Journal of Education for Students Placed at Risk (JESPAR)

ISSN: 1082-4669 (Print) 1532-7671 (Online) Journal homepage: www.tandfonline.com/journals/hjsp20

Tutoring Young Children at Risk of Developing Reading Difficulties: Effects of an Early Tutoring Intervention on Literacy Skills and Executive Functions

Julie K. Seerup, Jens Dietrichson, Vibeke M. Jensen & Paul Bingley

To cite this article: Julie K. Seerup, Jens Dietrichson, Vibeke M. Jensen & Paul Bingley (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 (JESPAR), 30:4, 365-395, DOI: 10.1080/10824669.2025.2468701

To link to this article: https://doi.org/10.1080/10824669.2025.2468701

JOURNAL OF EDUCATION FOR STUDENTS PLACED AT RISK (JESPAR)
2025, VOL. 30, NO. 4, 365–395
https://doi.org/10.1080/10824669.2025.2468701

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

Tutoring Young Children at Risk of Developing Reading Difficulties: Effects of an Early Tutoring Intervention on Literacy Skills and Executive Functions

Julie K. Seerup, Jens Dietrichson, Vibeke M. Jensen and Paul Bingley

VIVE – The Danish Center for Social Science Research

ABSTRACT

This study investigated the effects of an early literacy intervention on measures of literacy and executive functions in the first year of Danish primary school. The program targeted six- to seven-year-old children at risk of developing reading difficulties and made use of tutoring and multi-sensory methods. We conducted a randomized controlled trial in eight primary schools with a total sample of 72 students and measured the effects on three standardized measures of literacy: decoding, letter knowledge, and phonological awareness. We also examined the effects on five standardized measures related to executive functions: consistency of attention, inattention, impulse control, verbal memory, and working memory. The tutors wrote short diary notes for each tutoring session, which we used to examine implementation fidelity. The primary analysis indicates that the treatment group improved more than the control group on all but one of the outcome measures; however, the differences are not statistically significant. The exception to the improvements is verbal memory, for which the estimate is negative, small, and insignificant. Exploratory analyses indicate positive and significant effects on broader constructs of attention and literacy, and null effects on memory. The examination of implementation fidelity suggests that deviations from the program manual and contextual factors related to the level and progression of letter knowledge skills weakened the impact. Our results suggest testing the hypothesis that literacy tutoring improves executive functions in larger samples would be promising.

Introduction

Not being able to read proficiently has the potential to hinder learning in school and can have lasting impacts into adulthood (Gustafsson, 2016; Stanovich, 1986; Sulkunen & Malin, 2018). The Organization for Economic Co-operation and Development (OECD) estimates that approximately 20% of Danish 15-year-olds do not have the literacy skills necessary to be classified as *proficient readers*—a level which compares to the OECD average of 26% (OECD, 2023). These students likely encounter reading

CONTACT Julie K. Seerup jkp@vive.dk Herluf Trolles Gade 11, 1052 Copenhagen K, Denmark.
Supplemental data for this article can be accessed online at https://doi.org/10.1080/10824669.2025.2468701.

difficulties early in their school careers (Connor, 2016; Francis et al., 1996). Finding effective early literacy interventions for students who struggle with learning to read is therefore of much concern to education policymakers and researchers.

Literacy skills are associated with more general cognitive skills, such as executive functions (EF; Castles et al., 2018). EF are a set of skills used when directing behavior toward the attainment of a certain goal (Miyake et al., 2000; Zelazo, 2015). The core EF skills typically include inhibitory control, working memory, and cognitive flexibility, which are fundamental building blocks of more general cognitive and social-emotional skills thought to be important for learning (Diamond, 2013). Research shows that learning academic skills, such as reading, requires the use of EF skills (Castles et al., 2018; Clements et al., 2016; Peng & Kievit, 2020). Furthermore, research shows that EF are not only associated with the formation of academic skills (Barnes et al., 2022; Blair & Razza, 2007; Castles et al., 2018; Follmer, 2018; Jacob & Parkinson, 2015; Peng & Kievit, 2020), but also, for example, career success (Prince et al., 2007), marriage satisfaction (Eakin et al., 2004), and physical and mental health (Moffitt et al., 2011). Thus, improving EF may potentially improve a range of desirable outcomes throughout life.

A large number of interventions and commercial products are built on the premise that training cognitive skills improve academic skills; that is, they assume the causality runs from cognitive skills to academic skills (Simons et al., 2016). This is also the direction of the causal relationship that has been examined most, that is, that training EF will affect academic skills. However, despite a large number of studies, the evidence is not compelling that training which solely focus on building EF skills has substantial effects on academic skills (see e.g., Cortese et al., 2015; Jacob & Parkinson, 2015; Melby-Lervåg et al., 2016; Melby-Lervåg & Hulme, 2013; Rapport et al., 2013; Redick et al., 2015; Sala & Gobet, 2017; and Schwaighofer et al., 2015 for reviews). Nevertheless, the effects found in meta- analyses are not precisely estimated null effects and these reviews do not rule out small positive effects on academic skills.

Some studies have instead theorized that the causality between EF skills and academic skills is bi-directional. In other words, it is not only EF skills which can have an effect on academic skills, but training specific academic skills can also improve the development of EF skills (e.g., Castles et al., 2018; Clements et al., 2016; Connor, 2016; Fung et al., 2020; Goldstein, 1976; Peng & Kievit, 2020). In this study we test whether training a set of specific academic skills—literacy skills—improves students' EF skills.

There are at least three theoretical reasons to expect training academic skills to affect EF skills (Dietrichson et al., 2022). One is that cognitive skills such as EF develop through use; with repeated use, the neural circuits involved in the mental operation become more efficient (Zelazo, 2015). A second reason is exemplified by the review of Oberauer et al. (2018), who reported that prior knowledge positively affects performance in working memory tests. Manifestations of these effects include that prior learning improves chunking (i.e., combining items into larger sets), that known words are easier to remember than unknown, and that repetition improves performance (Oberauer et al., 2018). Taatgen (2013) and Gathercole et al. (2016) provided a third reason: that training one set of skills 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.

To sum up, since academic skills training involves the use of EF, and may increase domain-specific knowledge and create novel cognitive routines that can be applied to other tasks involving EF, training academic skills may improve EF skills. If the relationship between EF and academic skills is bi-directional (or runs from academic skills to EF), then interventions targeting specific academic skills may prove more valuable for children's general cognitive development than previously assumed. Further, finding an effect of early literacy training on EF would have both theoretical and practical implications for our understanding of what potential early literacy interventions hold.

Early literacy interventions

The literature on early literacy interventions is vast (for reviews, see e.g., Dietrichson et al., 2017, 2021, 2021; Fryer, 2017; Neitzel et al., 2022; Nickow et al., 2024; Scammacca et al., 2015; Slavin et al., 2011; Suggate, 2016; Wanzek et al., 2006, 2018). Some main conclusions can be drawn concerning which intervention formats work best. For example, a review by Dietrichson et al. (2021) concludes that tutoring is one of two instructional methods that are most commonly associated with large and robust effects for at-risk students. The other method is peer-assisted learning.

Tutoring, either one-to-one or in small groups, allows for individualized pace and support, frequent feedback from the tutor, and social interaction with a supportive role model (i.e., the tutor). These are features emphasized by pedagogical and psychological theories of learning (e.g., Bandura, 1986; Hattie & Timperley, 2007; Vygotsky, 1978). Tutoring provides ample opportunities for students to practice the regulation of behavior, concentration and attention, and involves the use of their working and short-term memory. Thus, tutoring programs share many advantageous features, which may, at least partly, explain why tutoring is generally quite effective (e.g., Juel, 1996).

However, the variation in tutoring effect sizes is large for reasons that are not well understood (Dietrichson et al., 2021; Seerup & Dietrichson, 2023; Wanzek et al., 2016). Potential reasons for the variation could be differences in study-level characteristics (e.g., study design, number of participants, implementation fidelity), treatment-level characteristics (e.g., program content, program form, type of control group, teacher experience and knowledge), or effect size-level characteristics (i.e., variation stemming from the measurement of more than one outcome, which includes test content and implementation problems) across studies (Seerup & Dietrichson, 2023). Most tutoring studies are from the US, but the heterogeneity is also present in studies from other countries. For example, the effects of early literacy interventions including tutoring and implemented in Scandinavian countries, range from small to very large (Bøg et al., 2021; Damsgaard et al., 2022, 2023; Fälth et al., 2015, 2020; Levlin & Nakeva von Mentzer, 2020; Lindström-Sandahl et al., 2023; Wolff, 2011, 2016). In these studies, as well as in the international literature, there is heterogeneity between studies and between outcomes within studies.

Knowledge about why effects differ across tutoring interventions and outcomes is important for the development of better programs, for targeting the right program to the right schools, and for providing the support schools need to implement programs well. Many countries and states are rolling out large-scale tutoring programs in the wake of the COVID-19 pandemic (e.g., Kraft & Falken, 2021; Staunton et al., 2022;

White et al., 2022). The success of these roll-outs depends on the effectiveness of the chosen programs, how well they fit the school contexts, and how well they are implemented. Thus, information that may help schools and policymakers choose between programs and implement them well is currently of particular importance.

Regarding the content of programs, early literacy interventions commonly focus on phonological awareness, phonics, and decoding. Phonological awareness and letter-sound knowledge are important building blocks of children's literacy development (e.g., Bus et al., 1995; Ehri et al., 2001; Snow et al., 1998; Snowling & Hulme, 2011) and are often seen as co-requisites for successful decoding, and therefore for reading acquisition (Share, 1995). Decoding skills are considered necessary for reading comprehension (e.g., Hoover & Tunmer, 2018) and are difficult for children to learn on their own (Castles et al., 2018). Share (1995) argues that proficient decoding may act as a self-teaching mechanism. Learning to decode early may facilitate memory representation of words and increase exposure to print, and therefore aid the acquisition of vocabulary, which may help with comprehension (Blachman et al., 2014). By helping students decipher the alphabetic code early on, programs that train decoding can provide an important stepping-stone to reading comprehension.

Potential effects of early literacy interventions on EF skills

The three aforementioned theoretical reasons to expect training academic skills to affect EF apply also to early literacy interventions. Literacy training clearly involves the use of EF (Castles et al., 2018). For example, to put letter sounds together into words, students have to use their working memory and inhibit impulses to focus on other, non-task-relevant, aspects of the environment. As literacy training may increase domain-specific knowledge, such knowledge is an additional channel through which interventions may affect working memory. Wolff (2014) found effects of a literacy tutoring intervention targeting decoding, phonemic awareness, fluency, and comprehension on Rapid Automated Naming (RAN), which is both a strong predictor of literacy development (e.g., Allor, 2002; Gray & Powell-Smith, 2024) and may be related to EF as some of the relationship between RAN and literacy skills seem to involve working memory (Arnell et al., 2009).

