015). With previous research identifying a strong association between EF and academic achievements (e.g., Jacob & Parkinson, 2015), educational researchers and policy makers have been right to take an interest in the training of these important cognitive skills within a school setting. However, researchers are yet to understand the nature of this association, and how it can be fully utilised for the benefit of the education system and ultimately its students.

The number of previous reviews of the effects of academic interventions on EF is small. Peng and Kievit (2020) reviewed evidence of a bidirectional relationship between academic and more general cognitive skills (including EF). Their results suggest that reading and mathematics skills predict cognitive skills and vice versa, that this bidirectional relationship is weaker for disadvantaged students, and that direct academic instruction can improve cognitive skills. The review did not include a meta-analysis. Their results provided motivation for including moderator analyses testing whether interventions have different effects depending on whether they mainly target at-risk students or not.

In a narrative review of interventions in preschool and early primary school, Clements et al. (2016) found more evidence of an association between executive functions and maths achievement than between executive functions and literacy or language achievement. Furthermore, they cited studies showing reciprocal associations between early numeracy and EF, but not between early literacy and EF. These findings motivated us to examine whether the content of the interventions moderate effect sizes.

A large number of recent reviews have examined the effects of physical activity interventions on EF skills (e.g., Alvarez-Bueno et al., 2017; Jylänki et al., 2022; Mavilidi et al., 2022; Muir et al., 2023; Norris et al., 2020; Peiris et al., 2022; Ruhland & Lange, 2021; Vasilopoulous et al., 2023). None of these reviews reported separate meta-analytic results for interventions combining physical activity with academic instruction on EF outcomes. Peiris et al. (2022) found no significant effect of physical activity breaks during lessons on a broader category of cognitive outcomes (which included EF), but included relatively few studies and it was not clear how much the interventions trained language, literacy, and math. In addition to physical activity interventions, Muir et al. (2023) included four math interventions, some of which trained EF directly, but did not meta-analyse these. Wang (2020) reviewed abacus training interventions and reported positive effects on EF, but did not conduct a meta-analysis. Our review included abacus training interventions and many other types of math interventions.

Less closely related, Ritchie and Tucker-Drob (2018) presented meta-analytic evidence that education can improve cognitive skills. Their meta-analysis found that an additional year of schooling increases IQ with 1 to 5 points. Ritchie and Tucker-Drob (2018) did not examine language, literacy or maths programs, or EF measures, and did not include preschool children. Stockard et al. (2018) reviewed Direct Instruction interventions for school-age children and found positive effects on IQ and cognitive skills measures. These measures were however not further defined in the review and it is unclear if any studies used similar measures as we included in our review. Reviews of targeted and universal preschool programs have found effects on, typically broad, cognitive skills measures but have not conducted analyses of EF skills (e.g., Dietrichson, Kristiansen, & Viinholt, 2020; Duncan et al., 2023; Duncan & Magnusson, 2013; van Huizen & Plantenga, 2018).

We are not aware of a previous meta-analysis examining the effects of language, literacy, and math interventions on EF skills. This review contributes with 1) a reversed perspective on the association between EF and academic achievement from most earlier reviews; 2) a thorough and comprehensive search, screening, and coding process, and a risk of bias assessment of included outcomes; 3) an examination of intervention studies only, estimating the effect of training specific academic skills on the development of EF; 4) inclusion of several types of EF as well as both language, literacy, and maths interventions; and 5) a meta-analysis on the above mentioned hypothesised relation between academic interventions and EF.

Objectives

Our main research question for this systematic review was: Do preschool and school-based interventions aimed at improving language, literacy, and/or mathematical skills increase children's and adolescents' executive functions?

As a secondary objective, we examined the following research question: How are the effects of language, literacy, and mathematics interventions on executive functions moderated by the subject of the intervention, child age or grade, the type of executive function measured, and the at-risk status of participants?

We pre-specified moderators regarding academic subject, child age, the type of executive function measured, and the at-risk status of participants, which corresponded to our confirmatory moderator analysis (see the *Subgroup analysis and investigation of* *heterogeneity*-section for definitions of these variables). In exploratory moderator analyses, we examined the association between effect sizes and other study characteristics. That is, our third research question was: How are the effects of language, literacy, and mathematics interventions on executive functions moderated by study characteristics?

Lastly, our fourth research question was: What are the effects of the included interventions on secondary outcomes (language, literacy, mathematical and other cognitive skills)?

The fourth research question was motivated by the risk that the included interventions may be ineffective regarding their primary aim, that is, to improve language, literacy, and/or mathematical skills. Specifically, if we were to find no effects on EF, ineffective interventions might explain the lack of improvement. Examining the effects on language, literacy, and maths skills might also tell us something about the relationship between these skills and EF. For example, if we were to find effects on executive functions despite observing no effects on language, literacy, and mathematical skills, such results would suggest that the effects on executive functions were less likely caused by interventions creating new cognitive routines useful for solving EF tasks or by improved domain-specific knowledge. Instead, the effects on executive functions would, in that case, be more likely attributed to the interventions directly involving the use of EF in the training of language, literacy, and maths. To examine this last hypothesis further, we also conducted a non-pre-registered, exploratory analyses in which we estimated the association between average effect sizes based on secondary outcomes with effect sizes based on EF tests.

As we required studies to have reported measures of executive functions, it is important to note that the interventions included in this review are unlikely to be representative of language, literacy, and maths interventions in general.

Methods

We followed the criteria for considering studies, the search methods for identification of studies, and the procedures for data collection and analysis described in our pre-registered protocol (Dietrichson et al., 2022). We comment on deviations from the protocol throughout and summarize them in the *Differences between protocol and review*-section.

Criteria for considering studies for this review

Types of studies

We included quantitative and experimental primary studies that examined the effects of preschool and school-based interventions. Eligible studies used a treatment-control group design, such as RCTs, in which the assignment to treatment was determined by a random sequence, QRCTs, in which the assignment to treatment was determined by means such as alternate allocation, person's birth date, the date of the week or month, case number, or alphabetical order, and QESs, in which the assignment to treatment occurred, for example, in the course of usual decisions, by a (non-random) researcher decision, or by a natural experiment (i.e., through some form of “natural” or administrative process, which is outside the researchers’ control).

Treatment-control studies needed to assign at least two ‘units’ (e.g., schools, classes, or children) to the treatment group and two units to the control group to be included. Treatment effects are difficult to separate from unit effects in studies with only one unit in either the treatment group or the control group. Effect sizes also had to satisfy specific risk of bias criteria before contributing to the data synthesis (for these criteria, see the *Assessment of risk of bias in included studies*-section). Studies in which all effect sizes were excluded from the data synthesis due to risk of bias criteria were still included in the review.

Control groups were defined as TAU conditions (including waiting list control groups), or placebo interventions. We excluded studies that only compared groups receiving different interventions, which were all hypothesized to improve academic or EF skills. In some studies that included an active control group and a TAU control group, the active control group received language, literacy, or maths training. We then included the contrast between the active and the TAU control group.

We excluded non-intervention studies, such as observational or descriptive studies, and qualitative study designs, as well as single-subject before-after designs, in which participants act as their own control group. Other reviews were not included in this synthesis, although we kept track of relevant reviews found in our search and used them where appropriate, for example, for citation tracking purposes (see the *Citation tracking*-section below for more information).

Only studies published in English, German, Danish, Swedish and Norwegian were eligible, due to language restrictions in the review team.

Types of participants

The eligible population samples for this review were children and adolescents attending either pre, primary, or secondary school (including high school). We included both normally achieving, not-at-risk students as well as those identified as at risk of academic difficulties because they were low-performing or educationally disadvantaged. Furthermore, we included clinical samples irrespective of the diagnosis. We did not restrict the type of preschools and schools, i.e., state, private, public, and boarding schools were all eligible for inclusion. Students could also attend both mainstream schools or special education schools.

We excluded interventions conducted in higher education (e.g., universities or professional development programs). We also excluded interventions conducted outside of the school year or school day, that is, summer schools and after school programs were not included, unless the intervention had a vital component embedded in the normal school day setting (we found no example of the latter).

Types of interventions

The review included primary studies examining language, literacy, and mathematical interventions implemented in preschool and school settings. Interventions varied in duration and methodology and did not have to be pre-validated in terms of having an effect on language, literacy, or maths skills. However, interventions had to constitute a condition different from TAU at the school or preschool. As TAU instruction might include components that improve EF skills without training language, literacy, or maths skills, including such conditions would risk introducing a confounding element into the analysis.

As an example of what we mean by being different from TAU, Araujo et al. (2016) examined the effects of teacher and classroom quality on, amongst others, measures of EF. They used an assignment rule where children were assigned to classrooms in a manner that was close to random. Although this intervention trained many of the skills we were interested in, no child received a different instruction than TAU in their school because of the intervention. The only aspect that differed was the “quality” of the teacher and peers in the classroom, and this study was therefore not included.

Interventions were implemented in a school or preschool setting and had to be administered by teachers, teaching assistants, or the like (including personnel hired by the researchers). Thus, interventions that solely increased the amount of homework or encouraged parents to train with their children at home or in other types of family care were not eligible.

Interventions were aimed at enhancing students’ academic achievement by training at least one specific mathematical, literacy or language skill. Interventions with an explicit focus on only improving students’ metacognitive strategies or study skills were not eligible for this review. In order to avoid confounding components, interventions with components that directly trained EF were not eligible for this review. It was the content of the training that determined whether an intervention directly trained EF, not the stated aims. For example, an intervention could aim to improve EF through early literacy training and be included, as long as there was no component that trained EF without involving early literacy training. Interventions that included multiple components all targeting specific academic skills were eligible (as long as none of the components targeted EF skills directly). Several included studies included multiple treatments, some of which were excluded because they trained EF directly.

A difficult case were interventions that included physical activity and interventions that included or targeted certain types of social-emotional skills. Intense physical activity like exercise is often thought to improve cognitive skills, especially in the long run (e.g., Donnelly et al., 2016). However, physical activity interventions involving language, literacy, or maths training are often relatively short, and there is an ongoing debate on how physical activity influences executive functions (e.g., Beck et al., 2016; Mavilidi et al., 2020). For example, Diamond and Ling (2016) argued that exercise only improves EF when it also involves training of EF skills whereas others have argued that it is improved cardiorespiratory fitness that in turn activates the areas of the brain that support EF (Hillman et al., 2019). As it did not seem certain that physical activity interventions affect EF through the physical activity but may, if there is any effect at all, do so through the training of language, literacy, and maths skills, we decided to include physical activity interventions. We present sensitivity analyses excluding these interventions in the *Results of the sensitivity analysis*-section.

Types of outcome measures

We included two types of outcome measures: measures of EF (our primary outcome) and measures of academic achievement (our secondary outcome). Studies had to include at least one primary outcome to be included.

Primary outcomes

The primary outcomes of this review were measures of EF skills. We included both direct measures obtained by use of cognitive tests, such as the Delis-Kaplan Executive Function System (Delis et al., 2004), Wechsler Intelligence Scale for Children (WISC; e.g., Wechsler, 2014), and the Automated Working Memory Assessment (Alloway et al., 2008), as well as indirect measures. Indirect measures are measures where for example teachers or parents rate the children. Examples included the Behaviour Rating Inventory of Executive Function (e.g., Gioia et al., 1996), the Behavioral and Emotional Rating Scale (Epstein & Sharma, 1998), and the Strengths and Weaknesses of ADHD Symptoms and Normal Behavior (Swanson, Schuck et al., 2012). We included measures of EF in a broad sense, meaning that both measures of working memory, inhibitory control, and cognitive flexibility were included, as well as measures of broader skill constructs like self-regulation, self-control, effortful control, planning, and problem solving that plausibly include the application of all three core EF skills (see the search strategy in Online Appendix A for examples of synonyms to these concepts).

Only studies applying pre-validated measures of EF were eligible for inclusion. That is, the measures had to be validated on other samples and not developed by researchers for the intervention at hand. Researcher-developed measures risk inflating effect sizes (e.g., Slavin & Madden, 2011). For a study to be eligible, we had to be able to confirm that the applied (sub)test of EF was in fact pre-validated. If a study applied a modified version of a pre-validated test, the study was also ineligible. This restriction was motivated by the fact that EF is a broad "umbrella term" referring to a number of different constructs that have changed over time. Earlier studies sometimes use the term executive functions to refer to constructs that we do not consider relevant today. Other reviews, such as Jacob and Parkinson (2015), handled this issue by simply excluding all papers published prior to 2000.

Working memory is a complex concept referring to a series of coherent brain functions related to memory. However, as with executive functions, the term is applied dissimilarly by researchers, and especially the distinction between "working memory" and "short-term memory" is disputed, as the two functions are closely related and somewhat overlapping. One example is tests of forward digit span, which are labelled both as tests of short-term memory and working memory in the literature (compare e.g., Boat et al., 2022; Ramani et al., 2017; Weiland & Yoshikawa, 2013). To account for the at times different use of "working memory", we also included studies using measures of short-term memory as the outcome measure.

Secondary outcomes

Secondary outcomes included measures of academic achievement such as standardised measures of specific numeracy, literacy, and oral language skills. We also coded additional cognitive measures, such as other scales than the working memory index from WISC (Wechsler, 2014). Our protocol stipulated that secondary outcomes had to be standardised in order to be included; however, a handful of studies reported exclusively non-standardised tests of the skills targeted in the intervention and in those cases, we chose to code the non-standardised measures as well. We tested whether the results were sensitive to this deviation from the protocol.

Duration of follow-up

We included both end-of-intervention tests and follow-up tests conducted after the end-of-intervention, regardless of the duration of follow-up. We describe how we analysed tests with different duration of follow-up in the *Data synthesis*-section.

Search methods for identification of studies

Relevant studies were identified through searches in electronic databases, hand search in specific targeted journals, governmental and grey literature repositories, internet search engines, citation tracking, and contact to international experts.

Electronic database searches

We searched the following databases up to April 2023 (platform in parenthesis): ERIC (EBSCO), PsycINFO (EBSCO), SocINDEX (EBSCO), Academic Search Premier (EBSCO), International Bibliography of the Social Sciences (ProQuest), Sociological Abstracts (ProQuest), Science Citation Index Expanded (Web of Science, Clarivate), Social Sciences Citation Index (Web of Science, Clarivate).

Description of search-string

The search string was based on the PICOS-model. Using that model, we identified five aspects of the topic and developed a search facet for each with relevant terms and synonyms. We did not apply a time or language limitation on the database searches. All of the five facets were searched as a title/abstract search. Some of the facets also utilized the subject terms, which varied according to each database thesaurus or index where applicable (the Web of Science-databases do not have a thesaurus). We included the full search documentation for all searches in Online Appendix A.

Searching other resources

This section describes our searches of other resources. We included documentation for these searches in Online Appendix A.

Hand search. We hand searched journals at the intersection between education and psychology, as we believed that they were most likely to include studies related to our review topic. The chosen journals furthermore focus on different parts of the included age range of children. We conducted hand searches of the following journals from late 2019 to late 2024:

Journal of Educational Psychology, Child Development, Contemporary Educational Psychology, Early Childhood Research Quarterly, and American Educational Research Journal.

Searches for unpublished literature in general. We split the search strategies in sub-sections for each type of unpublished literature. In general, most of the resources searched for this purpose included multiple types of literature and references. As an example, the resources listed to identify reports from national bibliographical resources also included working papers and dissertations, in addition to peer-reviewed references. A resource may therefore have been searched for multiple purposes, but for brevity, it is only listed once as a resource.

Search for dissertations. We searched EBSCO Open Dissertations (through EBSCO-host) for dissertations. Our protocol specified that we would also search ProQuest Dissertations & Theses Global. However, at the time of the search, we no longer had access to this database.

Search for working papers/conference proceedings. We searched the following resources for working papers/conference proceedings:

- European Educational Research Association (https://eera-ecer.de/)
- American Educational Research Association (https://www.aera.net/)
- PsyArXiv (https://osf.io/preprints/psyarxiv)
- Open Grey (http://www.opengrey.eu/)
- Google Scholar (https://scholar.google.com/)
- Google searches (https://www.google.com/)
- Social Science Research Network (https://www.ssrn.com/index.cfm/en/[1]).

Search for reports and non-US literature. We searched the following resources for reports and non-US literature:

- Research Portal Denmark (https://local.forskningsportal.dk/search/78730)
- SwePub (https://swepub.kb.se/form_extended.jsp)
- NORA (http://nora.openaccess.no/)
- CORE (https://core.ac.uk/)
- Best Evidence Encyclopedia (https://bestevidence.org/)

Search for systematic reviews. Reviews were identified in all previous databases and were searched for leads when we deemed them relevant. We also searched specifically for systematic reviews using the following resources**:**

- Campbell Systematic Reviews (https://onlinelibrary.wiley.com/search/advanced?publication=18911803&text1=)
- Cochrane Library (https://www.cochranelibrary.com/advanced-search)
- Centre for Reviews and Dissemination Database (https://www.crd.york.ac.uk/CRDWeb/)

Citation tracking. We used citation tracking methods to identify more relevant literature. We citation tracked all 79 included records and all relevant reviews (e.g., those mentioned in section *Why it is important to do this review* and those found in other parts of our search, in total 19 reviews) forward by using Google Scholar and backward by screening the reference lists. The latter was done by at least two members of the review team for each record. We furthermore citation tracked a number of primary records that were excluded but where we thought the content was close enough to our review that it might contain relevant leads.

Contact to experts. At the end of the search process (in December 2024), we contacted five international experts to identify unpublished and ongoing studies and studies we might have missed. We contacted authors that, for example, had written relevant reviews or had more than one included primary study and provided them with the inclusion criteria for the review along with the list of included studies. We received a reply from four authors.

Data collection and analysis

Selection of studies

The screening process for identifying relevant studies from the database searches were split in two overall stages: 1) screening based on title and abstract, and 2) screening based on full text. In order to ensure the quality of the screening process and reduce potential errors, we made use of independent double screening at both stages (Polanin et al., 2019; Stoll et al., 2019). The screeners were blinded to each other's work until comparing final judgements of the screened records. If the two screeners did not agree on the inclusion/exclusion of a specific record, then this reference was sent to one of the review authors for final judgement.

We conducted pilot screenings for each overall screening stage and for each screener. In the pilot screening based on title and abstract, the review team screened and compared 100 references. The team then discussed and resolved potential disagreements and uncertainties regarding the eligibility criteria. If the interrater agreement was above 90% in the pilot screening, then the rest of the references was screened. If the interrater agreement was below 90% in the first pilot, the review team members performed a second pilot screening in order to ensure sufficient reliability before the rest of the references were screened. At the full text stage of the screening process, the pilot consisted of 8-10 studies and was otherwise identical to the process described for first level. The review team met with regular intervals in all stages of the screening process in order to discuss uncertainties and minimize the potential for "coders drift" (Polanin et al., 2019). During the screening process, none of the review authors or review team members were blind to the authors, journals, or institutions responsible for the publication of eligible studies.

We used the machine learning (ML) functionality in EPPI Reviewer 4 to conduct priority screening in the title and abstract screening phase. We first screened 1,000 records on title and abstract using independent double screening. We then ranked the remaining records by the ML algorithm's probability of a record being included, and screened those with the highest probability first. We screened in batches and re-ranked the remaining records regularly.

Using priority screening had a dual purpose. First, we found relevant records earlier in the screening process, which meant that we could start the coding and the risk of bias assessment earlier and thereby speed up the completion of the review. Second, we found a large number of records in the database search (18,729). As stated in our protocol, we switched to one human screener when the inclusion rate became low. In the batch of records from 6,205 to 7,457, the inclusion rate was 1.3%. We deemed this to be low enough to start screening with one human screener because it was much lower than in previous batches (maximum inclusion rate was 8.1%) and because we had access to *AIscreenR* as the second screener (Vembye et al., 2025; Vembye & Olsen, 2025).

AIscreenR is an R package that uses OpenAI's GPT API models to conduct screening on title and abstract. Vembye et al. (2025) used a large number of systematic reviews to compare *AIscreenR*'s performance to human screeners. The results indicated that *AIscreenR* always performed on par with one human screener, and sometimes virtually as well as two human screeners. As the screening on title and abstract was comparatively difficult in our review, we developed and tested a prompt that was highly inclusive in the sense that recall was nearly perfect whereas specificity was lower (i.e., *AIscreenR* included nearly all records included by two human screeners, and many more). We also checked carefully that *AIscreenR* did not miss any records that, at that point, had been included after full text screening.

Of the 11,036 records for which we used one human screener and *AIscreenR*, 108 were included by the human screener. All these records were screened in full text (regardless of *AIscreenR*'s verdict). All records included in human single-screening and by *AIscreenR* were also screened in full text (31). All records excluded by the human screener but included by *AIscreenR* (846) were screened by the first author as the third screener on title and abstract. We similarly used *AIscreenR* as the second screener in the screening of the dissertations from EBSCO Open Dissertations.

AIscreenR was developed after our protocol was written but this deviation only served to improve the screening process (i.e., the protocol stated that we would single-screen once the inclusion rate was low enough, which surely would have been more error prone).

Data extraction and management

Two members of the review team extracted and coded data from the included studies. Any disagreements were resolved by discussion, or by involving a third coder. From all included studies, we extracted data on publication characteristics, study characteristics, participant characteristics, intervention characteristics, and outcome characteristics. If any disagreement or uncertainty emerged during the data extraction process which could not be resolved by the two study coders, a third reviewer (another of the review authors) with the appropriate expertise was consulted. However, we did not double-code studies that we could not use in the meta-analysis. We used EPPI Reviewer 4 and Microsoft Excel for data management and extraction.

Assessment of risk of bias in included studies

Two members of the review team, at least one of which was a review author, independently assessed the risk of bias for each eligible study outcome. If disagreements could not be solved by discussion, another review author was contacted for a final agreement.

For included non-randomised studies (QRCT and QES), we assessed the risk of bias for all included outcomes applying Cochrane's ROBINS-I tool (Sterne et al., 2016). For all included randomised studies, we assessed the risk of bias using a revised version of

Cochrane's risk of bias tool, ROB-2 (Elridge et al., 2016; Sterne et al., 2019). See our protocol (Dietrichson et al., 2022) for more information about the characteristics of each tool. In this section, we describe how we applied the tools.

In ROBINS-I, we rated every outcome on each domain as either having a 'low', 'moderate', 'serious', or 'critical' level of bias. If a study outcome received a 'critical' rating on at least one domain, it was considered too biased to provide useful evidence on the effects of the intervention. As a consequence, the outcome was excluded from the data synthesis. We did not continue the risk of bias assessment of an outcome measure if a domain was rated 'critical'. In each domain of the ROB-2 tool, we rated every outcome measure as either having 'low', 'some', or 'high' concerns. In both tools, we made an overall rating on the basis of the domain ratings, which was equal to the rating in the domain with the highest risk.

Both tools allow for an assessment of the overall bias direction for the assessed outcomes. This was typically very difficult to assess, and we do not discuss this direction further in the review. A further commonality is that both tools required pre-specification of the effect type that was assessed. We were interested in estimates that were closer to the effect of starting and adhering to the intervention (treatment on the treated, TOT) than the effect of assignment to the intervention (intention to treat, ITT). However, almost no study provided an appropriately estimated TOT estimate. Analyses were typically made on samples with attrition but where most or all remaining participants received the assigned treatment. Estimates were therefore typically neither pure ITTs nor pure TOTs, but these two estimands could often be expected to be quite close to one another in the included studies.

In the case of RCTs, where we found evidence that the randomisation had gone wrong or were no longer valid, we assessed the risk of bias of the outcome measures using ROBINS-I instead of ROB-2 (as recommended in Higgins et al., 2019). There was only one such example where we deemed the attrition to be so high for two outcomes that the randomisation was no longer valid. That is, the random assignment was, on its own, unlikely to produce unbiased estimates of the intervention effects.

Definition of critical confounders. ROBINS-I calls for a definition of critical confounders. In this review, we defined performance at baseline and age to be our critical confounders that deserved most attention in our risk of bias assessments. However, we also considered other important confounders such as students' socioeconomic status and gender as well as the risk of bias from unobservable confounders (i.e., defining critical confounders did not imply that other confounders were not considered).

Measures of treatment effect

We compared the intervention condition with the control condition on measures of EF, and on measures of academic achievement. All included studies used continuous outcome measures. We calculated the standardised mean difference (SMD) where possible, since the outcomes were measured and reported with a wide range of different scales. In order to correct for upward bias in small samples, we used the small sample bias-corrected Hedges' g in our analysis (Borenstein et al., 2009; Hedges, 1981; Lipsey & Wilson, 2001). We calculated g and its standard error (SE) as (Lipsey & Wilson, 2001, p. 47-49):

$$ES_g = \left(1 - \frac{3}{4N - 9}\right) \times \left(\frac{X_1 - X_2}{s_p}\right)$$

$$SE_g = \sqrt{\frac{N}{n_1 n_2} + \frac{{ES_g}^2}{2N}}$$

where $N = n_1 + n_2$ is the total sample size, X_1 - X_2 is an estimate of the mean difference between the treatment and control group, and s_p is the pooled standard deviation (SD) defined as:

$$S_p = \sqrt{\frac{(n_1 - 1){s_1}^2 + (n_2 - 1){s_2}^2}{(n_1 + n_2 - 2)}}$$

Here, s_1 and s_2 denote the unadjusted SDs of the treatment and control group. As recommended by Hedges et al. (2023), we included the small-sample adjustment when calculating effect sizes but not when calculating the standard errors/variances of the effect sizes.

In our primary analysis, we used the post-test SDs. Post-test SDs were reported more often than pre-test SDs and potentially also less subject to floor effects. In two studies, we could only use an SMD calculated with the control group SD (i.e., a Glass's delta). We included these studies in the analysis, as excluding them seemed to pose a greater risk of bias.

We used covariate-adjusted estimates of the mean difference whenever available (e.g., coefficients from linear regressions, adjusted means). Our protocol stated that we would use TOT estimates if available and test sensitivity to the inclusion of ITT estimates. There were so few studies with both ITTs and TOTs that the planned sensitivity analysis was subsumed into other sensitivity analyses (see section *Sensitivity analyses*).

Lastly, some studies reported only results for subgroups that were on a lower level than in other studies (and often not on the level of treatment assignment). Because effect sizes based on sub-groups omits the between-subgroup variation, standard deviations will be smaller and effect sizes mechanically larger. We therefore used the methods recommended by Wilson (2017) to aggregate the subgroup results to a level comparable to our other included studies.

Unit of analysis issues

Effect sizes and standard errors from studies with clustered assignment of treatment may be biased if the unit-of-analysis is the individual and an appropriate cluster adjustment is not used (Higgins & Green, 2011). We adjusted effect sizes by study using the methods suggested by Hedges (2007), Hedges et al. (2023), and What Works Clearinghouse (2021) and information about intra-cluster correlation coefficient (ICC) and realised cluster sizes. As for individually assigned treatments, we included a small-sample adjustment in the calculation of effect sizes but not in the calculation of standard errors/variances (Hedges et al., 2023).

When information about the ICC was not available, we adjusted the effect sizes using an ICC of 0.11. This value is approximately equal to the average of ICCs taken over grades from kindergarten to grade 12 and maths and reading tests in Hedges and Hedberg (2007; reported in Table 2 and 3, pp. 68–69, models with covariates). We tested whether the results were sensitive to this choice by using ICCs of 0 (the theoretical minimum) and 0.32 (the empirical maximum in the same two tables). We were not aware of similar evidence of typical ICCs for studies conducted in preschool, and used these ICCs also for preschool studies. We calculated the average cluster size by dividing the total sample size in a study by the number of clusters (classrooms or (pre)schools).

Another issue was studies that assigned treatment individually but where the treatment was received in groups. Adjusting for this issue was for several reasons much more challenging: First, it is not clear how to achieve an unbiased adjustment. For example, Weiss et al. (2016) indicate that both adjusting and not adjusting are likely to yield biased standard errors in primary studies (over- and underestimated, respectively). Second, information about group sizes was often imprecise or reported as intervals, and sometimes missing. Third, no study provided information about the relevant ICC for groups smaller than regular class sizes and we are not aware of estimates similar to Hedges and Hedberg's (2007) for small-to-medium group interventions.

For these reasons, and because our protocol did not specify an adjustment procedure for individually assigned group treatments, we used the following procedure. For treatments received in classes or larger groups in which the control group was also likely clustered in similar units, we adjusted the estimates using the same procedure outlined earlier in this section and included the adjusted effect sizes and variances in our primary analysis. For treatments received in groups smaller than classes in which the control group likely was not clustered in similar units, we included the unadjusted effect sizes and variances in our primary analysis. We tested the sensitivity to this choice by calculating a design effect equal to 1 + (group size – 1) × ICC (Higgins et al., 2019), and used this design effect to adjust the treatment group's sample size. As the ICC seems likely to be higher the smaller the group, we choose the empirical maximum from Hedges and Hedberg (2007; i.e., 0.32). We used the mid-point of intervals to estimate group sizes and imputed the mean across studies for studies with missing information. We then used this effective sample size in the calculation of variances.

Criteria for determination of independent findings

In a few cases, several records used the same sample of data. The records either reported estimates for different outcomes or exactly the same estimates for the same outcomes. We included only one estimate of the same outcome from each sample of data in the meta-analysis in order to avoid duplication.

Studies reported multiple and dependent effect sizes, sometimes of conceptually similar outcomes. We included all such outcomes in our analysis, and we describe how we took the dependency between effect sizes into account in the *Data synthesis*-section.

Dealing with missing data

Studies had to permit calculation of a numeric effect size for the outcomes to be eligible for inclusion in the meta-analysis. When studies had missing data, such as missing standard deviations or means, we tried to derive these from, for example, *F*, chi-squared, and *t*-statistics, and correlation coefficients using the methods suggested by Lipsey and Wilson (2001). When that was not possible, we contacted the corresponding author of the study.

We assessed missing data and attrition for the individual studies with ROB-2 and ROBINS-I, which both have specific domains focusing on biases arising from missing data (Sterne et al., 2016; Sterne et al., 2019).

Assessment of heterogeneity

Heterogeneity can stem from variation in effects within and between studies or from sampling errors in included studies. In this review, we assumed that within and between-study variation in effects would occur and therefore used a random-effects model in our main analysis. We assessed the degree of heterogeneity with the Q and the I^2 statistics, and the within (ω), between (τ), and total standard deviation ($\sigma = \sqrt{(\omega^2 + \tau^2)}$; Higgins et al., 2003; Pustejovsky, 2021). We calculated an overall I^2 statistic by summing the between and within-study variance components (see e.g., Viechtbauer, 2024). We also reported prediction intervals in order to examine how effects are dispersed. We calculated prediction intervals using the predict-function in metafor (see Viechtbauer, 2025).

Assessment of reporting biases

Reporting bias might refer to both publication bias and selective reporting of outcome data and results. We assessed bias from selective reporting of outcome data and results with ROB-2 and ROBINS-I.

To assess the extent of publication bias, we showed funnel plots and examined whether they were asymmetric (Higgins & Green, 2011). To formally test for asymmetry, we used a version of Egger’s test (Egger et al.,1997) suggested by Rodgers and Pustejovsky (2021) and further developed by Chen and Pustejovsky (2024). This version has lower Type I errors than the original version and is suited for dependent effect sizes and the meta-analysis models we used. Egger's test examines asymmetry by including a measure of effect size precision as a predictor in a meta-regression with effect sizes as the outcome variable. A significant coefficient on the precision measure is interpreted as evidence of asymmetry.

Following Rodgers and Pustejovsky (2021), we interpreted the rejection of the null hypothesis of no asymmetry in a one-sided test with significance level 0.05 as an indication of asymmetry.

Asymmetric funnel plots are not necessarily caused by publication bias (and publication bias does not necessarily cause asymmetry in a funnel plot). We therefore also tested how sensitive our results were to potential publication bias using the methods developed by Mathur and VanderWeele (2020).

Data synthesis

We conducted the data synthesis in the following steps: First, we presented descriptive summaries of the contextual, methodological, and outcome characteristics for the studies included in the data synthesis. Second, our main effects analysis reported a weighted average effect size comparing the results on all EF skills tests of children in the intervention groups with children in the control groups (corresponding to our first research question).

