ACLA WORKING PAPER SERIES

BACK TO SCHOOL: A META-ANALYSIS OF SCHOOL-BASED INTERVENTIONS TO REDUCE ABSENTEEISM

LAURENS DE CROES
NIENKE RUIJS
MELVIN VOOREN
STAN VERMEULEN
CHRIS VAN KLAVEREN
ILJA CORNELISZ

Working Paper 20251

http://www.acla.amsterdam/workingpapers-wp20251

AMSTERDAM CENTER FOR LEARNING ANALYTICS
van der Boechorststraat 7
1081 BT, Amsterdam
February 2025

Back to School: A Meta-Analysis of School-Based Interventions to Reduce Absenteeism

Laurens de Croes[1,2], Nienke Ruijs[1,3], Melvin Vooren[1], Stan Vermeulen[2,3], Chris van Klaveren[1] , and Ilja Cornelisz[1]

[1] *Amsterdam Center for Learning Analytics, Vrije Universiteit Amsterdam*

[2] *Dutch Inspectorate of Education*

[3] *Education Lab, Research Centre for Education and the Labour Market (ROA), School of Business and Economics, Maastricht University*

Author Note

Laurens de Croes https://orcid.org/0000-0002-7048-6259

Nienke Ruijs https://orcid.org/0000-0003-1944-5416

Melvin Vooren https://orcid.org/0000-0002-8043-7856

Stan Vermeulen https://orcid.org/0000-0002-0965-599X

Chris van Klaveren https://orcid.org/0000-0001-5975-4454

Ilja Cornelisz https://orcid.org/0000-0002-9948-7861

We have no conflicts of interest to disclose.

Correspondence concerning this article should be addressed to Laurens de Croes, Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081HV Amsterdam , The Netherlands. Email: l.x.de.croes@vu.nl

Abstract

School absenteeism is a pervasive issue affecting educational systems across the globe. Previous meta-analyses on interventions to reduce absenteeism have emphasized the need for caution in interpreting small positive or null effects due to significant heterogeneity and potential publication bias. Our study expands the empirical evidence-base by analyzing 65 studies (119 outcomes) from school-based interventions conducted across 16 countries between 1990 and 2024. Whereas a traditional random effects model yields significant overall beneficial effects (d = -.198) and intervention-specific effects (up to d = -.405), these findings become inconsistent or negligible when applying novel methodologies like weighted and iterative least squares. This sensitivity of the results illustrates the complexity of evaluating interventions across heterogeneous educational contexts. We advocate for acknowledging this heterogeneity by employing multiple meta-analytic approaches, each with distinct assumptions about effect-size distribution and publication bias, to enhance statistical power and accurately assess the effectiveness of interventions.

Keywords: reduce absenteeism, dropout prevention, school attendance, meta-analysis, schoolwide

Back to School: A Meta-Analysis of School-Based Interventions to Reduce Absenteeism

Globally, school absenteeism is associated with numerous long-term negative outcomes, including reduced well-being and impaired social-emotional development (Egger et al., 2003; Gubbels et al., 2019; Nik Jaafar et al., 2013) as well as lower academic achievement (Gershenson et al., 2017; Goodman, 2014; Gottfried, 2014; Kirksey, 2019; Lehr et al., 2004) and an increased risk of dropout (Balfanz & Byrnes, 2018; Gubbels et al., 2019). When students miss school, it therefore not only infringes on their legal right to education but also represents a public health concern.

For decades, educational policies worldwide have prioritized universal school attendance. Notable examples include the No Child Left Behind Act (NCLB) of 2001 and its successor, the Every Student Succeeds Act (ESSA) of 2015 in the United States, as well as the European Union's policy objective to reduce early school leaving to below 9%. Despite these sustained efforts, absenteeism rates remained relatively stable from 1990 (Kaufman et al., 2004) until the onset of the COVID-19 pandemic, which triggered a significant increase in absenteeism across various countries and educational levels (OECD, 2024).

Previous meta-analyses with a main focus on interventions aimed at reducing absenteeism (Eklund et al., 2020; Tanner-Smith & Wilson, 2013; Kremer et al., 2015; Maynard et al., 2013) vary considerably in scope, averaging 37 outcomes of which 24% are based on randomized controlled trials (RCTs). Collectively, they cover studies published between 1985 and 2018. Importantly, these meta-analyses primarily focus on the U.S. educational context and employ random effects (RE) models. Reported effect sizes range from non-significant results (g = .03, 95% confidence interval [CI] [-.02, .10]; Kremer et al., 2015) to beneficial effects (g = .46, 95% CI [.30, .62]; Maynard et al., 2013). However, these results are interpreted with great caution, due to small sample sizes (Kremer et al., 2015), the heterogeneous nature of the sample (Eklund et al., 2020) and potential sample or publication bias (Tanner-Smith & Wilson, 2013).

Building on the inconclusive evidence from previous meta-analyses, which relied on methodologies that were considered standard at the time, we include more recent evaluations and apply novel meta-analytic techniques with fewer restrictions on the assumptions imposed on the estimated unbiased effect. These methods differentially address the causes of potential heterogeneity and reveal the robustness of the estimates. Thus, our study has three main objectives: (a) to assess whether school-based interventions are effective in reducing absenteeism; (b) to evaluate whether this effect represents an unbiased effect size; and (c) to determine whether intervention, study, or contextual characteristics can explain the observed results. In doing so, this study aims to update and further enhance our understanding of effective interventions for reducing absenteeism in the inherently heterogeneous context of education.

The remainder of this paper is structured as follows: The next section outlines the theoretical background on school absenteeism, drawing on previous literature and meta-analyses. Building on this framework, we elaborate on our guiding principles. We then detail our methodology and analytical strategy, followed by the presentation of results. Finally, we conclude with a summary of the key findings and a discussion of their implications.

Measuring Absenteeism

Education is a fundamental right, as outlined in United Nation's Article 26 of the Universal Declaration of Human Rights, which mandates free and compulsory elementary education.[1]

[1] See: https://www.un.org/en/about-us/universal-declaration-of-human-rights

Many countries expand this mandate by extending the required age for education to meet minimum qualification standards. Absenteeism is then defined as the proportion of time a student is absent relative to the total amount of education offered, typically measured by the percentage of days absent (or attended) over a specific period. Consistent and accurate attendance records are thus crucial for reliable comparisons.

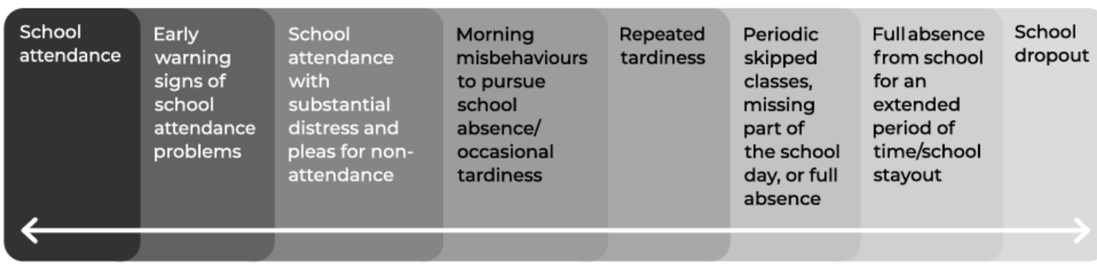
A key challenge in measuring absenteeism is the multitude of definitions and classifications. Absenteeism can be defined as 'excused' or 'unexcused' absenteeism, with the latter often regarded as more concerning (Gottfried, 2014). However, not all attendance records differentiate between these categories. Chronic absenteeism is typically defined as an absence rate of 10% or more within a given time frame (Gottfried, 2014; Heyne et al., 2020; Maynard et al., 2013; Simon et al., 2020). The classification of absenteeism varies widely, encompassing terms such as truancy, school refusal, withdrawal, or exclusion (Kreitz-Sandberg & Fredriksson, 2023), often depending on the underlying cause. For instance, school refusal is commonly associated with emotional distress (Heyne et al., 2020). Furthermore, multiple definitions often coexist within a single term. In a systematic review of 51 studies from eleven countries, Heyne et al. (2020) identified 28 distinct outcome constructs used to measure the effects of interventions targeting school refusal.