Literacy training may furthermore create new cognitive routines that are useful for other tasks that involve EF. For example, reading with a tutor may provide additional support for student to learn to regulate attention to the tasks at hand than what is prompted in the classroom context. Moreover, verbal processes seem to play an important part in the development and exercise of EF skills (e.g., Bishop et al., 2014; Luria, 1959; Vygotsky, 1978; Weiland et al., 2014). Literacy training may improve phonological, vocabulary, and comprehension skills (Morrison et al., 2019; Stanovich, 1986), and may thus affect EF skills through improved language skills. Melby-Lervåg and Hulme (2010) found that second graders improved serial and free recall memory skills, which are related to EF, after training phoneme awareness and vocabulary, but not after training rhymes.

A small number of empirical studies have tested the hypothesis that programs targeting literacy skills can improve EF skills. Hempenstall (2008) used a combined small-group (around 10 students in each group) and peer-assisted instruction approach for Australian struggling readers in grades four and five. Messer and Nash (2018) investigated the effects of an intervention combining computer-assisted and small-group instruction for struggling readers in grade two in the UK. The interventions in both studies targeted decoding and phonological awareness, and found positive, large, and statistically significant effects on measures corresponding to the targeted content (Hedges' *g* of 0.4–0.9). Both studies also found positive and significant effects on measures of working memory (Hedges' *g* of 0.3–0.4).

The program evaluated in our study, and, for example, programs based on the Orton-Gillingham approach for instructing students with dyslexia, use multi-sensory methods. It is possible that multi-sensory methods both help children overcome limitations of their EF skills, and simultaneously strengthen such skills. Multi-sensory stimulus, by activating the kinesthetic, visual, tactile, and auditory senses, allows information to enter the memory system through multiple processing channels and may thereby help overcome the capacity limitations of each individual channel (Clark & Paivio, 1991; Shams & Seitz, 2008).

For similar reasons, multi-sensory stimulus may help students overcome limitations in their working memory capacity (see Bøg et al., 2021, for a discussion). Training tools used to activate senses, such as figurines and fingerspelling, may furthermore activate episodic memory (the memory of autobiographical events; Tulving, 2002). They may also be more fun and motivating for students and thus help them to pay attention and inhibit non-relevant impulses. As overcoming limitations on EF skills enable students to accomplish more tasks and thus enable more practice, multi-sensory methods may also strengthen EF skills.

While the theoretical foundation for multi-sensory methods is well-established, the evidence base is less robust. Recent meta-analyses have not found compelling evidence that programs including multi-sensory methods improve literacy skills more than programs that do not include such methods (Hall et al., 2023; Stevens et al., 2021). However, Hall et al. (2023) included only five multi-sensory interventions and used study-level moderator analyses. As other study characteristics may confound the association between multi-sensory methods and outcomes, moderator analyses are unlikely to capture causal effects. Stevens et al. (2021) found insignificant effects in favor of Orton-Gillingham methods (effect sizes of 0.32 and 0.14 for foundational literacy skills and vocabulary/comprehension, respectively). These are relatively large effects, considering that the authors included comparisons with alternative treatments in their design and considering typical effect size ranges of school interventions (Kraft, 2020).

Because findings from these meta-analyses are mixed, and some programs that include multi-sensory methods have found large positive effects (e.g., Bøg et al., 2021; Ehri et al., 2007; Mesa et al., 2020), we believe further studies are needed. Furthermore, there are no previous studies testing the effect of a multi-sensory literacy program on EF skills.

The Læseklar program

In this study, we examined the effects of Læseklar (meaning "ready to read"): a multi-sensory and play-based literacy program, originating from Sweden (see Bøg et al., 2021 for a description of the Swedish version). The instruction method is tutoring, either on a 1:1 basis or with small groups of students. Læseklar targets students in primary school who either are at risk of developing reading difficulties or have already developed reading difficulties.

The program's overall aims are threefold. First, to make the students realize that words have both content (e.g., "lion" represents the animal) and form made up of a number of phonemes. Second, the program aims to help students grasp the connection between letters and sounds, i.e., grapheme-phoneme correspondence. Third, and most importantly, the program aims to show students how to decode simple words (i.e., to translate written letter combinations into sounds). The program consists of three different session types, with activities that change as the children progress. Although the sessions did not explicitly target EF, the use of working memory and short-term verbal memory, and students learning how to focus their attention on the tasks at hand were intrinsic to the type of activities.

Type 1 sessions

The first type of session aimed to help students understand that they can divide the audio stream from the spoken language into phonemes and learn how to differentiate the sounds from one another. The tutor used clay figurines representing 24 (Danish) single-letter grapheme-phoneme correspondences, and a box with a compartment for each letter. The tutor also taught the student the corresponding hand sign for each letter and used fingerspelling actively in the sessions. Fingerspelling provides multi-sensory stimulus and is common in literacy instruction for hearing-impaired or deaf children (Morere & Roberts, 2012; Shams & Seitz, 2008). Recent literacy studies using fingerspelling combined with other kinesthetic tools also show positive effects on letter knowledge and decoding-related skills (e.g., Damsgaard et al., 2022, 2023).

Type II and III sessions

In the second type of session, the tutor used printed images representing short words, which the student needed to spell out aloud. To aid their memory of grapheme-phoneme correspondences, the student could use fingerspelling. Using fingerspelling aimed to increase the chance of keeping two or three letter-sound combinations in memory long enough to read words. As young students with potential reading difficulties often have limited working memory capacity (e.g., Castles et al., 2018; Demetriou et al., 2014), aiding working memory in this way may be crucial.

The third type of session aimed to build decoding speed and fluency by practicing the decoding of short syllables with the help of flashcards. The task involved decoding as many syllables as possible during one minute. The task was repeated three times per session; each repetition within a session included the same syllables. The tutor registered the result on a chart to make improvements visible to the students and motivate them to practice more.

The second and third types of sessions were often used interchangeably, which increased the chance of students focusing on the task at hand. These sessions aimed to help students put the letter sounds together to form words and to do so fluently, that is, to build decoding skills.

The current study

The current study employed a pre-registered randomized controlled trial (RCT) and the intervention targeted at-risk students in their first year of primary school. The trial took place in the school year 2019/20, and our analysis included 72 students from eight Danish schools. The research question that guided the study was: what impact does Læseklar have on measures of literacy skills and executive functions?

This study adds to the literature on early literacy interventions by examining the effects on EF (attention, working memory, and verbal memory) and on literacy skills (decoding, letter knowledge, and phonological awareness). Another contribution is the adaptation of Læseklar to a new context. The intervention stems from Sweden, where it has been evaluated using a similar pre-registered RCT design. In the Swedish trial, the authors found large and positive effects of the program on measures of literacy skills but did not examine the effects on EF (Bøg et al., 2021). By embedding the program in a new context, we were able to explore what contextual factors might influence intervention effects (Aguilar, 2020; Cai et al., 2018). Furthermore, the current study was pre-registered, which has not been a standard practice within educational research trials (Dietrichson et al., 2021; Gehlbach & Robinson, 2018, 2021) and we are not aware of any pre-registered study of the effects of literacy training on EF.

Methods

Participants

In total, 13 schools, 26 tutors, and 117 kindergarten students were included in the trial. COVID-19 related attrition meant that only eight schools and 72 students participated in the whole trial (see section on *Attrition and Post-Test Sample* below for more information). We obtained an ethical approval on the 16th of September, 2019. Parents of eligible students were informed of the study and gave their consent prior to intervention, with the option to opt out by choosing “No, I do not want my child to participate in the study: Læseklar” on the consent form. In addition, parents and participating students had the possibility to opt out of the intervention at any time (no one did).

The research team recruited schools from a municipality (the municipalities are the Danish local education authorities) that had agreed to participate (nine schools) and through direct contact with school administrations (four schools). The intervention targeted students at risk of developing reading difficulties in kindergarten, which is the first year of primary school in Denmark (similar to kindergarten in the US).

The Læseklar program was prescriptive regarding how to carry out the tutoring sessions, that is, how the sessions should be structured and conducted. However, the program left the participating schools free to choose how to organize the intervention at their location in terms of who was responsible for carrying out the training, when and where to conduct the training, and so forth. In all cases, schools chose tutors who were kindergarten teachers employed at the schools, most of whom were responsible for managing one of the schools’ kindergarten classes. Information from interviews with teachers showed that the majority of the participating teachers had been employed and teaching kindergarten classes at the schools for several years (median: 12 years;

min-max 0.5–25 years). Of the interviewed teachers ($n = 22$) all but one held an early childhood education degree. The one teacher without this degree was trained as a social and healthcare assistant and had 11 years of work experience as a kindergarten teacher.

The number of tutors at each school ranged from one to three teachers. Most tutors were new to the program, and only a couple were already acquainted with it (as they had helped to develop the Danish version). Most tutors participated in a two-hour introductory course held by two research team members before the intervention started. All tutors had access to online instruction videos and a detailed written manual.

At the beginning of the school year, teachers selected the participating students. The teachers were instructed to choose the students whom they considered to be most in need of extra help learning to read. The teachers could use the pretests (described below in the section *Test Measures*) but were also free to use additional information to guide their selection. However, we required that the participating students were not able to decode any words. The students were, on average, 6.4 years old at the start of the intervention, and 42% were girls. There were between 2 and 13 tutored students in each school, and the tutors were responsible for tutoring between one and six students each.

Table 1 presents demographic data from the kindergarten cohorts at the participating schools compared to national averages. The table contains information on parental educational level (parents holding either a primary, secondary or tertiary education degree), ethnicity, and gender. In sum, the participating schools resemble the national average on the shown demographic measures except for ethnicity. The participating schools have a higher share of students with other ethnic origin than Danish (22% for participating schools vs. national average of 15%).

Attrition and post-test sample

The COVID-19 pandemic and subsequent national lockdowns in spring 2020 affected the current trial. During the first wave of the pandemic, schools were closed from the middle of March until late April 2020, after which restricted re-openings were allowed. The lockdown took effect while some schools were still training the treatment group, and for other schools, in the middle of post-testing, which meant that it was not possible to conduct any post-tests in four schools. At one school, we were able to perform almost all tests of literacy, but were unable to test EF. The lockdown and

Table 1. School level demographic data compared to national average (school year 2019/2020).

	National average	School average	
	Mean (%)	Mean (%)	Range (%)
Student characteristic			
Gender (male)	51	50	39–73
Ethnic origin other than Danish	15	22	9–50
Parental educational level			
Primary education	8	11	0–15
Secondary education	31	33	9–50
Tertiary education	60	61	36–92

Note. Ethnic origin denotes students who are either classified as immigrants or first-generation descendants hereof. Parents educational levels correspond to the following ISCED codes: Primary education = ISCED 0–2; secondary education = ISCED 3–4; tertiary education = ISCED 5–8.

following restrictions (distancing, small groups, short days, etc.) during the re-opening also meant that it was not possible for schools and tutors to resume intervention training or for the researchers to resume post-testing.