Pooling all EF tests may obscure heterogeneous effects across EF skills, but we think the pooled analysis is motivated by the substantial correlations and the conceptual overlap between EF skills (Diamond & Ling, 2016; Kassai et al., 2019; Miyake et al., 2000; Silva et al., 2022). For example, it is hard to conceive of a test that does not, to some degree, measure inhibitory control as testing a child without any inhibitory control skill would be difficult (see next section for moderator analyses that examine whether the effects differ between types of EF skill tests). We examined heterogeneity with a forest plot, prediction intervals, and heterogeneity statistics. We also reported the effects on tests of academic achievement and cognitive skills (corresponding to our fourth research question). Third, we conducted sensitivity analyses of the main effects. Fourth, we presented the results from confirmatory and exploratory moderator analyses (described in the *Subgroup analysis and investigation of heterogeneity*-section).

We performed all statistical analyses in R. The data used in the analysis and the analytic code are available at OSF. As we expected the results to be heterogeneous, we assumed a random-effects model. We used inverse-variance based weights. To estimate the overall effect size and heterogeneity statistics, we used the correlated-hierarchical effects model combined with robust-variance estimation (CHE-RVE) developed by Pustejovsky and Tipton (2022). This method allowed us to consider both dependencies between effect sizes that arise because the same sample is tested on different tests (‘correlated effects’) and because different samples are included in the same study (‘hierarchical effects’). This feature is an advantage over the original RVE method developed by Hedges et al. (2010). The original RVE procedure may furthermore have some disadvantages in terms of estimating heterogeneity parameters (see Tanner-Smith et al., 2016 for a discussion). We implemented the CHE-RVE model in three steps.

In step 1, we identified an appropriate working model based on the features of our sample, that is, whether there were correlated or hierarchical effects, or both. Both these types of dependencies were present in our case. In the primary outcome analysis, eight studies had both a correlated and hierarchical structure, nine had a purely hierarchical structure, 25 had a correlated structure, and only eight had neither type of dependency (i.e., only one primary outcome in one sample). The model requires a baseline value for the correlation between pairs of effect sizes from the same study (ρ) to be specified. We followed Pustejovsky and Tipton (2022) and used 0.6. We tested if our results were sensitive to lower (0.4) and higher (0.9) values.

In step 2, we used the working model to estimate meta-regressions with a combination of the *clubSandwich* (Pustejovsky, 2021) and *metafor* (Viechtbauer, 2010) packages for R. The dependent variable was effect sizes based on either EF skills tests or on our secondary outcomes. We used the *clubSandwich* package to specify the correlation structure between effect size estimates within studies. Then, we estimated the random-effects variance components, inverse-variance weight matrices, and the meta-regression coefficients using the restricted maximum likelihood (REML) procedure in the *metafor* package. With three exceptions, all studies measured effects within two months of the end of intervention. As stated in our protocol, we included an indicator for the three exceptions.

In step 3, we calculated confidence intervals based on the RVE standard errors obtained from the *clubSandwich* package. These standard errors were adjusted for small-sample bias as suggested by Tipton (2015) and Tipton and Pustejovsky (2015). We reported 95% confidence intervals for all analyses. As the results in Tanner-Smith and Tipton (2013) and Tipton (2015) suggest that standard errors from the RVE procedure are unreliable when the adjusted degrees of freedom are below four, we reported the degrees of freedom and mention in text when they were below or close to four.

In the main effects-analysis, we estimated the weighted average effect size and its confidence interval by using an intercept-only meta-regression where the coefficient on the intercept is the weighted average. We used this model to estimate the heterogeneity statistics as well. In the next section, we describe how we used moderator and subgroup analyses to examine the heterogeneity.

Subgroup analysis and investigation of heterogeneity

We conducted moderator analyses to identify the characteristics of interventions, participants, outcome measures, and studies that were possibly associated with smaller and larger effects on the EF outcomes (i.e., we did not conduct moderator analyses for our secondary outcomes). We used the CHE-RVE model and the same estimation procedure as in our main effects-analysis and report 95% confidence intervals for all regression parameters. We again included an indicator for follow-up outcomes in the moderator analyses.

Confirmatory moderator analysis. In our pre-registered, confirmatory moderator analysis, which corresponds to our second research question, we pooled all primary outcomes and included the following moderators:

Indicators for intervention content domain: We included an indicator equal to one if the intervention targeted maths content, and possibly other domains. That is, we contrasted any maths with interventions targeting literacy and language. We also show a specification where the maths indicator is equal to one for interventions targeting only maths, in which interventions that targeted maths and other content domains where in the contrast group. We chose this contrast because, as Clements et al. (2016) suggested, maths may have a more direct connection to executive functions and maths interventions may therefore have larger effect sizes. We planned to pool literacy and language interventions due to their overlap: On the one hand, language skills influence children's phonological awareness, which is a major component in the development of literacy skills (Anthony & Francis, 2005), and language comprehension is necessary for reading comprehension (Hoover & Gough, 1990). On the other hand, improved literacy skills may improve language skills (Morrison et al., 2019; Stanovich, 1986), and vocabulary interventions may affect language skills as well as for example reading comprehension (Hjetland et al., 2017; Rogde et al., 2019). We therefore expected substantial overlap in the targeted content domains among language and literacy interventions, an expectation that turned out to be correct.

Indicator for school setting and age: We included an indicator for interventions implemented in preschool (and in one study, both in preschool and in primary school). The contrast was interventions implemented in primary and secondary school. The discussion in the *How the intervention might work*-section indicated that the neural systems of younger children are more plastic. Effect sizes may therefore be larger for earlier interventions than later interventions.

Indicators for executive function outcome measures: We developed three moderators for outcome measures. An indicator for inhibitory control and attention outcomes, an indicator for short-term and working memory outcomes, and an indicator for cognitive flexibility outcomes. The reference group were outcomes based on measures of composite or broader EF skills, such as self-regulation, self-control, and planning and problem-solving that plausibly involve all core aspects of EF. Although the direction was not clear from our theoretical discussion (e.g., because we do not know which language, literacy, and maths skills transfer to what EF skill), the type of EF skill may moderate effect sizes.

An indicator for at-risk target groups: We included an indicator for interventions that mainly targeted a group of children who were experiencing, or were at risk of, academic difficulties. There had to be information that a majority of the participants were at risk. The reference group was interventions targeting a mix of at-risk and not-at-risk children (no study targeted high-achievers only), typically interventions where whole classrooms participated. Peng and Kievit (2020) found a weaker bidirectional relationship between EF skills and academic skills for at-risk than for not-at-risk students, motivating this contrast.

Because moderators may be correlated, we included all prespecified moderators in one meta-regression.

Exploratory analyses. To explore additional possible explanations of heterogeneity and examine the sensitivity of the confirmatory moderator analysis, we conducted exploratory moderator analyses. We added the following, not prespecified, moderators one-by-one to the final confirmatory model described above:

- *QES*: an indicator for QES.
- *Tutoring*: an indicator for interventions with adult-led small-group instruction using group sizes no larger than five students (e.g., Dietrichson et al., 2017; Nickow et al., 2024).

- *Physical activity*: an indicator for interventions that included physical activity during language, literacy, and maths instruction.
- *Duration*: a mean-centred, continuous variable measuring the duration of the intervention in units of 10 weeks.
- *Direct tests*: an indicator for tests conducted with the child present (in contrast to e.g., teacher or parent assessments of child skills).
- *Secondary average*: a mean-centred, continuous variable measuring the average effect size across secondary outcomes.

We added only one moderator at a time because the degrees of freedom where relatively low for some moderators in the confirmatory analysis model. Nearly all studies provided information for these moderators. A partial exception was the intervention duration, for which we used the intended duration for most and imputed the received duration in the few studies that did not contain information about the intended duration. Three studies did not provide any secondary outcomes and we imputed the mean in these cases.

We found only three studies that used a placebo control condition, and we refrained from including them in the moderator analyses. Similarly, we deemed that the number of studies with relevant variation was too small to include interaction variables (e.g., between intervention content and age).

Subgroup analysis: We also show results for more fine-grained definitions of subjects, type of EF tests, and instructional methods. We used single-variable subgroup analysis to estimate the weighted average effect size on end-of-intervention primary outcomes for interventions targeting only language, only literacy, and only mathematics, and for the following type of EF tests: inhibitory control and attention, working memory, short-term memory, cognitive flexibility, and composite EF tests. As mentioned, the definitions of what is short-term memory and working memory tests are not consistent in the literature. For our subgroup analysis, we used what we considered to be the majority opinion. In particular, we coded forward digit span and similar tests as short-term memory tests. Lastly, we estimated the weighted average effect size for four types of interventions, which were the most frequently occurring in our data: tutoring, physical activity, CAI, and curriculum and professional development. In all subgroup analyses, we used a similar CHE-RVE model as in the primary analysis.

Heterogeneity and effective sample size: Lastly, we examined whether the observed heterogeneity was reduced when we excluded studies with small effective sample sizes. The within and between-study heterogeneity measures take sampling error into account (i.e., they measure heterogeneity over and above sampling error). However, the sampling error may not capture sources of uncertainty such as outcome measurement error (Hedges, 1981), or baseline differences between the treatment and control group (Hedges, 1983). Even properly conducted RCTs are only balanced in expectation, and there may be chance bias–random baseline differences between treatment and control groups (Roberts & Torgerson, 1999).

Because both measurement error and baseline differences are more likely to create heterogeneity within and between studies with small sample sizes, examining the relation between the total heterogeneity and the effective sample size (i.e., sample size adjusted for clustering) may be informative. If there is less heterogeneity in large studies, it may be an indication that part of the heterogeneity is due to sources that are random and therefore not explainable.

Sensitivity analysis

In order to explore the sensitivity of the main effects on our primary outcome, we examined, as mentioned, sensitivity to our choices of ICC and ρ, and to individually assigned grouped treatments. We also examined sensitivity to outliers, how we calculated SMDs, risk of bias, and to removing one study or group of studies at a time by the methods described below.

Outliers. We examined the distribution of effect sizes for the presence of outliers. We then Winsorized the outliers to the nearest non-outlier value (e.g., Lipsey & Wilson, 2001), and re-ran the main effects-analysis.

Calculated standardised mean differences. We estimated SMDs with post-test SDs. We checked the sensitivity of our results to this choice by calculating alternative SMDs where the pre-test SDs were used. We also examined whether differences at baseline between the treatment and control group affected our results by calculating pre-test "effect sizes", where we used the pre-test means and SDs to calculate *g*, and then used these effect sizes as the dependent variables in the same specifications used for the main effects-analysis.

Two studies included only effect size information that allowed for the calculation of a Glass's delta. That is, an SMD where the standardisation is done with the control group SD. We examined if excluding these two studies changed our results.

Risk of bias. We examined sensitivity to risk of bias by removing effect sizes from the sample that received at least one rating of either 'high' or 'serious' risk of bias. We also present an analysis including only RCTs because QES have a higher risk of confounding.

Leave-one-out estimates. To make sure that our results were not driven by a single study, we re-estimated our main effects model repeatedly, leaving out each study from one estimation. We also estimated a model without physical activity interventions. As discussed in section *Types of interventions*, there was a risk that such interventions trained EF skills directly.

Differences between protocol and review

The protocol stipulated that secondary outcomes had to be standardised in order to be included (Dietrichson et al., 2022). However, a handful of studies reported exclusively non-standardised tests of the skills targeted in the intervention and in those cases, we chose to code the non-standardised measures as well and included them in our analysis of secondary outcomes.

We planned to search ProQuest Dissertations & Theses Global for dissertation. However, at the time of the search, we no longer had access to this database.

We used AIscreenR (Vembye et al., 2024; Vembye, 2025) as a second screener after switching to single-human screening, which was an extra safeguard not mentioned in our protocol.

The protocol stated that we would double-code included studies, but did not specify whether this meant included in the review or in the meta-analysis. To save resources, we did not double-code studies that we could not use in the meta-analysis (the risk of bias assessment of these studies were done by at least two review team members).

We did not specify in the protocol how to adjust studies that assigned treatment individually and where the treatment was received in groups. We used unadjusted effect sizes and variances in the primary analysis, and tested for sensitivity to this issue.

Lastly, the protocol stated that we would use TOT estimates if available and test sensitivity to the inclusion of ITT estimates. There were so few studies with both ITTs and TOTs that the planned sensitivity analysis was subsumed into other sensitivity analyses.

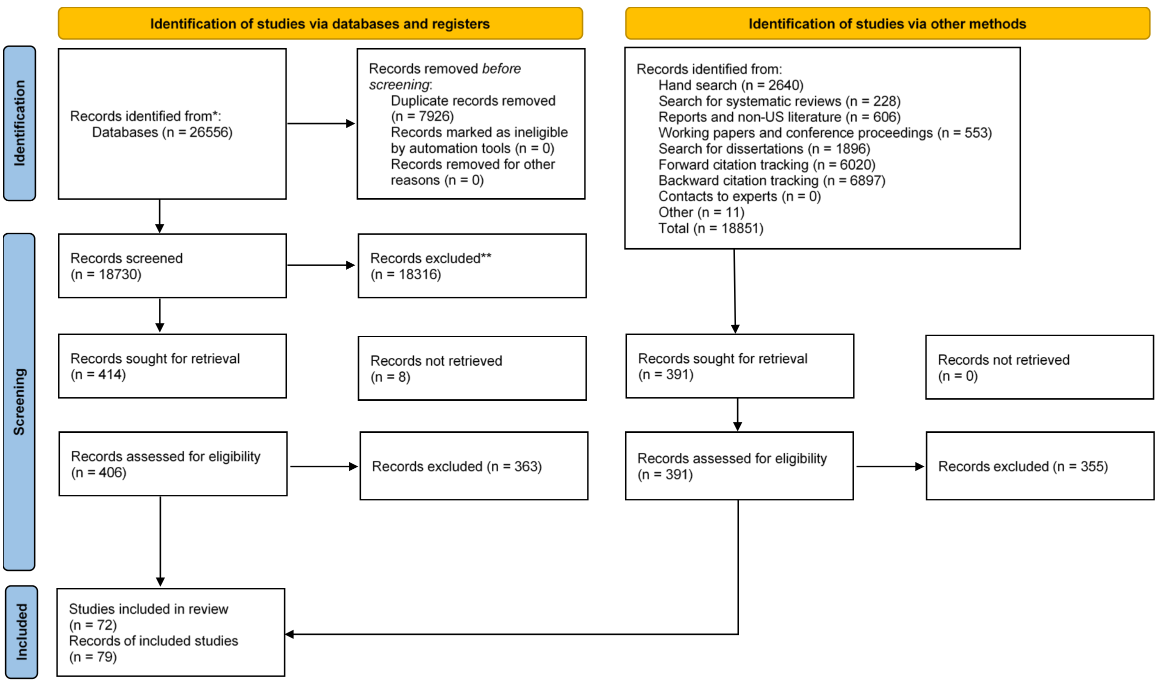
Results

Results of the search

Figure 1 describes the search and screening process using a PRISMA flow chart developed by Page et al. (2o21). The electronic database search yielded 18,730 potentially relevant records after removing 7,926 duplicates. The title and abstract screening of these records found 414 potentially relevant records. We were unable to retrieve the full texts for eight of these records. Of the remaining 406, we included 43 records.

The hand search, search for systematic reviews, reports and non-US literature, working papers and conference proceedings, search for dissertations, forward and backward citation tracking, contact to experts, and other sources (e.g., subscriptions to working paper lists) yielded in total 18,851 potentially relevant records. In particular, the citation tracking included many records (in total 12,917). Our protocol stated that we would put extra effort into this part of the search, which is one reason for the large number of records. However, the number is exaggerated in the sense that we did not remove duplicates (i.e., the number is the total number of citations for forward citation tracking and the total number of references for backward citation tracking). We included 391 of these records after screening on title and abstract, and were able to retrieve all of them in full text. We included 36 records after screening on full text.

FIGURE 1. *Flow chart of the search and screening process*



Thus, in total we screened 37,581 records on title and abstract, and 797 on full text. We included 79 records after full text screening. In the following, we refer to “studies” by which we mean clusters of effect sizes that on the one hand were unlikely to be independent from one another because they were obtained from the same intervention or same research study, and on the other hand were likely to be independent from other included effect sizes. A study can contain multiple treatments and be reported in multiple records. The 79 included records belonged to 72 studies.

Included studies

We coded all 72 studies. However, we were unable to include all of them in the meta-analysis. We lacked the information to calculate effect sizes in three studies and 18 studies were excluded because all outcomes received a critical risk of bias rating. We describe common reasons for giving a critical rating in the *Risk of bias in included studies*-section below. The rest of this section describes the 51 studies that were included in the meta-analysis.

Figure 2 shows the number of studies included in the meta-analysis by year of publication. The literature is quite recent: only two studies were published before 2000, and the majority has been published after 2015. Although dominated by studies from the United States (26), the studies were from a large and quite diverse set of countries: four from Italy, three from Denmark, two from Australia, China, Germany, Iran, the Netherlands, and Slovakia, and one each from Brazil, Lebanon, Norway, Spain, Sweden, and the United Kingdom.

FIGURE 2. *Number of studies included in the meta-analysis by year of publication*

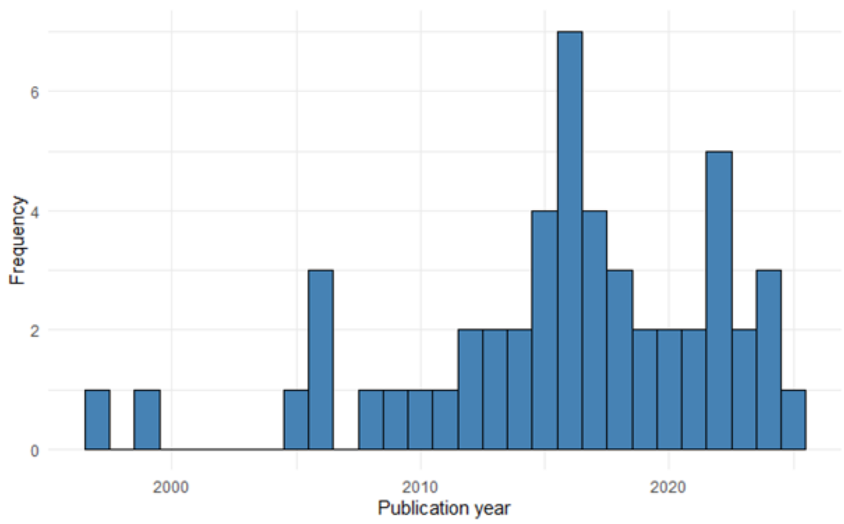


Table 1 presents descriptive statistics for the 504 effect sizes that we included in the meta-analysis. This number encompasses both primary and secondary outcomes, and the table shows statistics based on the effect sizes.

We characterize the included effect sizes by the sample characteristics (sample size, mean age, mean grade, and share of girls) and the variables that we included in our confirmatory and exploratory moderator analyses. Some studies lack information about mean age, mean grade, and share of girls from some of our studies (note that mean grade is missing by definition for studies set in preschool). For all other variables, we have complete information.

Most studies in our sample were small, with the median and mean sample size being 80 and 210.1, respectively. Interventions were often implemented in preschool or in the early school years: the median and mean age of the participants were both around seven years, 22% of effect sizes were from preschool interventions, and the median and mean grade in the school interventions were both around 2.5. There were slightly fewer girls in our sample, with the median and mean share being 47.8% and 47.2%.

A majority of effect sizes were from interventions that targeted at-risk children (68.5%). Relatively few effect sizes came from interventions that targeted only language (1.2%), whereas a larger proportion included language content alongside literacy or maths (in total 27.0%). Literacy was targeted alone in 20.4% of effect sizes and an additional 23.4% came from interventions that also targeted language or maths. Slightly more than half of the effect sizes came from maths-only interventions (51.4%) and 62.7% from interventions that included some maths content.

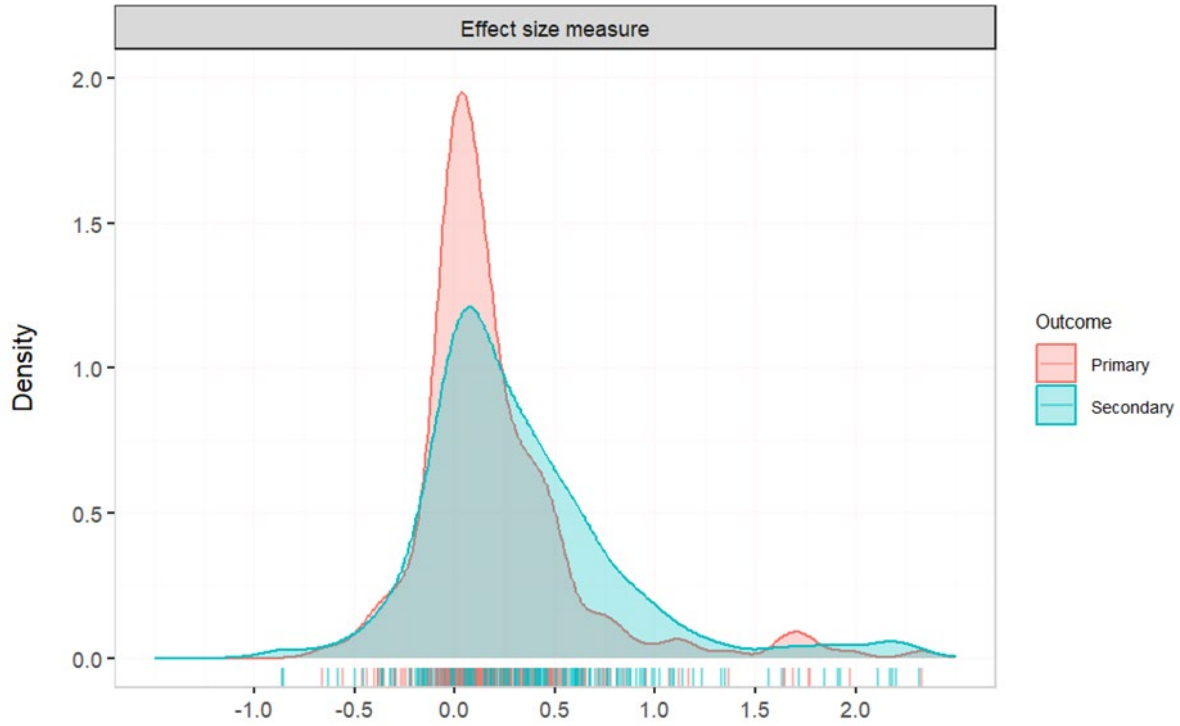
TABLE 1. *Descriptive statistics of effect sizes included in the meta-analysis*

Variable	Effect sizes	Min	Max	Median	Mean	SD
Sample size	504	20	2844	80	210.1	421.4
Mean age	441	3.38	13.07	6.89	7.335	2.188
Mean grade	395	0.00	7.50	2.50	2.516	1.734
Share of girls	483	13.00	70.00	47.80	47.106	8.231
Preschool	504	0	1	0	0.216	0.412
At risk	504	0	1	1	0.685	0.465
Language	504	0	1	0	0.012	0.109
Literacy	504	0	1	0	0.204	0.404
Language and literacy	504	0	1	0	0.157	0.364
Language and mathematics	504	0	1	0	0.036	0.186
Language, literacy and mathematics	504	0	1	0	0.077	0.267
Mathematics only	504	0	1	1	0.514	0.500
Any mathematics	504	0	1	1	0.627	0.484
Primary outcome	504	0	1	0	0.440	0.497
Inhibitory control/attention	504	0	1	0	0.151	0.358
Short-term and working memory	504	0	1	0	0.200	0.401
Cognitive flexibility	504	0	1	0	0.046	0.209
Composite	504	0	1	0	0.044	0.205
Direct measure	504	0	1	1	0.986	0.117
QES	504	0	1	0	0.062	0.240
Tutoring	504	0	1	1	0.583	0.493
Physical activity	504	0	1	0	0.131	0.338
Duration	504	3.00	113.10	15.00	22.180	20.619

Tests of EF, our primary outcome, constituted 44.0% of the effect sizes. Tests of short-term and working memory were most common (20.0% of the total number of effect sizes/45.5% of EF tests), followed by tests of inhibitory control and attention (15,1%/34.3%). Tests of cognitive flexibility (4.6%/10.5%) and tests of broader or composite EF skills (4.4%/10.0%) were comparatively rare. Almost all effect sizes were derived from direct measures (98.6%).

A large proportion of included studies were RCTs of tutoring interventions, only 6.2% of the effect sizes came from QES and 58.3% came from tutoring interventions. A second relatively common form of intervention was physical activity interventions, which constituted 13.1% of effect sizes. Lastly, the interventions varied substantially in terms of their duration, the shortest was just three weeks and the longest 113.1 weeks. The median and mean were 15.0 and 22.2 weeks, respectively. Figure 3 shows the estimated probability densities of effect sizes (*g*) from primary and secondary outcomes. Both densities indicate that there were more positive than negative effect sizes, but this tendency was much more visible for the secondary outcomes. The primary outcomes are centred slightly to the right of zero and there are few effect sizes below -0.5 and above 0.5. The negative part of the distribution looks similar for the secondary outcomes but there are many more effect sizes above 0.5 and relatively many are above 1.0 (note that we excluded a few outliers in terms of primary outcomes to make the figure more legible).

FIGURE 3. *Density plot of effect sizes from primary and secondary outcomes*



Excluded studies

Due to the large number of studies screened in full text, we were unable to describe all excluded studies. In this section, we exemplify how we applied the inclusion criteria by describing a selection of studies that met many of our inclusion criteria and that readers may expect to be included but which did not meet all inclusion criteria.

We included studies with a treatment-control group design that compared one or more treatments to a TAU or placebo condition with at least two units in each condition. We excluded studies that compared two alternative treatments. Examples include Barner et al. (2016), where the comparison group received three hours per week of supplementary maths practice, and Lawton (2016), where the comparison group received linguistic word building training. Because of language restrictions in the review team, only studies published in English, German, Danish, Swedish and Norwegian were eligible. Examples of studies excluded in the full text screening for language reasons were Herrera et al. (2007) and Toll and van Luit (2014; note that this does not mean that they met all other criteria).

Interventions had to target either language, literacy, mathematics, or combinations of these domains. We did not consider for example computational thinking and chess training to be mathematics interventions (e.g., Sala & Gobet, 2017; Scholz et al., 2008). Furthermore, interventions were excluded if they trained EF skills directly, and not through training language, literacy, and mathematics. Because regular preschool programs are likely to include activities that train EF skills directly, we excluded studies comparing children attending preschool to non-attending children (e.g., McCoy et al., 2017; Peisner-Feinberg et al., 2017). For similar reasons, we also excluded studies of mindfulness and yoga interventions (e.g., Bergen-Cico et al., 2015).

Studies had to include at least one pre-validated post-test of EF skills. Goldstein (1976) provided perhaps the first test of whether a reading intervention affected memory skills. However, the tests used by Goldstein 1976 were not pre-validated measures of executive functions, and we excluded the study. Stebbings et al. (2020) used the Fluid reasoning subtest of Woodcock-Johnson IV Tests of Cognitive Abilities. Although fluid reasoning is related to executive functions, it is a more general cognitive skill than executive functions (Diamond, 2013). We therefore excluded Stebbings et al. (2020).

Risk of bias in included studies

We excluded 18 studies from the synthesis because all outcomes received a critical rating in at least one domain of ROBINS-I. All of these studies were QES, that is, there was no RCT in which we assessed that the randomisation was no longer valid. As mentioned, we also excluded two outcomes from one study due to very high attrition but other outcomes from this study were included in the analysis. All except one of the QES received critical risk of bias in confounding (the exception received critical in the deviation bias domain). Common reasons for the critical rating were that there was little or no adjustment for confounders, and that the assignment to treatment was not clearly described and there were large imbalances at baseline. It is important to note that the ratings are assessments of the risk of bias, not actual bias, and that the ratings are not an assessment of a study's value. Studies can be very valuable, despite having too high risk of bias for our purposes.

In Figure 4 and 5, we used the R package *robvis* to visualize the risk of bias of the 221 primary outcomes included in the meta-analysis (McGuinness & Higgins, 2020; McGuinness & Kothe, 2019). There were few secondary outcomes that received a different judgement than the corresponding primary ones, so the table would look highly similar for secondary outcomes.

Figure 4 displays the risk of bias ratings in RCTs. No outcome received low risk of bias in all domains. The main reason was that only four studies had a pre-specified analysis plan, which is a necessary requirement for receiving low risk of bias in the selection of reported results-domain. Outcomes from two studies were rated high risk of bias in this domain due to indications in the published records that the reported outcomes were selected. That is, for most studies we found no indications of selective reporting, although this was of course difficult to assess given the lack of pre-specified analysis plans.

Most outcomes received some concerns in the measurement of the outcome-domain, which was mainly due to either lack of information about whether assessors were blind to treatment status or that the assessors knew treatment status, combined with a potential for such knowledge to influence the outcome assessment. The high risk of bias ratings in this domain were mostly because of a high risk of ceiling effects.

There were also relatively few outcomes that we rated low risk of bias in the randomisation process-domain. Most studies lacked a description of the random sequence generation. To receive a rating of some concerns, there also had to be signs of baseline imbalances and thus a risk that the randomisation had gone wrong. As most studies were small, it was however difficult to know whether baseline imbalances were due to chance or were a sign that the randomisation had gone wrong. The deviations from intended interventions and the missing outcome data are split relatively evenly between low risk of bias and higher ratings. The reasons for ratings higher than low in the former category were mainly connected to problems with implementation fidelity or lack of information about fidelity. In the latter domain, substantial and differential attrition, especially when coupled with no formal tests of differential attrition, yielded some concerns or high risk of bias ratings.