Despite variations in measurement, absenteeism consistently refers to missing school days, with a shared understanding that regular attendance is essential for student well-being (Egger et al., 2003; Gubbels et al., 2019; Nik Jaafar et al., 2013). As the number of missed school days increases, it may be classified as chronic absenteeism, and if this persists for several months, it can become a risk factor for school dropout. Dropout can be assessed through enrolment status, duration of absence, failure to graduate (Fernandes et al., 2024; Wang et al.,

2024) or leaving school before earning a high school credential (Kearney, 2008a). These concepts are interrelated not only in their operationalization but also in their progression, with absenteeism serving as a precursor for dropout. Figure 1 (adapted from Kearney et al., 2019) illustrates this spectrum approach, describing attendance issues along a continuum from full attendance to complete absence.

Figure 1

Spectrum of School Attendance and its Problems



Note. Adapted from *Reconciling Contemporary Approaches to School Attendance and School Absenteeism: Toward Promotion and Nimble Response, Global Policy Review and Implementation, and Future Adaptability (Part 1)* by Kearney et al., *Frontiers in Psychology, 10,* Article 2222. Copyright 2019 by the authors. Licensed under CC-BY 4.0. The figure has been adapted by converting it to grayscale, and simplifying the design.

The empirical literature shows that missing school leads to negative outcomes, such as impaired social-emotional development (Egger et al., 2003; Gubbels et al., 2019; Nik Jaafar et al., 2013) and reduced academic achievements (Gershenson et al., 2017; Goodman, 2014; Gottfried, 2009, 2015; Lehr et al., 2004; Nik Jaafar et al., 2013). Using a multivariate regression model, Hancock et al. (2013) estimate that each additional day of absence has a measurable negative impact on academic performance.

Second, attendance patterns are strong predictors of future attendance and are closely related to dropout (Kearney, 2002). Dropout, in turn, is associated with higher unemployment rates, poorer health outcomes, and an increased likelihood of incarceration (Belfield & Levin, 2007; Johansson, 2019; Wang et al., 2024). Due to the endogenous relationship, absenteeism and dropout share underlying risk factors discussed by Gubbels et al. (2019), which will be explored in the next paragraph. Absenteeism is thus a risk factor believed to negatively impact cognitive and socio-emotional outcomes and can serve as a precursor to future dropout, along with its associated negative long-term effects.

Underlying Risk Factors of Absenteeism

Gubbels et al. (2019) conducted a comprehensive meta-analysis of risk factors for school absenteeism and permanent dropout. Based on 75 studies conducted between 1978 and 2019 in Western countries, the researchers examined 781 potential risk factors for school absenteeism and 635 for dropout. These risk factors were classified into 44 risk domains, 28 of which showed significant effects on school absenteeism, encompassing family-related, school-related, and child-related factors. For dropout, 23 risk domains were found to have significant effects. Gubbels et al. (2019) concluded that the risk themes for school absenteeism and dropout are largely similar in nature.

Balkis et al. (2016) used structural equation modeling to identify predictors of absenteeism. Among other factors, they found previous absenteeism to be a significant predictor, emphasizing the cyclical nature of attendance issues. These findings align with Kearney’s (2008b) interdisciplinary model of school absenteeism, which posits that various factors influence problematic school attendance. These include student-, parent-, family-, peer-, school-, and community-level factors. Over time, these factors may lead to a progression from acute to chronic absenteeism, potentially resulting in permanent dropout (Gubbels et al., 2019).

Other reviews have focused on types of school absenteeism. For instance, Leduc et al. (2024) introduced an ecological framework for understanding school refusal through a systematic review of 15 studies, identifying 67 different factors. Their analysis highlighted the significance of psychological, social, and contextual factors in school refusal, thereby offering a nuanced perspective on the complex interplay among these factors.

Meta-analyses between 2010 and 2024 on Absenteeism

Tanner-Smith and Wilson (2013) synthesized 74 effect sizes from 50 quasi-experimental (QE) and 24 experimental (RCT) studies, conducted between 1985 and 2010. Using a random effects (RE) approach, the estimated effect size for studies with an RCT design suggests that dropout prevention programs significantly lower students' absenteeism (g = .23, 95% CI [.09, .38]). A separate bivariate moderator analysis revealed no significant relationship between intervention characteristics and absenteeism; however, programs were more effective among samples with a greater share of male participants (b = 1.22, p = .02, 95% CI [0.19, 2.25], β = .43). The mean effect size within program types was significant for two of the nine types of programs: supplemental academic training (g = .38; p= .03, 95% CI [.03, .73]) and vocational/employment-oriented programs (g = .84, p = .00, 95% CI [0.41, 1.28]).

In contrast, similar analysis on the QE studies yielded no overall significant mean effect but bivariate meta-regression highlighted differential beneficial impacts for younger participants (ages 11 – 13) compared to older participants (b = -.22, p = .009, 95% CI[-.35, -.06], β = -.40). Multi-service package programs (g = -.20, 95% CI [-.36, -.04) and mentoring/counseling programs (g = -.18, 95% CI [-.44, -.09]) had significant detrimental effects on reducing absenteeism.

The authors suggest that the significant results of the moderator analyses in the RCT and QE samples may be attributed to differences in the risk- and protective factors targeted by the various interventions (Tanner-Smith & Wilson, 2013). Based on the notably different mean effect sizes, inconsistent moderator effects, and the strongly suspected small study bias among the RCTs, the authors further concluded there is insufficient evidence to confirm the overall effectiveness of these interventions.

Maynard et al. (2013) included 16 published and unpublished studies conducted in the United States between 1990 and 2007, focusing on attendance as an outcome. Their RE model estimated a positive and significant effect size (g = .46, 95% CI [.30, .62]) indicating that truancy interventions effectively improve attendance among chronically truant students. Using bivariate meta-regression, they found no evidence of moderating effects. They noted that, despite the significant mean effect, the substantial heterogeneity of effect sizes and the small number of studies might have jeopardized their possibility of empirically detecting moderating effects.

Kremer et al. (2015) focused on after-school programs targeting at-risk primary and secondary school students. Their study included 16 effect sizes derived from 24 published and unpublished studies, six of which used an RCT design, conducted between 1980 and 2014 across five countries. Using a RE meta-analysis with robust variance estimation to account for dependencies among effect sizes, they found a non-statistically significant effect (g = .04, 95% CI [-.02, .10]).

Eklund et al. (2020) examined school-based interventions for elementary, middle, and high school students conducted between 2000 and 2018 in four countries. They used an RE analysis to account for the wide range of characteristics of the studies and lack of previous meta-analytic research and reported small but positive effects on reducing absenteeism (g = .26, 95% CI [.14,

.38]). However, given the small sample size and substantial heterogeneity between studies, they caution against overinterpreting the results and recommend adopting more rigorous methodologies in future research.

Despite differences in inclusion criteria, the aforementioned meta-analyses share five common features. First, these meta-analyses all classify interventions according to their underlying strategies, which include school-based interventions (Maynard et al., 2013), behavioral or academic approaches (Eklund et al., 2020), school/class restructuring, and cognitive-behavioral skills training (Tanner-Smith & Wilson, 2013). Second, they generally code population characteristics related to underlying absenteeism risks, student attributes, and study features. Third, while inclusion criteria for study design varied, all four meta-analyses incorporated quasi-experimental designs, which were defined as experimental studies lacking explicitly randomized control groups, rather than designs that exploit exogenous variation for causal estimation.

Fourth, they employ RE models to estimate the overall mean effect, which was considered the most accurate approach at the time of publication for addressing heterogeneity concerns. For moderator analyses, additional techniques, such as bivariate meta-analyses or robust variance estimation techniques are adopted (Kremer et al., 2015), although there is no consistent rationale for how these models control for publication bias. Publication bias is primarily addressed through visual inspection of funnel plots, supplemented by empirical tests and the inclusion of unpublished research. These methodological choices align with the Campbell Collaboration guidelines relevant to that time.[2]

Fifth, due to substantial heterogeneity and potential biases from publication or small sample sizes, these meta-analyses do not claim conclusive overall evidence for specific interventions. Even when significant effects are estimated, they are interpreted with caution, as potential publication bias, small sample bias, and substantial heterogeneity may have contributed to these findings.