In total, attrition was 38.5% of the total number of students, and the number of students we were able to post-test at least once was 72.

Research design

This trial used a waiting list control design. Figure 1 shows an overview of the trial design and randomization procedure. We first stratified by class to accommodate scheduling challenges for schools and then matched students in pairs and, if the number of students were uneven, one triple using their literacy pretest scores. In one case, a school did not need scheduling accommodation due to their classroom organization. In this case, we matched students in pairs and triples without stratifying by class first. We then randomly assigned one student in each matched pair to the treatment group and one to the control group. In the case of triples, we randomly determined whether one or two students would be assigned to the treatment group. The random number sequences used in the randomization were created using Stata's random number generator (StataCorp, 2017).

The stratification procedure implied that each school (and each pair/triple) was its own experiment. Hence, despite the attrition caused by the lockdown, the design retained the crucial advantage of RCTs, namely that the treatment and control groups are balanced in expectation. However, attrition reduced the statistical power, and the smaller sample size increased the risk of chance bias (e.g., Goldberg, 2019; Roberts & Torgerson, 1999).

In total, we randomized 59 students to treatment and 58 students to control, creating a sample of 117 students. Table 2 shows the pretest and post-test scores of the treatment and control group as well as age and gender. As mentioned, there was substantial school-level attrition due to the pandemic. Therefore, we also show the pretest results for the group of students whom we were able to post-test, in addition to the information for the full sample.

As Table 2 shows, there were no large imbalances between the treatment and control groups in the full sample. In the post-test sample, there is one significant difference in favor of the control group on one of the EF subtests (errors of omission; $ES = 0.59$, $p = 0.032$). As Goldberg (2019) described, large differences are likely to appear by chance in small samples and are not necessarily a sign that the randomization is broken. In line with this interpretation, the sign of the differences is not systematic in Table 2—some favor the control group, and others favor the treatment group.

We checked the robustness of the randomization procedure by running a logit regression with the treatment status as the dependent variable and the pretests as independent variables in order to see whether the pretest scores jointly predicted treatment assignment. We did not find any statistically significant indications hereof, neither for the full sample ($n = 111$, $chi^2(5) = 2.37$, $p = 0.796$) nor for the post-test sample ($n = 67$, $chi^2(5) = 8.59$, $p = 0.127$). Note that the samples in these tests were smaller than in Table 2 due to a few students lacking one or more pretests (due to, e.g., illness on the testing days).

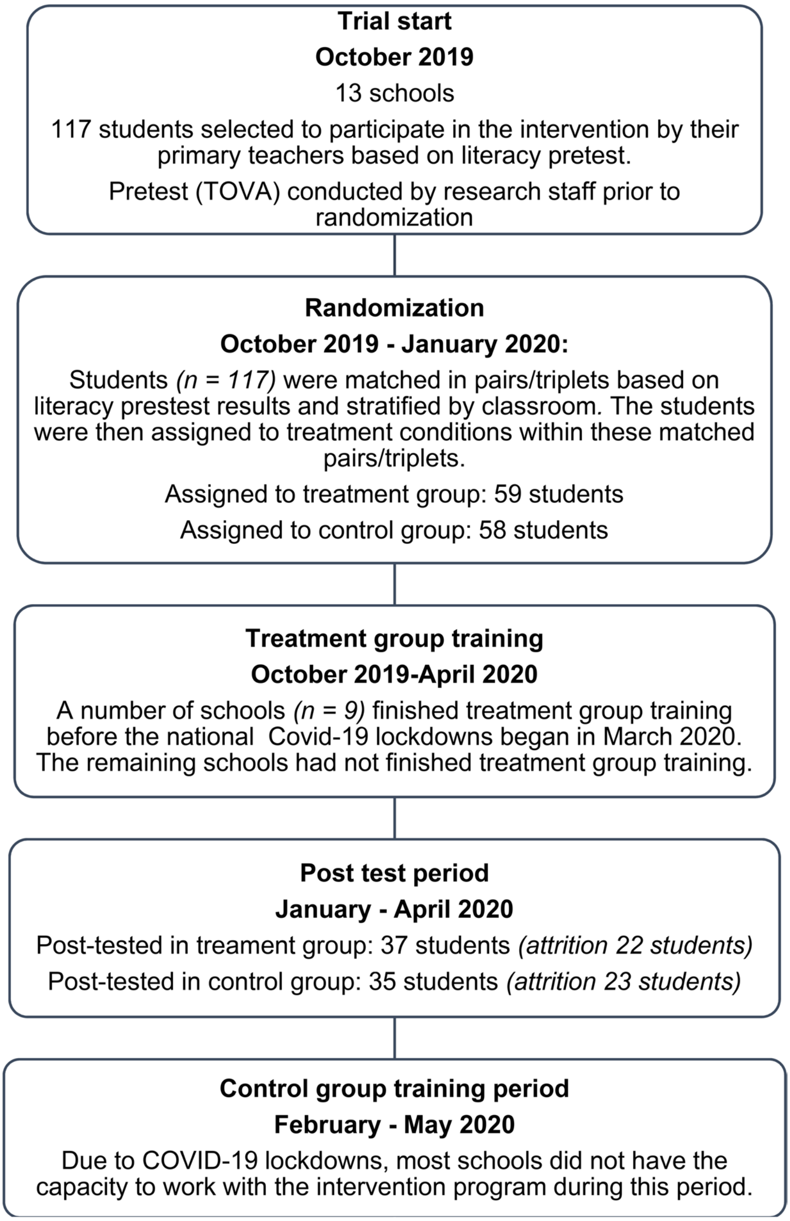

Figure 1. Flowchart of research design and plan.

Control group condition

The control group, the students who were to train with the program after the treatment group, only received the classroom's regular literacy instruction during the period the treatment group trained. One student assigned to the control group received training with the program before the treatment group had ended training, i.e., before the post-test. We take this circumstance into consideration in our sensitivity analyses.

Common for all school settings was an explicit focus on developing phonological awareness. Besides phonological awareness, teachers are required to have an explicit

Table 2. Summary statistics on student characteristics and outcome measures at pre- and post-test.

	Treatment			Control				
Variables	Mean	SD	*N*	Mean	SD	*N*	ES	*p*-Value
Pretest (full sample)								
Age	6.39	0.36	59	6.41	0.38	58	−0.05	0.728
Gender (male = 1)	0.61	0.49	59	0.55	0.50	58	0.08	0.526
Letter knowledge	7.71	4.46	59	7.78	3.92	58	−0.01	0.935
Phonological awareness	4.92	2.55	59	4.66	2.53	58	0.07	0.581
Commission errors	41.47	31.99	57	39.02	19.90	54	0.08	0.631
Omission errors	79.70	65.31	57	70.41	47.68	54	0.16	0.396
Response time variability	270.41	69.20	57	275.78	84.85	54	−0.08	0.715
Pretest (post-test sample)								
Age	6.36	0.35	37	6.43	0.35	35	−0.15	0.358
Gender (male = 1)	0.54	0.51	37	0.54	0.51	35	0.00	0.985
Letter knowledge	6.19	4.29	37	6.94	3.69	35	−0.14	0.428
Phonological awareness	4.62	2.58	37	3.86	2.16	35	0.23	0.174
Commission errors	41.28	35.50	36	35.74	16.76	31	0.17	0.367
Omission errors	87.61	72.05	36	54.03	38.98	31	0.59	0.032
Response time variability	270.03	75.30	36	260.98	64.53	31	0.16	0.643
Post-test								
Decoding	2.35	4.89	37	1.89	3.51	35	0.08	0.646
Letter knowledge	23.41	5.47	37	22.06	5.86	35	0.18	0.316
Phonological awareness	9.69	4.17	36	8.77	3.90	35	0.17	0.339
Commission errors	31.86	24.20	36	45.72	36.82	32	−0.46	0.068
Omission errors	83.47	64.20	36	62.31	43.08	32	0.41	0.120
Response time variability	254.70	65.87	36	270.70	83.73	32	−0.28	0.382
Working memory index	90.56	9.74	32	92.00	10.90	30	−0.11	0.586
Verbal memory	7.97	3.63	32	7.73	3.55	30	0.05	0.798

Note. ES are Hedges' *g* with small sample adjustment. *p*-Values are from two-sample tests assuming equal variance.

focus on developing skills related to conversation, storytelling, writing, and reading. Furthermore, the national guidelines provided by the Ministry of Education (2019) explicitly require that kindergarten students learn to recognize all letters of the alphabet (exceptions are q, w, x, and z, which are not commonly used in Danish).

School visits and interviews with the participating teachers showed that specific learning systems, instructional materials, and ways of organizing learning material for regular literacy instruction in the classroom varied between the schools and teachers (Seerup, 2023a). For example, some schools and teachers followed an unified learning system embedded throughout the school, whereas others were more likely to construct their own learning materials from various learning systems they had encountered throughout their careers.

Test measures

Pretest measures

Literacy tests. The students were pretested using age-corresponding versions of the tests of letter knowledge ("Capital letters") and phonological awareness ("Vowels") from the standardized test battery "Reading evaluation at entry-level" (Borstrøm & Klint Petersen, 2004). Staff at the school, most often the kindergarten teacher or the school's literacy coach, conducted these tests. Danish schools are mandated to conduct tests in kindergarten at the beginning of the school year, which meant that we could use the normal school procedures for our literacy pretests. The school staff had a great deal of experience in conducting the tests we used and were likely to be reliable testers (school staff were not allowed to conduct any of the post-tests in order to ensure adequate blinding). The test material indicated that reliability was adequate: Cronbach's $\alpha = 0.88$ for letter knowledge and $\alpha = 0.74$ for phonological awareness (Alinea, 2023; Petersen, 2005).

We exempted decoding skills from the pretest measures because the absence of decoding skills was an important program inclusion criterion; teachers were explicitly told that students who already could decode one or more words should not receive Læseklar. Furthermore, Danish students are not expected to be able to decode at this stage of primary school, and testing decoding is not part of the testing procedures at the start of kindergarten. For these reasons and because all students scored zero on a similar decoding pretest in the Swedish trial, in which teachers also selected the participating students (Bøg et al., 2021), we expected that all students would obtain a zero score on the decoding test.

Executive function pretest. In addition to the two measures of literacy skills, research staff pretested the students' attention and inhibitory control with the Test of Variables and Attention (TOVA). TOVA is a culture and language-free computerized test that is normed and can be used from when children are four years old (Leark et al., 2008). We used the visual version with both stimulus conditions: infrequent and frequent. We used three subscales: The first scale ("response time variability") measures the consistency of attention. The second scale is a measure of inattention ("errors of omission"), and the third scale measures impulsivity and inhibitory control ("errors of commission"). Pearson's correlation coefficients between stimulus conditions are high and yield an $r = 0.72$ (first half of the test) and $r = 0.70$ (second half of the test) for errors of omission, $r = 0.79$ (first half), and $r = 0.82$ (second half) for errors of commission; $r = 0.70$ (first half) and $r = 0.86$ (second half) for response time variability.