FIGURE 4. *Risk of bias of primary outcomes included in the meta-analysis from RCTs*

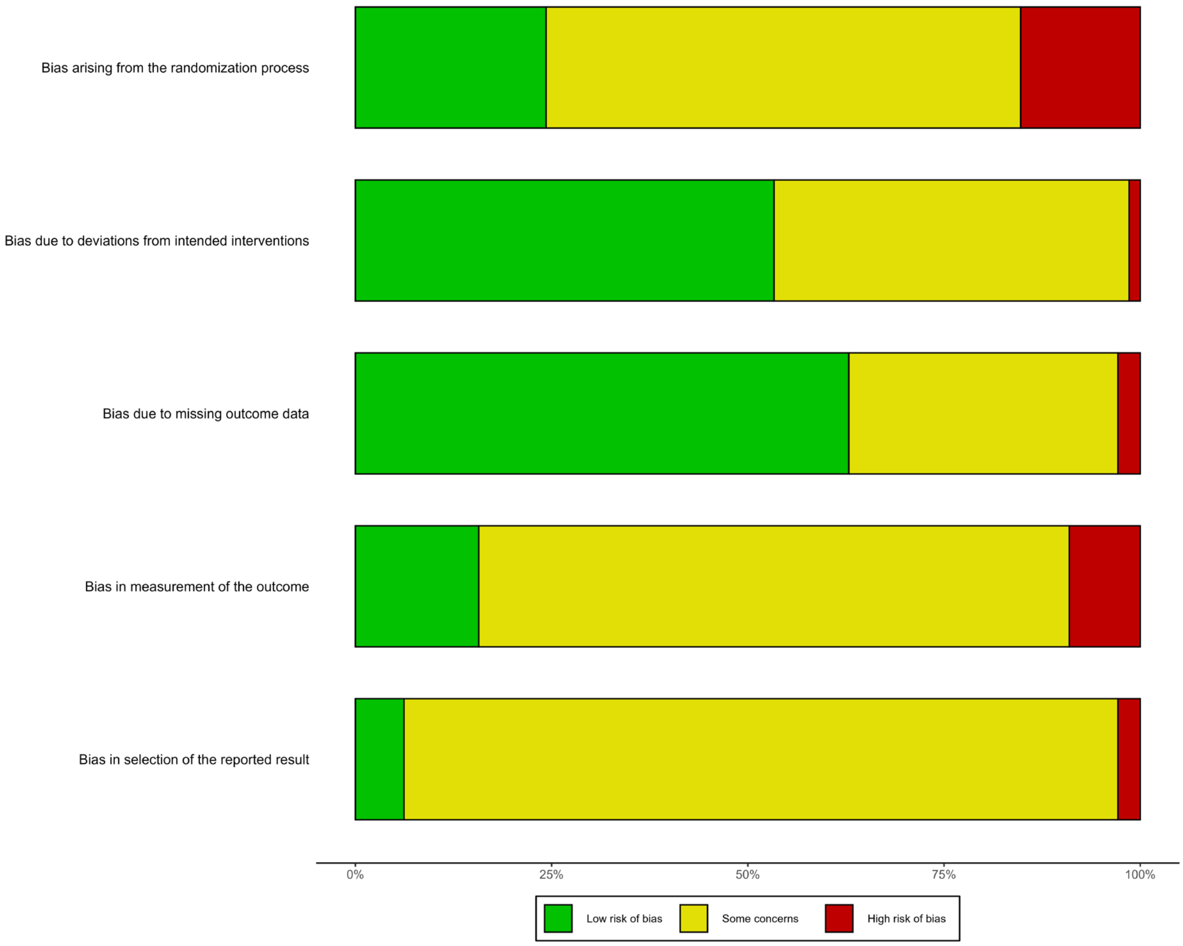
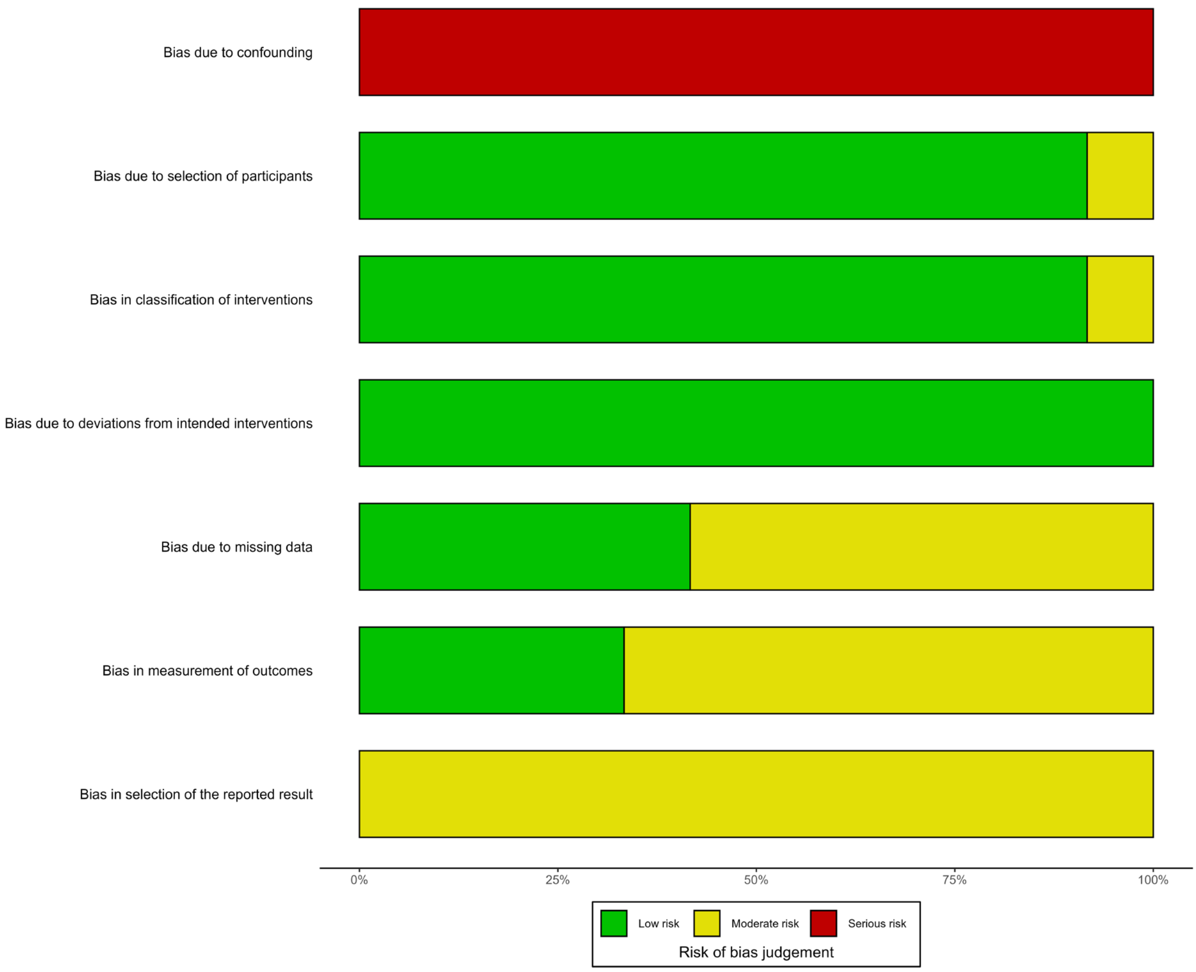


Figure 5 displays the risk of bias judgements in QES. All outcomes in the four QES that were included in the meta-analysis had serious risk of bias in the confounding-domain. There was also no example of a pre-specified analysis plan, but otherwise no indications of selective reporting. Risk of bias due to selection of participants, in the classification of interventions, and due to deviations from intended interventions were rare, whereas most outcomes had a moderate risk of bias due to missing data and in the measurement of outcomes (for similar reasons as the RCTs).

FIGURE 5. *Risk of bias of primary outcomes included in the meta-analysis from QES*



Synthesis of results

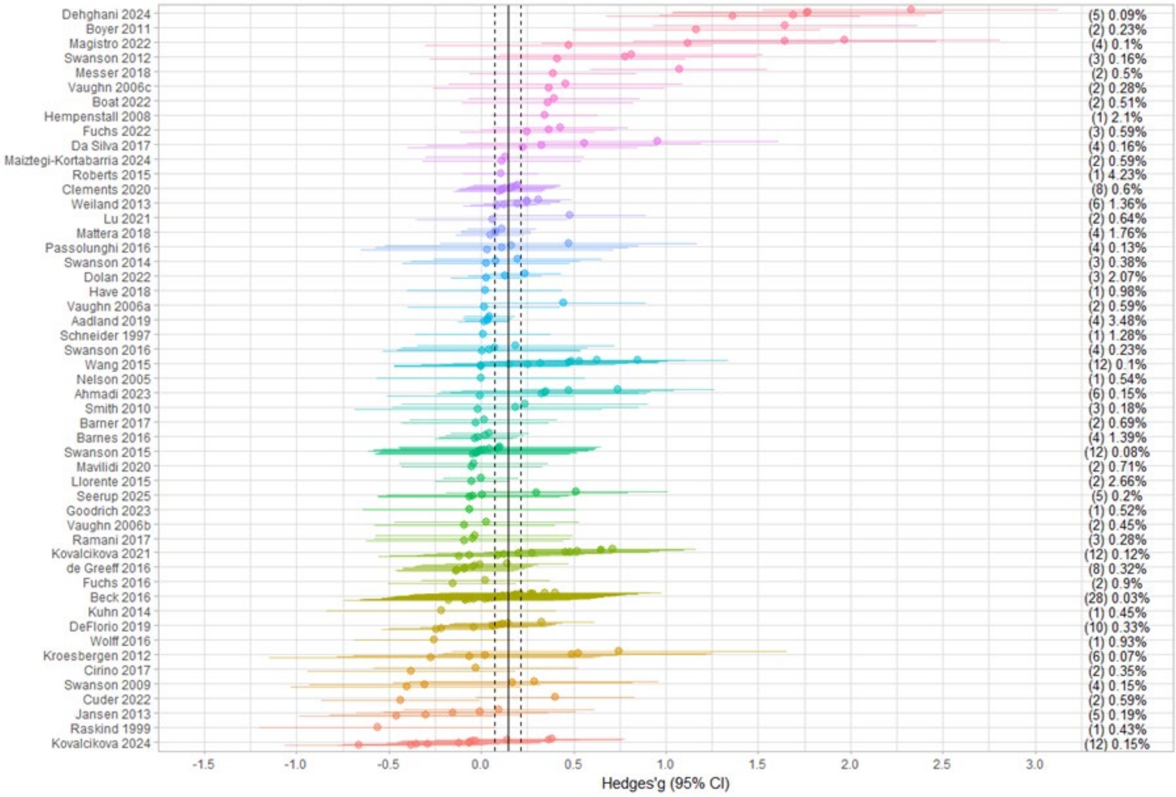
Main effects

We included 222 effect sizes nested in 51 studies in the main effects-analysis of primary outcomes. We found a statistically significant weighted average effect size of 0.14 (95% CI = [0.07, 0.22]) for outcomes measured at the end of intervention (henceforth, significant refers to statistical significance only). The indicator of follow-up outcomes was small and not significant (β = -0.01, 95% CI = [-0.34, 0.32]). As mentioned, there were only three studies with follow-up outcomes. This indicator was included in our analyses here (and in the moderator analyses) but because of the small sample size, it was difficult to interpret and we refrain from commenting on it in the following.

All absolute measures indicated substantial heterogeneity. The within (ω = 0.13), between (τ = 0.15), and total standard deviation (σ = 0.20) were large in relation to the average effect size. The Q statistic was strongly significant (Q = 458.7, p < 0.0001). The prediction interval for the end-of-intervention effect ranged between -0.26 and 0.55. However, the I^2 statistic, which measures how much of the total variance that is not due to sampling error and is a relative measure of heterogeneity (Borenstein et al., 2017), was relatively low: total I^2 was 54.0% (within-study I^2 = 24.0%, and between-study I^2 = 30.0%). That is, almost half of the total variance was estimated to be due to sampling error. A likely reason was that most studies had small sample sizes.

These heterogeneity patterns can be seen in the forest plot shown in Figure 6. The figure shows the effect sizes and 95% confidence intervals by study, arranged from the study with the largest effect size to the smallest. The solid line is the estimated average weighted effect size at the end of intervention and the dashed lines its 95% confidence interval. The label on the right side shows the number of effect sizes and the weight of each effect size in the main effects-analysis.

FIGURE 6. *Forest plot of primary outcomes*



Note: The figure shows all effect sizes and their 95% confidence intervals by study, the number of effect sizes in each study (in parenthesis), and the average weight of each effect size. We abbreviated the study references to the last name of the first author and the year of publication to make the figure more legible.

The heterogeneity, both between and within studies, is clearly visible in the figure. The figure also indicates that there might be some outliers, which we examined further in the next section. It is furthermore noteworthy that the confidence intervals are often wide and very few individual effect sizes are significant. In particular, only two of the negative effect sizes are significant. Because we were sometimes forced to use raw means instead of covariate-adjusted estimates to be able to calculate effect sizes, our coding of the studies may underestimate the precision of the estimates somewhat. Despite this underestimation, the forest plot indicates that many studies with small sample sizes were underpowered to find effects of similar magnitude to the average effect size.

In the analysis of secondary outcomes, we included 282 effect sizes nested in 48 studies (three studies did not report any secondary outcomes). We found a statistically significant weighted average effect size of 0.37 (95% CI = [0.24, 0.50]) for outcomes measured at the end of intervention. The indicator of follow-up outcomes was not significant (β = 0.08, 95% CI = [-1.01, 1.16]) and, as in the analysis of primary outcomes, based on only three studies. There was also large heterogeneity regarding the secondary outcomes. The within (ω = 0.30), between (τ = 0.34), and total standard deviation (σ = 0.46) indicated considerable heterogeneity of the effect sizes. The Q statistic was strongly significant (Q = 1,351.8, $p < 0.0001$). The prediction interval for the end-of-intervention effect ranged between -0.56 and 1.31. The I^2 statistic was larger than for the primary outcomes (82.7%).

The large effect size for end-of-intervention secondary outcomes was to some extent driven by the inclusion of non-standardised tests. Including an indicator for non-standardised secondary outcomes, we found a twice as large, albeit non-significant, effect for such tests (β = 0.37, 95% CI = [-0.33, 1.07]). However, it was only in six studies that reported exclusively non-standardised secondary outcomes that we were forced to use them, and the estimate for standardised secondary outcomes was still large and highly significant (β = 0.33, 95% CI = [0.22, 0.44]). Most secondary outcomes were tests of language, literacy, and maths. However, some were tests of non-EF cognitive skills (e.g., fluid intelligence). We included a moderator for such tests, which yielded a non-significant coefficient of -0.24 (95% CI = [-0.56, 0.08]). In this meta-regression, the average effect size for standardised tests of language, literacy, and maths was 0.34 and highly significant (95% CI = [0.23, 0.46]).

In sum, we found positive and significant effects of language, literacy, and maths interventions on end-of-intervention tests of EF skills, as well as on our secondary outcomes. The effects were strongly heterogeneous for both primary and secondary outcomes. In the next sections, we examine how sensitive the effects on primary outcomes are to changes to our specifications and to publication bias. We then examine whether we can explain some of the heterogeneity by including moderators in our meta-regressions, by examining subgroups, as well as the relation between heterogeneity and effective sample size.

Results of the sensitivity analyses

We examined how sensitive the estimate of the average effect on EF skills from the primary analysis was to our choices of ICC and ρ, adjusting for individually assigned group treatments, to outliers, how we calculated SMDs, risk of bias, and to removing one study or group of studies at a time.

Using different values of ICC and ρ left our results virtually unchanged. Despite using relatively extreme choices, 0 and 0.32, for studies that did not report information about the ICC, the average effect size remained close to the primary analysis, 0.16 (95% CI = [0.08, 0.25]) and 0.15 (95% CI = [0.07, 0.22]), respectively. We obtained an estimate of 0.15 (95% CI = [0.07, 0.23]) when we imputed 0.11 for all studies (also the ones with information about the ICC). The average effect size was 0.16 (95% CI = [0.08, 0.24]) with $\rho = 0.4$ and 0.13 with $\rho = 0.9$ (95% CI = [0.07, 0.19]). Adjusting for individually assigned group treatments increased the average effect size slightly to 0.18 (95% CI = [0.08, 0.28]).

We examined the distribution of effect sizes to find candidates for outliers. The distribution in Figure 6 indicated that there were mainly outliers among the positive effect sizes, where effect sizes around 1-1.5 and larger stand out. The studies finding effect sizes over 1.0 are Dehghani et al. (2024), Boyer and Ehri (2011), Magistro et al. (2022), and Messer and Nash (2018). In three of these studies, there were features that may explain the large outcomes. In Boyer and Ehri (2011), the control group score worsened from pre to post-test, which is uncommon. The standard deviations in Magistro et al. (2022) were calculated from the standard errors of the mean, which are reported to only one decimal point. As it makes a relatively big difference if the standard error is, for example, 0.05 or 0.14, the estimates from this study were uncertain. Dehghani et al. (2024) had the largest effect sizes in our sample and was a small study (20 students in each group). It was harder to find a reason for the one large value in Messer and Nash (2018; although that study was not large either, 45 and 33 students in the treatment and control group, respectively).

We therefore first Winsorized the three studies where we could find reasons for the outlier-status to the next largest effect size (i.e., the one in Messer & Nash, 2018). We then used a more data-driven approach and Winsorized effect sizes larger than three times the interquartile range (Pustejovsky et al. 2025), which also included studies with negative effects. The two procedures produced very similar results: the first yielded an estimate of 0.12 (95% CI = [0.06, 0.17]) and the second an estimate of 0.12 (95% CI = [0.07, 0.18]).

In the primary analysis, we estimated SMDs with post-test SDs. Using the pre-test SDs instead yielded an average effect of 0.19 (95% CI = [0.09, 0.28]; based on the 45 studies and 201 effect sizes for which we had information about the pre-test SD). Excluding the two studies for which we could only calculate Glass's delta did not change our results (β = 0.15, 95% CI = [0.07, 0.24]). We also examined whether differences at baseline between the treatment and control group might have affected our results. The "effect size" at pre-test was small and non-significant (β = 0.01, 95% CI = [-0.04, 0.06]; based on the 39 studies and 174 effect sizes for which we had the requisite information).

Removing outcomes with high or serious overall risk of bias left 33 studies and 149 effect sizes in the sample. Because all QES had a serious risk of bias in at least one domain, and no RCT had a low risk of bias in all domains, these effect sizes were from RCTs rated some concerns in at least one domain. The average effect size in this sample was slightly larger than in the primary analysis, 0.19 (95% CI = [0.07, 0.31]). We also removed outcomes with high risk of bias by domain in ROB-2 (as there were only four QES, we did not conduct this analysis for them). These removals changed the average effect size very little, it ranged from 0.13-0.16 and was in all cases still significant. Using only the 210 effect sizes from the

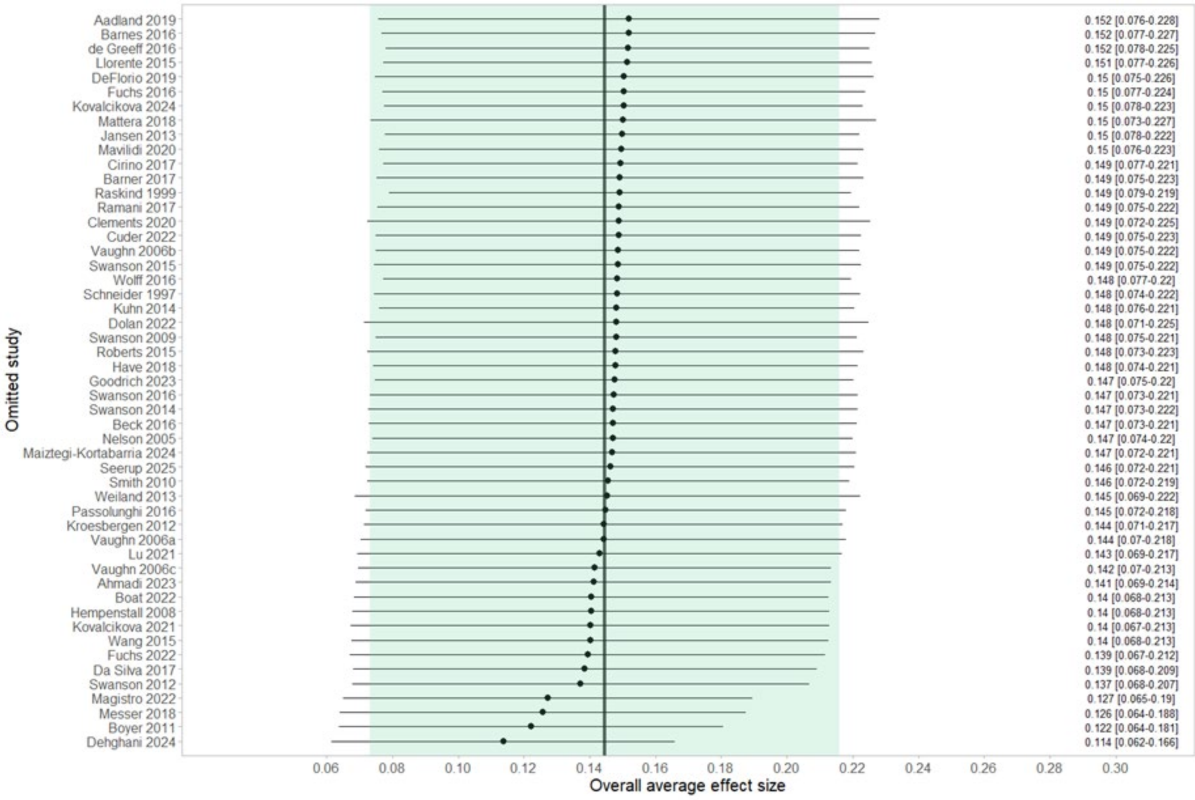
47 RCTs yielded a highly similar average effect size as in the primary analysis, 0.14 (95% CI = [0.06, 0.22]).

Figure 7 shows the results from an analysis where we leave out one study at a time, to ensure that our results are not driven by a single study. The solid line shows the average effect size at the end of intervention and the shaded green area shows the confidence interval from the primary analysis. The black dots and lines are the effect sizes and confidence intervals, re-estimated on a sample without the study mentioned in the left column. For most studies, we obtained a higher average effect size and it was only for three of the studies, which we discussed above in relation to outliers, that the estimate moved visibly downward. However, these estimates were still reasonably close to the primary estimate. Moreover, all estimates are significant and safely within the confidence interval from the primary analysis.

Lastly, we examined whether the average effect size was sensitive to our choice of including physical activity interventions. Excluding these interventions and re-estimating the primary analysis model on the remaining 43 studies and 168 effect sizes, we obtained an average effect size of 0.13 (95% CI = [0.07, 0.19]).

In sum, the average effect size estimate from the primary analysis was robust in all sensitivity analyses.

FIGURE 7. *Results of the leave-one-out analysis*



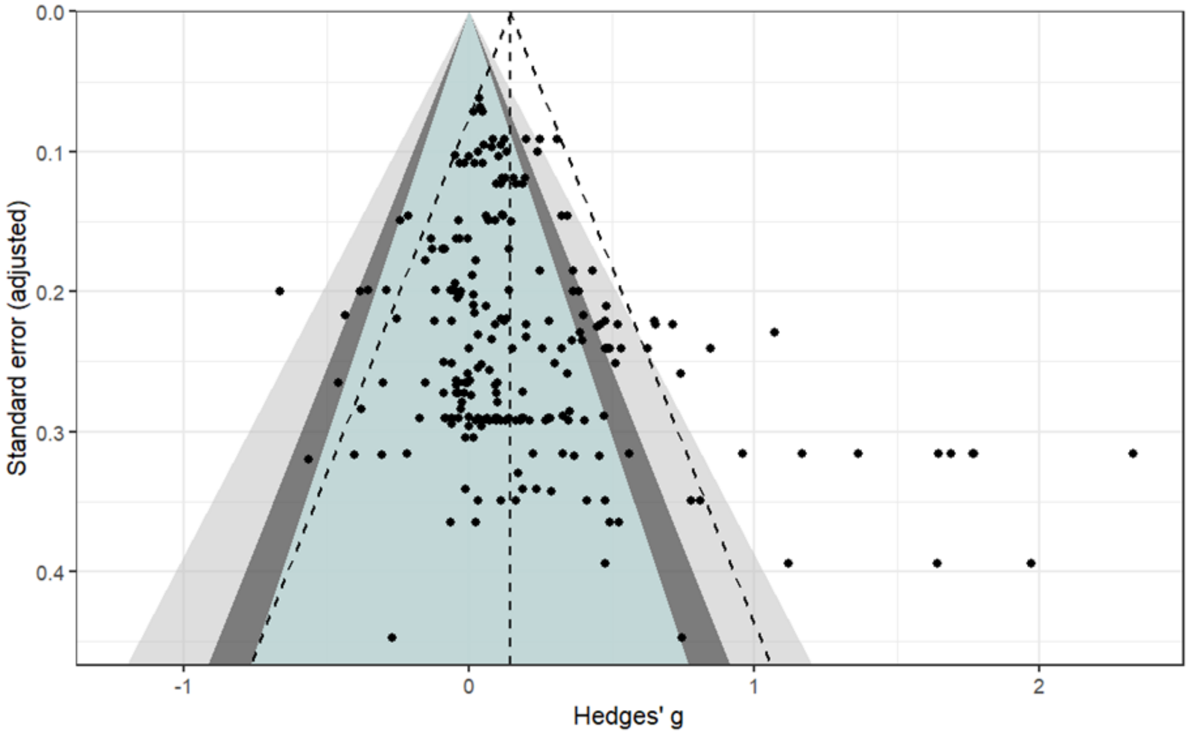
Note: The figure shows the leave-one-out estimates. The solid line shows the average effect size at the end of intervention and the shaded green area shows the confidence interval from the primary analysis. The black dots and lines are the effect sizes and confidence intervals, re-estimated on a sample without the study mentioned in the left column. We abbreviated the study references to the last name of first author and year of publication to make the figure more legible.

Assessment of publication bias

Figure 8 displays a funnel plot of the primary outcomes. The plot has the effect sizes on the x-axis, and the adjusted standard errors of the effect sizes on the y-axis (the adjusted standard errors are derived from the sample size only, see Chen & Pustejovsky, 2025). The dashed, straight line represents the average effect size from the primary analysis. The dashed diagonal lines represent 95% confidence limits around the average effect size. The shaded triangles show whether effect size estimates are significant at the 10, 5, and 1 percent level, respectively.

In line with the previous analysis, the funnel plot indicates that there was substantial heterogeneity in our data and that there are outliers. The plot is not strongly asymmetric, but the Egger Sandwich test rejects the null of no association between the adjusted standard errors and the effect size in a one-sided test at the 5% level (p = 0.023, one-sided test).

FIGURE 8. *Funnel plot of primary outcomes*



Note: The figure shows the individual effect sizes (Hedges' g) on the x-axis and the adjusted standard error on the y-axis. The dashed, straight line represents the average effect size from the primary analysis. The dashed diagonal lines represent 95% confidence limits around the average effect size. The shaded triangles show whether effect size estimates are significant on the 10, 5, and 1 percent level, respectively.

The association indicates that smaller studies have larger effects. Although such an association may arise for reasons that do not have anything to do with publication bias (see e.g., Dietrichson et al., 2021, for a discussion in relation to school interventions), it calls for a sensitivity analysis. We used the method suggested by Mathur and VanderWeele (2020), implemented using the R package PublicationBias (Braginsky et al., 2023). We first assumed that positive and significant studies were two times as likely to be found as other studies. This assumption yielded an adjusted average effect size of 0.11 (95% CI = 0.05, 0.17). We then used a version of Mathur and VanderWeele's (2020) "worst-case" scenario adapted to the CHE-RVE model. This scenario simply removes all positive and significant effect sizes. Also in the worst-case scenario, our estimated average effect size was positive and significant (β = 0.05, 95% CI = [0.02, 0.08]).

In sum, we found some indications of potential publication bias in this literature. However, our estimate of the average effect size remained positive and significant also under very unfavourable assumptions about the nature and extent of this publication bias. The funnel plot reaffirmed that the effect sizes we included were strongly heterogeneous. The next two sections examine if we can explain some of this heterogeneity with observable moderators.

Confirmatory moderator analysis

Table 2 presents the results from our pre-specified, confirmatory moderator analysis. Column 1 contains the primary analysis estimate for easy reference. In column 2, we included all the pre-specified moderators in one meta-regression. The intercept in this specification was larger than in the primary analysis (0.24), but not significant. It represents the estimate for end-of-intervention effect sizes from treatments that did not target maths, were set in school and not in preschool, and were derived from composite EF skills.

No moderator was significantly associated with effect sizes. The coefficients on the indicators for different types of EF skills test were all negative: the inhibitory control and attention-tests was relatively large (-0.18). While smaller, the indicators for short-term and working memory tests (-0.10) and for cognitive flexibility (-0.06) were also relatively substantial in relation to the overall weighted average effect size. In all cases, the total marginal effect (intercept plus coefficient) was positive.

TABLE 2. *Confirmatory moderator analysis*

Variable	(1)	df	(2)	df
Intercept	0.14	40.5	0.24	15.3
	[0.07, 0.22]		[-0.03, 0.5]	
Follow-up			-0.01	1.5
			[-0.28, 0.26]	
Maths			0.07	30.9
			[-0.15, 0.29]	
Preschool			-0.01	21.0
			[-0.22, 0.20]	
Inhibitory control			-0.18	4.0
			[-0.39, 0.03]	
Memory			-0.10	8.5
			[-0.29, 0.10]	
Cognitive flexibility			-0.06	3.6
			[-0.34, 0.23]	
At risk			-0.02	4.7
			[-0.17, 0.12]	
τ	0.15		0.19	
ω	0.13		0.12	
σ	0.20		0.23	

Note: The sample size for each specification is 222 effect sizes nested in 51 studies. 95% confidence intervals in brackets. The numbered columns show the coefficient estimates and the corresponding 95% confidence interval (in brackets below the estimate). The Df-column reports the small-sample adjusted Satterthwaite degrees of freedom.

The indicator for interventions targeting maths was positive (0.07). Using an indicator for targeting only math did not substantially change the results, the coefficient was then 0.08 (95% CI = [-0.12, 0.28]; not shown in table). The preschool indicator and the at-risk indicator were both close to zero.

Lastly, as can be seen by the estimates of ω, τ, and σ at the bottom of the table, the heterogeneity remained substantial.

Exploratory moderator analysis

Table 3 displays the results from the exploratory moderator analysis. We included the moderators from the confirmatory model and then added additional moderators one-by-one to preserve degrees of freedom. Overall, adding new moderators did not affect the moderators from the confirmatory model much. The coefficients are stable across specifications in Table 3.

With the exception of the study-level average across secondary outcomes, all associations between effect sizes and the added moderators were insignificant. A one standard deviation larger effect on secondary outcomes was associated with a 0.28 standard deviation larger effect on EF skills. Heterogeneity also remained substantial in all specifications, and it was only when we added the average across secondary outcomes that the total SD was visibly reduced (to 0.13).

However, the estimates of the indicators of QES (0.13), tutoring (0.08), physical activity (0.16), and direct tests (0.09) were quite large in relation to the weighted average effect. Also the estimate for duration (0.01), which was measured in 10-week units, seem large enough to potentially be important. That is, the statistical insignificance might be an issue of statistical power rather than substantive importance.

TABLE 3. *Exploratory moderator analysis*

Variable	(1)	df	(2)	df	(3)	df	(4)	df	(5)	df	(6)
Intercept	0.22	18.3	0.20	18.7	0.18	15.5	0.24	18.7	0.17	9.9	0.23
	[-0.05, 0.50]		[-0.04, 0.45]		[-0.15, 0.51]		[-0.03, 0.50]		[-0.11, 0.45]		[0.01, 0.45]
Follow-up	-0.01	1.3	-0.02	1.3	-0.01	1.3	-0.01	1.3	-0.01	1.3	0.01
	[-0.28, 0.26]		[-0.3, 0.27]		[-0.27, 0.25]		[-0.28, 0.26]		[-0.29, 0.27]		[-0.33, 0.34]
Maths	0.08	29.3	0.09	28.7	0.06	28.1	0.06	29.1	0.06	28.6	0.05
	[-0.16, 0.32]		[-0.13, 0.30]		[-0.16, 0.28]		[-0.16, 0.28]		[-0.17, 0.28]		[-0.10, 0.20]
Preschool	-0.04	17.8	-0.03	21.6	0.03	22.8	-0.01	21.5	-0.01	21.5	-0.05
	[-0.29, 0.2]		[-0.23, 0.18]		[-0.16, 0.22]		[-0.22, 0.19]		[-0.23, 0.20]		[-0.18, 0.07]
Inhibitory control	-0.18	5.4	-0.18	5.4	-0.19	5.3	-0.18	5.4	-0.19	5.2	-0.17
	[-0.39, 0.03]		[-0.39, 0.03]		[-0.4, 0.02]		[-0.39, 0.03]		[-0.41, 0.03]		[-0.37, 0.02]
Memory	-0.10	7.8	-0.10	7.8	-0.09	7.8	-0.09	7.9	-0.11	7.5	-0.09
	[-0.29, 0.09]		[-0.3, 0.10]		[-0.29, 0.10]		[-0.29, 0.10]		[-0.32, 0.1]		[-0.27, 0.09]
	-0.06	5.9	-0.06	5.9	-0.06	5.9	-0.06	5.9	-0.06	5.8	-0.06
	[-0.35, 0.23]		[-0.34, 0.23]		[-0.35, 0.23]		[-0.35, 0.23]		[-0.36, 0.23]		[-0.32, 0.20]
	-0.02	7.7	-0.04	7.3	0.02	4.5	-0.02	7.9	-0.02	7.4	0.01
	[-0.17, 0.13]		[-0.19, 0.12]		[-0.22, 0.26]		[-0.17, 0.12]		[-0.17, 0.13]		[-0.09, 0.1]
QES	0.13	4.0									
	[-0.19, 0.45]										
Tutoring			0.08	5.1							

			[-0.04, 0.21]								
Physical activity					0.16	12.8					
					[-0.31, 0.62]						
Duration							0.01	2.7			
							[-0.03, 0.04]				
Direct tests									0.09	2.2	
									[-0.14, 0.33]		
Secondary average											0.28
											[0.06, 0.50]
τ	0.19		0.19		0.21		0.19		0.20		0.04
ω	0.12		0.12		0.12		0.12		0.12		0.12
σ	0.23		0.23		0.24		0.22		0.23		0.13

Note: The sample size for each specification is 222 effect sizes nested in 51 studies. The numbered columns show the coefficient estimates and the corresponding 95% confidence interval (in brackets below the estimate). The Df-column reports the small-sample adjusted Satterthwaite degrees of freedom.

In Table 4, we explore the effects on primary outcomes across subgroups defined by subject, type of EF skill test, and intervention type. We restricted the analysis to end-of-intervention tests only.