Meta-analyses between 2010 and 2024 on Dropout

Wilson et al. (2011) examined general dropout prevention programs, analyzing 152 studies with 385 effect sizes and 317 independent samples (*n*). Using an RE model, they identified a significant impact on dropout rates, with a mean odds ratio of 1.72 (95% CI [1.56, 1.90]. Of the 12 program types, 11 (n = 144) showed positive effects, except for the type “other” (n = 8). Programs delivered in a classroom (n = 175) or mixed (n = 36) setting demonstrated significantly larger effects compared to a community setting (n = 34) (Wilson et al., 2011). However, due to overlapping settings and program types, the precise mechanisms behind these effects remains unclear. Participant characteristics had minimal influence on effectiveness, but less frequent programs were more effective (p <.10).

Performing a Peters’ regression test for publication bias indicates that small-study effects may influence these results (b = 16.80, p = .001). The reasons for these effects are difficult to pinpoint; they may involve selective outcome reporting, spurious inflation of effects in smaller

[2] See: https://www.campbellcollaboration.org/methods/standards/

studies, and sampling variation artifacts. Despite acknowledging the potential overestimation of the mean effect size due to small-study effects, Wilson et al. (2011) provide explorative evidence indicating that small studies do not substantially impact the mean effect sizes, as shown by the absence of trimmed/filled effect sizes (Duval & Tweedie, 2000). However, this conclusion is tentative given the significant heterogeneity observed (Q_E = 3065, *df* = 316–384, *p* <.05) which might violate the symmetric funnel plot assumption and affect the reliability of the trim-and-fill method.

Chappell et al. (2015) include 67 effect sizes from 35 high school programs in the United States, using experimental or quasi-experimental designs. Their RE model found a small overall effect on dropout (*d* = .15, 95% CI [.08, .21]). In their meta-regression, they estimate the effect size for each strategy, holding all other covariates constant. Eight of the ten strategies had significant effects, ranging from *d* = .11 (*p* = .00[3]) for academic support to *d* = .46 (*p* = .01) for behavioral intervention. Publication bias was addressed by using the *fail-safe N* method. This method calculates how many null-effect studies would need to be added to the analysis to bring the overall type I error for the overall mean effect above the typical threshold of 0.05. The results show that 1,677 studies would be needed. This method was only applied for the overall mean effect; publication bias was not accounted for in the subgroup analysis or meta-regression.

Wang et al. (2024) analyzed dropout interventions for students aged 4-17 in North America. Their fixed-effects model, including 16 randomized and 15 matched designs, revealed a marginally positive effect of dropout intervention programs (*g* = .19, 95% CI [.10, .28]). Sub-

---

[3] The confidence intervals for these adjusted estimates are not reported, and because the meta-regression model accounts for multiple covariates simultaneously, we cannot accurately reconstruct them.

group analysis and meta-regression indicated that behavioral (*k* = *6*, *b* = 0.20, 95% CI [-.16, .57]) and combined interventions (*k* = 6, *b* = 0.19, 95% CI [-.07, .46]) produced insignificantly larger effect sizes compared to the reference group, academic interventions (*k* = 9, *g* = .12, 95% CI [-.05, .30]). However, when accounting for sampling methods and covariates in a meta-regression model with dummies for each intervention strategy, interventions that included a behavioral strategy were found to yield significantly larger effect sizes (*b* = 0.26, *p* < .05, 95% CI [.01, .51]) than those without this strategy. Publication bias was addressed separately by visually inspecting the funnel plots and using an extended version of Egger's regression test. The latter – after controlling for moderators, suggested minimal evidence of publication bias (*t* = 1.23, *p* = .23).

The meta-analyses with a dropout focus include more studies and more effect sizes, since studies generally consider dropout and graduation rates simultaneously. Estimating the effect size using RE and meta-regression models yield promising results. Typically, publication bias is addressed as a separate concern in interpreting the estimate instead of directly incorporating it into the meta-analytic model for obtaining an unbiased estimate.

Synthesis of Characteristics of Effective Interventions

The current body of research provides insufficient evidence to conclusively identify specific characteristics of interventions or student populations that consistently contribute to the effectiveness of absenteeism reduction strategies. However, for dropout prevention interventions, preliminary and inconsistent evidence suggests a potential benefit from behavioral interventions (Chappell et al., 2015; Wang et al., 2024). Kearney et al. (2019) propose a multi-tiered approach for effective interventions, where the required intervention intensity increases along the continuum from full attendance to complete absence (see Figure 1).

Meta-analyses typically characterize intervention intensity through three primary dimensions: duration, underlying strategy or program type, and the selection criteria for the target population. This framework provides a structured approach to understanding and evaluating the components of interventions, although more research is needed to establish definitive links between these characteristics and intervention effectiveness.

This Study

Building on the inconclusive evidence from previous meta-analyses, which relied on methodologies that were considered standard at the time, we adopt a dual approach to enhance both the comprehensiveness and comparability of interventions across diverse educational contexts. We broaden our selection criteria to include a wider range of countries and extend the publication years covered. By focusing on potentially universal characteristics of school-based interventions and target populations, we aim to improve inference with respect to which educational contexts and students the observed results apply.

While expanding our scope increases heterogeneity, we complement common meta-analytic methods with innovative techniques that offer greater flexibility in assumptions about underlying mechanisms. Concurrently, we apply strict criteria for outcome measurement to ensure meaningful comparisons. Additionally, unlike previous meta-analyses, which broadly defined quasi-experimental studies as those lacking randomization, we focus exclusively on designs capable of credible causal inference under their identifying assumptions.

Methods

Principles Based on Previous Research

The findings from previous meta-analyses on effective interventions and risk factors related to absenteeism provide a solid foundation for formulating three principles to facilitate cross-country and temporal comparisons in our selection criteria, definitions, and coding process: school-based interventions, absence rate outcome measures, mechanisms and aims of intervention.

First, we focus on interventions occurring within school settings or under the school's primary responsibility. Interventions delivered outside of schools, such as community-based therapies, may depend on community characteristics and could undermine our assumption of universality. We acknowledge that school-based interventions might also depend on school characteristics and reflect broader socioeconomic and structural factors at the national level, potentially challenging our universality assumption. By including only school-based interventions, we assume a form of relative universality. This universality is contingent upon each country's specific implementation of the fundamental human right to education, making them contextually adaptable universal school-based interventions.

Second, in defining key concepts, our preferred outcome is the absence rate. By using proportion, we mitigate differences between countries in the total amount of education they offer. Our second preferred outcome is chronic absenteeism, which is less precise and typically assumes a threshold. We avoid predefined thresholds for chronic absenteeism, acknowledging that differences in these thresholds relate to specific country characteristics, yet they express a common desired reduction. Dropout is our least preferred outcome, as we view it as a less precise threshold at the extreme end of the absenteeism continuum (cf. Figure 1).

Third, to facilitate cross-country and temporal comparisons, we differentiate between the underlying mechanisms of interventions and their specific aims in addressing risks associated with the target population. We exclude interventions that focus solely on medical reasons for absenteeism, such as asthma, or those primarily addressing issues typically managed by other institutions, like drug addiction. Including such interventions would rely heavily on country-specific institutional characteristics, further complicating comparisons and contradicting our first principle: contextually adaptable universal school-based interventions.

Eligibility criteria and screening process

We developed a protocol based on our three guiding principles to screen titles and abstracts for potential relevance. Eligible studies had to meet specific criteria: they must be school-based interventions, targeting students aged 4-20 in a mainstream school setting without special admission criteria (i.e. regular primary and secondary school). This is measured at the student level using official attendance records.