Post-test measures

Research staff conducted all post-tests and were blind to treatment status. We conducted post-tests within one month of the treatment ending.

Literacy post-tests. At the post-tests, the students were tested on three outcomes related to literacy skills: decoding, letter knowledge, and phonological awareness. We measured decoding skills with a subtest from a standardized Danish test of dyslexia ("Elbro's wordlists"), which was developed by Gellert and Elbro (2016) and is widely used in Denmark. The test-retest reliability reported by Gellert and Elbro (2016) is 0.78. The test construction is similar to the decoding test used in the Swedish trial (Elwér et al., 2016), and to the commonly used Word Reading Efficiency subtest of the Test of Word Reading Efficiency (Torgesen et al., 1999). The main difference is that the latter two tests have time caps of 45 s for the total number of words read, whereas "Elbro's wordlists" have time caps of 10 s per word. The maximum score is also lower on the subtest we used (30 points), but this difference was inconsequential as no student scored at the maximum (the highest score was 27).

We used two age-corresponding subtests to measure letter knowledge and phonological awareness ("all letters" and "initial sounds and rhymes"). Both were from the same test battery, Reading evaluation at entry level, as used in the pretests (Borstrøm & Klint Petersen, 2004). The reported test reliability measures were $\alpha = 0.87$ for letter knowledge and $\alpha = 0.79$ for phonological awareness (Alinea, 2023).

Executive function post-tests. In addition to literacy skills, we measured three aspects of EF: attention/inhibitory control, working memory, and verbal memory. We assessed attention/inhibitory control with the same three subscales from TOVA as used in the pretest: response time variability, errors of omission, and errors of commission. Because the tests of working memory and verbal memory had to be conducted by testers with formal psychological or pedagogical training, they required more resources and we could only conduct them at post-test.

We assessed working memory with the Working Memory Index (WMI) from the Danish version of the Wechsler Preschool and Primary Scale of Intelligence—Fourth Edition (WPPSI-IV). The WMI is composed of two subtests: "Picture Memory" and "Zoo Locations". The Danish version of WPPSI-IV has been normed on a Scandinavian sample of 400 children, and age intervals span from 2:6 years to 7:7 years (Wechsler, 2014). In this normed sample, the test reliability shows a Cronbach's alpha of $\alpha = 0.82$ for Zoo Locations and $\alpha = 0.91$ for Picture Memory (Wechsler, 2014).

We assessed verbal memory with the "Paired Recall" subtest from the Danish version of Test of Memory and Learning 2 (TOMAL-2). TOMAL-2 is normed in the United States and can be used with children from age five (Reynolds & Voress, 2017). Cronbach's alpha is $\alpha = 0.87$ for the paired recall test for the subsample of children at the age of six, and the test-retest reliability for individuals aged 5–18 is 0.78 (Reynolds & Voress, 2017).

Implementation of the intervention

The first school began the intervention training in October 2019, and the last school began in January 2020. The training lasted around ten weeks (excluding school holidays). The first school finished the treatment group's training in January 2020 and the last schools were expected to finish in April 2020, had there been no lockdowns.

The tutors were asked to train the students when their class was engaged in activities other than literacy instruction, and while another teacher was present to manage the rest of the class. Tutors typically trained students 1:1, although one tutor trained 1:3, one tutor trained 1:2 and one tutor alternated between 1:1 and 1:2 training. Ten of the participating schools accepted a visit from one of the researchers, who also interviewed 22 of the 26 tutors about their experiences with the program after the intervention had ended (for further information about the qualitative research, see Seerup, 2023a, 2023b).

Implementation fidelity

All tutors were asked to keep a diary on each child, where dates, use of time, progress, and notes were recorded. The research team collected the diaries after post-testing and used them to gain knowledge about the intervention delivery. Table 3 shows the primary analysis sample's descriptive measures for the intervention training.

The program's recommended number of training sessions was between 30 and 35 sessions for a typical participant. If the student had grave difficulties, the number of training sessions could be higher (i.e., the number of sessions was dependent on the needs of the student). Table 3 shows that the average number of training sessions was approximately 25, which is 5–10 sessions fewer than intended or 0.5–1 session less per week than prescribed by the intervention guidelines, assuming a 10-week duration. The median of the total number of training sessions was 27 (the minimum was 7 and the maximum 42), which suggests a large variation in the total number of sessions. Tutors worked with the students in sessions lasting 15 min on average (ranging from 10 to 23 min) in each session, which corresponded well with the intervention guidelines.

The tutor and students worked with Type I session activities in 17 sessions on average (a tutoring session could include more than one type of activity). This meant that the tutors, on average, worked with session I activities in approximately 70% of the total sessions, which was higher than prescribed (the guidelines prescribe 30–40%). The average numbers of Type II and III session activities were 11 and five, respectively, which is below the number of those activities prescribed by the intervention manual.

Analysis plan

We present the results according to the pre-registered analysis plan, which is stored in the American Economic Association's registry for Randomized Controlled Trials with ID AEARCTR-0004803 [an anonymized analysis plan is included as Supplementary Material to this submission]. The pre-registered power analysis assumed 110 students, evenly divided between the treatment and control groups, with a statistical significance of 0.05, and power equal to 0.8. With a similar level of explained variance by covariates as in Bøg et al. (2021; R^2 around 0.56), the resulting minimum detectable effect (MDE) was around 0.4. This MDE is very close to the average of early literacy tutoring interventions targeting at-risk students and decoding in the first two years of primary school (Seerup & Dietrichson, 2023). It is less than half of the effect sizes found for decoding and letter knowledge and below the effect size for phonological awareness in Bøg et al. (2021).

Our primary pre-registered analysis estimated the program's intention-to-treat (ITT) effects using a linear regression framework. The specification included the treatment indicator as an independent variable together with baseline test scores and strata fixed effects as covariates, as recommended by Bruhn and McKenzie (2009):

Table 3. Descriptive measures of implementation for the treatment group.

Variables	Mean	SD	*N*	Min	Max
Total number of tutoring sessions	24.89	8.38	35	7	42
Session length (minutes)	14.88	4.13	22	9.8	22.5
Session 1	16.64	7.16	36	6	39
Session 2	11.13	4.85	31	4	24
Session 3	4.87	3.29	23	1	17

Note. It is possible to work with multiple session types in one tutoring session. Hence, the number of sessions 1, 2, and 3 does not sum up to the total number of sessions.

$$y_{ij,post} = \beta Treatment_{ij} + \lambda y_{ij,pre} + \mu_j + \varepsilon_{ij}, \tag{1}$$

where $y_{ij,post}$ is one of our eight outcome variables for student i in pair/triple j measured at post-test, and $Treatment_{ij}$ is an indicator for whether student i was randomized to the treatment group or not. The vector $\boldsymbol{y}_{ij,pre}$ denotes pretreatment scores from the test of letter knowledge, the test of phonological awareness, and the three tests of attention/inhibition. Parameters to be estimated are denoted by β and λ, μ_j is a pair/triple fixed effect, and ε_{ij} is an individual-level random error term. We examined the program's effects by testing whether β was different from zero. We also calculated Hedges' g with small sample bias correction for all outcomes (Hedges, 1981) to compare our effect sizes to those of other studies.

We reported on the pre-registered sensitivity analyses concerning descriptive balance checks and specifications estimating the short-term effects with and without covariates. Following our pre-registered analysis plan concerning exposure and imperfect compliance, we conducted an instrumental variables analysis in order to account for the control group student who received some training before the post-test (Angrist et al., 1996). Most of the additional pre-registered sensitivity and exploratory analyses were planned with a larger sample size in mind. We report them in Online Supplementary Appendix D but due to the smaller than expected sample size, we think they should be interpreted very cautiously and will therefore not give them much weight in our discussion of the results.

In addition to the main preregistered analyses, we plotted the distribution of scores against the number and type of sessions to examine how implementation fidelity was related to outcomes. We also conducted a non-preregistered exploratory analysis of three broader constructs—literacy, attention, and memory—in order to examine whether the positive but insignificant estimates were signs of illusory results (i.e., that the effects were not real; Gehlbach & Robinson, 2018), or a result of low statistical power.

In this analysis, we standardized all outcomes to have mean zero and standard deviation of one, then pooled the standardized outcomes by construct (giving each outcome the same weight). The literacy outcome contained the decoding, letter knowledge, and phonological awareness scores (i.e., three scores per student). The attention outcome contained response time variability, errors of omission, and errors of commission scores (three scores per student), and the memory outcome contained the WMI and verbal memory scores (two scores per student). We analyzed the pooled outcomes using a similar regression framework as in Equation (1) and clustered the standard errors by student to account for the dependence between the student's scores.

Results

This section presents the results of our pre-registered analyses and non-preregistered exploratory analyses.

Pre-registered analyses

Table 4 presents the results of the primary regression analysis of the program. The analyses showed no significant treatment effect estimates on any of the outcome

Table 4. Primary regression analyses on effects of treatment on outcome measures.

	(1) Decoding	(2) Letter knowledge	(3) Response time variability	(4) Working memory index	(5) Errors of omission	(6) Errors of commission	(7) Verbal memory	(8) Phonological awareness
Treatment	0.67	1.95	−22.66	0.04	3.42	−15.81*	−0.16	1.31
	(0.97)	(1.38)	(15.61)	(3.14)	(13.81)	(7.87)	(0.85)	(0.96)
Pretest scores								
Letter knowledge	0.05	0.37	−4.50	0.92	−4.99	−0.55	−0.11	0.04
	(0.26)	(0.39)	(3.59)	(0.76)	(6.04)	(1.72)	(0.18)	(0.29)
Response time variability	−0.01	−0.00	0.47*	−0.02	0.01	0.16	0.02*	−0.01
	(0.01)	(0.01)	(0.23)	(0.03)	(0.22)	(0.10)	(0.01)	(0.01)
Errors of omission	−0.00	−0.01	0.11	−0.05	0.46*	−0.03	−0.04***	0.00
	(0.01)	(0.01)	(0.27)	(0.03)	(0.26)	(0.09)	(0.01)	(0.01)
Errors of commission	0.04	−0.05	0.09	0.01	−0.03	0.28	0.07**	−0.02
	(0.04)	(0.04)	(0.61)	(0.09)	(0.52)	(0.23)	(0.03)	(0.02)
Phonological awareness	−1.01*	−0.27	3.52	−1.02	0.79	3.85	0.23	−0.34
	(0.52)	(0.76)	(6.87)	(1.24)	(6.18)	(3.16)	(0.43)	(0.59)
Constant	7.08**	27.08***	184.1	90.92***	−52.39	−18.15	8.36***	20.29**
	(3.01)	(5.3)	(65.84)	(10.27)	(73.57)	(22.55)	(2.80)	(9.10)
Observations	67	67	64	58	64	64	58	66
R-squared	0.82	0.76	0.85	0.79	0.82	0.87	0.79	0.84
Pair/Triplet fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Hedges' g	0.15	0.35	−0.30	0.00	0.06	−0.52	−0.04	0.32

Note. Outcome variables are post-test scores. Robust standard errors are shown in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

measures ($p > 0.05$). The point estimates are positive, indicating beneficial effects on all three literacy measures (decoding, $g = 0.15$; letter knowledge, $g = 0.35$; phonological awareness, $g = 0.32$). For the subtests of the TOVA, where negative estimates indicate beneficial effects, the point estimates indicate beneficial effects of treatment on response time variability ($g = -0.30$) and errors of commission ($g = -0.52$), but a small harmful effect on errors of omission ($g = 0.06$). The effect estimate of verbal memory indicates a small harmful effect of treatment ($g = -0.04$), and the estimate for working memory indicates no effect at all ($g = 0.00$).