TABLE 4. *Subgroup analysis*

Subgroup	***n***	***k***	***ES***	***95% CI***	***df***	***σ***
Language	2	5	0.13	[0.04, 0.21]	1.0	0.00
Literacy	10	21	0.17	[-0.18, 0.52]	8.8	0.45
Mathematics	26	143	0.17	[0.07, 0.28]	21.3	0.21
Inhibitory control	25	74	0.06	[0.00, 0.11]	12.8	0.04
Working memory	30	73	0.17	[0.07, 0.27]	23.6	0.17
Short-term memory	16	24	0.33	[0.09, 0.57]	14.7	0.39
Cognitive flexibility	7	23	0.14	[0.01, 0.26]	4.9	0.18
Composite	8	20	0.19	[-0.04, 0.42]	6.2	0.25
Tutoring	25	86	0.16	[0.06, 0.26]	18.4	0.19
Physical activity	8	54	0.33	[-0.17, 0.82]	6.9	0.55
CAI	5	14	0.15	[-0.36, 0.65]	3.9	0.37
Curricullum and PD	8	31	0.12	[0.05, 0.2]	4.6	0.08

Note: The number of studies is denoted *n*, and the number of effect sizes *k*. The estimates in the *ES* column and the 95% confidence intervals (in brackets) come from an intercept-only, CH-RVE model estimated on a subset of the data corresponding to the subgroup. The *Df*-column reports the small-sample adjusted Satterthwaite degrees of freedom, and *σ* is the total within and between study heterogeneity expressed in standard deviation units.

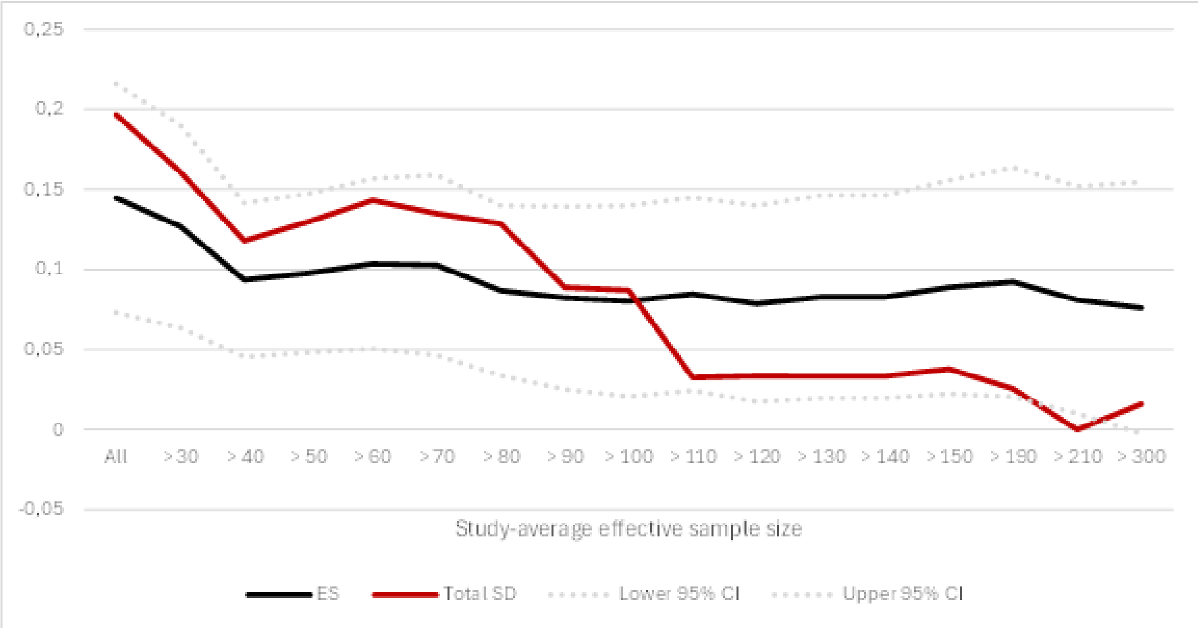
The weighted average effect sizes are positive for all single-subject interventions. However, the estimate for literacy is not significant (albeit equally large as the one for mathematics interventions). The language estimate is significant but based on two studies and should be interpreted cautiously. Similarly, all estimates are positive for the different EF skills test, and only the composite test is not significant. The estimate for inhibitory control and attention tests is smaller and the estimate for short-term memory tests is larger than the others. The estimates are again all positive for the intervention types, significantly so for tutoring, and curriculum and professional development interventions. Both physical activity and CAI were imprecisely estimated and exhibited substantial heterogeneity. This was the case for most subgroups.

Heterogeneity and effective sample size

Figure 9 displays the relations between the total within and between-study heterogeneity (σ), the effective sample size, and the magnitude and precision of the weighted average effect size among primary outcomes. The figure shows results from meta-regressions, which progressively exclude studies with small effective samples (the x-axis shows the cutoff used). On the y-axis, the weighted average effect size is shown in black, the 95% confidence interval of the effect estimate by the two grey dotted lines, and the total within and between-study heterogeneity in red.

The figure shows that the weighted average effect size declines to 0.1 or slightly below once very small studies are omitted. Thereafter, the effect size is stable. The confidence interval gets progressively wider but the change is modest and the weighted average effect size is significantly different from zero until we required the effective sample size to be more than 300. At that point, the number of included studies in the meta-regression is seven. The total heterogeneity is clearly larger than the weighted average effect size when the meta-regressions include all or most of the studies. However, in contrast to the weighted average effect size, the total heterogeneity is drastically reduced, starting when we include only effective sample sizes larger than 90. It is below 0.05 for effective sample sizes larger than 110. For large effective sample sizes (around 200 and over), the total heterogeneity is close to or at zero.

FIGURE 9. *Heterogeneity and effective sample size*



Note: The y-axis shows the estimated weighted effect size (black line), the total heterogeneity (red line), and the 95% confidence interval of the estimated weighted effect size in standard deviation units (dotted grey lines). The x-axis shows the cutoff for the effective sample size used in the meta-regression corresponding to the cutoff. Note that the x-axis does not have equal increments at large effective sample sizes. The reason is that no study had an effective sample between, e.g., 210 and 300.

The results provided support for the idea that a relatively large part of the heterogeneity found in our analyses was due to studies being small and therefore affected by random factors. Such factors are likely unrelated to the intervention and unlikely to be explainable. It is furthermore noteworthy that the reduction of the heterogeneity did not seem to be due to the setting, study designs, and interventions becoming homogeneous when effective sample sizes become large. For example, the eight studies included in the meta-regression of studies with effective sample size larger than 210 (when the estimated total heterogeneity was zero) included studies from the US, Lebanon, and Norway, set in preschool and school, that examined interventions including curriculum changes, coaching/professional development, tutoring and medium-group instruction, and physical activity programs, and targeting math, language and math, language and literacy, or all three skills.

Discussion

Summary of main results

Summarizing our findings by research question, we can answer the first question affirmatively. Preschool and school-based language, literacy, and mathematics interventions have robust and statistically significant average effects on measures of EF. We found a weighted average effect size of 0.14. The estimate was relatively precise, the 95% confidence interval ranged from 0.07 to 0.22.

To put the magnitude in perspective, Kassai et al. (2019) found an (insignificant) effect of 0.11 of treatments directly training specific EF skills on other, untrained EF skills. Although it is difficult to make fair comparisons of effect sizes across reviews, Kassai et al. (2019) included reasonably similar outcomes and study designs as we did (although with a larger proportion of active control groups). Less closely related, the magnitude is larger than the overall median effect size of 0.10 in a large set of RCTs with outcomes measured by standardised tests in mathematics and reading reported by Kraft (2020), and would be categorized as a medium effect size by Kraft's benchmarks. As one reason to implement interventions is to reduce skill inequalities between children, it may also be interesting to compare the effect to, for example, differences between high and low-SES groups. On similar EF tests to the ones we included, these differences ranged between 0.06 and 0.46 in Last et al. (2018) and Cuartas et al. (2022). These comparisons indicated that our average effect size was not only statistically significant, but also substantial.

We found large heterogeneity of the effect sizes. To answer our second research question, we used meta-regressions to examine how effect sizes varied across a set of pre-registered moderators: the subject of the intervention, whether the intervention was conducted in preschool or school, the type of EF measured, and the at-risk status of participants. Including these moderators did not reduce the heterogeneity. We found virtually no reductions of the heterogeneity statistics and no significant moderators in this confirmatory analysis, although in particular the estimate for inhibitory control and attention tests was large.

To answer our third research question, we conducted exploratory moderator analyses by adding (one by one) the following moderators to our confirmatory meta-regression specification: indicators for QES, tutoring interventions, physical activity interventions, direct tests, and mean-centred variables measuring intervention duration and the study-average across secondary outcomes. With the exception of study-average across secondary outcomes, none of these moderators were significantly associated with effect sizes. Larger effects on secondary outcomes were associated with larger effects on EF skills. Furthermore, only the inclusion of the study-average across secondary outcomes yielded a substantial reduction of heterogeneity. Our subgroup analysis indicated that the positive effects were broadly present across subgroups defined by subject, type of EF test, and intervention type, and that heterogeneity was typically substantial also within subgroups.

One reason for the substantial heterogeneity might be that many studies were small. Small studies may be more affected by random factors (e.g., measurement errors, baseline differences), which create variation that is captured by the heterogeneity statistics but are unlikely to be related to the intervention or be explainable by moderators. We found quite drastic reductions of the total within and between-study heterogeneity when we progressively restricted the sample to studies with large effective sample sizes. The estimates of the weighted average effect size were slightly reduced, but stayed close to 0.1 in all meta-regressions and retained statistical significance until we were down to very few studies. These results support the idea that important parts of the heterogeneity have random sources.

Our fourth research question concerned the effects of the interventions on secondary outcomes, mainly standardised tests of language, literacy, and mathematical skills but also some tests of other cognitive skills (e.g., intelligence tests, processing speed). We found a positive and strongly significant weighted average effect size of 0.37 (0.33 for standardised tests of language, literacy and maths skills). This effect was large in comparison to effects normally seen in educational interventions (Kraft, 2020), and comparable to the average effects of tutoring or small-group interventions (e.g., Dietrichson et al., 2017, Dietrichson, Filges et al., 2020, 2021; Kraft et al., 2024; Nickow et al., 2024). This result was not all too surprising since a majority of effect sizes in the meta-analysis came from tutoring interventions. However, that the included interventions had strong effects on what was typically the primary target of the intervention may mean that other educational interventions, which have less strong effects on academic skills, might not produce similar effects on EF.

Overall completeness and applicability of evidence

Although around half of our studies were from the United States, 14 other countries contributed studies to the meta-analysis. These countries were diverse in terms of geography. We included studies from Europe, the Middle East, East Asia, Oceania, and North and South America. The countries were high income, upper middle income, and lower middle income according to the World Bank (2025) classification. However, high income countries were clearly overrepresented and more research from other types of countries is needed. Furthermore, we only included studies published in English, German, Danish, Swedish and Norwegian due to language restrictions in the review team. Studies from countries where these languages are not used may therefore be underrepresented in our review. Nonetheless, we included studies from many countries with other first languages than the ones understood by the review team (Brazil, China, Iran, Italy, Lebanon, the Netherlands, Slovakia, and Spain).

We aimed to include interventions with participants in both preschool and school, without restricting either the type of participants or the type of interventions further. Our sample covered a broad range of participants in both preschool and school, but studies set in school and interventions including only students experiencing, or at risk of experiencing, academic difficulties were the majority. Although the effects were not significantly different across these characteristics, our results are less representative of preschool interventions and student groups that are not at risk, or that are high-achieving. We included many different types of interventions but the majority used some form of (adult-led) tutoring. Interventions involving physical activity were also relatively numerous. We found no evidence of significantly different effects of these two types of interventions, but our sample was likely not sufficiently large to estimate differences between intervention types with adequate statistical power.

In terms of intervention content, we included interventions targeting language, literacy, and maths. All content areas were represented in our review, but relatively few interventions targeted only language and a majority of studies included some form of maths training. We included a range of EF outcomes, which we bundled into four categories in the moderator analysis: tests of inhibitory control and attention, short-term and working memory, cognitive flexibility, and tests of broader or composite EF skills like self-regulation, emotional control, and planning. The first two categories were most common in our data.

Quality of the evidence

Our meta-analyses were based on a considerable amount of data, comprising 51 studies. Our main effects-estimates were precisely estimated and robust across all sensitivity analyses. We found some evidence of small-study effects, which may indicate publication bias but could also be a consequence of the heterogeneity or of large effects being easier to obtain in small studies. Regardless, our estimates were not sensitive to publication bias and were still positive and significant also in a "worst-case" scenario (Mathur & VanderWeele, 2020) where we removed all positive and significant effect sizes from our sample.

The sample contained a large proportion of RCTs. Although the effects were on average larger in the four QES we included in the meta-analysis, the difference was not significant and our results were not driven by the QES. The overall risk of bias rating was relatively high, and no outcome received a low risk of bias rating in all domains. The main reason for the RCT's was the lack of pre-specified analysis plans. The confounding domain was the main problem for QES, all received a serious risk of bias rating in this domain. However, when we confined the analysis to outcomes from RCTs with one or more ratings of some concerns, the effects were slightly larger than in the primary analysis and still highly significant.

We found very few studies that measured outcomes with follow-up tests after the end of intervention. Thus, we do not know if the effects will fade out, like the effects of most educational interventions (e.g., Bailey et al., 2020; Dietrichson et al., 2026, Hart et al., 2024).

Most studies in our sample were small. The median across effect sizes was only 80 children. Because some studies assigned treatment on the class or school level, even this number is an exaggeration of the effective sample size. Consequently, very few effect sizes were individually significant in our sample and almost all studies did not have adequate statistical power to find effects similar in magnitude to the estimated average effect size.

As expected with small samples sizes across studies, we observed considerable heterogeneity, much of which could be explained by sampling error. However, the within and between-study parts of the heterogeneity was also substantial and, for the most part, remained unexplained in our moderator analyses. Because we included many types of interventions and outcomes, within and between-study heterogeneity is, at least to some extent, to be expected. However, we found few strong moderators whereas our analysis of the relation between heterogeneity and effective sample size indicated that some of the heterogeneity might have random sources. Nevertheless, the inconsistency of effects and the lack of significant moderators meant that it was not possible to identify certain types of interventions that were more effective than others as to improving EF skills.

Potential biases in the review process

Our results rest on a large search and screening effort. Our database search included 18,729 records and we put a lot of effort into searching other sources, which included 18,850 records. Despite the comprehensiveness of the search and screening, there were issues that may have caused bias. The search string included a facet related to the outcomes (i.e., EF tests). We included this facet since pilot searches indicated that the number of hits would otherwise far exceed what the review team was able to handle and would have included a large proportion of irrelevant records. This restriction means that we risked missing relevant records because they did not mention the outcome in the title or abstract, or because they were not indexed with the relevant subject terms.

However, for the following reasons, we believe this risk to be sufficiently low to motivate the restriction: First, it is typically costly to test EF. For example, standardised tests of working memory often require that the tests are conducted by certified psychologists. Second, testing executive functions is relatively unusual and therefore often an important part of a study. Thus, for these two reasons, researchers may be more likely to mention these outcomes in the abstract compared to other types of tests. Finally, as mentioned earlier, we made extra efforts especially regarding backward and forward citation tracking, and we contacted several experts in the field, who had either conducted studies we included or written reviews about similar topics. Thus, we should only have missed studies that we did not find in the database search, that did not cite and were not cited by other relevant studies and reviews, and that were unknown to experts in the field.

Because of the proliferation of synonyms for EF and that the outcomes were typically not the primary outcomes of the studies, the screening and coding were complex. Furthermore, the review team included many persons. To mitigate the risk of bias from these features, we had regular and frequent meetings both during the screening and coding process. We used AIscreenR as a second screener after switching to single-human screening, which was an extra safeguard not mentioned in our protocol. Furthermore, especially at the start of the process, we checked full texts also of abstracts that did not clearly mention EF tests to make sure that they did not test EF anyway. The fact that reaching out to experts in the field did not yield any studies that we had not already screened may serve as some evidence that we managed to mitigate the risk of missing relevant studies.

To increase consistency of the coding and data extraction, the first and second author was one of at least two review team members that assessed and coded each study (except for Seerup et al., 2025, which was coded by a review team member and an external colleague, who were not co-authors of that study). The first author was one of at least two team members that extracted data from every study.

Our protocol specified that we would also search ProQuest Dissertations & Theses Global. However, at the time of the search, we no longer had access to this database. No dissertation, neither from the search of EBSCO Open Dissertations nor from the other electronic databases, which also included dissertations, was included in our meta-analysis. Thus, we believe that this deviation was unlikely to bias our results. Similarly, the number of potentially relevant studies that we could not retrieve was small (eight, less than one percent of the studies screened in full text). Potentially more problematic was the three studies that we could not include in the meta-analysis due to lack of information needed to calculate effect sizes. However, to move our primary estimate substantially, the effects in these studies would have to be outliers.

Agreements and disagreements with other studies or reviews

To the best of our knowledge, there is no other meta-analysis of the effects of language, literacy, and maths interventions on EF skills. That is, there is no direct comparison to the average effect size from our primary analysis. Narrative reviews like Clements et al. (2016) and Peng and Kievit (2020) have suggested effects in the direction that we found. The tendencies of stronger associations between maths and EF in Clements et al. (2016) and that effects were smaller for at-risk children in Peng and Kievit (2020) were not robustly replicated in our data. Maths interventions had larger effects than language and literacy interventions but the difference was not significant, and the at-risk indicator was typically negative but always close to zero.

Reviews of physical activity interventions, which often include components training academic skills, have found either null effects or positive effects on cognitive skills (Alvarez-Bueno et al., 2017; Jylänki et al., 2022; Mavilidi et al., 2022; Muir et al., 2023; Norris et al., 2020; Peiris et al., 2022; Ruhland & Lange, 2021; Vasilopoulous et al., 2023). None of these reviews included a meta-analysis of language, literacy, and maths interventions on EF outcomes. As the null effects might have been caused by lack of statistical power, these reviews did not contradict our results.

Our results that training academic skills on average improved EF were in line with meta-analyses and reviews finding effects of years of schooling and preschool programs on broader measures of cognitive skills (Dietrichson, Kristiansen, & Viinholt, 2020; Duncan et al., 2023; Duncan & Magnusson, 2013; Ritchie & Tucker-Drob, 2018; van Huizen & Plantenga, 2018). The results were also in line with positive effects of Direct Instruction interventions on cognitive skills (Stockard et al., 2018). That is, our results contribute to a large and growing body of evidence that domain-general cognitive skills like EF are malleable and are, on average, positively affected by educational interventions.

Implications for practice and policy

The main finding of this review was that preschool and school-based language, literacy, and mathematics interventions have positive, robust, and statistically significant average effects on measures of EF. A straightforward implication of our results is that preschools and schools should continue to conduct language, literacy, and mathematics interventions. The motivation to conduct such interventions and implement policies that support them is stronger than they would be, had the effects been confined to academic skills.

It is important to note that although we found many examples of negative effect sizes, only two of them were statistically significant. As such, we found no robust evidence of language, literacy and maths interventions having negative effects on EF. With many small studies and an average effect size of 0.14, it would be very surprising, for statistical reasons alone, if there were no negative estimates. Our examination of the relation between heterogeneity and effective sample sizes also indicated that a substantial part of the heterogeneity may have random sources. Furthermore, we are not aware of any theory predicting negative effects on EF when training academic skills. However, based solely on our review, we cannot definitively establish whether all negative effect size estimates stem from sampling error and additional random sources, or if some types of academic skill interventions could negatively affect EF.

The implications for what type of interventions that should be implemented, and whether they should include components that also, or only, train EF are less straightforward. In the absence of robust evidence of larger effects of any type of intervention on EF skills, implementing interventions that have large effects on language, literacy, or maths seem like a good choice – at least for target groups that have difficulties with those subjects. There is extensive and robust evidence that some interventions can substantially improve language, literacy, and maths (see e.g., Baye et al., 2019; Dietrichson et al., 2017, 2020a, 2021; Fryer, 2017; Neitzel et al., 2022; Pellegrini et al., 2021; Slavin et al., 2011, for reviews comparing intervention types). That the association was positive and significant between the study-level average effects on secondary outcomes and the effects on EF, supported this implication.

For preschools and schools wondering whether to include EF training or not in an intervention, our results have less bearing. Our results do not provide evidence of whether combining EF and academic skills training is better or worse than just training academic skills, or, for that matter, evidence against EF training in general. As discussed in the introduction, the evidence of transfer effects of EF training to academic skills is not robust. However, most studies of EF training, the vast majority of which have examined working memory training, have relied on domain-general tasks (Fuchs et al., 2016). Domain-specific EF training could be more fruitful and there may be synergies created by combining EF and academic skills training (e.g., Doebel, 2020; Fuchs et al., 2022; Peng & Swanson, 2022). Furthermore, the end-of-intervention effects of EF training on the trained EF skill are robust. For children whose primary difficulty is EF, an EF-only or EF and academic skills training program might be the best option but more research is needed on what program to give to which children and how to combine EF and academic skills training.

Implications for research

Our review has some implications for future research. The results were in line with theories predicting bidirectional or reciprocal effects between EF and academic skills (e.g., Castles et al., 2018; Clements et al., 2016; Connor, 2016; Peng & Kievit, 2020), or theories that emphasise the unidirectional relationship from academic skills to EF skills (Demetriou et al., 2014; Gathercole et al., 2019). The results also provide evidence against theories predicting unidirectional effects from basic or domain-general cognitive skills like EF to academic skills.

Our review did not identify the mechanisms through which academic skills training affect EF. However, some of our results may have bearing on the question of mechanisms. The included interventions differed greatly in their duration. We were not able to calculate the number of intervention hours for a relatively large proportion of interventions but they ranged from 1.6 hours to 640 hours for the ones with information, and were highly correlated with the duration (0.76). That the association between intervention duration and effect sizes was not significant might be therefore be evidence that the mechanism through which language, literacy, and maths interventions affect EF is not only that neurocognitive skills like EF develop when they are used. That is, if this was the only mechanism, we would have expected larger effects in longer interventions that include more practice and a stronger association between duration and effect sizes.

Mechanisms that posit that training academic skills create new cognitive routines or increase domain-specific knowledge has a more ambiguous relation to the duration but would likely imply a more positive effect on EF from interventions that increase academic skills more. We found an association in this direction. More studies that are able to identify mechanisms are needed and would be very valuable – both for the development of theory and better interventions.

We were unable to explain much of the heterogeneity between and within studies, and most of our moderators were not statistically significant. Several of the moderators were not precisely estimated null effects and it might be the case that some types of interventions have larger effects on EF skills, and that some EF skills are more responsive to academic skill training than others.

Most studies that we included in our analysis were RCTs and we believe that the random assignment of treatment and the controlled nature of RCTs are beneficial features for testing these types of questions. To learn more about which interventions that have effects on which EF skill, future studies could incorporate three additional features: First, pre-registered analysis plans. Second, larger samples to improve statistical power. Third, testing alternative treatments against one another in the same study and on the same outcomes. However, there may be a conflict between incorporating larger samples and keeping implementation fidelity high. Therefore, medium-sized studies with high fidelity that can later be meta-analysed would also be valuable, even if they are underpowered.

As discussed in the previous section, our review did not answer the question of which type of skill training – EF or academic skills – that should be prioritized, or whether this prioritization should differ depending on the target group. The preferred study design for answering such questions would be to include treatments that train academic skills only, EF skills only, and combinations of academic and EF skill training alongside a control condition. Such multi-treatment designs are obviously difficult to implement (Fuchs et al., 2022 is the one example we are aware of), especially if sample sizes should be large enough to estimate differences with high precision. But, as mentioned, small studies may have a fidelity advantage and can be combined using meta-analysis, and studies including fewer contrasts would also move the literature forward (see Ahmadi et al., 2023; Barnes et al., 2016; Clements et al., 2020; Dolan et al., 2022; Fuchs et al., 2016; Goodrich et al., 2023; Kroesbergen et al., 2012; Kuhn & Holling, 2014; Passolunghi et al., 2016; and Ramani et al., 2017, for examples of studies that contrast an academic skills training condition to an EF training condition, or to a combination of EF and academic skills training).

References

References to included studies

Aadland K. N., Ommundsen Y., Anderssen, S. A., Brønnick, K. S., Moe, V. F., Resaland, G. K., Skrede, T., Stavnsbo, M., & Aadland E. (2019). Effects of the Active Smarter Kids (ASK) physical activity school-based intervention on executive functions: A cluster-randomized controlled trial. *Scandinavian Journal of Educational Research*, *63*(2), 214–228.

Ahmadi, A., Chuang, S. S., McClelland, M., Gonzales, C. R., & Beh-Pajooh, A. (2023). Executive functioning and early math skills in young children at risk for mathematical difficulties: Evaluation of intervention efficacy and transfer effects. *Early Education and Development*, 1–28. https://doi.org/10.1080/10409289.2023.2298166

Ansah, E. E. (2022). *Assessing the visuospatial working memory of school children undergoing abacus training* [Doctoral dissertation, University of Ghana]. Retrieved from http://ugspace.ug.edu.gh.

Aragon, E., Menacho, I., Navarro, J. I., & Aguilar, M. (2024). Teaching strategies, cognitive factors and mathematics. *Heliyon*, *10*, e29831.

Aragon-Mendizabal, E. L., Canto-Lopez, M. D. C., Marchena-Consejero, E., Navarro-Guzman, J. I., & Aguilar-Villagran, M. (2017). Cognitive profile in learning mathematics with open calculation based on numbers algorithm. *Revista de Psicodidactica*, *22*(1), 54-59. https://doi.org/10.1387/RevPsicodidact.16396

Balan, A., & Green, J. (2019). Effekten av fysisk aktivitet i matematikundervisningen. *Forskning Om Undervisning Och Lärande*, *7*(3), 6–27.

Barner, D., Athanasopoulou, A., Chua, J., Lewis, M., Marchanda, E., Schneider, R., & Frank, M. (2017). A one-year classroom-randomized trial of mental abacus instruction for first- and second-grade students. *Journal of Numerical Cognition*, *3*(3), 540–558.

Barnes, M. A., Klein, A., Swank, P., Starkey, P., McCandliss, B., Flynn, K., Zucker, T., Huang, C-W., Fall, A-M., & Roberts, G. (2016). Effects of tutorial interventions in mathematics and attention for low-performing preschool children. *Journal of Research on Educational Effectiveness*, *9*(4), 577–606.

Beck, M. M., Lind, R. R., Geertsen, S. S., Ritz, C., Lundbye-Jensen, J., & Wienecke, J. (2016). Motor-enriched learning activities can improve mathematical performance in preadolescent children. *Frontiers in Human Neuroscience*, *10*, Article 645.

Bhavya, R. L., Yerrabelli, D., & Philip, R. A. (2022). Assessment and comparison of cognitive function tests in abacus trained and untrained students aged 8-12 years in the South-Indian population. *Biomedicine*, *42*(1), 127–133. https://doi.org/10.51248/.v42i1.44

Boat, R., Cooper, S. B., Carlevaro, F., Magno, F., Bardaglio, G., Musella, G., & Magistro, D. (2022). 16 weeks of physically active mathematics and English language lessons improves cognitive function and gross motor skills in children aged 8-9 years. *International Journal of Environmental Research and Public Health*, *19*(24), 16751. https://doi.org/10.3390/ijerph192416751

Boyer, N. & Ehri, L. C. (2011). Contribution of phonemic segmentation instruction with letters and articulation pictures to word reading and spelling in beginners. *Scientific Studies of Reading*, *15*(5), 440–470. https://doi.org/10.1080/10888438.2010.520778

Cirino, P. T., Miciak, J., Gerst, E., Barnes, M. A., Vaughn, S., Child, A., & Huston-Warren, E. (2017). Executive function, self-regulated learning, and reading comprehension: A training study. *Journal of Learning Disabilities*, *50*(4), 450–467. https://doi.org/10.1177/0022219415618497

Clements, D. H., Sarama, J., Layzer, C., Unlu, F., & Fesler, L. (2020). Effects on mathematics and executive function of a mathematics and play intervention versus mathematics alone. *Journal for Research in Mathematics Education*, *51*(3), 301–333.

Cuder, A., Vidoz, M., De Vita, C., Pellizzoni, S., & Passolunghi, M. C. (2022). Numerical training videos and early numerical achievement: A study on 3-year-old preschoolers. *Brain Sciences*, *12*, 88. https://doi.org/10.3390/brainsci12010088

Cui, J., Xiao, R., Ma, M., Yuan, L., Kodash, R. C., & Zhou, X. (2022). Children skilled in mental abacus show enhanced non-symbolic number sense. *Current Psychology*, *41*, 2053–2066.

Da Silva, E. R., Dos Santos Baldin, M., & Dos Santos, F. H. (2017). Cognitive effects of numeracy musical training in Brazilian preschool children: A prospective pilot study. *Psychology & Neuroscience*, *10*(3), 281–296. https://doi.org/10.1037/pne0000098

de Greeff, J. W., Hartman, E., Mullender-Wijnsma, M. J., Bosker, R. J., Doolaard, S., & Visscher, C. (2016). Long-term effects of physically active academic lessons on physical fitness and executive functions in primary school children. *Health Education Research*, *31*(2), 185–194. https://doi.org/10.1093/her/cyv102

DeFlorio, L., Klein, A., Starkey, P., Swank, P. R., Taylor, H. B., Halliday, S. E., Beliakoff, A., & Mulcahy, C. (2019). A study of the developing relations between self-regulation and mathematical knowledge in the context of an early math intervention. *Early Childhood Research Quarterly*, *46*, 33–48. https://doi.org/10.1016/j.ecresq.2018.06.008

Dehghani, Y., Hoseini, F. S., & Jamshidi, F. (2024). The effects of EASY Minds program on working memory and selective attention in students with math learning disabilities. *International Journal of Disability, Development and Education*, *71*(4), 620–632. https://doi.org/10.1080/1034912X.2022.2140782

Dolan, C. T., Kim, H. Y., Brown, L., Gjicali, K., Borsani, S., El Houchaimi, S., & Aber, J. L. (2022). Supporting Syrian refugee children's academic and social-emotional learning in national education systems: A cluster randomized controlled trial of nonformal remedial support and mindfulness programs in Lebanon. *American Educational Research Journal*, *59*(3), 419–460. https://doi.org/10.3102/00028312211062911

Fuchs, D., Elleman, A. M., Fuchs, L. S., Peng, P., Kearns, D. M., Compton, D. L., Gilbert, J. K., & Patton III, S. (2016). *A randomized control trial of explicit instruction with and without cognitive training to strengthen the reading comprehension of poor readers in first grade*. Unpublished manuscript.

Fuchs, L. S., Fuchs, D., Sterba, S. K., Barnes, M. A., Seethaler, P. M., & Changas, P. (2022). Building word-problem solving and working memory capacity: A randomized controlled trial comparing three intervention approaches. *Journal of Educational Psychology*, *114*(7), 1633–1653. https://doi.org/10.1037/edu0000752

Goodrich, J. M., Peng, P., Bohaty, J., Leiva, S., & Thayer, L. (2023). Embedding executive function training into early literacy instruction for dual language learners: A pilot study. *Journal of Speech, Language, and Hearing Research*, *66*, 573–588. https://doi.org/10.1044/2022_JSLHR-22-00253

Harris, S. (2010). *Early intervention for poverty-stricken children: A study of preschoolers receiving jumpstart* [Doctoral dissertation, Illinois State University]. ProQuest Information & Learning, 3528245.