Outcome measures must directly express the absenteeism (cf. Figure 1) rather than proxy indicators, such as graduation rates. Proxies for absenteeism are excluded because they could emphasize the role of other student, school, or country-specific characteristics, thereby jeopardizing the generality of the results.

Given our assumption of the universal nature of absenteeism and to account for differences across educational contexts and time periods, we impose strict research design criteria. To ensure the scientific quality of the studies included, we limit our selection to peer-reviewed studies employing experimental designs with randomized control and treatment groups (RCTs) or quasi-experimental designs utilizing differences-in-differences (DiD) or regression-discontinuity (RD) approaches for causal estimation. Conditional on that they meet underlying (identifying) assumptions, these methodologies are generally considered credible causal designs (Kraft & Novicoff, 2024). Finally, we only include studies published in English due to accessibility and interpretation constraints.

Search procedure and terms

We conducted an electronic search of abstracts using APA PsycInfo, APA PsycArticles, ERIC, and Psychology and Behavioral Sciences Collection databases for studies published between January 1990 and July 2024. This timeframe was selected because 1990 marks a significant global shift in educational policy and practice. This shift was highlighted by the launch of the Education for All (EFA) initiative in Jomtien, Thailand, which aimed to expand access to basic education and improve educational quality across countries. Correspondingly, this period can be characterized by an increased emphasis on educational outcomes over inputs, with many countries adopting standards, assessments, and accountability systems to enhance student performance and school accountability (Olssen et al., 2024). The search terms included keywords for the intervention (e.g., "program"), the context (e.g., "school"), the outcome (e.g., "absent*"), and design (e.g., "effect"). Within each category, terms were separated by OR, while the categories themselves were separated by AND (see full details of the search strategy and terms in the supplementary materials, available in the online version of this article).

As the flowchart in Figure S1 (online only) shows, after double-screening 5,597 titles and abstracts, we assessed 389 full-text articles for eligibility. Sixty-five studies met our inclusion criteria and were included. To ensure coding reliability, a random sample of fifty percent of these studies was double-coded, with the process continuing until an intercoder agreement level of 0.90 was obtained. The subsequent subsection provides a detailed account of the coding process and the extraction of 119 effect sizes.

Data appraisal Coding Characteristics Design, Intervention and Population

Intervention Characteristics

Based on previous meta-analyses, we developed a comprehensive typology that included nine program types for coding the interventions. Based on their underlying strategies, we grouped these nine program types into four primary intervention categories: (1) aid programs, which address socioeconomic disparities by providing additional resources; (2) support networks, which offer counseling and academic or personal development support; (3) school curriculum and learning environment, which involve restructuring the curriculum or enhancing the learning environment; and (4) package programs, which combine elements of the strategies 1-3. Table 1 provides an overview of these categories, their prevalence within our sample, and the associated strategies.

Table 1

Intervention Types, Descriptions and Frequencies

Intervention type (and tier)	**Program type**	**Description**	**Study** *(k)*
Aid programs (tier 1)	*Strategies that mitigate social-economic background by (conditional) supplying extra sources and monitoring attendance*		*10*
	(conditional) cash	A (conditional) financial incentive, e.g. free books or a fee	6
	breakfast	Providing free breakfast	4
Support networks (tier 2 or 3)	*Strategies that provide (network) support or counselling for academic and personal (development) issues, for example improving self-esteem*		*24*
	Coach	An internal or external (trained) counsellor with a focus on academic and personal goals	5
	Mentor	An internal or external (trained) mentor with a focus on academic and personal goals	8
	Peer	A peer to provide general support in academic and personal goals	1
	Family	Organising family involvement and support to achieve academic and personal goals	10
School Curriculum and	*Strategies that restructure the school curriculum or support, based on the idea that students school behavior is influenced by the school structure and behavior of school staff*		*20*

Learning Environments strategies (tier 1)			
	School wide program	School-wide support strategies and specific interventions for students at risk for example smaller classes	10
	Teaching time	Changes in the offered hours of education or starting time of the school and changing in the curriculum, for example social training	10
[1, 2]	*Package*	*Programs that implement two or more strategies, often a combination of for example monitoring student behavior and organizing family involvement*	*11*

Note. Type of tier is based on the multidimensional multi-tiered system of support by Kearney and Childs (2023).

Using the Multidimensional Multi-Tiered System of Support (Kearney & Childs, 2023), we categorize the underlying strategies into three tiers: Tier 1 refers to universal prevention practices aimed at promoting school attendance and improving school climate; Tier 2 includes selective or secondary interventions addressing emerging attendance problems; and Tier 3 encompasses intensive interventions or tertiary prevention practices targeting severe attendance issues.

To further assess the intensity of the intervention and the severity of absenteeism (Kearney et al., 2019) we recorded temporal aspects such as the initiation year, conclusion year, duration in months, and the time elapsed between intervention commencement and outcome measurement. We created binary indicators, indicating the levels at which the intervention is aiming to reduce absenteeism: student, teacher, parent, or classroom. Additionally, we categorized continuous variables, such as duration, into broader categories (e.g., interventions lasting more or less than one year). A detailed overview of all extracted variables, their possible values, and the categorization can be found in the supplementary material (Table S1, S2, S3, S4 Figure S1 and S2).

Participant characteristics

We coded participant characteristics; for schools, there is a dummy indicating whether the target school was selected based on risks. For the target population, based on Gubbels et al. (2019) we define six risk factors: family context, school history, behavioral problems, emotional problems, cognitive problems, and mild handicap. This stratification allows controlling for these variables in subsequent analyses.

Design characteristics

For each study, we coded the design type: RCT or QE (with subcategories DiD or RD), country of origin, publication date, level of randomization, total sample size, and the sizes of control and treatment groups (see supplementary material for a detailed overview of possible values). For QE studies, we additionally documented the type of exogenous variation employed for generating internally valid effect estimates.

Effect Size Extraction

We employed Cohen's *d*, a standardized mean difference, as our effect size measure. This measure is calculated by dividing the difference in sample means between the treatment and control groups by the pooled standard deviation. However, these statistics are not always reported, particularly in QE studies. To ensure comparability and account for potential structural biases inherent in these study designs, we adopted a structured approach to effect size extraction. For RCTs, we prioritized using sample means and standard deviation for effect size calculation, applying appropriate transformations when necessary (Sánchez-Meca et al., 2003). Table 2 illustrates the distribution of outcome measures and the methods of effect size extraction.

Table *2*

Distribution and Extraction of Outcome Variables by Order of Importance

Outcome	**All**	**1ps**	**Unique**	means and SD		*t*-value		Reported Cohen's *d*		Trans. ES		*p* label	
				R	Q	R	Q*	R*	Q	R	Q	R	Q
1. Rate: days (un)attended/days offered*	99	49	99	61	-	5	21	7	-	-	3	-	2
Absenteeism rate	*58*	*33*	*58*	*43*	-	*5*	*5*	*2*	-	-	*3*	-	-
Attendance rate	*41*	*16*	*41*	*18*	-	-	*16*	*5*	-	-		-	*2*
2. Did (not) meet threshold	5	4	4	3	-	-	-	-	-	1		-	-
3. Did (not) drop out	19	12	16	5	-	-	9	-	-	2		-	-
Total	123	65	119	69	-	5	30	7	-	3	3	-	2

Note. SD = standard deviation, R = RCT and Q = QE. All = total outcomes, 1ps = 1, outcome per study based on priority, - = zero, Trans. ES = effect sizes transformed to Cohen's *d* when originally reported differently. Unique = selected unique outcomes in terms of sample and time. The right panel shows the ES calculation method in order of preference for our selected outcomes.

QE designs presented unique challenges due to varying model specifications, covariate adjustments, and limited reported data. We primarily extracted *t*-values from reported standard errors (n = 23), confidence intervals (n = 6), or extracted *p*-values (n = 3) to calculate standardized mean differences. We converted three effect sizes, and for two outcomes lacking *t*-statistics, we approximated it using the reported significance levels, employing the most conservative approach, potentially resulting in an underestimate of the effect size.