Performing the regression analyses with no covariates or only pair/triple fixed effects neither changed the results in any substantial way nor yielded any significant results of treatment; see Tables A1 and A2 in the Appendix A for detailed information on the results of these analyses.

For our pre-registered analysis concerning exposure and imperfect compliance, we did not find any substantial change in the results when we considered the one student in the control group who actually received the treatment (see Table A3 in the Appendix A for the results).

Exploratory analyses

Using linear regression, Figure 2 shows the relation between the decoding scores (Y-axis) and the number and type of sessions (X-axis). The solid line shows the relation of the decoding score with the total number of sessions, while the dashed line shows the relation of the decoding scores with the number of Type II and II sessions. The dotted horizontal line defines the mean score for the number of words decoded by the treatment group.

Figure 2 illustrates two main points. First, there is a general (and expected) tendency for students to score better on the decoding test, the more sessions they receive.

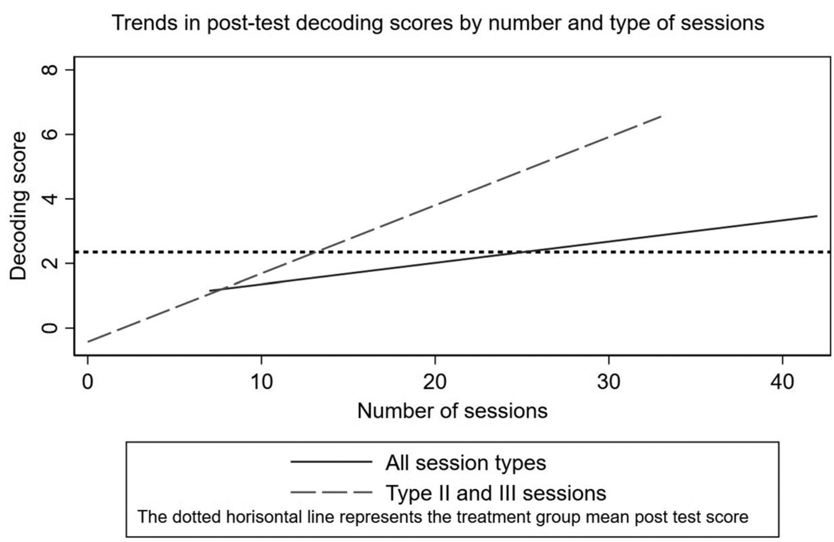

Figure 2. Trends in post-test decoding scores by number and type of sessions.

Second, the relation between the decoding scores and the number of type II and III sessions seems stronger than that between the decoding scores and the total number of sessions. Students who received under 10 Type II or III sessions score below the treatment group mean. All students scoring above the treatment group mean received more than 12 sessions with either Type II or III sessions activities.

The non-preregistered analysis of the three broader constructs of literacy, attention, and memory resulted in a positive and significant effect for literacy ($\beta = 0.26$, $p = 0.015$), a positive and significant effect for attention ($\beta = 0.25$, $p = 0.047$), and a small, negative, and insignificant effect on memory ($\beta = -0.02$, $p = 0.909$).

Discussion

In this study, we examined the effects of early supplementary literacy tutoring on Danish students at risk of reading difficulties in the first year of primary school. We examined the effects on measures of literacy skills and EF. The primary analysis showed mainly positive but statistically insignificant effects for both types of measures. Despite the statistically insignificant effects, we think the results of this paper hold a number of contributions to the field of early literacy tutoring and its potential effects on EF. Below, we discuss these contributions, first in terms of the EF results and then the literacy results. We focus the latter discussion on contextual differences that may explain the smaller effect estimates found in this study compared to a Swedish trial of the same program (Bøg et al., 2021).

The effects of early literacy tutoring on executive functions

One of the main hypotheses that motivated this study is that training specific academic skills will also train and improve EF. The causal relationship between academic skills and EF has mostly been examined in the opposite direction, i.e., that training EF will have an effect on academic skills. This literature has found little robust evidence that training EF skills substantially improves academic skills (Cortese et al., 2015; Jacob & Parkinson, 2015; Melby-Lervåg et al., 2016; Melby-Lervåg & Hulme, 2013; Redick et al., 2015; Sala & Gobet, 2017; Schwaighofer et al., 2015 for reviews). Furthermore, in a recent meta-analysis, Kassai et al. (2019) found significant effects of training working memory, inhibitory control, or cognitive flexibility on the trained component but smaller and insignificant effects on non-trained components of EF. That is, training specific components of EF might not even transfer to other EF skills.

In our primary analysis, we found no statistically significant effects on EF of the intervention program that specifically targeted decoding skills. Other studies that examined the effects of interventions targeting decoding skills on measures of working memory have found statistically significant improvements (Hempenstall, 2008; Messer & Nash, 2018). One potential reason for the lack of significant effects in our study was the loss of large parts of our sample due to COVID-19 school lockdowns, and the resulting decrease in statistical power. In exploratory analyses, we, therefore, analyzed broader constructs of attention and memory (i.e., we analyzed the subtests of attention/inhibition and verbal/working memory jointly) and found a substantial ($g = 0.25$) and significant effect on the broader construct of attention, but not on the broader construct of memory. This analysis was not pre-registered and should be interpreted cautiously. However, the analysis indicated that there may be beneficial effects of Læseklar on children's attention skills.

In contrast to Hempenstall (2008) and Messer and Nash (2018), we found no effect on memory-related tests. Like our intervention, Hempenstall (2008) and Messer and Nash (2018) targeted struggling students, focused on training decoding, and had school staff delivering the intervention. Salient differences include a longer program duration (7–10 months) than Læseklar (10 wk), and that the participants were older (children aged 7–9 years) than the participants in our intervention (6.4 years).

The effects found in meta-analyses of working memory training programs on academic skills typically find positive, small, and statistically insignificant effects (Cortese et al., 2015; Melby-Lervåg et al., 2016; Melby-Lervåg & Hulme, 2013; Rapport et al., 2013; Sala & Gobet, 2017; Schwaighofer et al., 2015). If the small effects are real, they might provide important theoretical insights into the relationship between working memory and literacy skills, although the effects may be too small to justify programs focused solely on training working memory skills isolated from academic skills. Indeed, we think it is too early to completely rule out small effects of working memory training—the most studied component of EF—on literacy skills, and more domain-specific approaches that links EF and literacy training may yield larger effects (Peng & Swanson, 2022).

In sum, larger intervention studies of the relationship between literacy training and working memory and other executive functions are needed.

The effect of early literacy tutoring on literacy measures

The estimates for our literacy measures (decoding $g = 0.15$; letter knowledge $g = 0.35$, phonological awareness $g = 0.32$) were not statistically significant. They were also much smaller than those presented in the evaluation of the Swedish program by Bøg et al. (2021), who found positive and significant effect sizes for phonological awareness ($g = 0.56$), letter knowledge ($g = 1.07$), and decoding ($g = 1.03$). Although we did not obtain the same results as in the Swedish study, we still believe that the Danish Læseklar intervention has potential. Following Kraft (2020), our study's effect sizes are medium in relation to decoding and large in relation to letter knowledge and phonological awareness. If we compare our decoding results to the decoding subgroup analysis presented by Dietrichson et al. (2021), then our estimated effect sizes fall within their 95% confidence level ($g = 0.305$, CI = [0.145, 0.465]). The decoding estimate is less than half of the average effect of early literacy tutoring interventions on literacy measures ($g = 0.433$) in the first two years of primary school in the meta-analysis reported in Seerup and Dietrichson (2023), whereas the letter knowledge and phonological awareness estimates are slightly below this average.

In the exploratory analysis using a broader literacy construct (i.e., a combination of the three outcome measures), we found a significant and positive effect. Although not definitive, the results of the exploratory analyses of broader constructs of both literacy and EF skills indicate that the insignificant estimates were a result of low statistical power rather than the estimated effects being illusory.

Differences and similarities between the Danish and Swedish study contexts

The effects found in Bøg et al. (2021) were unusually large (compare e.g., the reviews of Dietrichson et al.; 2021, Nickow et al., 2024; Seerup & Dietrichson, 2023). Although we—as indicated by our original power analysis—did not expect to replicate their results exactly in Denmark, discussing contextual similarities and differences might still help explain some of the discrepant findings from the Danish and the Swedish studies. Below we discuss similarities and differences between the two study contexts related to research design, language, baseline literacy skills and progression, as well as implementation fidelity.

Research design. Both studies were pre-registered with an a priori analysis plan. The research design in both studies was similar with students matched into pairs or triples within schools prior to randomization. The methods used to estimate the effects were also very similar and so were the literacy outcome measures. The Swedish study included students in first grade in addition to kindergarten and a larger proportion of tutors were special education teachers (there were no special education teachers in the Danish study), but the effects were similar across both grades and types of tutors. These study characteristics are thus unlikely to explain the effect size differences.

Language. A salient difference is that the studies were conducted in different languages. The rates of students with reading difficulties are typically higher in opaque orthographies, like English, compared to more transparent orthographies, like Finnish or German (Castles et al., 2018; Landerl et al., 2013). Instruction in phonological awareness may therefore be less important in more transparent orthographies (see e.g., Furnes & Samuelsson, 2010; Landerl et al., 2013, for discussions). However, the transparency of Danish and Swedish is reasonably similar, both are relatively opaque languages although less so than English (Borleffs et al., 2017). Spoken Danish seems to be more difficult to learn than Swedish (Bleses et al., 2011; Trecca et al., 2021). However, these difficulties pertain to vocabulary and grammar, not phonological development (Trecca et al., 2021). Thus, while we cannot rule out that language differences contributed to the smaller effects in the Danish study, they do not appear to be the most likely explanation.