Have, M., Nielsen, J. H., Ernst, M. T., Gejl, A. K., Fredens, K., Grøntved, A., & Kristensen, P. L. (2018). Classroom-based physical activity improves children's math achievement – A randomized controlled trial. *PLoS ONE*, *13*(12), e0208787. https://doi.org/10.1371/journal.pone.0208787

Hempenstall, K. (2008). Corrective Reading: An evidence-based remedial reading intervention. *Australasian Journal of Special Education*, *32*(1), 23–54. https://doi.org/10.1017/S1030011200025744
Holder-Spriggs, J. (2015). *Expressive writing interventions for children and young people: A systematic review and exploration of the literature* [Doctoral dissertation, University of Southampton]. EBSCO Open Dissertations. Retrieved from https://eprints.soton.ac.uk/389518/
Hu, Y., Geng, F., Tao, L., Hu, N., Du, F., Fu, K., & Chen, F. (2011). Enhanced white matter tracts integrity in children with abacus training. *Human Brain Mapping*, *32*, 10–21. https://doi.org/10.1002/hbm.20996
Jansen, B. R. J., De Lange, E., & Van der Molen, M. J. (2013). Math practice and its influence on math skills and executive functions in adolescents with mild to borderline intellectual disability. *Research in Developmental Disabilities*, *34*(5), 1815–1824. https://doi.org/10.1016/j.ridd.2013.02.022
Jimenez-Parra, J. F., Belando-Pedreno, N., & Valero-Valenzuela, A. (2023). The effects of the ACTIVE VALUES program on psychosocial aspects and executive functions. *International Journal of Environmental Research and Publich Health*, *20*(1), 595. https://doi.org/10.3390/ijerph20010595
Korkman, M., & Peltomaa, A. K. (1993). Preventive treatment of dyslexia by a preschool training program for children with language impairments. *Journal of Clinical Child Psychology*, 22(2), 277–287. https://doi.org/10.1207/s15374424jccp2202_13
Kovalcikova, I., Veerbeek, J., Vogelaar, B., Pridavkova, A., Ferjencik, J., Simcikova, E., & Tomkova, B. (2021). Domain-specific stimulation of executive functioning in low performing students with a Roma background: Cognitive potential of mathematics. *Education Sciences*, *11*, 285. https://doi.org/10.3390/educsci11060285
Kovalcikova, I., Veerbeek, J., Vogelaar, B., Klimovic, M., & Gogova, E. (2024). Tracing progress in children's executive functioning and language abilities related to reading comprehension via ExeFun-READ intervention. *Education Sciences*, *14*, 237. https://doi.org/10.3390/educsci14030237
Krajewski, K., Nieding, G., & Schneider, W. (2008). Kurz- und langfristige effekte mathematischer frühförderung im kindergarten durch das programm 'mengen, zählen, zahlen'. *Zeitschrift Für Entwicklungspsychologie Und Pädagogische Psychologie*, *40*(3), 135–146. https://doi.org/10.1026/0049-8637.40.3.135
Kroesbergen, E., Van't Noordende, J. E., & Kolkman, M. E. (2012). Number sense in low-performing kindergarten children: Effects of a working memory and an early math training. In Breznitz, Z., Rubinsten, O., Molfese, V., & Molfese, D. (eds.), *Reading, Writing, Mathematics and the Developing Brain: Listening to Many Voices* (pp. 295–313). Dordrecht: Springer Science + Business Media B. V.
Kuhn, J-T., & Holling, H. (2014). Number sense or working memory? The effect of two computer-based trainings on mathematical skills in elementary school. *Advances in Cognitive Psychology*, *10*(2), 59–67. https://doi.org/10.5709/acp-0157-2
Kyttälä, M., Kanerva, K., & Kroesbergen, E. (2015). Training counting skills and working memory in preschool. *Scandinavian Journal of Psychology*, *56*, 363–370. https://doi.org/10.1111/sjop.12221
Lee, Y-S., Lu, M-J., & Ko, H-P. (2007). Effects of skill training on working memory capacity. *Learning and Instruction*, *17*, 336–344. https://doi.org/doi:10.1016/j.learninstruc.2007.02.010
Llorente, C., Pasnik, S., Moorthy, S., Hupert, N., Rosenfeld, D., & Gerard, S. (2015). *Preschool teachers can use a PBS KIDS Transmedia Curriculum Supplement to support young children's mathematics learning: Results of a randomized controlled trial*.

Conference paper, Society for Research on Educational Effectiveness, retrieved from https://eric.ed.gov/?id=ED562360.

Lonigan, C. J., & Phillips, B. M. (2012). *Comparing skills-focused and self-regulation focused preschool curricula: Impacts on academic and self-regulatory skills*. Conference paper, Society for Research on Educational Effectiveness, retrieved from https://eric.ed.gov/?id=ED530178.

Lu, Y., Ma, M., Chen, G., & Zhou, X. (2021). Can abacus course eradicate developmental dyscalculia. *Psychology in the Schools*, *58*(2), 235–251. https://doi.org/10.1002/pits.22441

Magistro, D., Cooper, S. B., Carlevaro, F., Marchetti, I., Magno, F., Bardaglio, G., & Musella, G. (2022). Two years of physically active mathematics lessons enhance cognitive function and gross motor skills in primary school children. *Psychology of Sport & Exercise*, *63*, 102254. https://doi.org/10.1016/j.psychsport.2022.102254

Maiztegi-Kortabarria, J., Arribas-Galarraga, S., Luis-de Cos, I., Espoz-Lazo, S., & Valdivia-Moral P. (2024). Effect of an active break intervention on attention, concentration, academic performance, and self-concept in compulsory secondary education. *European Journal of Investigation in Health, Psychology and Education*, *14*, 447–462. https://doi.org/10.3390/ejihpe14030030

Mattera, S. K., Jacob, R., & Morris, P. A. (2018). *Strengthening children's math skills with enhanced instruction: The impacts of making Pre-K Count and High 5s on kindergarten outcomes*. MDRC. Retrieved from https://www.mdrc.org/work/publications/strengthening-childrens-math-skills-enhanced-instruction

Mavilidi, M. F., Lubans, D. R., Miller, A., Eather, N., Morgan, P. J., Lonsdale, C., Noetel, M., Karayanidis, F., Shaw, K., & Riley, N. (2020). Impact of the 'Thinking while Moving in English' intervention on primary school children's academic outcomes and physical activity: A cluster randomised controlled trial. *International Journal of Educational Research*, *102*, 101592. https://doi.org/10.1016/j.ijer.2020.101592

McCormick, M. P., Mattera S. K., Maier, M. F., Xia, S., Jacob, R., & Morris, P. A. (2022). Different settings, different patterns of impacts: Effects of a pre-K math intervention in a mixed-delivery system. *Early Childhood Research Quarterly*, *58*, 136–154. https://doi.org/10.1016/j.ecresq.2021.08.005

Messer, D. & Nash, G. (2018). An evaluation of the effectiveness of a computer-assisted reading intervention. *Journal of Research in Reading*, *41*(1), 140–158. https://doi.org/10.1111/1467-9817.12107

Na K-S., Lee, I. S., Park, J-H., Jung, H-Y., & Ryu, J-H. (2015). Association between abacus training and improvement in response inhibition: A case-control study. *Clinical Psychopharmacology and Neuroscience*, *13*(2), 163–167. http://dx.doi.org/10.9758/cpn.2015.13.2.163

Nelson, J. R., Stage. S. A., Epstein, M. H., & Pierce, C. D. (2005). Effects of a prereading intervention on the literacy and social skills of children. *Exceptional Children*, *72*(1), 29–45. https://doi.org/10.1177/001440290507200102

Pasnik, S., & Llorente, C. (2013). *Preschool teachers can use a PBS KIDS Transmedia Curriculum Supplement to support young children's mathematics learning: Results of a randomized controlled trial*. Education Development Center and SRI International. Retrieved from https://eric.ed.gov/?id=ED594258.

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

Ramani, G. B., Jaeggi, S. M., Dauberta, E. N., & Buschkuehl, M. (2017). Domain-specific and domain-general Training to improve kindergarten children's mathematics. *Journal of Numerical Cognition*, *3*(2), 468–495. http://dx.doi.org/10.5964/jnc.v3i2.31

Raskind, M. H., & Higgins, E. L. (1999). Speaking to read: The effects of speech recognition technology on the reading and spelling performance of children with learning disabilities. *Annals of Dyslexia*, *49*, 251–281. https://doi.org/10.1007/s11881-999-0026-9

Resaland, G. K., Aadland, E., Moe, V. F., Aadland, K. N., Skrede, T., Stavnsbo, M., Suominena L., Steene-Johannessen, J., Glosvik, Ø., Andersen, J. R., Kvalheim, O. M., Engelsrud, G., Andersen, L. B., Holme, I. M., Ommundsen, Y., Kriemler, S., van Mechelen, W., McKay H. A., Ekelund, U., & Anderssen, S. A. (2016). Effects of physical activity on schoolchildren's academic performance: The Active Smarter Kids (ASK) cluster-randomized controlled trial. *Preventive Medicine*, *91*, 322–328. http://dx.doi.org/10.1016/j.ypmed.2016.09.005

Roberts, G., Rane, S., Fall, A-M., Denton, C. A., Fletcher, J. M., & Vaughn, S. (2015). The impact of intensive reading intervention on level of attention in middle school students. *Journal of Clinical Child and Adolescent Psychology*, *44*(6), 942–953. https://doi.org/10.1080/15374416.2014.913251

Rosenfeld, D., Dominguez, X., Llorente, C., Pasnik, S., Moorthy, S., Huperta, N., Gerard, S., & Vidiksis, R. (2019). A curriculum supplement that integrates transmedia to promote earlymath learning: A randomized controlled trial of a PBS KIDS intervention. *Early Childhood Research Quarterly*, *49*, 241–253. https://doi.org/10.1016/j.ecresq.2019.07.003

Rowe, K. J., & Rowe, K. S. (1999). Investigating the relationship between students' attentive-inattentive behaviors in the classroom and their literacy progress. *International Journal of Educational Research*, *31*(1–2), 1–137. https://doi.org/10.1016/S0883-0355(98)00068-8

Schneider, W., Kuspert, P., Roth, E., Vise, M., & Marx, H. (1997). Short-and long-term effects of training phonological awareness in kindergarten: Evidence from two German studies. *Journal of Experimental Child Psychology*, *66*(3), 311–340. https://doi.org/10.1006/jecp.1997.2384

Schneider, W., Vise, M., Reimers, P., & Blaesser, B. (1994). Effects of a phonological awareness training on the acquisition of literacy. *Zeitschrift für Pädagogische Psychologie*, *8*(3–4), 177–188.

Seerup, J. K., Dietrichson, J., Jensen, V. M., & Bingley, P. (2023). Tutoring young children at risk of developing reading difficulties: Effects of an early tutoring intervention on literacy skills and executive functions. In Seerup, J. K. *Ready to read: Exploring school-based tutoring interventions for struggling readers*, Dissertation, Department of Sociology, University of Copenhagen.

Seerup, J. K., Dietrichson, J., Jensen, V. M., & Bingley, P. (2025). Tutoring young children at risk of developing reading difficulties: Effects of an early tutoring intervention on literacy skills and executive functions. *Journal of Education for Students Placed at Risk*, online access. https://doi.org/10.1080/10824669.2025.2468701

Smith, H. (2010). *The effects of a drama-based language intervention on the development of theory of mind and executive function in urban kindergarten children* [Doctoral dissertation, Georgia State University]. ProQuest Information & Learning. https://scholarworks.gsu.edu/cps_diss/54/

Swanson, H. L. (2014). Does cognitive strategy training on word problems compensate for working memory capacity in children with math difficulties? *Journal of Educational Psychology*, *106*(3), 831–848. https://doi.org/10.1037/a0035838

Swanson, H. L. (2015). Cognitive strategy interventions improve word problem solving and working memory in children with math disabilities. *Frontiers in Psychology*, *6*, 1099. https://doi.org/10.3389/fpsyg.2015.01099

Swanson, H. L. (2016). Word problem solving, working memory and serious math difficulties: Do cognitive strategies really make a difference? *Journal of Applied Research in Memory and Cognition*, *5*, 368–383. https://doi.org/10.1016/j.jarmac.2016.04.012

Swanson, H. L., Moran, A. S., Bocian, K., Lussier, C., & Zheng, X. (2012). Generative strategies, working memory, and word problem solving accuracy in children at risk for math disabilities. *Learning Disability Quarterly*, *36*(4), 203–214. https://doi.org/10.1177/0731948712464034

Swanson, H. L., & O'Connor R. (2009). The role of working memory and fluency practice on the reading comprehension of students who are dysfluent readers. *Journal of Learning Disabilities*, *42*(6), 548–575. https://doi.org/10.1177/0022219409338742

Tam, I. O. L., & Leung, C. (2019). Evaluation of the effectiveness of a literacy intervention programme on enhancing learning outcomes for secondary students with dyslexia in Hong Kong. *Dyslexia*, *25*(3), 296–317. https://doi.org/10.1002/dys.1626

Vaughn, S., Mathes, P., Linan-Thompson, S., Cirino, P., Carlson, C., Pollard-Durodola, S., Cardenas-Hagan, E., & Francis, D. J. (2006c). Effectiveness of an English intervention for first-grade English language learners at risk for reading problems. *Elementary School Journal*, *107*(2), 153–180.

Vaughn, S., Linan-Thompson, S., Mathes, P. G., Cirino, P. T., Carlson, C. D., Pollard-Durodola S. D., Cardenas-Hagan, E., & Francis, D. J. (2006b). Effectiveness of Spanish intervention for first-grade English language learners at risk for reading difficulties. *Journal of Learning Disabilities*, *39*(1), 56–73. https://doi.org/10.1177/00222194060390010601

Vaughn, S., Cirino, P., Linan-Thompson, S., Mathes, P., Carlson, C., Cardenas-Hagan, E., Pollard-Durodola, S., Fletcher, J. & Francis, D. J. (2006a). Effectiveness of a Spanish and an English intervention for English-language learners at risk for reading problems. *American Educational Research Journal*, *43*(3), 449–487. https://doi.org/10.3102/00028312043003449

Vazou, S., Long, K., Lakes, K. D., & Whalen, N. L. (2021). "Walkabouts" integrated physical activities from preschool to second grade: Feasibility and effect on classroom engagement. *Child & Youth Care Forum*, *50*, 39–55. https://doi.org/10.1007/s10566-020-09563-4

Wang, C., Geng, F., Yao, Y., Weng, J., Hu, Y., & Chen, F. (2015). Abacus training affects math and task switching abilities and modulates their relationships in Chinese children. *PLoS ONE*, *10*(10), e0139930. https://doi.org/10.1371/journal.pone.0139930

Wang, C., Weng, J., Yao, Y., Dong, S., Liu, Y., & Chen, F. (2017). Effect of abacus training on executive function development and underlying neural correlates in Chinese children. *Human Brain Mapping*, *38*(10), 5234–5249. https://doi.org/10.1002/hbm.23728

Wang, C., Xu, T., Geng, F., Hu, Y., Wang, Y., Liu, H., & Chen, F. (2019). Training on abacus-based mental calculation enhances visuospatial working memory in children. *Journal of Neuroscience*, *39*(33), 6439–6448. https://doi.org/10.1523/JNEUROSCI.3195-18.2019

Weiland, C., & Yoshikawa, H. (2013). Impacts of a prekindergarten program on children's mathematics, language, literacy, executive function, and emotional skills. *Child Development*, *84*(6), 2112–2130. https://doi.org/10.1111/cdev.12099

Wolff, U. (2016). Effects of a randomized reading intervention study aimed at 9-year-olds: A 5-year follow-up. *Dyslexia*, *22*(2), 85–100. https://doi.org/10.1002/dys.1529

References to excluded studies

Barner, D., Alvarez, G., Sullivan, J., Brooks, N., Srinivasan, M., & Frank, M. C. (2016). Learning mathematics in a visuospatial format: A randomized, controlled trial of mental abacus instruction. *Child Development*, *87*(4), 1146–1158. https://doi.org/10.1111/cdev.12515

Bergen-Cico, D., Razza, R., & Timmins, A. (2015). Fostering self-regulation through curriculum infusion of mindful yoga: A pilot study of efficacy and feasibility. *Journal of Child and Family Studies*, *24*, 3448–3461. https://doi.org/10.1007/s10826-015-0146-2

Herrera, L., Defior, S., & Lorenzo, O. (2007). Educational intervention in phonological awareness in preliterate children with Spanish and Tamazight as first language: Comparison of two training programmes. *Infancia y Aprendizaje*, *30*(1), 39–54. https://doi.org/10.1174/021037007779849718

Lawton, T. (2016). Improving dorsal stream function in dyslexics by training figure/ground motion discrimination improves attention, reading fluency, and working memory. *Frontiers in Human Neuroscience*, *10*, 397. https://doi.org/10.3389/fnhum.2016.00397

McCoy, D. C., Zuilkowski, S. S., Yoshikawa, H., & Fink, G. (2017). Early childhood care and education and school readiness in Zambia. *Journal of Research on Educational Effectiveness*, *10*(3), 482–506. https://doi.org/10.1080/19345747.2016.1250850

Peisner-Feinberg, E. S., Mokrova, I. L., & Anderson, T. L. (2017). *Effects of participation in the North Carolina pre-kindergarten program at the end of kindergarten: 2015-2016 statewide evaluation*. Chapel Hill, NC: The University of North Carolina, FPG Child Development Institute.

Sala, G., & Gobet, F. (2017). Does chess instruction improve mathematical problem-solving ability? Two experimental studies with an active control group. *Learning & Behavior*, *45*, 414–421. https://doi.org/10.3758/s13420-017-0280-3

Scholz, M., Niesch, H., Steffen, O., Ernst, B., Loeffler, M., Witruk, E., & Schwarz, H. (2008). Impact of chess training on mathematics performance and concentration ability of children with learning disabilities. *International Journal of Special Education*, *23*(3), 138–148.

Toll, S. W. M., & van Luit, J. E. H. (2014). Structural support for kindergartners lagging behind in early numeracy. *Pedagogische Studien*, *91*(2), 82–96.

Additional references

Ahmed, S. F., Kuhfeld, M., Watts, T. W., Davis-Kean, P. E., & Vandell, D. L. (2021). Preschool executive function and adult outcomes: A developmental cascade model. *Developmental Psychology*, *57*(12), 2234–2249. https://doi.org/10.1037/dev0001270

Allan, N. P., Hume, L. E., Allan, D. M., Farrington, A. L., & Lonigan, C. J. (2014). Relations between inhibitory control and the development of academic skills in preschool and kindergarten: A meta-analysis. *Developmental Psychology*, *50*(10), 2368–2379. https://doi.org/10.1037/a0037493

Alloway, T. P., Gathercole, S. E., Kirkwood, H., & Elliott, J. (2008). Evaluating the validity of the automated working memory assessment. *Educational Psychology*, *28*(7), 725-734. https://doi.org/10.1080/01443410802243828

Alvarez-Bueno, C., Pesce, C., Cavero-Redondo, I., Sanchez-Lopez, M., Martínez-Hortelano, J. A., & Martinez-Vizcaino, V. (2017). The effect of physical activity interventions on children's cognition and metacognition: A systematic review and meta-analysis. *Journal of the American Academy of Child & Adolescent Psychiatry*, *56*(9), 729–738. https://doi.org/10.1016/j.jaac.2017.06.012

Anthony, J. L., & Francis, D. J. (2005). Development of phonological awareness. *Current Directions in Psychological Science*, *14*(5), 255–259. https://doi.org/10.1111/j.0963-7214.2005.00376.x

Araujo, M. C., Carneiro, P., Cruz-Aguayo, Y., & Schady, N. (2016). Teacher quality and learning outcomes in kindergarten. *Quarterly Journal of Economics*, *131*(3), 1415–1453. https://doi.org/10.1093/qje/qjw01

Au, J., Sheehan, E., Tsai, N., Duncan, G. J., Buschkuehl, M. & Jaeggi, S. M. (2015). Improving fluid intelligence with training on working memory: A meta-analysis. *Psychonomic Bulletin & Review*, *22*(2) 366–377. https://doi.org/10.3758/s13423-014-0699-x

Bailey, D. H., Duncan, G. J., Cunha, F., Foorman, B. R., & Yeager, D. S. (2020). Persistence and fade-out of educational-intervention effects: Mechanisms and potential solutions. *Psychological Science in the Public Interest*, *21*(2), 55–97. https://doi.org/10.1177/1529100620915848

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin*, *128*(4), 612–637. https://doi.org/10.1037/0033-2909.128.4.612

Baye, A., Inns, A., Lake, C., & Slavin, R. E. (2019). A synthesis of quantitative research on reading programs for secondary students. *Reading Research Quarterly*, *54*(2), 133–166. https://doi.org/10.1002/rrq.229

Berger, E. M., Fehr, E., Hermes, H., Schunk, D., & Winkel, K. (2025). The impact of working-memory training on children's cognitive and noncognitive Skills. *Journal of Political Economy*, *133*(2), 492–521. https://doi.org//10.1086/732884

Bishop, D. V. M., Nation, K., & Patterson, K. (2014). When words fail us: Insights into language processing from developmental and acquired disorders. *Philosophical Transactions of the Royal Society B*, *369*, 20120403. https://doi.org/10.1098/rstb.2012.0403

Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development*, *78*(2), 647–663. https://doi.org/10.1111/j.1467-8624.2007.01019.x

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Wiley & Sons Ltd.

Borenstein, M., Higgins, J. P. T., Hedges, L. V., & Rothstein, H. R. (2017). Basics of meta-analysis: I-squared is not an absolute measure of heterogeneity. *Research Synthesis Methods*, *8*(1), 5–18. https://doi.org/10.1002/jrsm.1230

Bradley, R. H., & Corwyn Robert, F. (2002). Socioeconomic status and child development. *Annual Review of Psychology*, *53*(1), 371–399. https://doi.org/10.1146/annurev.psych.53.100901.135233

Braginsky, M., Mathur, M., VanderWeele, T. J., & Solymos, P. (2023). *PublicationBias: Sensitivity analysis for publication bias in meta-analyses*. Retrieved from https://CRAN.R-project.org/package=PublicationBias.

Brunner, M., Keller, L., Stallasch, S. E., Kretschmann, J., Hasl, A., Preckel, F., & Hedges, L. V. (2021). *Meta-analyzing individual participant data from educational large-scale assessments: Methodological guidance and applications*. http://psyarxiv.com/t79vk/2021

Bus, A. G., Van Ijzendoorn, M. H., & Pellegrini, A. D. (1995). Joint book reading makes for success in learning to read: A meta-analysis on intergenerational transmission of literacy. *Review of Educational Research*, *65*(1), 1–21. https://doi.org/10.3102/00346543065001001

Bøg, M., Dietrichson, J., & Isaksson, A. A. (2021). A multi-sensory tutoring program for students at-risk of reading difficulties: Evidence from a field experiment. *Journal of Educational Research*, *114*(3), 233–251. https://doi.org/10.1080/00220671.2021.1902254

Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. *Psychological Science in the Public Interest*, *19*(2), 93–51. https://doi.org/10.1177/1529100618772271

Chen, M., & Pustejovsky, J. E. (2025). Adapting methods for correcting selective reporting bias in meta-analysis of dependent effect sizes. Psychological Methods, Advance online publica-tion. https://doi.org/10.1037/met0000773Cheung, A. C. K., & Slavin, R. E. (2016). How methodological features affect effect sizes in education. *Educational Researcher*, *45*(5), 283–292. https://doi.org/10.3102/0013189X16656615

Clements, D. H., Sarama, J., & Germeroth, C. (2016). Learning executive function and early mathematics: Directions of causal relations. *Early Childhood Research Quarterly*, *36*, 79–90. https://doi.org/10.1016/j.ecresq.2015.12.009

Connor, C. M. (2016). A lattice model of the development of reading comprehension. *Child Development Perspectives*, *10*(4), 269–274. https://doi.org/10.1111/cdep.12200

Connor, C. M., Day, S. L., Phillips, B., Sparapani, N., Ingebrand, S. W., Mclean, L., Barrus, A., & Kaschak, M. P. (2016). Reciprocal effects of self-regulation, semantic knowledge, and reading comprehension in early elementary school. *Child Development*, *87*(6), 1813–1824. https://doi.org/10.1111/cdev.12570

Cortese, S., Ferrin, M., Brandeis, D., Buitelaar, J., Daley, D., Dittmann, R.W., Holtmann, M., Santosh, P., Stevenson, J., Stringaris, A., Zuddas, A., Sonuga-Barke, E. J., & European ADHD Guidelines Group. (2015). Cognitive training for attention-deficit/hyperactivity disorder: Meta-analysis of clinical and neuropsychological outcomes from randomized controlled trials. *Journal of the American Academy of Child and Adolescent Psychiatry*, *54*(3), 164–174. https://doi.org/10.1016/j.jaac.2014.12.010

Cragg, L., & Nation, K. (2010). Language and the development of cognitive control. *Topics in Cognitive Science*, *2*(4), 631–642. https://doi.org/10.1111/j.1756-8765.2009.01080.x

Cuartas, J., Hanno, E., Lesaux, N. K., & Jones, S. M. (2022). Executive function, self-regulation skills, behaviors, and socioeconomic status in early childhood. *PLoS One*, *17*(11), e0277013. https://doi.org/10.1371/journal.pone.0277013

Delis, D. C., Kramer, J. H., Kaplan, E., & Holdknack, J. (2004). Reliability and validity of the Delis-Kaplan executive function system: An update. *Journal of the International Neuropsychological Society*, *10*(2), 301–303. https://doi.org/10.1017/S1355617704102191

Demetriou, A., Spanoudis, G., & Shayer, M. (2014). Inference, reconceptualization, insight, and efficiency along intellectual growth: A general theory. *Enfence*, *3*(3), 365–396. https://doi.org/10.3917/enf1.143.0365

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*, 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Diamond, A. (2014). Want to optimize executive functions and academic outcomes? Simple, just nourish the human spirit. In D. Sera (Ed.) *Minnesota Symposium in Child Psychology* (Vol. 37). John Wiley & Sons Inc. https://doi.org/10.1002/9781118732373.ch7

Diamond, A., & Ling, D. S. (2016). Conclusions about interventions, programs, and approaches for improving executive functions that appear justified and those that, despite much hype, do not. *Developmental Cognitive Neuroscience*, *18*, 34–48. https://doi.org/10.1016/j.dcn.2015.11.005

Dietrichson, J., Bhatnagar, R., Filges, T., Vembye, M. H. (2026). *On the mechanisms of fade-out: Meta-analytic evidence from school interventions*. Unpublished manuscript.

Dietrichson, J., Bøg, M., Filges, T., & Klint Jørgensen, A.-M. (2017). Academic interventions for elementary and middle school students with low socioeconomic status: A

systematic review and meta-analysis. *Review of Educational Research*, *87*(2), 243–282. https://doi.org/10.3102/0034654316687036
Dietrichson, J., Filges, T., Klokker, R. R., Viinholt, B. C. A., Bøg, M., & Jensen, U. H. (2020). Targeted school-based interventions for improving reading and mathematics for students with, or at risk of, academic difficulties in grades 7-12: A systematic review. *Campbell Systematic Reviews*, *16*(2), e1081. https://doi.org/10.1002/cl2.1081
Dietrichson, J., Kristiansen, I. L., & Viinholt, B. A. (2020). Universal preschool programs and long-term child outcomes: A systematic review. *Journal of Economic Surveys*, *34*(5), 1007–1043. https://doi.org/10.1111/joes.12382
Dietrichson, J., Filges, T., Seerup, J. K., Klokker, R. H., Viinholt, B. A., Bøg, M., & Eiberg, M. (2021). Targeted school-based interventions for improving reading and mathematics for students with, or at-risk of, academic difficulties in grades K-6: A systematic review. *Campbell Systematic Reviews*, *17*(2), e1152. https://doi.org/10.1002/cl2.1152
Dietrichson, J., Thomsen, M. K., Seerup, J. K., Strandby, M. W., Viinholt, B. C. A., & Bengtsen, E. (2022). PROTOCOL: School-based language, math, and reading interventions for executive functions in children and adolescents: A systematic review. *Campbell Systematic Reviews*, *18*(3), e1262. https://doi.org/10.1002/cl2.1262
Doebel, S. (2020). Rethinking executive function and its development. *Perspectives on Psychological Science*, *15*(4), 942-956. https://doi.org/10.1177/1745691620904771
Doebel, S., & Zelazo, P. D. (2016). Seeing conflict and engaging control: Experience with contrastive language benefits executive function in preschoolers. *Cognition*, *157*, 219–226. https://doi.org/10.1016/j.cognition.2016.09.010
Donnelly, J. E., Hillman, C. H., Castelli, D., Etnier, J. L., Lee, S., Tomporowski, P., ... & Szabo-Reed, A. N. (2016). Physical activity, fitness, cognitive function, and academic achievement in children: A systematic review. *Medicine and Science in Sports and Exercise*, *48*(6), 1197–1222. https://doi.org/10.1249/MSS.0000000000000901
Duncan, G., Kalil, A., Mogstad, M., & Rege, M. (2023). Investing in early childhood development in preschool and at home. In Hanushek, E. A., Machin, S., Woessmann, L. (eds.) *Handbook of the Economics of Education*, *6*, 1–91. https://doi.org/10.1016/bs.hesedu.2022.11.005
Duncan, G. J., & Magnuson, M. (2013). Investing in preschool programs. *Journal of Economic Perspectives*, *27*(2), 109–132. https://doi.org/10.1257/jep.27.2.109
Eakin, L., Minde, K., Hechtman, L., Ochs, E., Krane, E., Bouffard, R., Greenfield, B., & Looper, K. (2004). The marital and family functioning of adults with ADHD and their spouses. *Journal of Attention Disorders*, *8*(1), 1–10. https://doi.org/10.1177/108705470400800101
Egger, M., Smith, G. D. S., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, *315*(7109), 629–634. https://doi.org/10.1136/bmj.315.7109.629
Emslander, V., & Scherer, R. (2022). The relation between executive functions and math intelligence in preschool children: A systematic review and meta-analysis. *Psychological Bulletin*, *148*(5-6), 337–369. https://doi.org/10.1037/bul0000369
Elridge, S., Campbell, M., Campbell, M., Dahota, A., Giruadeau, B., Higgins, J., Reeves, B., & Siegfried, N. (2016). *Revised Cochrane risk of bias tool for randomized trials (RoB 2.0). Additional considerations for cluster-randomized trials (pp. 1–17)*. Retrieved from https://www.riskofbias.info/welcome/rob-2-0-tool
Epstein, M. H., & Sharma, J. M. (1998). *Behavioral and emotional rating scale: A strength-based approach to assessment*. Austin, TX: PRO-ED. https://doi.org/10.1177/073724770002500304

Esping-Andersen, G., Garfinkel, I., Han, W. J., Magnuson, K., Wagner, S., & Waldfogel, J. (2012). Child care and school performance in Denmark and the United States. *Children and Youth Services Review*, *34*(3), 576–589. https://doi.org/10.1016/j.childyouth.2011.10.010
Fryer, R. G. (2017). The production of human capital in developed countries: Evidence from 196 randomized field experiments. In *Handbook of Economic Field Experiments* (Vol. 2, pp. 95-322). North-Holland. https://doi.org/10.1016/bs.hefe.2016.08.006
Follmer, D. J. (2018). Executive function and reading comprehension: A meta-analytic review. *Educational Psychologist*, *53*(1), 42–60. https://doi.org/10.1080/00461520.2017.1309295
Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin*, *134*(1), 31–60. https://doi.org/10.1037/0033-2909.134.1.31
Gathercole, E., Dunning, D. L., Holmes, J., & Norris, D. (2019). Working memory training involves learning new skills. *Journal of Memory and Language*, *105*, 19–42. https://doi.org/10.1016/j.jml.2018.10.003
Gioia, G. A., Andrewes, K., & Isquith, P. K. (1996). *Behavior rating inventory of executive function-preschool version (BRIEF-P)*. Psychological Assessment Resources.
Glennerster, R., & Takavarasha, K. (2013). *Running randomized evaluations: A practical guide*. Princeton University Press. https://doi.org/10.1515/9781400848447
Goldstein, D. M. (1976). Cognitive-linguistic functioning and learning to read in preschoolers. *Journal of Educational Psychology*, *68*(6), 680–688. https://doi.org/10.1037/0022-0663.68.6.680
Golinkoff, R. M., Hoff, E., Rowe, M. L., Tamis-LeMonda, C. S., & Hirsh-Pasek, K. (2019). Language matters: Denying the existence of the 30-million-word gap has serious consequences. *Child Development*, *90*(3), 985–992. https://doi.org/10.1111/cdev.13128
Gooch, D., Thompson, P., Nash, H. M., Snowling, M. J., & Hulme, C. (2016). The development of executive function and language skills in the early school years. *Journal of Child Psychology and Psychiatry*, *57*(2), 180–187. https://doi.org/10.1111/jcpp.12458
Hart, E. R., Bailey, D. H., Luo, S., Sengupta, P., & Watts, T. W. (2024). Fadeout and persistence of intervention impacts on social-emotional and cognitive skills in children and adolescents: A meta-analytic review of randomized controlled trials. *Psychological Bulletin*, *150*(10), 1207–1236. https://doi.org/10.1037/bul0000450
Hart, B., & Risley, T. R. (2003). The early catastrophe: The 30-million word gap by age 3. *American Educator*, *27*(1), 4–9.
Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, *6*(2), 107–128. https://doi.org/10.3102/10769986006002107
Hedges, L. V. (2007). Effect sizes in cluster-randomized designs. *Journal of Educational and Behavioral Statistics*, *32*(4), 341–370. https://doi.org/10.3102/1076998606298043
Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, *1*(1), 39–65. https://doi.org/10.1002/jrsm.5
Hedges, L. V., Tipton, E., Zejnullahi, R., & Diaz, K. G. (2023). Effect sizes in ANCOVA and difference-in-differences designs. British Journal of Mathematical and Statistical Psychology, 76(2), 259-282. https://doi.org/10.1111/bmsp.12296
Hedges, L. V., & Hedberg, E. C. (2007). Intraclass correlation values for planning group-randomized trials in education. *Educational Evaluation and Policy Analysis*, *29*(1), 60–87. https://doi.org/10.3102/0162373707299706

Hernández, M. M., Eisenberg, N., Valiente, C., Spinrad, T. L., Johns, S. K., Berger, R. H., Silva, K. M., Diaz, A., Gal-Szabo, D. E., Thompson, M. S., & Southworth, J. (2018). Self-regulation and academic measures across the early elementary school grades: Examining longitudinal and bidirectional associations. *Early Education and Development*, *29*(7), 914–938. https://doi.org/10.1080/10409289.2018.1496722

Higgins J. P. T, Eldridge S., & Li, T. (2019). Chapter 23: Including variants on randomized trials [last updated October 2019]. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane Handbook for Systematic Reviews of Interventions version 6.5. Cochrane, 2024*. Available from www.training.cochrane.org/handbook.