To address potential biases inherent in different study designs and reporting practices, we introduced a variable identifying the risk of bias for all outcomes not directly extracted using our prioritized method within each design type (n = 7). We ranked outcome types by precision, with rates based on days considered the most precise and threshold or dropout measures ranked as less precise. All effect sizes were calculated so that a negative value indicates a reduction in absenteeism—our desired outcome. This direction was the most commonly observed, as shown in Table 2.

For studies reporting multiple outcomes (e.g., at different time points), we coded each outcome accordingly and accounted for within-study correlation by clustering multiple outcomes at the study level. This structured approach to effect size extraction allows for systematic comparison across studies. When multiple outcomes are based on the same time point and sample, we select the priority outcome as defined in Table 2.

Estimating mean effect sizes with a random effects model

Given the heterogeneity across studies—including differences in intervention types, years, and countries—we posited the existence of multiple true effect sizes rather than a single one. To address this heterogeneity and answer whether school-based interventions effectively reduce absenteeism (research question a), we first applied a traditional and commonly-used RE analysis. This model pools effect sizes using inverse-variance weighting, where weights are calculated as $(\frac{1}{(v_i + \tau^2)})$, with v_i representing within-study sampling error, and τ^2 representing between-study heterogeneity. The RE model assumes that true effect sizes are normally distributed around an overall mean effect size and independent of sampling errors (Stanley et al., 2022). This implies exchangeability: true effects are randomly assigned to studies along with heterogeneity and independently of sampling errors. However, two key challenges arise in educational contexts: precision bias and systematic heterogeneity. As for precision bias, larger studies with more precise estimates may systematically differ from smaller studies in ways that affect effect sizes. If this occurs, inverse-variance weights may introduce bias rather than circumvent it. As for systematic heterogeneity, if this exists it cannot be accounted for by the RE model as it assumes that unobserved heterogeneity is independent from study-level covariates. Yet, particularly in educational research, contextual factors – such as intervention type or study design – may violate this assumption. To address these limitations and test whether RE estimates are unbiased in educational contexts (research question b), we examined potential precision bias in the next subsection.

Heterogeneity and sample size: addressing precision bias

To assess the potential bias of the RE model, we applied three models within a generalized least squares regression framework: Unrestricted Weighted Least Squares (UWLS), Precision Effect Test (PET), and Precision Effect Estimate with Standard Error (PEESE). All three models represent the effect size of study *i* as the Cohen's *d* effect size divided by its standard error (SE) ($t_{i=} \frac{ES_i}{SE_i}$). The models differ in their assumptions about how to estimate the treatment effect while controlling for standard error differences, as illustrated by their respective equations:

(1) $$t_i = \alpha_0 \cdot \left(\frac{1}{SE_i}\right) + \mathrm{v}_i$$ UWLS

(2) $$t_i = \beta_0 + \alpha_0 \cdot \left(\frac{1}{SE_i}\right) + \mathrm{v}_i$$ PET

(3) $$t_i = \beta_0 + \alpha_0 \cdot SE^2 + \mathrm{v}_i$$ PEESE

The UWLS equation, introduced by Stanley and Doucouliagos (2017), regresses the standardized effect size of study *i* (t_i) on precision (*1/*SE_i) without inclusion of an intercept. The model thus assumes that the observed variation in effect sizes can be explained solely by differences in precision.

The PET equation extends UWLS by including a model constant, such that the estimation model becomes more flexible. The term $\alpha_0 \cdot \left(\frac{1}{SE_i}\right)$ captures the systematic bias related to study precision and β_0 provides the treatment effect estimate while controlling for precision differences. Finally, the PEESE equation modifies PET by replacing $\left(\frac{1}{SE_i}\right)$ with $SE^2{}_i$ as a predictor. This quadratic specification allows for a more flexible quadratic functional form between precision and effect.

These models are also considered when addressing publication bias because they control for the systematic relationship between sample size and observed effect size. Yet, sampling variability bias may also arise from other sources of heterogeneity, such as differences in interventions or educational contexts that systematically influence sample size and, consequently, precision.

Heterogeneity and sensitivity: the role of sample characteristics

By applying the RE model alongside UWLS, PET, and PEESE, we derive estimates that correct for heterogeneity caused by sampling variability bias. However, another common approach to addressing heterogeneity bias is to assume that it arises from the composition of the dataset itself. To test this hypothesis, we restricted the sample based on predefined structural differences in study characteristics that might bias the estimates. Alternatively, we implemented a bootstrap aggregation approach to address potential noise without relying on predefined assumptions about which sample differences contribute to heterogeneity.

1. Sample restricted to RCT: Given the substantial differences in sample sizes between randomized controlled trials (RCTs) and quasi-experimental (QE) studies, we test if the estimation results differ when only including studies that use RCT design; a similar

distinction was made by Tanner-Smith and Wilson (2013). Our sample of QE studies is too small to detect meaningful differences.

2. Sample excluding outliers: We define outliers using a method proposed by Viechtbauer and Cheung (2010), which considers a study an outlier if the confidence intervals falls outside the confidence interval of the overall effect size. However, removing outliers can artificially push estimates toward significance due to the reliance of methods like PET, PEESE, and meta-regression on precision and effect size. As Hunter and Schmidt (2004) caution, it is almost impossible to distinguish between large sampling errors and true outliers, and removing studies with extreme effects should be treated as part of sensitivity analyses rather than routine practice.
3. Bootstrap Aggregation: This ensemble learning method is commonly used to reduce variance in noisy datasets. By resampling subsets of data and averaging the resulting estimates, this approach minimizes sensitivity to specific data structures. Unlike outlier exclusion, bootstrapping provides a more robust way to address potential noise without introducing bias toward significance. It allows us to account for potential heterogeneity stemming from varying sources, such as differences in study design or outlier influence. The bootstrapped mean provides a more robust estimate of the overall effect size while offering insight into how heterogeneous samples affect pooled results. We specified tolerance levels for convergence at 0.005 for both mean and standard deviation, ensuring stability over a minimum of 2,000 iterations with at least 20 stable iterations.

These methods collectively provide a deeper understanding of the credibility of the overall mean effect size and potential sources of heterogeneity bias. However, it is important to note that these approaches do not explicitly address publication bias.

Heterogeneity: Publication bias

Publication selection bias remains a significant concern in meta-analyses, especially when included studies are restricted to only peer-reviewed publications. To address this issue and enable comparison, we start with the classic funnel plot visualization (Eklund et al., 2020; Maynard et al., 2013; Tanner-Smith & Wilson, 2013). Following Stanley and Doucouliagos (2010), we construct an inverted funnel plot, placing the standard error (SE) on the vertical axis and flipping the usual axes. This allows us to view and interpret the graph as a regression relationship. Grey contour shading centered around zero highlights areas where null effects would not reject the null hypothesis.

While visual inspection provides initial insights, it is typically complemented by more objective assessments. These methods have become standard methods in the meta-analytic literature, including Egger's regression (Tanner-Smith & Wilson, 2013), the addition of a *fail-safe N* (Eklund et al., 2020), and exploratory trim and fill methods (Tanner-Smith & Wilson, 2013). Intuitively, these methods aim to detect patterns suggesting bias, such as asymmetry in effect sizes (Egger's regression), calculate how many unpublished studies with null effects would nullify the observed results (*fail-safe N*), or estimate and adjust for potentially missing studies to provide a more balanced effect size (trim-and-fill).

Currently, a PET-PEESE approach is considered one of the most reliable methods for capturing publication bias, but has limited power when effect sizes are small or when heterogeneity is high, with a recommended cutoff at 80% I^2 for reliable use (Stanley, 2017). To address these limitations and tackle a form of publication bias less dependent on precision, we applied the Weighted Iterative Least Squares (WILS) method, introduced by Stanley and Doucouliagos (2024). WILS is a novel technique that iteratively removes studies with inflated effect sizes resulting from publication selection bias. This specific bias, known as Excess Statistical Significance (ESS), occurs when observed effects in meta-analysis appear more significant than they truly are due to selective reporting. WILS targets studies contributing most to ESS by using statistical power calculations based on the known distribution of effect sizes in meta-analyses.