Baseline literacy skills and progression. Differences related to the participants' baseline literacy skills and progress may be more promising explanations for the different effect sizes. In relation to letter knowledge, Danish students knew, on average, more letters (7–8 letters in the full sample, 6–7 letters in the post-test sample) at the start of the intervention than their Swedish peers (around 5 letters prior to intervention). Furthermore, the control group progressed on average 15.1 letters compared to 5.5 in the Swedish trial. A higher starting level and much faster progress will likely result in smaller effects, especially on a naturally bounded measure such as letter knowledge.

Thus, one explanation for the discrepancy between studies is that the students in the Danish study were simply not as struggling as the students in the Swedish study, and learned more letters during regular classroom instruction. Most students in Denmark and Sweden attended the universal preschool systems prior to starting primary school. Differences in how Danish versus Swedish children are phonologically stimulated in these preschool contexts might influence how they respond to treatment.

As far we know, there is no unified approach for early literacy instruction in Sweden or Denmark. However, most of the schools in the Swedish study used versions of the "Bornholm model" in their primary classroom instruction, which focuses on phonological awareness (Lundberg et al., 1988). The higher level and faster progress in Denmark may indicate that both the Danish preschool system and classroom instruction in kindergarten have a stronger emphasis on letter and letter-sound knowledge than their Swedish counterparts. In support of this hypothesis, the Swedish national curriculum mentions that kindergarten instruction should include letter knowledge but, in contrast to the Danish curriculum, does not mention knowing the letters as a target or requirement (National Agency for Education, 2024).

However, the smaller effects on the decoding test cannot be explained by the fact that students struggled less and progressed faster on letter knowledge. On the contrary, a higher level of letter knowledge ought to make it easier to learn decoding. In line with this argument, the treatment group's average score was considerably lower in the Danish trial than in the Swedish trial (2.4 versus 6.6 words), and the control group's average score was slightly higher (1.9 versus 1.3 words).

Implementation fidelity. Differences in implementation fidelity may have played a part in the differences found between the Danish and Swedish studies. Overall, the average number of sessions in our study was lower than in the Swedish trial (25 compared to 33). In contrast to the Swedish trial, the current trial monitored the implementation process closely. The training diaries showed that there were more Type I sessions than expected (on average, 17 sessions). Type I sessions in the program focused on letter-sound correspondences, whereas Type II and III sessions focused on decoding of simple words and nonsense syllables. The program manual recommended that tutors trained with Type I sessions until the student knew around 12–15 letter sounds, before moving on to Type II and III sessions (expected 10–12 Type I sessions).

If we consider that the average number of letters known in the treatment group prior to intervention was 6.2 and the progress was fast, then there might have been an over-emphasis on Type I activities relative to Type II and III. Training past the recommended 12–15 letter sounds before moving on to Type II and III sessions was also seen in some of the training diaries. As Figure 2 also suggests, the number and mix of the different types of sessions might matter. Therefore, a plausible explanation for the small effects on decoding compared to the effects of the Swedish trial might be that the mix of sessions conducted in the Danish setting was not optimal and that the Danish students received, on average, fewer sessions.

Two possible reasons why implementation fidelity differed stand out. First, in the Swedish study, the program developer conducted one training session for school staff prior to the intervention. School staff had the opportunity to see the program developer work with the intervention material and were able to ask questions about the material and method in person. This session may have resulted in greater fidelity than in the Danish context, where the training of school staff primarily was given in the form of instruction videos and a written manual (which were also available for the Swedish tutors).

Second, the selection of participating schools differed between the study contexts. In the Swedish study, the majority of schools were contacted directly and school administrations and relevant teachers decided on participation. In the Danish case, the majority of schools were recruited through a municipality, and only a few schools were contacted directly. Unfortunately, we could only post-test one of the schools we contacted directly. Although the Danish school administrations and teachers that were recruited through the municipality were given the option not to participate, it is still more likely that the schools in Sweden were of a type that had the resources and motivation to benefit from the program compared to those in Denmark. Furthermore, the recruitment process may have affected the level of buy-in from school staff, who were responsible for implementing the Læseklar intervention.

Summary. As mentioned previously in the paper, there are many suggested reasons why effect sizes vary between and within studies, especially in relation to implementation (Seerup & Dietrichson, 2023). Our combination of similar research designs and detailed data on implementation and contextual factors have allowed us to develop and contribute with more concise hypotheses concerning the reasons why we observe differences in effect sizes. As for letter knowledge, it seems plausible that the higher baseline level and the faster progress of the control group explain the differences. Concerning decoding, it seems that implementation fidelity, in particular the mix of training sessions, is a likely explanation of the differences we observe.

Limitations

The most important limitation of this study is the substantial attrition due to the lockdown in the first wave of the COVID-19 pandemic in 2020. The national lockdown in Denmark and the subsequent school restrictions made it impossible for the research team to conduct post-tests for 39% of the initial sample. This attrition considerably reduced the statistical power. However, because of the stratified research design, the randomization was still valid in the schools that could be post-tested. As the stratification made each school a mini-experiment, the effect estimates are unbiased in expectation also when whole schools could not be post-tested. The results, although underpowered, therefore still provide valuable information—for example for meta-analysis, in which chance bias is likely to average out across studies.

Conclusions and future directions

In this paper we examined the effects of Læseklar, an early literacy tutoring program targeting students in kindergarten who were considered at risk of developing reading difficulties. We used an RCT to examine the effects of the program on literacy skills and executive functions. We measured literacy skills by testing constructs of decoding, letter knowledge and phonological awareness, and executive functions by three subtests related to attention/inhibition, a test of verbal memory, and a test of working memory. Two main contributions stand out.

Firstly, this study contributes with one brick in the building of a strong research literature on the effects of literacy interventions on executive functions. In our pre-registered primary analysis, we found positive but insignificant effect estimates on the three literacy tests and mixed and insignificant effect estimates on the tests of executive functions. In non-pre-registered exploratory analyses, we examined three broader constructs and found positive and significant effects on literacy and attention, and null effects on memory. Thus, our results suggest testing the hypothesis that literacy instruction improves executive functions, particularly attention, in larger samples would be promising. Larger samples would also allow for moderator analyses of the heterogeneity of treatment effects across baseline levels of EF and literacy skills. To test whether the relation between EF and literacy skills is bi-directional, including treatments that train only EF and the combination of EF and literacy would be interesting. Additionally, focusing future research on early literacy interventions where contextually relevant EF training is embedded in the content would be of interest as well (see e.g., Fuchs et al., 2022 for an example in the areas of working memory and mathematics).

Secondly, the contrast between our results and the results in a Swedish trial of the same program provides lessons for educational policymakers and schools, and hypotheses for future research. Two different hypotheses for why the Danish intervention did not have the same effects as the Swedish regarding letter knowledge and decoding stand out. Concerning the difference in letter knowledge, Danish students had a slightly better starting point compared to their Swedish peers. More importantly, the Danish control students progressed considerably more, likely due to a stronger emphasis on letter knowledge during regular instruction in Denmark. These results indicate that targeting students most in need of help and content that receives less emphasis in regular instruction may yield larger effects.

Concerning the difference in decoding effects, we believe that the lack of implementation fidelity likely explains the smaller effects on decoding in Denmark compared to Sweden. Our results indicated an over-emphasis on Type I sessions and too few Type II and III sessions in the Danish trial. Providing adequate support for implementation or targeting schools with adequate implementation resources might, therefore, be as important as the chosen programs for the success of large-scale roll-outs of tutoring programs. Further examinations of how much implementation quality and dosage, as well as the severity of student difficulties and targeted content, can explain the large variation in effect sizes across tutoring programs would clearly be valuable.

Acknowledgements

The authors of this article would like to thank Professor Carole Torgerson and Professor Björn Öckert for providing helpful comments on the study's analysis plan.

Ethical approval

Before the trial began, the research project gained an IRB approval from VIVE—the Danish Center for Social Science Research (case 301309) and pre-registered the randomized trial in the American Economic Association's registry for randomized controlled trials (ID AEARCTR-0004803).

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

TrygFonden funded the research [grant 128499].

ORCID

Julie K. Seerup http://orcid.org/0000-0002-2523-7456
Jens Dietrichson http://orcid.org/0000-0003-4178-7682

Data availability statement

We do not have permission from the participants to share the data publicly but the data can be made available for replication purposes from the corresponding author upon reasonable request.

References

Aguilar, M. S. (2020). Replication studies in mathematics education: What kind of questions would be productive to explore? *International Journal of Science and Mathematics Education*, *18*(S1), 37–50. https://doi.org/10.1007/s10763-020-10069-7

Alinea. (2023). *Baggrundsoplysninger om prøverne i Læseevaluering på begyndertrinnet*. https://www.alinea.dk/laeseevaluering-laerervejledning (last accessed 2023 March 22).

Allor, J. H. (2002). The relationships of phonemic awareness and rapid naming to reading development. *Learning Disability Quarterly*, *25*(1), 47–57. https://doi.org/10.2307/1511190

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, *91*(434), 444–455. https://doi.org/10.1080/01621459.1996.10476902

Arnell, K. M., Joanisse, M. F., Klein, R. M., Busseri, M. A., & Tannock, R. (2009). Decomposing the relation between Rapid Automatized Naming (RAN) and reading ability. *Revue Canadienne De Psychologie Experimentale* [Canadian Journal of Experimental Psychology], *63*(3), 173–184. https://doi.org/10.1037/a0015721

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice-Hall.

Barnes, Z. T., Boedeker, P., Cartwright, K. B., & Zhang, B. (2022). Socioeconomic status and early reading achievement: How working memory and cognitive flexibility mediate the relation in low-achieving and typically developing K to first grade students. *Journal of Research in Reading*, *45*(2), 204–222. https://doi.org/10.1111/1467-9817.12398

Blachman, B. A., Schatschneider, C., Fletcher, J. M., Murray, M. S., Munger, K. A., & Vaughn, M. G. (2014). Intensive reading remediation in grade 2 or 3: Are there effects a decade later? *Journal of Educational Psychology*, *106*(1), 46–57. https://doi.org/10.1037/a0033663

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 of London. Series B, Biological Sciences*, *369*(1634), 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

Bleses, D., Basbøll, H., & Vach, W. (2011). Is Danish difficult to acquire? Evidence from Nordic past-tense studies. *Language and Cognitive Processes*, *26*(8), 1193–1231. https://doi.org/10.1080/01690965.2010.515107

Borleffs, E., Maassen, B. A., Lyytinen, H., & Zwarts, F. (2017). Measuring orthographic transparency and morphological-syllabic complexity in alphabetic orthographies: A narrative review. *Reading and Writing*, *30*(8), 1617–1638. https://doi.org/10.1007/s11145-017-9741-5

Bøg, M., Dietrichson, J., & Isaksson, A. A. (2021). A multi-sensory tutoring program for students at-risk of reading difficulties. *Journal of Educational Research*, *115*(3), 233–251.