Higgins, J. P. T., & Green, S. (2011). *Cochrane handbook for systematic reviews of interventions*. Wiley Blackwell.

Higgins, J. P. T., Savovic, M., Page, J. & Sterne, J. A. C. (Eds.). (2019). *Revised Cochrane risk-of-bias tool for randomized trials (RoB 2): Detailed guidance*. Retrieved from http://www.riskofbias.info

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ*, *327*(7414), 557–560. https://doi.org/10.1136/bmj.327.7414.557

Hillman, C. H., McAuley, E., Erickson, K. I., Liu-Ambrose, T., & Kramer, A. F. (2019). On mindful and mindless physical activity and executive function: A response to Diamond and Ling (2016). *Developmental Cognitive Neuroscience*, *37*. https://doi.org/10.1016/j.dcn.2018.01.006

Hjetland, H. N., Brinchmann, E. I., Ronny, R., & Melby-Lervåg, M. (2017). Preschool predictors of later reading comprehension ability: A systematic review. *Campbell Systematic Reviews*, *13*(1), 1–155. https://doi.org/10.4073/csr.2017.14

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing*, *2*, 127–160. https://doi.org/10.1007/BF00401799

Jacob, R., & Parkinson, J. (2015). The potential for school-based interventions that target executive function to improve academic achievement: A review. *Review of Educational Research*, *85*(4), 512–552. https://doi.org/10.3102/0034654314561338

James, W. (1890). *Principles of psychology* (Vol. 1). Dover Publications Inc.

Jones, A., Atkinson, J., Marshall, C., Botting, N., St Clair, M. C., & Morgan, G. (2020). Expressive vocabulary predicts nonverbal executive function: A 2-year longitudinal study of deaf and hearing children. *Child Development*, *91*(2), e400–e414. https://doi.org/10.1111/cdev.13226

Jurado, M. B., & Rosselli, M. (2007). The elusive nature of executive functions: a review of our current understanding. *Neuropsychology Review*, *17*(3), 213–233. https://doi.org/10.1007/s11065-007-9040-z

Jylänki, P., Mbay, T., Hakkarainen, A., Sääkslahti, A., & Aunio, P. (2022). The effects of motor skill and physical activity interventions on preschoolers' cognitive and academic skills: A systematic review. *Preventive Medicine*, *155*, 106948. https://doi.org/10.1016/j.ypmed.2021.106948

Kassai, R., Futo, J., Demetrovics, Z., & Takacs, Z. K. (2019). A meta-analysis of the experimental evidence on the near-and far-transfer effects among children’s executive function skills. *Psychological Bulletin*, *145*(2), 165–188. https://doi.org/10.1037/bul0000180

Katz, B., Shah, P., & Meyer, D. E. (2018). How to play 20 questions with nature and lose: Reflections on 100 years of brain-training research. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(40), 9897–9904. https://doi.org/10.1073/pnas.1617102114

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, *49*(4), 241-253. https://doi.org/10.3102/0013189X20912798
Kraft, M. A., Schueler, B. E., & Falken, G. (2024). *What impacts should we expect from tutoring at scale? Exploring meta-analytic generalizability* (EdWorkingPaper No. 24-1031) Annenberg Institute for School Reform at Brown University. https://doi.org/10.26300/zygj-m525
Last, B. S., Lawson, G. M., Breiner, K., Steinberg, L., & Farah, M. J. (2018). Childhood socioeconomic status and executive function in childhood and beyond. *PloS One*, *13*(8), e0202964. https://doi.org/10.1371/journal.pone.0202964
Lawson, G. M., Hook, C. J., & Farah, M. J. (2018). A meta-analysis of the relationship between socioeconomic status and executive function performance among children. *Developmental Science*, *21*(2), 1–22. https://doi.org/10.1111/desc.12529
Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Sage Publications.
Luria, A. R. (1959). The directive function of speech in development and dissolution. *WORD*, *15*(2), 341–352. https://doi.org/10.1080/00437956.1959.11659703
Mathur, M. B., & VanderWeele, T. J. (2020). Sensitivity analysis for publication bias in meta-analyses. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *69*(5), 1091–1119. https://doi.org/10.1111/rssc.12440
Mavilidi, M. F., Pesce, C., Benzing, V., Schmidt, M., Paas, F., Okely, A. D., & Vazou, S. (2022). Meta-analysis of movement-based interventions to aid academic and behavioral outcomes: A taxonomy of relevance and integration. *Educational Research Review*, *37*, 100478. https://doi.org/10.1016/j.edurev.2022.100478
McClelland, M. M., & Cameron, C. E. (2012). Self-regulation in early childhood: Improving conceptual clarity and developing ecologically valid measures. *Child Development Perspectives*, *6*(2), 136–142. https://doi.org/10.1111/j.1750-8606.2011.00191.x
McClelland, M. M., & Cameron, C. E. (2019). Developing together: The role of executive function and motor skills in children's early academic lives. *Early Childhood Research Quarterly*, *46*, 142–151. https://doi.org/10.1016/j.ecresq.2018.03.014
McGuinness, L. A., & Higgins, J. P. (2021). Risk-of-bias VISualization (robvis): an R package and Shiny web app for visualizing risk-of-bias assessments. *Research Synthesis Methods*, *12*(1), 55-61. https://doi.org/10.1002/jrsm.1411
McGuinness, L. A., & Kothe, E. (2019). robvis: Visualize the Results of Risk-of-Bias (ROB) Assessments. Retrieved from https://CRAN.R-project.org/package=robvis.
Melby-Lervåg, M., & Hulme, C. (2010). Serial and free recall in children can be improved by training: evidence for the importance of phonological and semantic representations in immediate memory tasks. *Psychological Science*, *21*(11), 1694–1700. https://doi.org/10.1177/0956797610385355
Melby-Lervåg, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology*, *49*(2), 270–291. https://doi.org/10.1037/a0028228
Melby-Lervåg, M., Redick, T. S., & Hulme, C. (2016). Working memory training does not improve performance on measures of intelligence or other measures of far transfer: Evidence from a meta-analytic review. *Perspectives on Psychological Science*, *11*(4), 512-534. https://doi.org/10.1177/1745691616635612
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
Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., Houts, R., Poulton, R., Roberts, B. W., Ross, S., Sears, M. R., Thomson, W. M., & Caspi, A.

(2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(7), 2693–2698. https://doi.org/10.1073/pnas.1010076108

Morgan, P. L., Farkas, G., Hillemeier, M. M., & Maczuga, S. (2012). Are minority children disproportionately represented in early intervention and early childhood special education? *Educational Researcher*, *41*(9), 339–351. https://doi.org/10.3102/0013189X12459678

Morrison, F. J., Kim, M. H., Connor, C. M., & Grammer, J. K. (2019). The causal impact of schooling on children's development: Lessons for developmental science. *Current Directions in Psychological Science*, *28*(5), 441–449. https://doi.org/10.1177/0963721419855661

Muir, R. A., Howard, S. J., & Kervin, L. (2023). Interventions and approaches targeting early self-regulation or executive functioning in preschools: A systematic review. *Educational Psychology Review*, *35*(1), article 27. https://doi.org/10.1007/s10648-023-09740-6

Munakata, Y., & Michaelson, L. E. (2021). Executive functions in social context: Implications for conceptualizing, measuring, and supporting developmental trajectories. *Annual Review of Developmental Psychology*, *3*(1), 139–163. https://doi.org/10.1146/annurev-devpsych-121318-085005

Neitzel, A. J., Lake, C., Pellegrini, M., & Slavin, R. E. (2022). A synthesis of quantitative research on programs for struggling readers in elementary schools. *Reading Research Quarterly*, *57*(1), 149–179. https://doi.org/10.1002/rrq.379

Nickow, A., Oreopoulos, P., & Quan, V. (2024). The promise of tutoring for PreK–12 learning: A systematic review and meta-analysis of the experimental evidence. *American Educational Research Journal*, *61*(1), 74–107. https://doi.org/10.3102/0002831223120

Nigg, J. T., Sibley, M. H., Thapar, A., & Karalunas, S. L. (2020). Development of ADHD: Etiology, heterogeneity, and early life course. *Annual Review of Developmental Psychology*, *2*(1), 559–583. https://doi.org/10.1146/annurev-devpsych-060320-093413

Norris, E., van Steen, T., Direito, A., & Stamatakis, E. (2020). Physically active lessons in schools: A systematic review and meta-analysis of effects on physical activity, educational, health and cognition outcomes. *British Journal of Sports Medicine*, *54*(14), 826–838. https://doi.org/10.1136/bjsports-2018-100502

Oberauer, K., Lewandowsky, S., Awh, E., Brown, G., Conway, A., Cowan, N., Donkin, C., Farrell, S., Hitch, G. J., Hurlstone, M. J., Ma, W. J., Morey, C. C., Nee, D. E., Schweppe, J., Vergauwe, E., & Ward, G. (2018). Benchmarks for models of short-term and working memory. *Psychological Bulletin*, *144*(9), 885–958. https://doi.org/10.1037/bul0000153

OECD (2020). *Public spending on education (indicator)*. Retrieved from https://data.oecd.org/eduresource/public-spending-on-education.htm

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*, n71. https://doi.org/10.1136/bmj.n71

Peiris, D. L. I. H. K., Duan, Y., Vandelanotte, C., Liang, W., Yang, M., & Baker, J. S. (2022). Effects of in-classroom physical activity breaks on children’s academic performance, cognition, health behaviours and health outcomes: a systematic review and meta-analysis of randomised controlled trials. *International Journal of Environmental Research and Public Health*, *19*(15), 9479. https://doi.org/10.3390/ijerph19159479

Pellegrini, M., Lake, C., Neitzel, A., & Slavin, R. E. (2021). Effective programs in elementary mathematics: A meta-analysis. *Aera Open*, *7*(1), 1–29. https://doi.org/10.1177/23328584209862

Peng, P., Barnes, M., Wang, C., Wang, W., Li, S., Swanson, H. L., Dardick, W., & Tao, S. (2018). A meta-analysis on the relation between reading and working memory. *Psychological Bulletin*, *144*(1), 48–76. https://doi.org/10.1037/bul0000124

Peng, P., & Miller, A. C. (2016). Does attention training work? A selective meta-analysis to explore the effects of attention training and moderators. *Learning and Individual Differences*, *45*, 77-87. https://doi.org/10.1016/j.lindif.2015.11.012
Peng, P., Namkung, J., Barnes, M., & Sun, C. (2016). 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*(4), 455–473. https://doi.org/10.1037/edu0000079
Peng, P., & Kievit, R. A. (2020). The development of academic achievement and cognitive abilities: A bidirectional perspective. *Child Development Perspectives*, *14*(1), 15–20. https://doi.org/10.1111/cdep.12352
Peng, P., & Swanson, H. L. (2022). The domain-specific approach of working memory training. *Developmental Review*, *65*, 101035. https://doi.org/10.1016/j.dr.2022.101035
Polanin, J. R., Pigott, T. D., Espelage, D. L., & Grotpeter, J. K. (2019). Best practice guidelines for abstract screening large-evidence systematic reviews and meta-analyses. *Research Synthesis Methods*, *10*(3), 330–342. https://doi.org/10.1002/jrsm.1354
Prince, M., Patel, V., Saxena, S., Maj, M., Maselko, J., Phillips, M. R., & Rahman, A. (2007). No health without mental health. *Lancet*, *370*(9590), 859–877. 10.1016/S0140-6736(07)61238-0
Pustejovsky, J. E. (2021). *clubSandwich*. Retrieved from https://CRAN.R-project.org/package=clubSandwich
Pustejovsky, J. E., & Tipton, E. (2022). Meta-analysis with robust variance estimation: Expanding the range of working models. *Prevention Science*, *23*, 425–438. https://doi.org/10.1007/s11121-021-01246-3
Pustejovsky, J. E., Zhang, J., & Tipton, E. (2025). *A preliminary data analysis workflow for meta-analysis of dependent effect sizes*. MetaArXiv, https://doi.org/10.31222/osf.io/vfsqx_v1
Rapport, M. D., Orban, S. A., Kofler, M. J., & Friedman, L. M. (2013). Do programs designed to train working memory, other executive functions, and attention benefit children with ADHD? A meta-analytic review of cognitive, academic, and behavioral outcomes. *Clinical Psychology Review*, *33*(8), 1237–1252. https://doi.org/10.1016/j.cpr.2013.08.005
Redick, T. S., Shipstead, Z., Wiemers, E. A., Melby-Lervåg, M., & Hulme, C. (2015). What's working in working memory training? An educational perspective. *Educational Psychology Review*, *27*(4), 617–633. https://doi.org/10.1007/s10648-015-9314-6
Ritchie, S. J., & Tucker-Drob, E. M. (2018). How much does education improve intelligence? A meta-analysis. *Psychological Science*, *29*(8), 1358–1369. https://doi.org/10.1177/0956797618774253
Roberts, G. J., Dumas, D. G., McNeish, D., & Coté, B. (2022). Understanding the dynamics of dosage response: A nonlinear meta-analysis of recent reading interventions. *Review of Educational Research*, *92*(2), 209–248. https://doi.org/10.3102/00346543211051423
Rodgers, M. A., & Pustejovsky, J. E. (2021). Evaluating meta-analytic methods to detect selective reporting in the presence of dependent effect sizes. *Psychological Methods*, *26*(2), 141–160. https://doi.org/10.1037/met0000300
Rogde, K., Hagen, Å. M., Melby-Lervåg, M., & Lervåg, A. (2019). The effect of linguistic comprehension instruction on generalized language and reading comprehension skills: A systematic review. *Campbell Systematic Reviews*, *15*(4), e1059. https://doi.org/10.1002/cl2.1059
Romeo, R. R., Flournoy, J. C., McLaughlin, K. A., & Lengua, L. J. (2022). Language development as a mechanism linking socioeconomic status to executive functioning

development in preschool. *Developmental Science*, *25*(5), e13227. https://doi.org/10.1111/desc.13227
Ruhland, S., & Lange, K. W. (2021). Effect of classroom-based physical activity interventions on attention and on-task behavior in schoolchildren: A systematic review. *Sports Medicine and Health Science*, *3*(3), 125–133. https://doi.org/10.1016/j.smhs.2021.08.003
Sala, G., & Gobet, F. (2017). Working memory training in typically developing children: A meta-analysis of the available evidence. *Developmental Psychology*, *53*(4), 671–685. https://doi.org/10.1037/dev0000265
Schunk, D., Berger, E. M., Hermes, H., Winkel, K., & Fehr, E. (2022). Teaching self-regulation. *Nature Human Behaviour*, *6*(12), 1680–1690. https://doi.org/10.1038/s41562-022-01449-w
Schwaighofer, M., Fisher, F., & Bühner, M. (2015). Does working memory training transfer? A meta-analysis including training conditions as moderators. *Educational Psychologist*, *50*(2), 138–166. https://doi.org/10.1080/00461520.2015.1036274
Shipstead, Z., Redick, T. S., & Engle, R. W. (2012). Is working memory training effective? *Psychological Bulletin*, *138*(4), 628–654. https://doi.org/10.1037/a0027473
Silva, C., Sousa-Gomes, V., Fávero, M., Oliveira-Lopes, S., Merendeiro, C. S., Oliveira, J., & Moreira, D. (2022). Assessment of preschool-age executive functions: A systematic review. *Clinical Psychology & Psychotherapy*, *29*(4), 1374–1391. https://doi.org/10.1002/cpp.2718
Simons, D. J., Boot, W. R., Charness, N., Gathercole, S. E., Chabris, C. F., Hambrick, D. Z., & Stine-Morrow, E. A. (2016). Do "brain-training" programs work? *Psychological Science in the Public Interest*, *17*(3), 103–186. https://doi.org/10.1177/1529100616661983
Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research*, *75*(3), 417–453. https://doi.org/10.3102/00346543075003417
Slates, S. L., Alexander, K. L., Entwisle, D. R., & Olson, L. S. (2012). Counteracting summer slide: Social capital resources within socioeconomically disadvantaged families. Journal of *Education for Students Placed at Risk*, *17*(3), 165–185. https://doi.org/10.1080/10824669.2012.688171
Slavin, R., & Madden, N. A. (2011). Measures inherent to treatments in program effectiveness reviews. *Journal of Research on Educational Effectiveness*, *4*(4), 370–380. https://doi.org/10.1080/19345747.2011.558986
Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, *XXI*(4), 360–407. https://doi.org/10.1177/0022057409189001-204
Stebbings, J. A., & Kline, E. (2020). Academic and cognitive remediation for students with learning disabilities: A comparison between Orton-Gillingham and NILD Educational Therapy. *International Journal for Research in Learning Disabilities*, *4*(2), 15–34. https://doi.org/10.28987/ijrld.4.2.15
Sterne, J. A. C., Hernán, M. A., Reeves, B. C., Savovic, J., Berkman, N. D., Viswanathan, M., Henry, D., Altman, D. G., Ansari, M. T., Boutron, I., Carpenter, J. R., Chan, A.-W., Churchill, R., Deeks, J. J., Hróbjartsson, A., Kirkham, J., Jüni, P., Loke, Y. K., Pigott, T. D., …, & Higgins, J. P. T. (2016). Risk of bias in non-randomized studies of interventions (ROBINS-I): Detailed guidance. *BMJ*, *355*. https://doi.org/10.1136/bmj.i4919
Sterne, J., Savović, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., Cates, C. J., Cheng, H.-Y., Corbett, M. S., Eldridge, S. M., Emberson, J. R., Hernán, M. A., Hopewell, S., Hróbjartsson, A., Junqueira, D. R., Jüni, P., Kirkham, J. J., Lasserson, T., Li, T., …

Higgins, J. P. T. (2019). RoB 2: A revised tool for assessing risk of bias in randomised trials. *BMJ*, *366*, l4898. https://doi.org/10.1136/bmj.l4898
Stockard, J., Wood, T. W., Coughlin, C., & Rasplica Khoury, C. (2018). The effectiveness of direct instruction curricula: A meta-analysis of a half century of research. *Review of Educational Research*, *88*(4), 479–507. https://doi.org/10.3102/0034654317751919
Stoll, C. R. T., Izadi, S., Fowler, S., Green, P., Suls, J., & Colditz, G. A. (2019). The value of a second reviewer for study selection in systematic reviews. *Research Synthesis Methods*, *10*(4), 539–545. https://doi.org/10.1002/jrsm.1369
Swanson, J. M., Schuck, S., Porter, M. M., Carlson, C., Hartman, C. A., Sergeant, J., … Wigal, T. (2012). Categorical and dimensional definitions and evaluations of symptoms of ADHD: History of the SNAP and SWAN rating scales. *International Journal of Educational Psychological Assessment*, *10*(1), 51–70.
Taatgen, N. A. (2013). The nature and transfer of cognitive skills. *Psychological Review*, *120*(3), 439–471. https://doi.org/10.1037/a0033138
Tanner-Smith, E., & Tipton, E. (2013). Robust variance estimation with dependent effect sizes: practical considerations including a software tutorial in Stata and SPSS. *Research Synthesis Methods*, *5*(1), 13–30. https://doi.org/10.1002/jrsm.1091
Tanner-Smith, E., Tipton, E., & Polanin, J. R. (2016). Handling complex meta-analytic data structures using robust variance estimates: A tutorial in R. *Journal of Developmental and Life-Course Criminology*, *2*(1), 85–112. https://doi.org/10.1007/s40865-016-0026-5
Tipton, E. (2015). Small sample adjustments for robust variance estimation with meta-regression. *Psychological Methods*, *20*(3), 375–393. https://doi.org/10.1037/met0000011
Tipton, E., & Pustejovsky, J. E. (2015). Small-sample adjustments for tests of moderators and model fit using robust variance estimation in meta-regression. *Journal of Educational and Behavioral Statistics*, *40*(6), 604–634. https://doi.org/10.3102/1076998615606099
Valcan, D. S., Davis, H., & Pino-Pasternak, D. (2018). Parental behaviours predicting early childhood executive functions: A meta-analysis. *Educational Psychology Review*, *30*, 607–649. https://doi.org/10.1007/s10648-017-9411-9
Van Der Maas, H. L. J., Dolan, C. V., Grasman, R. P. P. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. J. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review*, *113*(4), 842–861. https://doi.org/10.1037/0033-295X.113.4.842
van Huizen, T., & Plantenga, J. (2018). Do children benefit from universal early childhood education and care? A meta-analysis of evidence from natural experiments. *Economics of Education Review*, *66*, 206–222. https://doi.org/10.1016/j.econedurev.2018.08.001
Vasilopoulos, F., Jeffrey, H., Wu, Y., & Dumontheil, I. (2023). Multi-level meta-analysis of physical activity interventions during childhood: Effects of physical activity on cognition and academic achievement. *Educational Psychology Review*, *35*(2), 59. https://doi.org/10.1007/s10648-023-09760-2
Vembye, M. H., & Olsen, T. (2025). *AIscreenR: AI Screening Tools in R for Systematic Reviewing.* https://CRAN.R-project.org/package=AIscreenR
Vembye, M. H., Christensen, J., Mølgaard, A. B., & Schytt, F. L. W. (2025). Generative pretrained transformer models can function as highly reliable second screeners of titles and abstracts in systematic reviews: A proof of concept and common guidelines. *Psychological Methods*, online access. https://doi.org/10.1037/met0000769
Viechtbauer, W. (2010). Conducting meta-analysis in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03
Viechtbauer, W. (2024). *I2 for multilevel and multivariate models*. https://www.metafor-project.org/doku.php/tips:i2_multilevel_multivariate. Accessed 2025-03-18.

Viechtbauer, W. (2025). *Predicted values for 'rma' objects*. https://wviechtb.github.io/metafor/reference/predict.rma.html#references-1. Accessed 2025-03-24.
Vygotsky, L. S. (1980). *Mind in society: The development of higher psychological processes*. Harvard University Press.
Wang, C. (2020). A review of the effects of abacus training on cognitive functions and neural systems in humans. *Frontiers of Neuroscience*, *14*, article 913. https://doi.org/10.3389/fnins.2020.00913
Wanzek, J., Vaughn, S., Scammaca, N. K., Metz, K., Murray, C. S., Roberts, G., & Danielson, L. (2013). Extensive reading interventions for students with reading difficulties after grade 3. *Review of Educational Research*, *83*, 163–195. https://doi.org/10.3102/0034654313477212
Wanzek, J., Vaughn, S., Wexler, J., Swanson, E. A., Edmonds, M., & Kim, A.-H. (2006). A synthesis of spelling and reading interventions and their effects on the spelling outcomes of students with LD. *Journal of Learning Disabilities*, *39*, 528–543. https://doi.org/10.1177/00222194060390060501
Wechsler, D. (2014). *The Wechsler intelligence scale for children* (5th ed.). Bloomington: The Psychological Corporation.
Weiland, C., Barata, M. C., & Yoshikawa, H. (2014). The co-occurring development of executive function skills and receptive vocabulary in preschool-aged children: A look at the direction of the developmental pathways. *Infant and Child Development*, *23*(1), 4‑21. https://doi.org/10.1002/icd.1829
Weiss, M. J., Lockwood, J. R., & McCaffrey, D. F. (2016). Estimating the standard error of the impact estimator in individually randomized trials with clustering. *Journal of Research on Educational Effectiveness*, *9*(3), 421-444. https://doi.org/10.1080/19345747.2015.1086911
What Works Clearinghouse (2021). *Supplement to the What Works Clearinghouse Procedures Handbook, Version 4.1*. National Center for Education Evaluation at IES, U.S. Department of Education.
Woodworth, R. S., & Thorndike, E. (1901). The influence of improvement in one mental function upon the efficiency of other functions (i). *Psychological Review*, *8*(3), 247–261. https://doi.org/10.1037/h0074898
World Bank (2025). *World Bank country classifications by income level for 2024-2025*. https://blogs.worldbank.org/en/opendata/world-bank-country-classifications-by-income-level-for-2024-2025. Accessed 2025-03-22.
Zelazo, P. D. (2015). Executive function: Reflection, iterative reprocessing, complexity, and the developing brain. *Developmental Review*, *38*, 55–68. https://doi.org/10.1016/j.dr.2015.07.001
Zelazo, P. D., & Carlson, S. M. (2020). The neurodevelopment of executive function skills: Implications for academic achievement gaps. *Psychology & Neuroscience*, *13*(3), 273–298. https://doi.org/10.1037/pne0000208.
Zelazo, P. D., & Frye, D. (1998). Cognitive complexity and control: II. The development of executive function in childhood. *Current Directions in Psychological Science*, *7*(4), 121–126. https://doi.org/10.1111/1467-8721.ep10774761
Zhang, Z., & Peng, P. (2023). Longitudinal reciprocal relations among reading, executive function, and social-emotional skills: Maybe not for all. *Journal of Educational Psychology*, *115*(3), 475–501. https://doi.org/10.1037/edu0000787.

Online Appendix A: Search documentation

This appendix provides documentation of our searches. First, we provide the full searches and results per electronic database. Second, we provide documentation for the searches of other sources: hand searches of relevant journals; searches for dissertations, working papers/conference proceedings, reports and non-US literature, and systematic reviews; backward and forward citation tracking; and contacts to experts in the field.

Electronic database search

ERIC

(EBSCO)

(1800 - current)

Searched 14.4.2023

Expanders - Apply equivalent subjects
Search modes - Boolean/Phrase

Search History

#	Query	Results
S23	S3 AND S6 AND S14 AND S18 AND S22	5,821
S22	S19 OR S20 OR S21	230,025
S21	DE ("Randomized Controlled Trials" OR "Quasiexperimental Design")	6,133
S20	AB (trial* OR experiment* OR intervention* OR treatment* OR “control group*” OR “compar* group*” OR randomiz* OR randomis* OR “random-assign*” OR “random* assign* OR “quasi-experiment*” OR “quasiexperiment*” OR “instrumental variable*" OR “regression discontinuity” OR “difference-in-difference*” OR “differences-in-difference*” OR “difference* in difference*” OR “event stud*” OR “event-stud*” OR matching OR “propensity score” OR case-control* OR “case control*”)	219,325
S19	TI (trial* OR experiment* OR intervention* OR treatment* OR “control group*” OR “compar* group*” OR randomiz* OR randomis* OR “random-assign*” OR “random* assign* OR “quasi-experiment*” OR “quasiexperiment*” OR “instrumental variable*" OR “regression discontinuity” OR “difference-in-difference*” OR “differences-in-difference*” OR “difference* in difference*” OR “event stud*” OR “event-stud*” OR matching OR “propensity score” OR case-control* OR “case control*”)	45,810
S18	S15 OR S16 OR S17	135,983
S17	DE ("Executive Function" OR "Inhibition" OR "Short Term Memory" OR "Self Management" OR "Attention" OR "Concept Formation" OR "Goal Orientation")	45,190

S16	AB ("executive function*" OR “impulse control” OR “emotion* control” OR “emotion* regulat*” OR “flexible thinking” OR “working memory” OR “working-memory” OR “self monitor*” OR self-monitor* OR “task initiation” OR ((“problem solving” OR problem-solving) AND (task OR test OR score)) OR "concept formation" OR "goal orientation" OR goal-orientation OR (planning* AND prioriti* AND (organis* OR organiz*)) OR "self regulat*" OR self-regulat* OR attention* OR inhibition* OR “inhibitory control” OR “short-term memory” OR “short term memory” OR “immediate memory” OR “verbal memory” OR “nonverbal memory” OR “delayed recall” OR delayed-recall OR “free-recall” OR “free recall” OR “serial-recall” OR “serial recall” OR “associative recall” OR associative-recall OR cogniti* flex* OR cogniti* refl* OR cogniti* control OR shifting OR set-shifting OR “effortful control” OR “self-control” OR “self control” OR “adaptable thinking” OR “task switch* OR “self management” OR self-management)	104,958
S15	TI ("executive function*" OR “impulse control” OR “emotion* control” OR “emotion* regulat*” OR “flexible thinking” OR “working memory” OR “working-memory” OR “self monitor*” OR self-monitor* OR “task initiation” OR ((“problem solving” OR problem-solving) AND (task OR test OR score)) OR "concept formation" OR "goal orientation" OR goal-orientation OR (planning* AND prioriti* AND (organis* OR organiz*)) OR "self regulat*" OR self-regulat* OR attention* OR inhibition* OR “inhibitory control” OR “short-term memory” OR “short term memory” OR “immediate memory” OR “verbal memory” OR “nonverbal memory” OR “delayed recall” OR delayed-recall OR “free-recall” OR “free recall” OR “serial-recall” OR “serial recall” OR “associative recall” OR associative-recall OR cogniti* flex* OR cogniti* refl* OR cogniti* control OR shifting OR set-shifting OR “effortful control” OR “self-control” OR “self control” OR “adaptable thinking” OR “task switch* OR “self management” OR self-management)	17,074
S14	S7 OR S8 OR S9 OR S10 OR S11 OR S12 OR S13	803,780
S13	DE (“Early Childhood Education” OR “Kindergarten” OR “Preschool Education” OR “Preschools” OR “Primary Education” OR “Elementary Education” OR “Elementary Schools” OR “Secondary Education” OR “Secondary Schools” OR “Grade 1” OR “Grade 2” OR “Grade 3” OR “Grade 4” OR “Grade 5” OR “Grade 6” OR “Grade 7” OR “Grade 8” OR “Grade 9” OR “Grade 10” OR “Grade 11” OR “Grade 12”)	419,342
S12	AB childhood N1 (education OR program* OR care OR initiativ* OR development*)	13,463
S11	TI childhood N1 (education OR program* OR care OR initiativ* OR development*)	5,335
S10	AB (“primary education” OR “secondary education” OR school* OR preschool* OR pre-school* OR kindergart* OR childcare OR "child* care" OR daycare OR “day care” OR pre-primar* OR “pre primar*” OR "early education" OR pre-K OR “pre K” OR prekindergart* OR pre-kindergart* OR nurser* OR “reception class”)	595,039
S9	TI (“primary education” OR “secondary education” OR school* OR preschool* OR pre-school* OR kindergart* OR childcare OR "child* care" OR daycare OR “day care” OR pre-primar* OR “pre primar*” OR "early education" OR pre-K OR “pre K” OR prekindergart* OR pre-kindergart* OR nurser* OR “reception class”)	235,289