The WILS process starts by calculating the effect size using UWLS along with other key statistics from the sample distribution, such as the total variance, which is the sum of the within-study variance (SE^2) and between-study heterogeneity (τ^2), accounting for the estimated heterogeneity variance. These statistics are used to estimate whether study *i* would publish a significant result under random significance assumptions. The cumulative standard normal distribution, (Φ), with a 1.96 cutoff for a 0.05 significance level, is used to determine this probability. Equation 4 summarizes this process, where $E_{\mathrm{sig},j}$ represents the sum of retrospective power Z for *k* studies, calculated for each study (*i*) in each iteration (*j*).

$$E_{\mathrm{sig},j} = \sum_{\{i=1\}}^{k_j}\left[1 - \Phi\left(Z_{i,j}\right)\right] \text{ where } Z_{i,j} = \frac{\left(1.96SE_i - \left|UWLS_j\right|\right)}{\sqrt{\left\{SE_i^2 + \tau_j^2\right\}}} \qquad (4)$$

ESS is computed by subtracting the total expected significance in iteration j ($\mathrm{E}_{\mathrm{sig},j}$) from the total observed number of statistically significant results. In each iteration *j*, studies contributing most to ESS are identified and removed. In the first iteration, studies are selected for removal based on their weighted residuals, which reflect deviations from the initial UWLS estimates. From the second iteration onward, ESS is recalculated based on the newly computed UWLS for the remaining studies, and those contributing the most to it are removed based on these updated calculations. The number of studies removed in each iteration corresponds to the

ESS value, rounded to the nearest integer. This iterative process continues until ESS approaches zero or until the remaining number of studies becomes too small (k = 2) for reliable analysis.

We adapt the procedure outlined by Stanley and Doucouliagos (2024) to suit the characteristics of our sample. Specifically, we use the absolute values of residuals to select studies for removal in the first iteration, while maintaining the conservative calculation of $Z_{i,j}$ based on the absolute value of UWLS ($|UWLS_j|$) in subsequent iterations. By focusing on the magnitude of deviations rather than their direction, we aim to identify studies with the most inflated effect sizes—whether positive or negative—without being influenced by the sign of the residuals. This method helps to address overestimated effect sizes, refining the estimate and reducing bias. In the next round, when calculating ESS, we address both overestimated and underestimated positive and negative effects, further refining the estimate by mitigating the influence of inflated or suppressed effect sizes. Using $Z_{i,j}$ ensures that the influence of negative ESS values, associated with insignificant results, is mitigated. This prioritizes significant positive effects and reduces the influence of insignificant negative ones, leading to a more accurate and robust final effect size estimate. As a robustness check, we also present results in the supplemental material (Table S5), following the exact procedure outlined by Stanley and Doucouliagos (2024). Additionally, we include an alternative version where all negative ESS values are set to zero prior to performing the remaining calculations, providing a sensitivity analysis of the model's assumptions.

Meta-regression: Intervention Type

Meta-regression, which is a combination of meta-analytic principles with the regression principles of predicting study effects using study-level covariates (Sutton & Higgins, 2008). To explore if the type of intervention contributes to the effect size estimate, we incorporate dummy variables for the four intervention types in a meta-regression approach. We run the RE and PET-PEESE models without an intercept, allowing coefficients to represent estimated mean effect sizes for each intervention type and apply a pairwise comparison using precision-weighted averaging for the WILS approach. Finally, to isolate the effect of each intervention type, we conduct separate regressions using individual dummy variables. We present comparative outcomes of the PET or PEESE, RE, and WILS models, both including and excluding outliers, to provide a robust analysis of intervention effectiveness.

To facilitate comparisons both within this study and across meta analyses, we report the confidence intervals for all estimates. For the WILS estimates of the overall mean, we additionally provide the standard error to enable a more intuitive comparison of precision across various sensitivity checks. Statistical significance is indicated with asterisks in the tables. To assess heterogeneity with RE in models based on the generalized least squares regression framework we calculated the I^2 statistic (Higgins et al., 2024), which quantifies the degree of between-study heterogeneity. Specifically, this statistic represents the proportion of variation across studies attributable to heterogeneity rather than random chance.

Results

Intervention Characteristics

In Table 3, we provide an overview of the key characteristics of the interventions from the 65 included studies, encompassing 119 outcomes. The temporal distribution shows that 13.8% of interventions started in the 1990s, which increased to 32.3% of interventions from the 2000s

(32.3%) and even 49.2% from the 2010s; a finding consistent across all age groups.[4] This indicates a potential bias towards more recent studies. Geographically, the data is skewed toward North American interventions, comprising 75.4% of the total. The proportions of intervention types are relatively balanced: aid programs (15.4%), package interventions (16.9%), school curriculum and learning environments (30.8%), and support networks (36.9%). The distribution of interventions varies across age groups, with aid programs being more prevalent (44.4%) when targeting all ages (4–20).

The sample is about evenly divided between interventions targeting schools with no specific risk (50.8%) and those targeting at-risk schools (49.2%). However, in the ages 4–20, interventions mostly target at-risk schools (77.8%). When the target population is selected based on risk, it is relatively often based on school attendance history (33.8%), followed by family socio-economic background (27.7%). Duration-wise, most interventions (63.1%) last for up to one year, with patterns consistent across all age groups. Having examined the intervention characteristics, we now turn to the design features of the studies in our sample.

Table 3

Intervention Characteristic by Age Group per Study

Characteristics	age 4 - 20 (*n* =9)	Age 4 - 12 (*n* = 23)	Age 12 - 20 (*n* = 32)	Overall (*n* = 65)
Decennium started				
1990s	0 (0%)	2 (8.7%)	7 (21.2%)	9 (13.8%)
2000s	3 (33.3%)	9 (39.1%)	9 (27.3%)	21 (32.3%)
2010s	5 (55.6%)	12 (52.2%)	15 (45.5%)	32 (49.2%)
2020s	1 (11.1%)	0 (0%)	2 (6.1%)	3 (4.6%)
Region				
North America	3 (33.3%)	19 (82.6%)	27 (81.8%)	49 (75.4%)
Western Europe	1 (11.1%)	4 (17.4%)	3 (9.1%)	8 (12.3%)

[4] The proportion of studies from 2020s is currently 4.6% and will likely grow in the coming years as new studies keep being published, with the final tally yet unknown.

Characteristics	age 4 - 20 (*n* =9)	Age 4 - 12 (*n* = 23)	Age 12 - 20 (*n* = 32)	Overall (*n* = 65)
Oceania	2 (22.2%)	0 (0%)	1 (3.0%)	3 (4.6%)
Other	3 (33.3%)	0 (0%)	2 (6.1%)	5 (7.7%)
Intervention type				
Aid programs	4 (44.4%)	1 (4.3%)	5 (15.2%)	10 (15.4%)
Package	2 (22.2%)	3 (13.0%)	6 (18.2%)	11 (16.9%)
School Curriculum and Learning Environments	1 (11.1%)	9 (39.1%)	10 (30.3%)	20 (30.8%)
Support networks	2 (22.2%)	10 (43.5%)	12 (36.4%)	24 (36.9%)
Target schools				
No risk	2 (22.2%)	13 (56.5%)	18 (54.5%)	33 (50.8%)
Risk	7 (77.8%)	10 (43.5%)	15 (45.5%)	32 (49.2%)
Target population				
Emotional or behavioral problems	1 (11.1%)	4 (17.4%)	6 (18.2%)	11 (16.9%)
Family socio-economic background	3 (33.3%)	6 (26.1%)	9 (27.3%)	18 (27.7%)
School attendance history	3 (33.3%)	9 (39.1%)	10 (30.3%)	22 (33.8%)
No risk	2 (22.2%)	4 (17.4%)	8 (24.2%)	14 (21.5%)
Duration cat				
Up to 1 year	7 (77.8%)	14 (60.9%)	20 (60.6%)	41 (63.1%)
More than 1 year	2 (22.2%)	9 (39.1%)	13 (39.4%)	24 (36.9%)

Note. North America: USA, Hawaii, Canada, Western Europe: Wales Portugal, Norway, Netherlands, Denmark, Belgium.