Borstrøm, I., & Klint Petersen, D. (2004). *Læsetræning på begyndertrinnet*. Alinea.
Bruhn, M., & McKenzie, D. (2009). In pursuit of balance: Randomization in practice in development field experiments. *American Economic Journal: Applied Economics*, *1*(4), 200–232. https://doi.org/10.1257/app.1.4.200
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
Cai, J., Morris, A., Hohensee, C., Hwang, S., Robison, V., & Hiebert, J. (2018). The role of replication studies in educational research. *Journal for Research in Mathematics Education*, *49*(1), 2–8. https://doi.org/10.5951/jresematheduc.49.1.0002
Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. *Psychological Science in the Public Interest*, *19*(1), 5–51. https://doi.org/10.1177/1529100618772271
Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review*, *3*(3), 149–210. https://doi.org/10.1007/BF01320076
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
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: Metaanalysis 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
Damsgaard, L., Nielsen, A. M. V., Gejl, A. K., Malling, A. S. B., Jensen, S. K., & Wienecke, J. (2022). Effects of 8 weeks with embodied learning on 5–6-year-old Danish children's pre-reading skills and word reading skills: The PLAYMORE project, DK. *Educational Psychology Review*, *34*(3), 1709–1737. https://doi.org/10.1007/s10648-022-09671-8
Damsgaard, L., Nielsen, A.-M. V., Topor, M. K., Hansen, R. A., Jensen, S. K., Markers, R. L., Gejl, A. K., Malling, A. S. B., & Wienecke, J. (2023). Embodied learning activities focusing on letter-sound knowledge increase spelling performance in 1st grade children with low and high reading ability. *Educational Psychology Review*, *35*(3), 74. https://doi.org/10.1007/s10648-023-09791-9
Demetriou, A., Spanoudis, G., & Shayer, M. (2014). Inference, reconceptualization, insight, and efficiency along intellectual growth: A general theory. *Enfance*, *3*(3), 365–396. https://doi.org/10.3917/enf1.143.0365
Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*(1), 135–168. https://doi.org/10.1146/annurev-psych-113011-143750
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., Seerup, J. K., Klokker, R. H., Viinholt, B. C. 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 7–12: A systematic review. *Campbell Systematic Reviews*, *17*(2), e1081. https://doi.org/10.1002/cl2.1081
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 grade 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. 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

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

Ehri, L. C., Dreyer, L. G., Flugman, B., & Gross, A. (2007). Reading Rescue: An effective tutoring intervention model for language-minority students who are struggling readers in first grade. *American Educational Research Journal*, *44*(2), 414–448. https://doi.org/10.3102/0002831207302175

Ehri, L., Nunes, S. R., Willows, D. M., Valeska, Schuster, B., Yaghoub-Zadeh, Z., & Shanahan, T. (2001). Phonemic awareness instruction helps children learn to read: Evidence from the National Reading Panel's meta-analysis. *Reading Research Quarterly*, *36*(3), 250–287. https://doi.org/10.1598/RRQ.36.3.2

Elwér, Å., Fridolfsson, I., Samuelsson, S., & Wiklund, C. (2016). *LäSt – Test i läsförståelse, läsning och stavning för åk 1-6*. Stockholm: Hogrefe Psykologiförlaget AB.

Fälth, L., Nilvius, C., & Anvegård, E. (2015). Intensive reading with reading lists: An intervention study. *Creative Education*, *06*(22), 2403–2409. https://doi.org/10.4236/ce.2015.622246

Fälth, L., Svensson, E., & Ström, A. (2020). Intensive phonological training with articulation – an intervention study to boost pupils' word decoding in grade 1. *Journal of Cognitive Education and Psychology*, *19*(2), 161–171. https://doi.org/10.1891/JCEP-D-20-00015

Follmer, J. D. (2018). Executive function and reading comprehension: A meta-analytic review. *Educational Psychologist*, *53*(1), 42–60. https://doi.org/10.1080/00461520.2017.1309295

Francis, D. J., Shaywitz, S. E., Stuebing, K. K., Shaywitz, B. A., & Fletcher, J. M. (1996). Developmental lag versus deficit models of reading disability: A longitudinal, individual growth curves analysis. *Journal of Educational Psychology*, *88*(1), 3–17. https://doi.org/10.1037/0022-0663.88.1.3

Fryer, R. G. (2017). The production of human capital in developed countries: Evidence from 196 randomized field experiments. In A. Banerjee and E. Duflo (Eds.), *Handbook of field experiments* (vol. 2, pp. 95–322). Elsevier.

Fung, W. K., Chung, K. K. H., & Lam, C. B. (2020). Executive functioning and word reading in Hong Kong Chinese children: A 1-year longitudinal perspective. *Journal of Research in Reading*, *43*(3), 382–393. https://doi.org/10.1111/1467-9817.12302

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

Furnes, B., & Samuelsson, S. (2010). Predicting reading and spelling difficulties in transparent and opaque orthographies: A comparison between Scandinavian and US/Australian children. *Dyslexia*, *16*(2), 119–142. https://doi.org/10.1002/dys.401

Gathercole, S. E., Dunning, D. L., Holmes, J., & Norris, D. (2016). 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

Gehlbach, H., & Robinson, C. D. (2018). Mitigating illusory results through preregistration in education. *Journal of Research on Educational Effectiveness*, *11*(2), 296–315. https://doi.org/10.1080/19345747.2017.1387950

Gehlbach, H., & Robinson, C. D. (2021). From old school to open science: The implications of new research norms for educational psychology and beyond. *Educational Psychologist*, *56*(2), 79–89. https://doi.org/10.1080/00461520.2021.1898961

Gellert, A. S., & Elbro, C. (2016). *Rapport om projekt vedrørende tidlig identifikation af elever i risiko for udvikling af alvorlige afkodningsvanskeligheder (herunder ordblindhed)*. Centre for Reading Research.

Goldberg, M. H. (2019). How often does random assignment fail? Estimates and recommendations. *Journal of Environmental Psychology*, *66*, 101351. https://doi.org/10.1016/j.jenvp.2019.101351

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

Gray, J. S., & Powell-Smith, K. A. (2024). Rapid automatized naming: What it is, what it is not, and why it matters. *Annals of Dyslexia*. Advance online publication. https://doi.org/10.1007/s11881-024-00312-z

Gustafsson, J. E. (2016). Lasting effects of quality of schooling: Evidence from PISA and PIAAC. *Intelligence*, *57*, 66–72. https://doi.org/10.1016/j.intell.2016.05.004

Hall, C., Dahl-Leonard, K., Cho, E., Solari, E. J., Capin, P., Conner, C. L., Henry, A. R., Cook, L., Hayes, L., Vargas, I., Richmond, C. L., & Kehoe, K. F. (2023). Forty years of reading intervention research for elementary students with or at risk for dyslexia: A systematic review and meta-analysis. *Reading Research Quarterly*, *58*(2), 285–312. https://doi.org/10.1002/rrq.477

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112. https://doi.org/10.3102/003465430298487

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

Hempenstall, K. (2008). Corrective reading: An evidence-based remedial reading intervention. *Australasian Journal of Special Education*, *32*(1), 23–54. https://doi.org/10.1080/10300110701839949

Hoover, W. A., & Tunmer, W. E. (2018). The simple view of reading: Three assessments of its adequacy. *Remedial and Special Education*, *39*(5), 304–312. https://doi.org/10.1177/0741932518773154

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

Juel, C. (1996). What makes literacy tutoring effective? *Reading Research Quarterly*, *31*(3), 268–289. https://doi.org/10.1598/RRQ.31.3.3

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

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., & Falken, G. T. (2021). *A blueprint for scaling tutoring across public schools* (No. 20). Annenberg Institute for School Reform at Brown University.

Landerl, K., Ramus, F., Moll, K., Lyytinen, H., Leppänen, P. H. T., Lohvansuu, K., O'Donovan, M., Williams, J., Bartling, J., Bruder, J., Kunze, S., Neuhoff, N., Tóth, D., Honbolygó, F., Csépe, V., Bogliotti, C., Iannuzzi, S., Chaix, Y., Démonet, J.-F., … Schulte-Körne, G. (2013). Predictors of developmental dyslexia in European orthographies with varying complexity. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *54*(6), 686–694. https://doi.org/10.1111/jcpp.12029

Leark, R. A., Greenberg, L. M., Kindschi, C. L., Dupuy, T. R., & Hughes, S. J. (2008). *TOVA professional manual*. TOVA Company.

Leark, R. A., Greenberg, L. M., Kindschi, C. L., Dupuy, T. R., & Hughes, S. J. (2017). *TOVA professional manual*. TOVA Company.

Levlin, M., & Nakeva von Mentzer, C. (2020). An evaluation of systematized phonics on reading proficiency in Swedish second grade poor readers: Effects on pseudoword and sight word reading skills. *Dyslexia*, *26*(4), 427–441. https://doi.org/10.1002/dys.1669

Lindström-Sandahl, H., Elwér, Å., Samuelsson, S., & Danielsson, H. (2023). Effects of a phonics intervention in a randomized controlled study in Swedish second-grade students at risk of reading difficulties. *Dyslexia*, *29*(4), 290–311. https://doi.org/10.1002/dys.1751

Lundberg, I., Frost, J., & Petersen, O. (1988). Effects of an extensive program for stimulating phonological awareness in preschool children. *Reading Research Quarterly*, *23*(3), 263–284. https://doi.org/10.1598/RRQ.23.3.1

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
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
Mesa, C., Newbury, D. F., Nash, M., Clarke, P., Esposito, R., Elliott, L., De Barbieri, Z., Fernández, M. A., Villanueva, P., Hulme, C., & Snowling, M. J. (2020). The effects of reading and language intervention on literacy skills in children in a remote community: An exploratory randomized controlled trial. *International Journal of Educational Research*, *100*, 101535. https://doi.org/10.1016/j.ijer.2020.101535
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
Ministry of Education. (2019). Børnehaveklassen – Faghæfte 2019. https://emu.dk/sites/default/files/2020-09/GSK_fagh%C3%A6fte_b%C3%B8rnehaveklassen_2020.pdf, accessed 2024-07-04.
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*, *108*(7), 2693–2698. https://doi.org/10.1073/pnas.1010076108
Morere, D. A., & Roberts, R. (2012). Fingerspelling. In D. Morere, & T. Allen, T. (Eds.), *Assessing literacy in deaf individuals: Neurocognitive measurement and predictors* (pp. 179–189). Springer.
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
National Agency for Education. (2024). Läroplan för grundskolan, förskoleklassen och fritidshemmet – Lgr 22. https://www.skolverket.se/undervisning/forskoleklassen/laroplan-for-forskoleklassen (accessed 2024 July 4).
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/10.3102/00028312231208687
Oberauer, K., Lewandowsky, S., Awh, E., Brown, G. D. A., 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 (2023). *PISA 2022 results (volume I): The state of learning and equity in education.* OECD Publishing.
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

Petersen, D. K. (2005). Distinkthedstræning i børnehaveklassen. *Dysleksi*, *3*, 16–19.