S8	AB ((grade* N1 (1 OR 2 OR 3 OR 4 OR 5 OR 6 OR 7 OR 8 OR 9 OR 10 OR 11 OR 12)) OR ((first OR second OR third OR fourth OR fifth OR sixth OR seventh OR eighth OR ninth OR tenth OR eleventh OR twelfth) N1 grade*))	101,246
S7	TI ((grade* N1 (1 OR 2 OR 3 OR 4 OR 5 OR 6 OR 7 OR 8 OR 9 OR 10 OR 11 OR 12)) OR ((first OR second OR third OR fourth OR fifth OR sixth OR seventh OR eighth OR ninth OR tenth OR eleventh OR twelfth) N1 grade*))	19,533
S6	S4 OR S5	711,860
S5	AB (reading* OR math* OR languag* OR literac* OR numerac* OR number* OR geometr* OR algebra* OR fraction* OR operation* OR arithmetic* OR addition* OR subtraction* OR multiplication* OR division* OR statistics* OR probability* OR calculus* OR combinatoric* OR computation* OR calculation* OR counting* OR “word problem*” OR “word-problem*” OR measurement* OR comprehension* OR decod* OR “word identification*” OR “word-identification” OR fluency OR phonic* OR “phon* aware*” OR phonem* OR fluency OR spelling OR vocabulary OR alphabetic* OR letter* "print aware*" OR "sound discrim*" OR "rhyme detect*" OR blending OR segmentation OR grammar OR syntax OR syntactic OR morpholog*)	684,652
S4	TI (reading* OR math* OR languag* OR literac* OR numerac* OR number* OR geometr* OR algebra* OR fraction* OR operation* OR arithmetic* OR addition* OR subtraction* OR multiplication* OR division* OR statistics* OR probability* OR calculus* OR combinatoric* OR computation* OR calculation* OR counting* OR “word problem*” OR “word-problem*” OR measurement* OR comprehension* OR decod* OR “word identification*” OR “word-identification” OR fluency OR phonic* OR “phon* aware*” OR phonem* OR fluency OR spelling OR vocabulary OR alphabetic* OR letter* "print aware*" OR "sound discrim*" OR "rhyme detect*" OR blending OR segmentation OR grammar OR syntax OR syntactic OR morpholog*)	231,988
S3	S1 OR S2	1,083,142
S2	AB (student* OR pupil* OR child* OR toddler* OR youth* OR adolescen* OR teenage* OR young*)	1,055,363
S1	TI (student* OR pupil* OR child* OR toddler* OR youth* OR adolescen* OR teenage* OR young*)	405,320

APA PsycInfo

(EBSCO)

(1800 - current)

Searched 26.4.2023

Expanders - Apply equivalent subjects
Search modes - Boolean/Phrase

Search History

#	**Query**	**Results**
S23	S3 AND S6 AND S14 AND S18 AND S22	8,496
S22	S19 OR S20 OR S21	1,599,133
S21	DE "Randomized Controlled Trials" OR DE "Quasi Experimental Methods"	1,471
S20	AB (trial* OR experiment* OR intervention* OR treatment* OR "control group*" OR "compar* group*" OR randomiz* OR randomis* OR "random-assign*" OR "random* assign*" OR "quasi-experiment*" OR "quasiexperiment*" OR "instrumental variable"* OR "regression discontinuity" OR "difference-in-difference*" OR "differences-in-difference*" OR "difference* in difference*" OR "event stud*" OR "event-stud*" OR matching OR "propensity score" OR "case-control*" OR "case control*")	1,551,033
S19	TI (trial* OR experiment* OR intervention* OR treatment* OR "control group*" OR "compar* group*" OR randomiz* OR randomis* OR "random-assign*" OR "random* assign*" OR "quasi-experiment*" OR "quasiexperiment*" OR "instrumental variable"* OR "regression discontinuity" OR "difference-in-difference*" OR "differences-in-difference*" OR "difference* in difference*" OR "event stud*" OR "event-stud*" OR matching OR "propensity score" OR "case-control*" OR "case control*")	332,656
S18	S15 OR S16 OR S17	606,059
S17	((((((((DE "Executive Function" OR DE "Cognitive Control" OR DE "Set Shifting" OR DE "Task Switching") OR (DE "Attention")) OR (DE "Short Term Memory")) OR (DE "Proactive Inhibition")) OR (DE "Self-Management")) OR (DE "Concept Formation")) OR (DE "Goal Orientation")) OR (DE "Free Recall" OR DE "Serial Recall")) OR (DE "Verbal Memory")) OR (MM "Self-Control")	176,089
S16	AB ("executive function*" OR "impulse control" OR "emotion* control" OR "emotion* regulat*" OR "flexible thinking" OR "working memory" OR "working-memory" OR "self monitor*" OR self-monitor* OR "task initiation" OR (("problem solving" OR "problem-solving") AND (task OR test OR score)) OR "concept formation" OR "goal orientation" OR "goal-orientation" OR (planning* AND prioriti* AND (organis* OR organiz*)) OR "self regulat*" OR "self-regulat*" OR attention* OR inhibition* OR "inhibitory control" OR "short-term memory" OR "short term memory" OR "immediate memory" OR "verbal memory" OR "nonverbal memory" OR "delayed recall" OR "delayed-recall" OR "free-recall" OR "free recall" OR "serial-recall" OR "serial recall" OR "associative recall" OR "associative-recall" OR cogniti* flex* OR cogniti* refl* OR cogniti* control OR shifting OR "set-shifting" OR "effortful control" OR "self-control" OR "self control" OR "adaptable thinking" OR "task switch*" OR "self management" OR "self-management")	534,277
S15	TI ("executive function*" OR "impulse control" OR "emotion* control" OR "emotion* regulat*" OR "flexible thinking" OR "working memory" OR "working-memory" OR "self monitor*" OR self-monitor* OR "task initiation" OR (("problem solving" OR "problem-solving") AND (task OR test OR score)) OR "concept formation" OR "goal orientation" OR "goal-orientation" OR (planning* AND prioriti* AND (organis* OR organiz*)) OR "self regulat*" OR "self-regulat*" OR attention* OR inhibition* OR	129,775

	“inhibitory control” OR “short-term memory” OR “short term memory” OR “immediate memory” OR “verbal memory” OR “nonverbal memory” OR “delayed recall” OR "delayed-recall" OR “free-recall” OR “free recall” OR “serial-recall” OR “serial recall” OR “associative recall” OR "associative-recall" OR cogniti* flex* OR cogniti* refl* OR cogniti* control OR shifting OR "set-shifting" OR “effortful control” OR “self-control” OR “self control” OR “adaptable thinking” OR “task switch*" OR “self management” OR "self-management")	
S14	S7 OR S8 OR S9 OR S10 OR S11 OR S12 OR S13	545,368
S13	(((DE "Middle Schools" OR DE "Secondary Education" OR DE "Elementary Education" OR DE "Elementary Schools") OR (DE "Kindergarten Students" OR DE "Nursery School Students" OR DE "Preschool Students")) OR (DE "Preschool Education")) OR (DE "Primary School Students")	51,350
S12	AB childhood N1 (education OR program* OR care OR initiativ* OR development*)	9,828
S11	TI childhood N1 (education OR program* OR care OR initiativ* OR development*)	3,377
S10	AB (“primary education” OR “secondary education” OR school* OR preschool* OR pre-school* OR kindergart* OR childcare OR "child* care" OR daycare OR “day care” OR pre-primar* OR “pre primar*” OR "early education" OR pre-K OR “pre K” OR prekindergart* OR pre-kindergart* OR nurser* OR “reception class”)	462,051
S9	TI (“primary education” OR “secondary education” OR school* OR preschool* OR pre-school* OR kindergart* OR childcare OR "child* care" OR daycare OR “day care” OR pre-primar* OR “pre primar*” OR "early education" OR pre-K OR “pre K” OR prekindergart* OR pre-kindergart* OR nurser* OR “reception class”)	182,068
S8	AB ((grade* N1 (1 OR 2 OR 3 OR 4 OR 5 OR 6 OR 7 OR 8 OR 9 OR 10 OR 11 OR 12)) OR ((first OR second OR third OR fourth OR fifth OR sixth OR seventh OR eighth OR ninth OR tenth OR eleventh OR twelfth) N1 grade*))	83,004
S7	TI ((grade* N1 (1 OR 2 OR 3 OR 4 OR 5 OR 6 OR 7 OR 8 OR 9 OR 10 OR 11 OR 12)) OR ((first OR second OR third OR fourth OR fifth OR sixth OR seventh OR eighth OR ninth OR tenth OR eleventh OR twelfth) N1 grade*))	16,101
S6	S4 OR S5	1,581,309
S5	AB (reading* OR math* OR languag* OR literac* OR numerac* OR number* OR geometr* OR algebra* OR fraction* OR operation* OR arithmetic* OR addition* OR subtraction* OR multiplication* OR division* OR statistics* OR probability* OR calculus* OR combinatoric* OR computation* OR calculation* OR counting* OR “word problem*” OR “word-problem*” OR measurement* OR comprehension* OR decod* OR “word identification*” OR “word-identification” OR fluency OR phonic* OR “phon* aware*” OR phonem* OR fluency OR spelling OR vocabulary OR alphabetic* OR letter* "print aware*" OR "sound discrim*" OR "rhyme detect*" OR blending OR segmentation OR grammar OR syntax OR syntactic OR morpholog*)	1,539,847
S4	TI (reading* OR math* OR languag* OR literac* OR numerac* OR number* OR geometr* OR algebra* OR fraction* OR operation* OR arithmetic* OR addition* OR	247,297

	subtraction* OR multiplication* OR division* OR statistics* OR probability* OR calculus* OR combinatoric* OR computation* OR calculation* OR counting* OR “word problem*” OR “word-problem*” OR measurement* OR comprehension* OR decod* OR “word identification*” OR “word-identification” OR fluency OR phonic* OR “phon* aware*” OR phonem* OR fluency OR spelling OR vocabulary OR alphabetic* OR letter* "print aware*" OR "sound discrim*" OR "rhyme detect*" OR blending OR segmentation OR grammar OR syntax OR syntactic OR morpholog*)	
S3	S1 OR S2	1,584,803
S2	AB (student* OR pupil* OR child* OR toddler* OR youth* OR adolescen* OR teenage* OR young*)	1,495,895
S1	TI (student* OR pupil* OR child* OR toddler* OR youth* OR adolescen* OR teenage* OR young*)	763,817

SocINDEX

(EBSCO)

(1908 - current)

Searched 24.4.2023

Expanders - Apply equivalent subjects
Search modes - Boolean/Phrase

Search History

#	Query	Results
S23	S3 AND S6 AND S14 AND S18 AND S22	791
S22	S19 OR S20 OR S21	288,202
S21	DE "RANDOMIZED controlled trials" OR DE "REGRESSION discontinuity design" OR DE "CASE-control method"	3,624
S20	AB (trial* OR experiment* OR intervention* OR treatment* OR “control group*” OR “compar* group*” OR randomiz* OR randomis* OR “random-assign*” OR “random* assign*" OR “quasi-experiment*” OR “quasiexperiment*” OR “instrumental variable*" OR “regression discontinuity” OR “difference-in-difference*” OR “differences-in-difference*” OR “difference* in difference*” OR “event stud*” OR “event-stud*” OR matching OR “propensity score” OR case-control* OR “case control*”)	271,138
S19	TI (trial* OR experiment* OR intervention* OR treatment* OR “control group*” OR “compar* group*” OR randomiz* OR randomis* OR “random-assign*” OR “random* assign*" OR “quasi-experiment*” OR “quasiexperiment*” OR “instrumental variable*" OR “regression discontinuity” OR “difference-in-difference*” OR “differences-in-difference*” OR “difference* in difference*” OR “event stud*” OR “event-stud*” OR matching OR “propensity score” OR case-control* OR “case control*”)	68,736
S18	S15 OR S16 OR S17	112,104

S17	DE "ATTENTION" OR DE "EXECUTIVE function" OR DE "SELF-control" OR DE "SELF-management (Psychology)"	5,771
S16	AB ("executive function*" OR “impulse control” OR “emotion* control” OR “emotion* regulat*” OR “flexible thinking” OR “working memory” OR “working-memory” OR “self monitor*” OR self-monitor* OR “task initiation” OR ((“problem solving” OR problem-solving) AND (task OR test OR score)) OR "concept formation" OR "goal orientation" OR goal-orientation OR (planning* AND prioriti* AND (organis* OR organiz*)) OR "self regulat*" OR self-regulat* OR attention* OR inhibition* OR “inhibitory control” OR “short-term memory” OR “short term memory” OR “immediate memory” OR “verbal memory” OR “nonverbal memory” OR “delayed recall” OR delayed-recall OR “free-recall” OR “free recall” OR “serial-recall” OR “serial recall” OR “associative recall” OR associative-recall OR cogniti* flex* OR cogniti* refl* OR cogniti* control OR shifting OR set-shifting OR “effortful control” OR “self-control” OR “self control” OR “adaptable thinking” OR “task switch* OR “self management” OR self-management)	108,007
S15	TI ("executive function*" OR “impulse control” OR “emotion* control” OR “emotion* regulat*” OR “flexible thinking” OR “working memory” OR “working-memory” OR “self monitor*” OR self-monitor* OR “task initiation” OR ((“problem solving” OR problem-solving) AND (task OR test OR score)) OR "concept formation" OR "goal orientation" OR goal-orientation OR (planning* AND prioriti* AND (organis* OR organiz*)) OR "self regulat*" OR self-regulat* OR attention* OR inhibition* OR “inhibitory control” OR “short-term memory” OR “short term memory” OR “immediate memory” OR “verbal memory” OR “nonverbal memory” OR “delayed recall” OR delayed-recall OR “free-recall” OR “free recall” OR “serial-recall” OR “serial recall” OR “associative recall” OR associative-recall OR cogniti* flex* OR cogniti* refl* OR cogniti* control OR shifting OR set-shifting OR “effortful control” OR “self-control” OR “self control” OR “adaptable thinking” OR “task switch*" OR “self management” OR self-management)	10,223
S14	S7 OR S8 OR S9 OR S10 OR S11 OR S12 OR S13	233,441
S13	(DE "PRESCHOOL education" OR DE "PRESCHOOLS" OR DE "PRIMARY education" OR DE "EARLY childhood education" OR DE "ELEMENTARY education" OR DE "ELEMENTARY schools" OR DE "KINDERGARTEN") OR (DE "SECONDARY education")	9,625
S12	AB childhood N1 (education OR program* OR care OR initiativ* OR development*)	1,895
S11	TI childhood N1 (education OR program* OR care OR initiativ* OR development*)	639
S10	AB (“primary education” OR “secondary education” OR school* OR preschool* OR pre-school* OR kindergart* OR childcare OR "child* care" OR daycare OR “day care” OR pre-primar* OR “pre primar*” OR "early education" OR pre-K OR “pre K” OR prekindergart* OR pre-kindergart* OR nurser* OR “reception class”)	217,251
S9	TI (“primary education” OR “secondary education” OR school* OR preschool* OR pre-school* OR kindergart* OR childcare OR "child* care" OR daycare OR “day care” OR pre-primar* OR “pre primar*” OR "early education" OR pre-K OR “pre K” OR prekindergart* OR pre-kindergart* OR nurser* OR “reception class”)	70,914
S8	AB ((grade* N1 (1 OR 2 OR 3 OR 4 OR 5 OR 6 OR 7 OR 8 OR 9 OR 10 OR 11 OR 12)) OR ((first OR second OR third OR fourth OR fifth OR sixth OR seventh OR eighth OR ninth OR tenth OR eleventh OR twelfth) N1 grade*))	14,170
S7	TI ((grade* N1 (1 OR 2 OR 3 OR 4 OR 5 OR 6 OR 7 OR 8 OR 9 OR 10 OR 11 OR 12)) OR ((first OR second OR third OR fourth OR fifth OR sixth OR seventh OR eighth OR ninth OR tenth OR eleventh OR twelfth) N1 grade*))	1,371
S6	S4 OR S5	478,140

S5	AB (reading* OR math* OR languag* OR literac* OR numerac* OR number* OR geometr* OR algebra* OR fraction* OR operation* OR arithmetic* OR addition* OR subtraction* OR multiplication* OR division* OR statistics* OR probability* OR calculus* OR combinatoric* OR computation* OR calculation* OR counting* OR “word problem*” OR “word-problem*” OR measurement* OR comprehension* OR decod* OR “word identification*” OR “word-identification” OR fluency OR phonic* OR “phon* aware*” OR phonem* OR fluency OR spelling OR vocabulary OR alphabetic* OR letter* "print aware*" OR "sound discrim*" OR "rhyme detect*" OR blending OR segmentation OR grammar OR syntax OR syntactic OR morpholog*)	462,399
S4	TI (reading* OR math* OR languag* OR literac* OR numerac* OR number* OR geometr* OR algebra* OR fraction* OR operation* OR arithmetic* OR addition* OR subtraction* OR multiplication* OR division* OR statistics* OR probability* OR calculus* OR combinatoric* OR computation* OR calculation* OR counting* OR “word problem*” OR “word-problem*” OR measurement* OR comprehension* OR decod* OR “word identification*” OR “word-identification” OR fluency OR phonic* OR “phon* aware*” OR phonem* OR fluency OR spelling OR vocabulary OR alphabetic* OR letter* "print aware*" OR "sound discrim*" OR "rhyme detect*" OR blending OR segmentation OR grammar OR syntax OR syntactic OR morpholog*)	56,067
S3	S1 OR S2	543,009
S2	AB (student* OR pupil* OR child* OR toddler* OR youth* OR adolescen* OR teenage* OR young*)	514,096
S1	TI (student* OR pupil* OR child* OR toddler* OR youth* OR adolescen* OR teenage* OR young*)	236,879

Academic Search Premier

(EBSCO)

(1956 - current)

Searched 25.4.2023

Expanders - Apply equivalent subjects
Search modes - Boolean/Phrase

Search History

#	Query	Results
S23	S3 AND S6 AND S14 AND S18 AND S22	4,935
S22	S19 OR S20 OR S21	6,870,261
S21	DE "RANDOMIZED controlled trials" OR DE "PROPENSITY score matching" OR DE "CASE-control method" OR DE "REGRESSION discontinuity design" OR DE "EXPERIMENTAL design"	197,723
S20	AB (trial* OR experiment* OR intervention* OR treatment* OR “control group*” OR “compar* group*” OR randomiz* OR randomis* OR “random-assign*” OR “random* assign*" OR “quasi-experiment*” OR “quasiexperiment*” OR “instrumental variable*" OR “regression discontinuity” OR “difference-in-difference*” OR “differences-in-difference*” OR “difference* in difference*” OR “event stud*” OR “event-stud*” OR matching OR “propensity score” OR "case-control*" OR “case control*”)	6,524,120

S19	TI (trial* OR experiment* OR intervention* OR treatment* OR “control group*” OR “compar* group*” OR randomiz* OR randomis* OR “random-assign*” OR “random* assign*" OR “quasi-experiment*” OR “quasiexperiment*” OR “instrumental variable*" OR “regression discontinuity” OR “difference-in-difference*” OR “differences-in-difference*” OR “difference* in difference*” OR “event stud*” OR “event-stud*” OR matching OR “propensity score” OR "case-control*" OR “case control*”)	1,252,221
S18	S15 OR S16 OR S17	1,306,831
S17	DE "ATTENTION" OR DE "EXECUTIVE function" OR DE "SHORT-term memory" OR DE "RESPONSE inhibition" OR DE "CONCEPTS"	74,813
S16	AB ("executive function*" OR “impulse control” OR “emotion* control” OR “emotion* regulat*” OR “flexible thinking” OR “working memory” OR “working-memory” OR “self monitor*” OR self-monitor* OR “task initiation” OR ((“problem solving” OR "problem-solving") AND (task OR test OR score)) OR "concept formation" OR "goal orientation" OR "goal-orientation" OR (planning* AND prioriti* AND (organis* OR organiz*)) OR "self regulat*" OR "self-regulat*" OR attention* OR inhibition* OR “inhibitory control” OR “short-term memory” OR “short term memory” OR “immediate memory” OR “verbal memory” OR “nonverbal memory” OR “delayed recall” OR "delayed-recall" OR “free-recall” OR “free recall” OR “serial-recall” OR “serial recall” OR “associative recall” OR "associative-recall" OR cogniti* flex* OR cogniti* refl* OR cogniti* control OR shifting OR "set-shifting" OR “effortful control” OR “self-control” OR “self control” OR “adaptable thinking” OR “task switch*" OR “self management” OR "self-management")	1,230,956
S15	TI ("executive function*" OR “impulse control” OR “emotion* control” OR “emotion* regulat*” OR “flexible thinking” OR “working memory” OR “working-memory” OR “self monitor*” OR self-monitor* OR “task initiation” OR ((“problem solving” OR "problem-solving") AND (task OR test OR score)) OR "concept formation" OR "goal orientation" OR "goal-orientation" OR (planning* AND prioriti* AND (organis* OR organiz*)) OR "self regulat*" OR "self-regulat*" OR attention* OR inhibition* OR “inhibitory control” OR “short-term memory” OR “short term memory” OR “immediate memory” OR “verbal memory” OR “nonverbal memory” OR “delayed recall” OR "delayed-recall" OR “free-recall” OR “free recall” OR “serial-recall” OR “serial recall” OR “associative recall” OR "associative-recall" OR cogniti* flex* OR cogniti* refl* OR cogniti* control OR shifting OR "set-shifting" OR “effortful control” OR “self-control” OR “self control” OR “adaptable thinking” OR “task switch*" OR “self management” OR "self-management")	200,266
S14	S7 OR S8 OR S9 OR S10 OR S11 OR S12 OR S13	1,074,459
S13	((((DE "SECONDARY education") OR (DE "PRIMARY education" OR DE "EARLY childhood education" OR DE "FIRST grade (Education)" OR DE "FOURTH grade (Education)" OR DE "SECOND grade (Education)" OR DE "THIRD grade (Education)" OR DE "ELEMENTARY education" OR DE "ELEMENTARY schools" OR DE "KINDERGARTEN")) OR (DE "PRESCHOOL education")) OR (DE "TWELFTH grade (Education)" OR DE "NINTH grade (Education)" OR DE "FIFTH grade (Education)" OR DE "ELEVENTH grade (Education)" OR DE "EIGHTH grade (Education)" OR DE "TENTH grade (Education)" OR DE "SIXTH grade (Education)" OR DE "SEVENTH grade (Education)")) OR (DE "PRESCHOOLS")	68,555
S12	AB childhood N1 (education OR program* OR care OR initiativ* OR development*)	13,409
S11	TI childhood N1 (education OR program* OR care OR initiativ* OR development*)	3,821
S10	AB (“primary education” OR “secondary education” OR school* OR preschool* OR pre-school* OR kindergart* OR childcare OR "child* care" OR daycare OR “day	907,357

	care” OR pre-primar* OR “pre primar*” OR "early education" OR pre-K OR “pre K” OR prekindergart* OR pre-kindergart* OR nurser* OR “reception class”)	
S9	TI (“primary education” OR “secondary education” OR school* OR preschool* OR pre-school* OR kindergart* OR childcare OR "child* care" OR daycare OR “day care” OR pre-primar* OR “pre primar*” OR "early education" OR pre-K OR “pre K” OR prekindergart* OR pre-kindergart* OR nurser* OR “reception class”)	287,638
S8	AB ((grade* N1 (1 OR 2 OR 3 OR 4 OR 5 OR 6 OR 7 OR 8 OR 9 OR 10 OR 11 OR 12)) OR ((first OR second OR third OR fourth OR fifth OR sixth OR seventh OR eighth OR ninth OR tenth OR eleventh OR twelfth) N1 grade*))	112,805
S7	TI ((grade* N1 (1 OR 2 OR 3 OR 4 OR 5 OR 6 OR 7 OR 8 OR 9 OR 10 OR 11 OR 12)) OR ((first OR second OR third OR fourth OR fifth OR sixth OR seventh OR eighth OR ninth OR tenth OR eleventh OR twelfth) N1 grade*))	17,743
S6	S4 OR S5	9,694,187
S5	AB (reading* OR math* OR languag* OR literac* OR numerac* OR number* OR geometr* OR algebra* OR fraction* OR operation* OR arithmetic* OR addition* OR subtraction* OR multiplication* OR division* OR statistics* OR probability* OR calculus* OR combinatoric* OR computation* OR calculation* OR counting* OR “word problem*” OR “word-problem*” OR measurement* OR comprehension* OR decod* OR “word identification*” OR “word-identification” OR fluency OR phonic* OR “phon* aware*” OR phonem* OR fluency OR spelling OR vocabulary OR alphabetic* OR letter* "print aware*" OR "sound discrim*" OR "rhyme detect*" OR blending OR segmentation OR grammar OR syntax OR syntactic OR morpholog*)	9,416,872
S4	TI (reading* OR math* OR languag* OR literac* OR numerac* OR number* OR geometr* OR algebra* OR fraction* OR operation* OR arithmetic* OR addition* OR subtraction* OR multiplication* OR division* OR statistics* OR probability* OR calculus* OR combinatoric* OR computation* OR calculation* OR counting* OR “word problem*” OR “word-problem*” OR measurement* OR comprehension* OR decod* OR “word identification*” OR “word-identification” OR fluency OR phonic* OR “phon* aware*” OR phonem* OR fluency OR spelling OR vocabulary OR alphabetic* OR letter* "print aware*" OR "sound discrim*" OR "rhyme detect*" OR blending OR segmentation OR grammar OR syntax OR syntactic OR morpholog*)	1,235,840
S3	S1 OR S2	2,970,074
S2	AB (student* OR pupil* OR child* OR toddler* OR youth* OR adolescen* OR teenage* OR young*)	2,750,215
S1	TI (student* OR pupil* OR child* OR toddler* OR youth* OR adolescen* OR teenage* OR young*)	1,110,255

International Bibliography of the Social Sciences (IBSS)

(ProQuest)

Coverage: 1951 - current

Searched on April 26 2023

Search Strategy

Set#	Searched for	Results
S1	title(student* OR pupil* OR child* OR toddler* OR youth* OR adolescen* OR teenage* OR young*)	173019

S2	abstract(student* OR pupil* OR child* OR toddler* OR youth* OR adolescen* OR teenage* OR young*)	294872
S3	S1 OR S2	354358
S4	title(reading* OR math* OR languag* OR literac* OR numerac* OR number* OR geometr* OR algebra* OR fraction* OR operation* OR arithmetic* OR addition* OR subtraction* OR multiplication* OR division* OR statistics* OR probability* OR calculus* OR combinatoric* OR computation* OR calculation* OR counting* OR “word problem*” OR “word-problem*” OR measurement* OR comprehension* OR decod* OR “word identification*” OR “word-identification” OR fluency OR phonic* OR “phon* aware*” OR phonem* OR fluency OR spelling OR vocabulary OR alphabetic* OR letter* "print aware*" OR "sound discrim*" OR "rhyme detect*" OR blending OR segmentation OR grammar OR syntax OR syntactic OR morpholog*)	128378
S5	abstract(reading* OR math* OR languag* OR literac* OR numerac* OR number* OR geometr* OR algebra* OR fraction* OR operation* OR arithmetic* OR addition* OR subtraction* OR multiplication* OR division* OR statistics* OR probability* OR calculus* OR combinatoric* OR computation* OR calculation* OR counting* OR "word problem*" OR "word-problem*" OR measurement* OR comprehension* OR decod* OR "word identification*" OR "word-identification" OR fluency OR phonic* OR "phon* aware*" OR phonem* OR fluency OR spelling OR vocabulary OR alphabetic* OR letter* "print aware*" OR "sound discrim*" OR "rhyme detect*" OR blending OR segmentation OR grammar OR syntax OR syntactic OR morpholog*)	592384
S6	S4 OR S5	669407
S7	title(((grade* NEAR/1 (1 OR 2 OR 3 OR 4 OR 5 OR 6 OR 7 OR 8 OR 9 OR 10 OR 11 OR 12)) OR ((first OR second OR third OR fourth OR fifth OR sixth OR seventh OR eighth OR ninth OR tenth OR eleventh OR twelfth) NEAR/1 grade*)))	790
S8	abstract(((grade* NEAR/1 (1 OR 2 OR 3 OR 4 OR 5 OR 6 OR 7 OR 8 OR 9 OR 10 OR 11 OR 12)) OR ((first OR second OR third OR fourth OR fifth OR sixth OR seventh OR eighth OR ninth OR tenth OR eleventh OR twelfth) NEAR/1 grade*)))	6322
S9	title(“primary education” OR “secondary education” OR school* OR preschool* OR pre-school* OR kindergart* OR childcare OR "child* care" OR daycare OR “day care” OR pre-primar* OR “pre primar*” OR "early education" OR pre-K OR “pre K” OR prekindergart* OR pre-kindergart* OR nurser* OR “reception class”)	54177
S10	abstract("primary education" OR "secondary education" OR school* OR preschool* OR pre-school* OR kindergart* OR childcare OR "child* care" OR daycare OR "day care" OR pre-primar* OR "pre primar*" OR "early education" OR pre-K OR "pre K" OR prekindergart* OR pre-kindergart* OR nurser* OR "reception class")	109126
S11	title(childhood NEAR/1 (education OR program* OR care OR initiativ* OR development*))	944
S12	abstract(childhood NEAR/1 (education OR program* OR care OR initiativ* OR development*))	1804
S13	MAINSUBJECT.EXACT("Early childhood education") OR MAINSUBJECT.EXACT("Elementary education") OR MAINSUBJECT.EXACT("Kindergarten") OR MAINSUBJECT.EXACT("Preschool education") OR MAINSUBJECT.EXACT("Elementary schools") OR MAINSUBJECT.EXACT("Secondary education") OR MAINSUBJECT.EXACT("Secondary schools")	19954
S14	S7 OR S8 OR S9 OR S10 OR S11 OR S12 OR S13	136450

S15	title("executive function*" OR “impulse control” OR “emotion* control” OR “emotion* regulat*” OR “flexible thinking” OR “working memory” OR “working-memory” OR “self monitor*” OR self-monitor* OR “task initiation” OR ((“problem solving” OR problem-solving) AND (task OR test OR score)) OR "concept formation" OR "goal orientation" OR goal-orientation OR (planning* AND prioriti* AND (organis* OR organiz*)) OR "self regulat*" OR self-regulat* OR attention* OR inhibition* OR “inhibitory control” OR “short-term memory” OR “short term memory” OR “immediate memory” OR “verbal memory” OR “nonverbal memory” OR “delayed recall” OR delayed-recall OR “free-recall” OR “free recall” OR “serial-recall” OR “serial recall” OR “associative recall” OR associative-recall OR cogniti* flex* OR cogniti* refl* OR cogniti* control OR shifting OR set-shifting OR “effortful control” OR “self-control” OR “self control” OR “adaptable thinking” OR “task switch*" OR “self management” OR self-management)	16530
S16	abstract("executive function*" OR "impulse control" OR "emotion* control" OR "emotion* regulat*" OR "flexible thinking" OR "working memory" OR "working-memory" OR "self monitor*" OR self-monitor* OR "task initiation" OR (("problem solving" OR problem-solving) AND (task OR test OR score)) OR "concept formation" OR "goal orientation" OR goal-orientation OR (planning* AND prioriti* AND (organis* OR organiz*)) OR "self regulat*" OR self-regulat* OR attention* OR inhibition* OR "inhibitory control" OR "short-term memory" OR "short term memory" OR "immediate memory" OR "verbal memory" OR "nonverbal memory" OR "delayed recall" OR delayed-recall OR "free-recall" OR "free recall" OR "serial-recall" OR "serial recall" OR "associative recall" OR associative-recall OR cogniti* flex* OR cogniti* refl* OR cogniti* control OR shifting OR set-shifting OR "effortful control" OR "self-control" OR "self control" OR "adaptable thinking" OR "task switch*" OR "self management" OR self-management)	161994
S17	MAINSUBJECT.EXACT("Inhibition") OR MAINSUBJECT.EXACT("Concept formation") OR MAINSUBJECT.EXACT("Short term memory") OR MAINSUBJECT.EXACT("Attention") OR MAINSUBJECT.EXACT("Selfmanagement") OR MAINSUBJECT.EXACT("Executive function")	9438
S18	S15 OR S16 OR S17	170979
S19	title(trial* OR experiment* OR intervention* OR treatment* OR “control group*” OR “compar* group*” OR randomiz* OR randomis* OR “random-assign*” OR “random* assign*" OR “quasi-experiment*” OR “quasiexperiment*” OR “instrumental variable*" OR “regression discontinuity” OR “difference-in-difference*” OR “differences-in-difference*” OR “difference* in difference*” OR “event stud*” OR “event-stud*” OR matching OR “propensity score” OR case-control* OR “case control*”)	77232
S20	abstract(trial* OR experiment* OR intervention* OR treatment* OR "control group*" OR "compar* group*" OR randomiz* OR randomis* OR "random-assign*" OR "random* assign*" OR "quasi-experiment*" OR "quasiexperiment*" OR "instrumental variable*" OR "regression discontinuity" OR "difference-in-difference*" OR "differences-in-difference*" OR "difference* in difference*" OR "event stud*" OR "event-stud*" OR matching OR "propensity score" OR case-control* OR "case control*")	262001
S21	MAINSUBJECT.EXACT("Case controlled studies") OR MAINSUBJECT.EXACT("Quasi-experimental methods") OR MAINSUBJECT.EXACT("Clinical trials") OR MAINSUBJECT.EXACT("Control groups")	5831
S22	S19 OR S20 OR S21	294438
S23	S3 AND S6 AND S14 AND S18 AND S22	629