Oceania = New Zeeland, Australia, Other = Brazil, Pakistan Sub-Sahara Afrika, Nigeria, Kenya

Design characteristics

Table 4 summarizes the design characteristics, comparing quasi-experimental (QE) designs with randomized controlled trials (RCTs). As anticipated, QE designs exhibit substantially larger average sample sizes (mean: 49,700) compared to RCTs (mean: 2,270). Overall, the total sample comprises 1,931,169 observations derived from 119 outcomes. Randomization at the student level is the most common approach (61.3%), followed by randomization at the school level (34.5%). Randomization at the classroom-level (1.7%) and county- or household-level randomization (2.5%) are relatively rare.

Regarding intervention levels, multi-level approaches are prominent. A significant proportion of studies (32.8%) target three or four levels simultaneously (e.g., student, teacher, parents, and/or classroom), while 31.1% focus on two levels. Single-level interventions targeting students alone account for 26.9% of the studies, with other single-level interventions making up 9.2% of the sample.

Table 4

Design Characteristics per Outcome

Design characteristics	QE (N=35)	RCT (N=84)	Overall (N=119)
Level of randomization			
student	19 (54.3%)	54 (64.3%)	73 (61.3%)
classroom	0 (0%)	2 (2.4%)	2 (1.7%)
school	14 (40.0%)	27 (32.1%)	41 (34.5%)
county or household	2 (5.7%)	1 (1.2%)	3 (2.5%)
Units of randomization			
Mean (SD)	21800 (57200)	1680 (9020)	7580 (33000)
Median [Min, Max]	6770 [258, 335000]	102 [2.00, 82200]	249[2.00, 335000]
Students			
Mean (SD)	49700 (85400)	2270 (9050)	16200 (51300)
Median [Min, Max]	8560[258, 416000]	431[14.0, 82200]	981[14.0, 416000]
Total students	1740634	190535	1931169
Level			
student	6 (17.1%)	26 (31.0%)	32 (26.9%)
other single-levels	5 (14.3%)	6 (7.1%)	11 (9.2%)
two-levels	10 (28.6%)	27 (32.1%)	37 (31.1%)
three or four-levels	14 (40.0%)	25 (29.8%)	39 (32.8%)

Note. Level refers to the following (combination) of levels: student, teacher, parents, classroom

Mean effect and sample size bias

To address the research questions—(a) whether school-based interventions are effective in reducing absenteeism and (b) whether this estimate represents an unbiased mean effect size—we first estimated the overall mean effect size using the traditional RE model. Subsequently, we compared this estimate with results from three models with differing assumptions about precision and the estimated mean: Unrestricted Weighted Least Squares (UWLS), Precision Effect Test (PET), and Precision Effect Estimate with Standard Error (PEESE). The results for each model are shown in Table 5. It shows the Cohen's *d* effect size, the 95% confidence intervals, the I^2 heterogeneity metrics, and bootstrapped effect sizes with confidence intervals.

Table 5

Mean Effect Estimates using RE, UWLS, PET and PEESE

Model	ES	95% CI	I^2	BS ES	BS 95% CI
RE	-0.198**	[-0.306, -0.089]	88	-0.181**	[-0.248, -0.088]
UWLS	-0.006	[-0.013, 0.001]	88	-0.006**	[-0.013, -0.002]
PET	0.007	[-0.001, 0.014]	99	0.007	[0.001, 0.016]
PEESE	-0.004	[-0.010, 0.003]	99	-0.004	[-0.010, 0.001]

Note: ES = effect size Cohen's *d*, BS = bootstrapped. Outcomes are clustered at the study level to account for potential correlations between multiple outcomes from the same study. For all models: studies (*k)* = 65 with *n* = 119 estimates.

$^*p < .05$; $^{**}p < .01$; $^{***}p < .001$.

The RE model produced a point estimate of *d* = -.198 (95% CI [-.306, -.089]), which aligns with previous meta-analyses using RE models that reported significant medium-sized effects (Eklund et al., 2020; Maynard et al., 2013; Tanner-Smith & Wilson, 2013). However, models that account for systematic relationships between sample variation and effect size yield insignificant results. The UWLS estimate, *d* = -.006, 95% CI: [-.013, .001], is smaller and more precise than the RE estimate. Similarly, the PET yielded *d* = .007 (95% CI: [-.001, .014]) while the PEESE yielded *d* = -.004, 95% CI[-0.010, 0.003], both of which include zero in their respective confidence intervals.

Our data exhibit substantial heterogeneity, with I^2 ranging from 88% in the UWLS and RE to 99% in PET and PEESE. The latter include the (squared) standard error as a predictor, which directly alters the residual variance; hence, it affects Cochran's *Q* statistic and, consequently, I^2. This heterogeneity likely inflated the RE model's effect size due to its assumption that heterogeneity is random. These findings suggest that the assumptions underlying the RE model might well be violated in this context. On the other hand, the PET-PEESE estimates are known to perform poorly under conditions of high heterogeneity ($I^2 > 80\%$; Stanley & Doucouliagos, 2017). We note that the bootstrapped PET estimate (*d* = .007, 95% CI [.001, .016]) suggests that interventions can on average increase absenteeism albeit very negligible in size—a somewhat counterintuitive but not a meaningful effect. Given these considerations, UWLS appears to provide a more reliable estimate under these conditions.

Sensitivity checks: heterogeneity by composition

To examine how sample composition affects heterogeneity and bias in effect size estimates, we conducted sensitivity analyses by restricting the sample based on predefined structural differences and applying bootstrap aggregation techniques. The results of these analyses are summarized in Table 6, which includes bootstrapped estimates for the sensitivity samples. For reference, the bootstrapped results for the main sample are detailed in Table 5.

Table 6

Sensitivity (Bootstrap) Checks with Restricted RCT and No Outliers Samples

Sensitivity check	ES	95% CI	I^2	BS ES	BS 95% CI
RCT only [k = 48, n = 84]					
RE	-0.262**	[-0.416, -0.109]	89	-0.243**	[-0.331, -0.122]
UWLS	-0.047	[-0.108, 0.015]	89	-0.048*	[-0.089, -0.015]
PET	0.032	[-0.045, 0.109]	96	0.033	[-0.018, 0.086]
PEESE	-0.027	[-0.085, 0.032]	96	-0.027	[-0.067, 0.008]
No outliers [k = 47, n = 77]					
RE	-0.138***	[-0.171, -0.105]	53	-0.146***	[-0.176, -0.116]
UWLS	-0.115***	[-0.151, -0.080]	53	-0.117***	[-0.152, -0.090]
PET	-0.079**	[-0.126, -0.032]	79	-0.082***	[-0.132, -0.046]
PEESE	-0.105***	[-0.143, -0.067]	79	-0.107***	[-0.143, -0.079]

Note. *k* = the number of studies, *n* = the number of outcomes, ES = effect size Cohen's *d*, BS = bootstrapped.

*$p < .05$; **$p < .01$; ***$p < .001$

Sample restricted to RCTs

When limiting the analysis to studies using RCT designs (48 studies with 84 outcomes), only the RE model produces a significant effect size ($d = -.262$, 95% CI: [-.416, -.109]). However, these confidence intervals are wide compared to the other models. Notably, the bootstrapped estimate of UWLS also becomes significant (*d* = −.048, 95% CI: [-.089, -.015]), indicating that study design might influence effect size estimates and that UWLS provides a more conservative yet robust estimate under these conditions.

Sample excluding outliers

After removing studies identified as outliers based on Viechtbauer and Cheung's (2010) method (47 studies with 77 outcomes), all effect size estimates across models become significant, including the bootstrapped estimates. For example, the RE model yielded an effect size of *d* = −.138, (95% CI: [-.171, -.105]), while UWLS showed *d* = −.115, (95% CI: [-.151, -.080]). These results suggest that removing outliers reduces noise and stabilizes estimates, with bootstrapped confidence intervals remaining consistent across models.