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. https://doi.org/10.1016/S0140-6736(07)61238-0

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

Reynolds, C. R., & Voress, J. K. (2017). *Test af Hukommelse og Indlæring 2. udgave – Vejledning (Test of memory and learning second edition – examiner's maunal)*. Hogrefe Psykologisk Forlag.

Roberts, C., & Torgerson, D. J. (1999). Baseline imbalance in randomised controlled trials. *BMJ*, *319*(7203), 185–185. https://doi.org/10.1136/bmj.319.7203.185

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

Seerup, J. K. (2023a). Balancing tutoring with classroom life: Identifying teachers' negotiations and adaptations of school-based tutoring. In Seerup (Ed.), Ready to read: Exploring school-based tutoring interventions for struggling readers [Doctoral dissertation]. University of Copenhagen.

Seerup, J. K. (2023b). Emotion work in tutoring: How teachers shape, navigate and expand emotions of tutoring. In Seerup (Ed.), Ready to read: Exploring school-based tutoring interventions for struggling readers [Doctoral dissertation]. University of Copenhagen.

Seerup, J. K., & Dietrichson, J. (2023). Heterogeneity in school-based tutoring interventions targeting at-risk students in reading. An exploratory meta-analysis. In Seerup (Ed.), Ready to read: Exploring school-based tutoring interventions for struggling readers [Doctoral dissertation]. University of Copenhagen.

Scammacca, N. K., Roberts, G., Vaughn, S., & Stuebing, K. K. (2015). A meta-analysis of interventions for struggling readers in grades 4-12: 1980-2011. *Journal of Learning Disabilities*, *48*(4), 369–390. https://doi.org/10.1177/0022219413504995

Schwaighofer, M., Fischer, 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

Shams, L., & Seitz, A. (2008). Benefits of multisensory learning. *Trends in Cognitive Sciences*, *12*(11), 411–417. https://doi.org/10.1016/j.tics.2008.07.006

Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition*, *55*(2), 151–218. https://doi.org/10.1016/0010-0277(94)00645-2

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

Slavin, R. E., Lake, C., Davis, S., & Madden, N. (2011). Effective programs for struggling readers: A best-evidence synthesis. *Educational Research Review*, *6*(1), 1–26. https://doi.org/10.1016/j.edurev.2010.07.002

Snow, C., Burns, S., & Griffin, P, the Committee on the Prevention of Reading Difficulties in Young Children, National Research Council. (1998). *Preventing reading difficulties in young children*. The National Academies Press.

Snowling, M., & Hulme, C. (2011). Evidence-based interventions for reading and language difficulties: Creating a virtuous circle. *The British Journal of Educational Psychology*, *81*(Pt 1), 1–23. https://doi.org/10.1111/j.2044-8279.2010.02014.x

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, *21*(4), 360–407. https://doi.org/10.1598/RRQ.21.4.1

StataCorp. (2017). *Stata statistical software: Release 15*. StataCorp LLC.

Staunton, R., Worth, J., Lynch, S., & Lucas, M. (2022). *Evaluation of the national tutoring programme year 2 study plan* (Technical report). Government Social Research.

Stevens, E. A., Austin, C., Moore, C., Scammacca, N., Boucher, A. N., & Vaughn, S. (2021). Current state of the evidence: Examining the effects of Orton-Gillingham reading interventions for students with or at-risk of word-level reading disabilities. *Exceptional Children*, *87*(4), 397–417. https://doi.org/10.1177/0014402921993406

Suggate, S. P. (2016). A meta-analysis of the long-term effects of phonemic awareness, phonics, fluency, and reading comprehension interventions. *Journal of Learning Disabilities*, *49*(1), 77–96. https://doi.org/10.1177/0022219414528540

Sulkunen, S., & Malin, A. (2018). Literacy, age and recentness of education among Nordic adults. *Scandinavian Journal of Educational Research*, *62*(6), 929–948. https://doi.org/10.1080/00313831.2017.1324898

Taatgen, N. A. (2013). The nature and transfer of cognitive skills. *Psychological Review*, *120*(3), 439–471. https://doi.org/10.1037/a0033138

Trecca, F., Tylén, K., Højen, A., & Christiansen, M. H. (2021). Danish as a window onto language processing and learning. *Language Learning*, *71*(3), 799–833. https://doi.org/10.1111/lang.12450

Torgesen, J. K., Rashotte, C. A., & Wagner, R. K. (1999). *TOWRE: Test of word reading efficiency*. Pro-ed.

Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology*, *53*(1), 1–25. https://doi.org/10.1146/annurev.psych.53.100901.135114

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard university press.

Wanzek, J., Stevens, E. A., Williams, K. J., Scammacca, N., Vaughn, S., & Sargent, K. (2018). Current evidence on the effects of intensive early reading interventions. *Journal of Learning Disabilities*, *51*(6), 612–624. https://doi.org/10.1177/0022219418775110

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*(6), 528–543. https://doi.org/10.1177/00222194060390060501

Wanzek, J., Vaughn, S., Scammacca, N., Gatlin, B., Walker, M. A., & Capin, P. (2016). Meta-analyses of the effects of tier 2 type reading interventions in grades K-3. *Educational psychology review*, *28*, 551–576. https://doi.org/10.1007/s10648-015-9321-7

Wechsler, D. (2014). *WPPSI-IV Wechsler Presschool and Primary Scale of Intelligence – fourth edition. Manual PART 1* (Danish version). Pearson Education, Inc.

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

White, S., Groom-Thomas, L., & Loeb, S. (2022). Undertaking complex but effective instructional supports for students: A systematic review of research on high-impact tutoring planning and implementation (No. 22). Annenberg Institute for School Reform at Brown University.

Wolff, U. (2011). Effects of a randomised reading intervention study: An application of structural equation modelling. *Dyslexia*, *17*(4), 295–311. https://doi.org/10.1002/dys.438

Wolff, U. (2014). RAN as a predictor of reading skills, and vice versa: Results from a randomised reading intervention. *Annals of Dyslexia*, *64*(2), 151–165. https://doi.org/10.1007/s11881-014-0091-6

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

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

Appendix A

Table A1. Primary regression analyses with no covariates.

	(1) Decoding	(2) Letter knowledge	(3) Response time variability	(4) Working Memory Index	(5) Errors of omission	(6) Errors of commission	(7) Verbal memory	(8) Phonological awareness
Treatment	0.47	1.35	−16.00	−1.44	21.16	−13.86*	0.24	0.92
	(1.00)	(1.34)	(18.42)	(2.63)	(13.14)	(7.65)	(0.91)	(0.96)
Constant	1.89***	22.06***	270.70***	92.00***	62.31***	45.72***	7.73***	8.77***
	(0.59)	(0.99)	(14.79)	(1.99)	(7.61)	(6.50)	(0.65)	(0.66)
Observations	72	72	68	62	68	68	62	71
R-squared	0.00	0.01	0.01	0.00	0.04	0.05	0.00	0.01

Note. Outcome variables are post-test scores. Robust standard errors are in parentheses. ***p<0.01; *p<0.1.

Table A2. Primary regression analysis with pair/triple fixed effects and no covariates.

	(1) Decoding	(2) Letter knowledge	(3) Response time variability	(4) Working memory index	(5) Errors of omission	(6) Errors of commission	(7) Verbal memory	(8) Phonological awareness
Treatment	0.31	1.17	−25.16	−2.00	19.74	−13.77*	0.23	0.88
	(0.72)	(1.06)	(16.51)	(2.05)	(12.34)	(6.92)	(0.89)	(0.67)
Constant	6.69***	27.83***	252.20***	98.00***	60.26***	22.77***	9.77***	13.12***
	(0.72)	(1.06)	(16.51)	(2.05)	(12.34)	(6.92)	(0.89)	(0.67)
Observations	72	72	68	62	68	68	62	71
R-squared	0.77	0.71	0.69	0.72	0.69	0.69	0.57	0.79
Pair/Triple fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	No	No	No	No	No	No	No	No

Note. Outcome variables are post-test scores. Robust standard errors are in parentheses. ***$p < 0.01$; *$p < 0.1$.

Table A3. Analysis of local average treatment effects.

	(1) Decoding	(2) Letter knowledge	(3) Response time variability	(4) Working memory index	(5) Errors of omission	(6) Errors of com-mission	(7) Verbal memory index	(8) Phonological awareness
Received treatment	0.69	2.00	−23.80	0.05	3.59	−16.60*	−0.16	1.35
	(1.01)	(1.42)	(16.69)	(3.26)	(14.43)	(8.30)	(0.88)	(0.97)
Pretest scores								
Letter knowledge	0.07	0.43	−5.37	0.92	−4.86	−1.16	−0.12	0.08
	(0.28)	(0.38)	(3.86)	(0.83)	(6.09)	(1.81)	(0.19)	(0.29)
Phonological awareness	−0.98*	−0.18	2.54	−1.02	0.94	3.16	0.22	−0.28
	(0.52)	(0.73)	(6.88)	(1.15)	(6.19)	(2.93)	(0.43)	(0.58)
Response time variability	−0.01	0.00	0.46*	−0.02	0.02	0.15	0.02*	−0.01
	(0.01)	(0.01)	(0.23)	(0.03)	(0.22)	(0.09)	(0.01)	(0.01)
Errors of omission	0.00	−0.01	0.11	−0.05	0.46*	−0.03	−0.04***	0.00
	(0.01)	(0.01)	(0.27)	(0.03)	(0.26)	(.09)	(0.01)	(0.01)
Errors of commission	0.04	−0.05	0.12	0.01	−0.03	0.30	0.07**	−0.02
	(0.04)	(0.04)	(0.64)	(0.09)	(0.53)	(0.23)	(0.03)	(0.02)
Observations	67	67	64	58	64	64	58	66
R-squared	0.24	0.22	0.48	0.27	0.40	0.50	0.54	0.17
First stage *F*-test	1391.70	1391.70	758.86	999.60	758.86	758.86	999.60	1248.88

Note. The table reports results from a two-stage least squares estimation where the randomly assigned treatment variable is the instrument for receiving treatment. Outcome variables are post-test scores. The first stage *F*-test row reports the results from the test of the excluded instrument in the first stage regression, i.e., if the treatment variable predicts receiving the program. The *F*-statistic differs across the outcome variables because the sample sizes differ but is always very large, which indicates a relevant instrument. Robust standard errors are in parentheses. ***p<.001; **$p < 0.05$; *$p < 0.1$.