Sociological Abstracts

(ProQuest)

Coverage: 1952 - current

Searched on April 26 2023

Search Strategy

Set#	Searched for	Results
S1	title(student* OR pupil* OR child* OR toddler* OR youth* OR adolescen* OR teenage* OR young*)	241743
S2	abstract(student* OR pupil* OR child* OR toddler* OR youth* OR adolescen* OR teenage* OR young*)	453864
S3	S1 OR S2	490333
S4	title(reading* OR math* OR languag* OR literac* OR numerac* OR number* OR geometr* OR algebra* OR fraction* OR operation* OR arithmetic* OR addition* OR subtraction* OR multiplication* OR division* OR statistics* OR probability* OR calculus* OR combinatoric* OR computation* OR calculation* OR counting* OR “word problem*” OR “word-problem*” OR measurement* OR comprehension* OR decod* OR “word identification*” OR “word-identification” OR fluency OR phonic* OR “phon* aware*” OR phonem* OR fluency OR spelling OR vocabulary OR alphabetic* OR letter* "print aware*" OR "sound discrim*" OR "rhyme detect*" OR blending OR segmentation OR grammar OR syntax OR syntactic OR morpholog*)	49511
S5	abstract(reading* OR math* OR languag* OR literac* OR numerac* OR number* OR geometr* OR algebra* OR fraction* OR operation* OR arithmetic* OR addition* OR subtraction* OR multiplication* OR division* OR statistics* OR probability* OR calculus* OR combinatoric* OR computation* OR calculation* OR counting* OR "word problem*" OR "word-problem*" OR measurement* OR comprehension* OR decod* OR "word identification*" OR "word-identification" OR fluency OR phonic* OR "phon* aware*" OR phonem* OR fluency OR spelling OR vocabulary OR alphabetic* OR letter* "print aware*" OR "sound discrim*" OR "rhyme detect*" OR blending OR segmentation OR grammar OR syntax OR syntactic OR morpholog*)	416147
S6	S4 OR S5	434648
S7	title(((grade* NEAR/1 (1 OR 2 OR 3 OR 4 OR 5 OR 6 OR 7 OR 8 OR 9 OR 10 OR 11 OR 12)) OR ((first OR second OR third OR fourth OR fifth OR sixth OR seventh OR eighth OR ninth OR tenth OR eleventh OR twelfth) NEAR/1 grade*)))	844
S8	abstract(((grade* NEAR/1 (1 OR 2 OR 3 OR 4 OR 5 OR 6 OR 7 OR 8 OR 9 OR 10 OR 11 OR 12)) OR ((first OR second OR third OR fourth OR fifth OR sixth OR seventh OR eighth OR ninth OR tenth OR eleventh OR twelfth) NEAR/1 grade*)))	10850
S9	title(“primary education” OR “secondary education” OR school* OR preschool* OR pre-school* OR kindergart* OR childcare OR "child* care" OR daycare OR “day care” OR pre-primar* OR “pre primar*” OR "early education" OR pre-K OR “pre K” OR prekindergart* OR pre-kindergart* OR nurser* OR “reception class”)	54452
S10	abstract("primary education" OR "secondary education" OR school* OR preschool* OR pre-school* OR kindergart* OR childcare OR "child* care" OR daycare OR "day care" OR pre-primar* OR "pre	142077

	primar*" OR "early education" OR pre-K OR "pre K" OR prekindergart* OR pre-kindergart* OR nurser* OR "reception class")	
S11	title(childhood NEAR/1 (education OR program* OR care OR initiativ* OR development*))	655
S12	abstract(childhood NEAR/1 (education OR program* OR care OR initiativ* OR development*))	1837
S13	MAINSUBJECT.EXACT("Elementary Education") OR MAINSUBJECT.EXACT("Secondary Schools") OR MAINSUBJECT.EXACT("Secondary Education") OR MAINSUBJECT.EXACT("Elementary Schools") OR MAINSUBJECT.EXACT("Preschool Education") OR MAINSUBJECT.EXACT("Kindergarten") OR MAINSUBJECT.EXACT("Primary Education")	15626
S14	S7 OR S8 OR S9 OR S10 OR S11 OR S12 OR S13	160745
S15	title("executive function*" OR “impulse control” OR “emotion* control” OR “emotion* regulat*” OR “flexible thinking” OR “working memory” OR “working-memory” OR “self monitor*” OR self-monitor* OR “task initiation” OR ((“problem solving” OR problem-solving) AND (task OR test OR score)) OR "concept formation" OR "goal orientation" OR goal-orientation OR (planning* AND prioriti* AND (organis* OR organiz*)) OR "self regulat*" OR self-regulat* OR attention* OR inhibition* OR “inhibitory control” OR “short-term memory” OR “short term memory” OR “immediate memory” OR “verbal memory” OR “nonverbal memory” OR “delayed recall” OR delayed-recall OR “free-recall” OR “free recall” OR “serial-recall” OR “serial recall” OR “associative recall” OR associative-recall OR cogniti* flex* OR cogniti* refl* OR cogniti* control OR shifting OR set-shifting OR “effortful control” OR “self-control” OR “self control” OR “adaptable thinking” OR “task switch*" OR “self management” OR self-management)	9016
S16	abstract("executive function*" OR "impulse control" OR "emotion* control" OR "emotion* regulat*" OR "flexible thinking" OR "working memory" OR "working-memory" OR "self monitor*" OR self-monitor* OR "task initiation" OR (("problem solving" OR problem-solving) AND (task OR test OR score)) OR "concept formation" OR "goal orientation" OR goal-orientation OR (planning* AND prioriti* AND (organis* OR organiz*)) OR "self regulat*" OR self-regulat* OR attention* OR inhibition* OR "inhibitory control" OR "short-term memory" OR "short term memory" OR "immediate memory" OR "verbal memory" OR "nonverbal memory" OR "delayed recall" OR delayed-recall OR "free-recall" OR "free recall" OR "serial-recall" OR "serial recall" OR "associative recall" OR associative-recall OR cogniti* flex* OR cogniti* refl* OR cogniti* control OR shifting OR set-shifting OR "effortful control" OR "self-control" OR "self control" OR "adaptable thinking" OR "task switch*" OR "self management" OR self-management)	132528
S17	MAINSUBJECT.EXACT("Attention") OR MAINSUBJECT.EXACT("Concept Formation")	3579
S18	S15 OR S16 OR S17	137080
S19	title(trial* OR experiment* OR intervention* OR treatment* OR “control group*” OR “compar* group*” OR randomiz* OR randomis* OR “random-assign*” OR “random* assign*" OR “quasi-experiment*” OR “quasiexperiment*” OR “instrumental variable*" OR “regression discontinuity” OR “difference-in-difference*” OR “differences-in-difference*” OR “difference* in difference*” OR “event stud*” OR “event-stud*” OR matching OR “propensity score” OR case-control* OR “case control*”)	57603
S20	abstract(trial* OR experiment* OR intervention* OR treatment* OR "control group*" OR "compar* group*" OR randomiz* OR randomis* OR "random-assign*" OR "random* assign*" OR "quasi-experiment*" OR "quasiexperiment*" OR "instrumental variable*" OR "regression discontinuity" OR "difference-in-difference*" OR "differences-in-difference*" OR "difference* in difference*" OR	241101

	"event stud*" OR "event-stud*" OR matching OR "propensity score" OR case-control* OR "case control*")	
S21	S19 OR S20	254869
S22	(S1 OR S2) AND (S4 OR S5) AND (S7 OR S8 OR S9 OR S10 OR S11 OR S12 OR S13) AND (S15 OR S16 OR S17) AND (S19 OR S20)	670

Science Citation Index Expanded

(1900 – Present)

Social Sciences Citation Index

(1956 – Present)

Searched 11.5.2023

#	Query	Results
8	#1 AND #2 AND #5 AND #6 AND #7	5,283
7	TS=(trial* OR experiment* OR intervention* OR “control group*” OR “compar* group*” OR randomiz* OR randomis* OR “random-assign*” OR “random* assign*" OR “quasi-experiment*” OR “quasiexperiment*” OR “instrumental variable"* OR “regression discontinuity” OR “difference-in-difference*” OR “differences-in-difference*” OR “difference* in difference*” OR “event stud*” OR “event-stud*” OR “propensity score” OR "case-control*" OR “case control*”)	8,520,092
6	TS=("executive function*" OR “impulse control” OR “emotion* control” OR “emotion* regulat*” OR “flexible thinking” OR “working memory” OR “self monitor*” OR self-monitor* OR “task initiation” OR ((“problem solving” OR "problem-solving") AND (task OR test OR score)) OR "concept formation" OR "goal orientation" OR "goal-orientation" OR (planning* AND prioriti* AND (organis* OR organiz*)) OR "self regulat*" OR "self-regulat*" OR attention* OR inhibition* OR “inhibitory control” OR “short-term memory” OR “short term memory” OR “immediate memory” OR “verbal memory” OR “nonverbal memory” OR “delayed recall” OR "delayed-recall" OR “free-recall” OR “free recall” OR “serial-recall” OR “serial recall” OR “associative recall” OR "associative-recall" OR cogniti* flex* OR cogniti* refl* OR cogniti* control OR shifting OR "set-shifting" OR “effortful control” OR “self-control” OR “self control” OR “adaptable thinking” OR “task switch*" OR “self management” OR "self-management")	3,487,355
5	#3 OR #4	311,736
4	TS=(childhood NEAR/1 (education OR program* OR care OR initiativ* OR development*))	13,114
3	TS=("first grade" OR "second grade" OR "third grade" OR "fourth grade" OR "fifth grade" OR "sixth grade" OR "seventh grade" OR "eighth grade" OR "ninth grade" OR "tenth grade" OR "eleventh grade" OR "twelfth grade" OR "grade 1" OR "grade 2" OR "grade 3" OR "grade 4" OR "grade 5" OR "grade 6" OR "grade 7" OR "gra de 8" OR "grade 9" OR "grade 10" OR "grade 11" OR "grade 12" OR "1st grade" OR "2nd grade" OR "3rd grade" OR "4th grade" OR "5th grade" OR "6th grade" OR "7t h grade" OR "8th grade" OR "9th grade" OR "10th grade" OR "11th grade" OR "12th grade" OR "elementary school" OR "elementary level" OR "elementary education" OR "primary school" OR	302,945

	"primary schools" OR preschool* OR "pre-school*" OR "junior high school" OR "junior high schools" OR "lower secondary school" OR "lower secondary schools" OR "middle school" OR "middle schools" OR “primary education” OR “secondary education” OR kindergart* OR childcare OR "child* care" OR daycare OR “day care” OR "pre-primar*" OR “pre primar*” OR "early education" OR pre-K OR “pre K” OR prekindergart* OR pre-kindergart* OR nurser* OR “reception class”)	
2	TS=(reading* OR math* OR languag* OR literac* OR numerac* OR ”number sense” OR ”number relations” OR ”number lines” OR ”number understanding” OR “number knowledge” OR “whole number” OR geometr* OR algebra* OR fraction* OR "operations" OR arithmetic* OR addition OR subtraction* OR multiplication* OR statistics* OR "probability" OR calculus* OR combinatoric* OR computation* OR calculation* OR counting* OR “word problem*” OR measurement* OR comprehension* OR decod* OR “word identification*” OR fluency OR phonic* OR “phon* aware*” OR phonem* OR fluency OR spelling OR vocabulary OR alphabetic* OR letter* "print aware*" OR "sound discrim*" OR "rhyme detect*" OR blending OR segmentation OR grammar OR syntax OR syntactic OR morpholog*)	12,103,830
1	TS=(student* OR pupil* OR child* OR toddler* OR youth* OR adolescen* OR teenage* OR young*)	3,271,023

Searching other resources

Hand search

- Journal of Educational Psychology (volumes 111(6) – 116(4)): 455 records screened on title/abstract, 20 in full text, 0 included.
- Child Development (volumes 90(4) – 94(4)): 978 records screened on title/abstract, 13 in full text, 0 included.
- Contemporary Educational Psychology (issues 59 – 79): 376 records screened on title/abstract, 11 in full text, 0 included.
- Early Childhood Research Quarterly (volumes 50 – 69): 607 records screened on title/abstract, 32 in full text, 0 included.
- American Educational Research Journal (volumes 56(6) – 61(5)): 224 records screened on title/abstract, 6 in full text, 1 included.

OpenDissertations

(1933 – present)
EBSCOhost
Expanders - Apply equivalent subjects
Search modes - Boolean/Phrase
Searched 19/03/2024

Search History

#	Query	Results
S24	S3 AND S7 AND S15 AND S19 AND S23	1,896
S23	S20 OR S21 OR S22	399,353
S22	KW (trial* OR experiment* OR intervention* OR treatment* OR ‘control group*’ OR ‘compar* group*’ OR randomiz* OR randomis* OR ‘random-assign*’ OR ‘random* assign* OR ‘quasi-experiment*’ OR ‘quasiexperiment*’ OR ‘instrumental variable* OR ‘regression	23,845

	discontinuity' OR 'difference-in-difference*' OR 'differences-in-difference*' OR 'difference* in difference*' OR 'event stud*' OR 'eventstud*' OR matching OR 'propensity score' OR case-control* OR 'case control*'))	
S21	AB (trial* OR experiment* OR intervention* OR treatment* OR 'control group*' OR 'compar* group*' OR randomiz* OR randomis* OR 'random-assign*' OR 'random* assign* OR 'quasi-experiment*' OR 'quasiexperiment*' OR 'instrumental variable* OR 'regression discontinuity' OR 'difference-in-difference*' OR 'differences-in-difference*' OR 'difference* in difference*' OR 'event stud*' OR 'eventstud*' OR matching OR 'propensity score' OR case-control* OR 'case control*'))	368,527
S20	TI (trial* OR experiment* OR intervention* OR treatment* OR 'control group*' OR 'compar* group*' OR randomiz* OR randomis* OR 'random-assign*' OR 'random* assign* OR 'quasi-experiment*' OR 'quasiexperiment*' OR 'instrumental variable* OR 'regression discontinuity' OR 'difference-in-difference*' OR 'differences-in-difference*' OR 'difference* in difference*' OR 'event stud*' OR 'eventstud*' OR matching OR 'propensity score' OR case-control* OR 'case control*'))	55,908
S19	S16 OR S17 OR S18	94,164
S18	KW ((('executive function*' OR 'impulse control' OR 'emotion* control' OR 'emotion* regulat*' OR 'flexible thinking' OR 'working memory' OR 'working-memory' OR 'self monitor*' OR self-monitor* OR 'task initiation' OR (('problem solving' OR problem-solving) AND (task OR test OR score)) OR 'concept formation' OR 'goal orientation' OR goal-orientation OR (planning* AND prioriti* AND (organis* OR organiz*)) OR 'self regulat*' OR self-regulat* OR attention* OR inhibition* OR 'inhibitory control' OR 'short-term memory' OR 'short term memory' OR 'immediate memory' OR 'verbal memory' OR 'nonverbal memory' OR 'delayed recall' OR delayed-recall OR 'freerecall' OR 'free recall' OR 'serial-recall' OR 'serial recall' OR 'associative recall' OR associative-recall OR cogniti* flex* OR cogniti* refl* OR cogniti* control OR shifting OR set-shifting OR 'effortful control' OR 'self-control' OR 'self control' OR 'adaptable thinking' OR 'task switch* OR 'self management' OR self-management))	7,132
S17	AB ((('executive function*' OR 'impulse control' OR 'emotion* control' OR 'emotion* regulat*' OR 'flexible thinking' OR 'working memory' OR 'working-memory' OR 'self monitor*' OR self-monitor* OR 'task initiation' OR (('problem solving' OR problem-solving) AND (task OR test OR score)) OR 'concept formation' OR 'goal orientation' OR goal-orientation OR (planning* AND prioriti* AND (organis* OR organiz*)) OR 'self regulat*' OR self-regulat* OR attention* OR inhibition* OR 'inhibitory control' OR 'short-term memory' OR 'short term memory' OR 'immediate memory' OR 'verbal memory' OR 'nonverbal memory' OR 'delayed recall' OR delayed-recall OR 'freerecall' OR 'free recall' OR 'serial-recall' OR 'serial recall' OR 'associative recall' OR associative-recall OR cogniti* flex* OR cogniti* refl* OR cogniti* control OR shifting OR set-shifting OR 'effortful control' OR 'self-control' OR 'self control' OR 'adaptable thinking' OR 'task switch* OR 'self management' OR self-management))	87,331
S16	TI ((('executive function*' OR 'impulse control' OR 'emotion* control' OR 'emotion* regulat*' OR 'flexible thinking' OR 'working memory' OR 'working-memory' OR 'self monitor*' OR self-monitor* OR 'task initiation' OR (('problem solving' OR problem-solving) AND (task OR test OR score)) OR 'concept formation' OR 'goal orientation' OR goal-orientation OR (planning* AND prioriti* AND (organis* OR organiz*)) OR 'self regulat*' OR self-regulat* OR attention* OR inhibition* OR 'inhibitory control' OR 'short-term memory' OR 'short term memory' OR 'immediate memory' OR 'verbal memory' OR 'nonverbal memory' OR 'delayed recall' OR delayed-recall OR 'freerecall' OR 'free recall' OR 'serial-recall' OR 'serial recall' OR 'associative recall' OR associative-recall OR cogniti* flex* OR cogniti* refl* OR cogniti* control OR shifting OR set-shifting OR 'effortful control' OR 'self-control' OR 'self control' OR 'adaptable thinking' OR 'task switch* OR 'self management' OR self-management))	11,623
S15	S8 OR S9 OR S10 OR S11 OR S12 OR S13 OR S14	95,613

S14	KW childhood N1 (education OR program* OR care OR initiativ* OR development*)	1,289
S13	AB childhood N1 (education OR program* OR care OR initiativ* OR development*)	959
S12	TI childhood N1 (education OR program* OR care OR initiativ* OR development*)	375
S11	KW ((‘primary education’ OR ‘secondary education’ OR school* OR preschool* OR pre-school* OR kindergart* OR childcare OR ‘child* care’ OR daycare OR ‘day care’ OR pre-primar* OR ‘pre primar*’ OR ‘early education’ OR pre-K OR ‘pre K’ OR prekindergart* OR prekindergart* OR nurser* OR ‘reception class’)	23,053
S10	(TI ((‘primary education’ OR ‘secondary education’ OR school* OR preschool* OR pre-school* OR kindergart* OR childcare OR ‘child* care’ OR daycare OR ‘day care’ OR pre-primar* OR ‘pre primar*’ OR ‘early education’ OR pre-K OR ‘pre K’ OR prekindergart* OR prekindergart* OR nurser* OR ‘reception class’) OR (AB ((‘primary education’ OR ‘secondary education’ OR school* OR preschool* OR pre-school* OR kindergart* OR childcare OR ‘child* care’ OR daycare OR ‘day care’ OR pre-primar* OR ‘pre primar*’ OR ‘early education’ OR pre-K OR ‘pre K’ OR prekindergart* OR prekindergart* OR nurser* OR ‘reception class’))	83,984
S9	KW ((grade* N1 (1 OR 2 OR 3 OR 4 OR 5 OR 6 OR 7 OR 8 OR 9 OR 10 OR 11 OR 12)) OR ((first OR second OR third OR fourth OR fifth OR sixth OR seventh OR eighth OR ninth OR tenth OR eleventh OR twelfth) N1 grade*))	660
S8	(TI ((grade* N1 (1 OR 2 OR 3 OR 4 OR 5 OR 6 OR 7 OR 8 OR 9 OR 10 OR 11 OR 12)) OR ((first OR second OR third OR fourth OR fifth OR sixth OR seventh OR eighth OR ninth OR tenth OR eleventh OR twelfth) N1 grade*)) OR (AB ((grade* N1 (1 OR 2 OR 3 OR 4 OR 5 OR 6 OR 7 OR 8 OR 9 OR 10 OR 11 OR 12)) OR ((first OR second OR third OR fourth OR fifth OR sixth OR seventh OR eighth OR ninth OR tenth OR eleventh OR twelfth) N1 grade*))))	12,272
S7	S4 OR S5 OR S6	559,827
S6	KW (reading* OR math* OR languag* OR literac* OR numerac* OR number* OR geometr* OR algebra* OR fraction* OR operation* OR arithmetic* OR addition* OR subtraction* OR multiplication* OR division* OR statistics* OR probability* OR calculus* OR combinatoric* OR computation* OR calculation* OR counting* OR ‘word problem*’ OR ‘word-problem*’ OR measurement* OR comprehension* OR decod* OR ‘word identification*’ OR ‘word-identification’ OR fluency OR phonic* OR ‘phon* aware*’ OR phonem* OR fluency OR spelling OR vocabulary OR alphabetic* OR letter* ‘print aware*’ OR ‘sound discrim*’ OR ‘rhyme detect*’ OR blending OR segmentation OR grammar OR syntax OR syntactic OR morpholog*)	75,207
S5	AB (reading* OR math* OR languag* OR literac* OR numerac* OR number* OR geometr* OR algebra* OR fraction* OR operation* OR arithmetic* OR addition* OR subtraction* OR multiplication* OR division* OR statistics* OR probability* OR calculus* OR combinatoric* OR computation* OR calculation* OR counting* OR ‘word problem*’ OR ‘word-problem*’ OR measurement* OR comprehension* OR decod* OR ‘word identification*’ OR ‘word-identification’ OR fluency OR phonic* OR ‘phon* aware*’ OR phonem* OR fluency OR spelling OR vocabulary OR alphabetic* OR letter* ‘print aware*’ OR ‘sound discrim*’ OR ‘rhyme detect*’ OR blending OR segmentation OR grammar OR syntax OR syntactic OR morpholog*)	495,273
S4	TI (reading* OR math* OR languag* OR literac* OR numerac* OR number* OR geometr* OR algebra* OR fraction* OR operation* OR arithmetic* OR addition* OR subtraction* OR multiplication* OR division* OR statistics* OR probability* OR calculus* OR combinatoric* OR computation* OR calculation* OR counting* OR ‘word problem*’ OR ‘word-problem*’ OR measurement* OR comprehension* OR decod* OR ‘word identification*’ OR ‘word-identification’ OR fluency OR phonic* OR ‘phon* aware*’ OR phonem* OR fluency OR spelling OR vocabulary OR alphabetic* OR letter* ‘print aware*’ OR ‘sound discrim*’ OR ‘rhyme detect*’ OR blending OR segmentation OR grammar OR syntax OR syntactic OR morpholog*)	83,419

S3	S1 OR S2	180,258
S2	KW (student* OR pupil* OR child* OR toddler* OR youth* OR adolescen* OR teenage* OR young*)	40,227
S1	(TI (student* OR pupil* OR child* OR toddler* OR youth* OR adolescen* OR teenage* OR young*) OR AB (student* OR pupil* OR child* OR toddler* OR youth* OR adolescen* OR teenage* OR young*))	173,551

Working papers/conference proceedings

European Educational Research Association
Date: 2024-09-27
Search site: https://eera-ecer.de/ecer-programmes
Search mode: Anywhere and any conference
Search terms: Executive function
Hits: 53
Screened first level: 53
Included first level: 1
Included second level: 0

American Educational Research Association
Date: 2024-09-27
Search site: https://www.aera.net/Search-Results?Search=search
Search mode: Sort after relevance
Search terms: Executive function~ AND (math~ OR language OR literacy) AND (trial OR intervention)
Hits: 213
Screened first level: 100
Included first level: 0
Included second level: 0

PsyArXiv
Date: 2024-09-17
Search site: https://osf.io/preprints/psyarxix
Search mode: Sort after relevance
Search terms: Executive function
Hits: 4516
Screened first level: 100
Included first level: 1
Included second level: 0

Open Grey
Date: 2024-09-27
Search site: opengrey.eu
Search mode: Sort after relevance
Search terms: executive functions intervention math literacy language
Hits: 100000
Screened first level: 100

Included first level: 1
Included second level: 1

Google
Date: 2024-10-08
Search site: google.com
Search mode: Incognito
Search terms: math OR language OR literacy AND trial OR intervention AND Executive function* OR cognit*
Hits: 1030000000
Screened first level: 100
Included first level: 11
Included second level: 1

Google Scholar
Date: 2024-09-30
Search site: scholar.google.se
Search mode: Incognito (sort after relevance)
Search terms: (‘Executive Function’ OR ‘Inhibition’ OR ‘Short Term Memory’ OR ‘Self Management’ OR ‘Attention’ OR ‘Concept Formation’ OR ‘Goal Orientation’ OR ‘Working memory’) AND (reading* OR math* OR languag* OR literac* OR numerac* OR number* OR geometr* OR algebra* OR fraction*) AND (‘primary education’ OR ‘secondary education’ OR school* OR preschool* OR pre-school* OR kindergart* OR childcare OR ‘child* care’ OR daycare OR ‘day care’ OR pre-primar* OR ‘pre primar*’ OR ‘early education’ OR pre-K OR ‘pre K’ OR pre kindergarten* OR pre- kindergarten* OR nurser* OR ‘reception class’) AND (student* OR pupil* OR child* OR toddler* OR youth* OR adolescen* OR teenage* OR young*) AND (trial* OR experiment* OR intervention* OR treatment* OR ‘control group*’ OR ‘compar* group*’ OR randomiz* OR randomis* OR ‘random-assign*’ OR ‘random* assign* OR ‘quasi-experiment*’ OR ‘quasiexperiment*’)
Hits: 4210
Screened first level: 100
Included first level: 0
Included second level: 0

Social Science Research Network
Date: 2024-10-01
Search site: https://www.ssrn.com/index.cfm/en/
Search mode: Advanced search in title, abstract & keywords
Search terms: Executive function math; Executive function language; Executive function literacy
Hits: 48
Screened first level: 48
Included first level: 0
Included second level: 0

Reports and non-US literature

Danish National Research Database

Date: 2024-03-13
Search site: https://local.forskningsportal.dk/search/1
Search mode: no choice of mode.
Search terms: (Eksekutive funktioner OR Hukommelse OR Tilbagekaldelse OR Følelseskontrol OR Arbejdshukommelse OR Selvmonitorering OR Målorientering OR Selvregulering OR Opmærksomhed OR Genkaldelse) AND (læsning OR matematik OR sprog OR læsefærdigheder OR regnefærdigheder OR tal OR geometri OR algebra OR brøker OR fonologisk OR grammatik OR subtraktion) AND (studerende* OR elev* OR barn* OR børn* OR barndom* OR ungdom* OR teenager* OR ung*)
Hits: 257
Screened first level: 257
Included first level: 6
Included second level: 0

SwePub

Date: 2024-03-27
Search site: https://swepub.kb.se/form_extended.jsp
Search mode: utökad sökning
Search terms: (exekutiva funktioner* OR impulskontroll* OR arbetsminne* OR emotionsreglering* OR målorientering* OR självreglering OR uppmärksamhet* OR inhibering* OR korttidsminne* OR "verbalt minne" OR "icke-verbal minne" OR återkall* OR "kognitiv flexibilitet" OR "kognitiv kontroll" OR självkontroll*) AND (elev* OR barn* OR småbarn* OR tonåring* OR ung*) AND (läs* OR matemati* OR språk* OR literac* OR numeri* OR tal* OR geometr* OR algebra* OR bråk* OR aritmeti* OR addition* OR subtraktion* OR multiplikation* OR division* OR statistik* OR sannolikhet* OR kombinatorik* OR beräkn* OR kalkyl* OR räkn* OR mätning* OR läsförståelse* OR avkod* OR läsflyt* OR fonetik* OR fonem* OR fonolog* OR stav* OR vokabulär OR ordforråd* OR alfabet* OR bokstav* OR bokstäv* OR skrift* OR rim* OR segmentering* OR grammati* OR syntax* OR syntak* OR morfolog*) AND (försök* OR experiment* OR intervention* OR insats* OR kontrollgrupp* OR jämförelsegrupp* OR randomis* OR slumpmässig* OR kvasi-experiment* OR kvasiexperiment* OR instrument* OR regression* OR matchning*)
Hits: 118
Screened first level: 118
Included first level: 4
Included second level: 1

NORA

Date: 2024-03-08
Search site: https://nora.openaccess.no
Search mode: no choice of mode
Search terms: Eksekutive funksjoner; Oppmerksomhet; Mål
Hits: 84
Screened first level: 84

Included first level: 0
Included second level: 0

CORE
Date: 2024-04-15
Search site: https://core.ac.uk/
Search mode: sort by relevance
Search terms: (reading OR math OR language OR literacy OR ”phonological awareness”) AND (”primary education” OR ”secondary education” OR school OR preschool OR kindergarten OR student OR pupil OR child OR toddler OR youth OR adolescent OR adolescence OR teenage OR young) AND (”executive function” OR control OR regulation OR ”working memory” OR self-monitor OR self-regulate OR attention OR inhibition OR “short-term memory” OR recall OR ”self control” OR self-management OR ”attention” OR ”Goal Orientation”) AND (experiment OR intervention OR treatment OR ”control group” OR ”compare group” OR randomize OR randomized OR ”random-assign” OR “random assign” OR quasi-experiment OR case-control OR ”case control”)
Hits: 174759
Screened first level: 100
Included first level: 1
Included second level: 0

Best Evidence Encyclopedia
Date: 2024-03-19
Search site: https://bestevidence.org/
Search mode: no overall search function
Search terms: Executive; Regula*; Goal; Memory; Recall; Impulse; Emotion; Cognit*; and we looked through article references and tables for records that might be relevant.
Hits: 47
Screened first level: 47
Included first level: 2
Included second level: 0

Systematic reviews
Cochrane Library
Date: 2024-04-17
Search site: https://www.cochranelibrary.com/advanced-search
Search mode: searched for "title abstract keyword" adding search lines in the normal search tab, and sorted for relevance
Search terms: used facets S16, S10, S5, S21, and S20 from electronic database search.
Hits: 822
Screened first level: 100
Included first level: 0
Included second level: 0

Campbell Systematic Reviews
Date: 2024-04-26

Search site: https://onlinelibrary.wiley.com/search/advanced?publication=18911803&text1=
Search mode: advanced search
Search terms: used facet S16 from electronic database search.
Hits: 28
Screened first level: 28
Included first level: 0
Included second level: 0

Centre for Reviews and Dissemination Databases
Date: 2024-04-24
Search site: https://www.crd.york.ac.uk/CRDWeb/
Search mode: advanced search
Search terms: (Executive Function OR Inhibition OR Short Term Memory OR Self Management OR Attention OR Concept Formation OR Goal Orientation) AND (reading* OR math* OR languag* OR literac* OR numerac* OR number*) AND (student* OR pupil* OR child* OR toddler* OR youth* OR adolescen* OR teenage* OR young*) AND (student* OR pupil* OR child* OR toddler* OR youth* OR adolescen* OR teenage* OR young*)
Hits: 400
Screened first level: 100
Included first level: 0
Included second level: 0

Backward citation tracking
We backward citation tracked all 79 included records and all relevant reviews (e.g., those mentioned in section *Why it is important to do this review* and those found in other parts of our search, in total 19 reviews) by double-screening the reference lists. We furthermore backward citation tracked a number of primary records that were excluded but where we thought the content was close enough to our review that it might contain relevant leads. In total, we screened 6,897 references (the total number of references in the reference lists) on title and sometimes abstract, screened 162 in full text, and included 7 records.

Forward citation tracking
We forward citation tracked all 79 included records and all relevant reviews (e.g., those mentioned in section *Why it is important to do this review* and those found in other parts of our search, in total 19 reviews) using Google Scholar. We furthermore citation tracked a number of primary records that were excluded but where we thought the content was close enough to our review that it might contain relevant leads. In total, we screened 6,020 references (the total number of references in the reference lists) on title and sometimes abstract, screened 59 in full text, and included 13 records.

Contacts to experts
At the end of the search process (in December 2024), we contacted five international experts to identify unpublished and ongoing studies and studies we might have missed. We contacted authors that, for example, had written relevant reviews or had more than one included primary study and provided them with the inclusion criteria for the review along with the list

of included studies. We received a reply from four authors. There were no new records suggested that we had not screened before.