Bootstrap aggregation

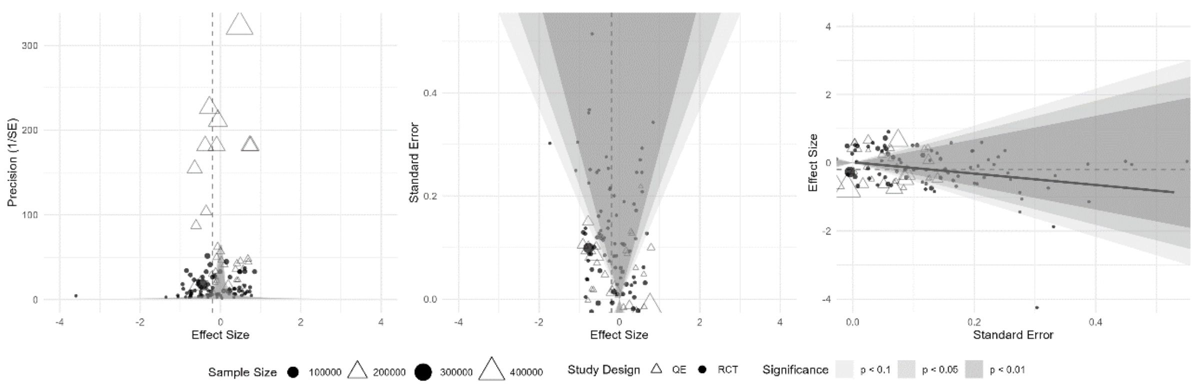
This method revealed how sensitive estimates are to sample composition without relying on predefined assumptions about heterogeneity sources. In our main sample, the RE model was the only one whose bootstrapped mean changed (*d* = −.181, 95% CI: [-.248, -.088]). For UWLS, the upper bound of the 95% confidence interval shifted slightly from [.001] to [-.002], rendering its effect size (*d* = −.006) significant but negligible in magnitude. Across all sensitivity checks, RE estimates consistently produce larger and less precise effects compared to UWLS, PET, and PEESE models—highlighting potential biases in the RE approach when heterogeneity is high or when larger studies systematically differ from smaller ones.

Publication bias

In this subsection, we address another source of bias that might be interacting with sources of heterogeneity: publication bias. The three-panel visualization in Figure 2 reveals several key patterns regarding potential publication bias. Effect sizes from quasi-experimental designs exhibit high precision, clustering predominantly around zero, with study sample sizes ranging from 250 to 450,000 participants. The contour-enhanced funnel plot (middle panel) demonstrates that most studies fall within significance thresholds, with a notable concentration of quasi-experimental studies in the high-precision region. The regression line (right panel) yields an intercept close to zero, while the slope indicates a systematic relationship where smaller studies tend to report larger effect sizes, which is a common finding in education research (Slavin & Smith, 2009).

Figure 2

(Inverted) funnel Plots for Assessing Publication Biases in total sample



Note. Grey contour shading centered at zero highlighted areas where null effects would not reject the null hypothesis. To increase visibility, we used jitter to add random noise. Specifically, jitter was applied with a horizontal width of 0.1 and a vertical height of 0.8, ensuring that overlapping points were distinguishable while preserving the overall data structure

To explicitly address publication bias, Table 7 shows the results of the WILS procedure, which iteratively removes sources of overrepresentation in statistically significant results.

Table 7

Controlling for publication bias: WILS estimates and procedure

WILS	**WILS procedure [BS]**			**Estimates**		**Bootstrap estimates**	
sample	**it**	***k***	***n***	**ES (SE)**	**95% CI**	**ES (SE)**	**95% CI**
Total	4 [4]	43 [43]	78 [78]	-0.002 (0.002)	[-0.005, 0.002]	-0.002 (0.003)	[-0.008, 0.004]
No outliers	8 [8]	35 [35]	57 [57]	-0,115*** (0.011)	[-0.136, -0.094]	-0.108*** (0.020)	[-0.148, -0.070]
RCT only	6 [6]	35 [35]	58 [57]	-0.055*** (0.013)	[-0.081, -0.029]	-0.041 (0.029)	[-0.098, 0.016]

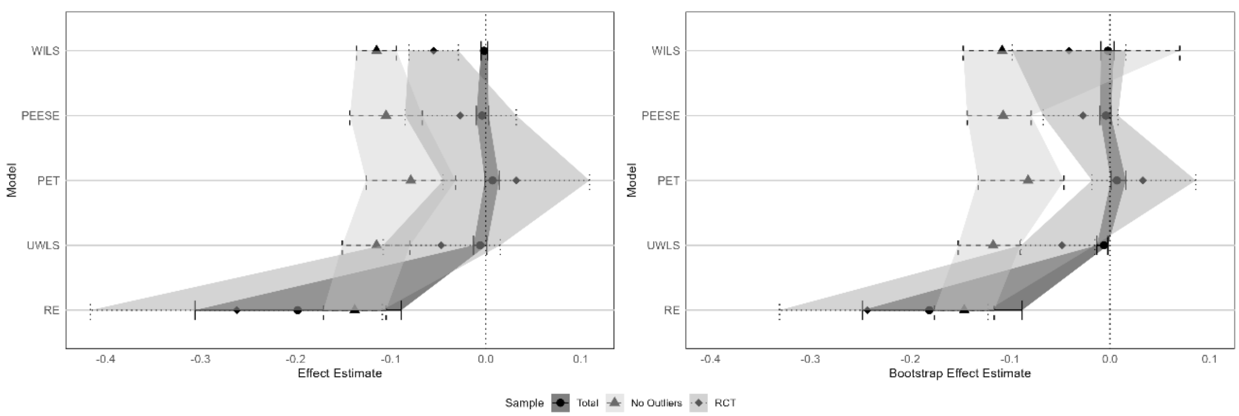
Note. ES = effect size Cohen's *d*, *it* = the number of iterations, *k* = the number of studies, *n* = the number of outcomes. The numbers in square brackets next to each statistic indicate the frequency of bootstrapping for that specific statistic. Standard errors (SE) are reported in parentheses. *p <.05; **p <.01; ***p < .001.

In the main sample, four iterations reduced the ESS to below zero, leaving a sample with 78 outcomes from 43 studies – resulting in a statistically insignificant final effect size estimate of $d = -.002$ (95% CI: [-.005, .002]), with a SE of 0.002. The restricted samples differ and display statistically significant estimates: the sample without outliers with $d = -.115$ (95% CI: [-.136, -.094]) and SE = 0.011 after eight iterations based on 57 outcomes and 35 studies. The RCT-only sample shows $d = -.055$ (95% CI: [-.081, -.029]) with a SE of 0.013 after six iterations. Both restricted samples make the effect size estimate statistically significant, but with a (relatively) larger standard error. In the bootstrapped estimates, iteration counts and remaining studies are nearly identical across samples. However, the stability of the estimate and SE diverges between the main and restricted samples. In the restricted samples, the SE increases in the RCT-only sample, rendering the estimate statistically insignificant. In contrast, the estimate for the no-outlier sample remains statistically significant despite an increased SE, while in the total sample, the SE increases only slightly from 0.001 to 0.003.

In order to facilitate inference with respect to whether interventions are effective across the different model specifications, the estimates of RE, UWLS, PET, PEESE, and WILS are shown on the y-axis in Figure 3, along with their effect estimates on the x-axis. The bootstrapped estimates are shown in the right panel. Figure 3 indicates that RE estimates look particularly inflated, followed by UWLS allowing nothing but perfect precision, hence shrinking the estimates towards zero. PET and PEESE take a more tolerant approach by allowing sample bias but controlling for it, yet their point estimates for the different samples also vary quite a bit. WILS appears consistent with stable point estimates but reveals uncertainty when outliers are absent in bootstrapped scenarios. Despite model differences in sensitivity to sample composition, they share one thing: apart from RE, the main sample effect size estimates are consistently and precisely centered around zero.

Figure 3

Comparison Mean Effects across Meta-analytic Models and Sensitivity Checks



Note. Points represent estimated effects; error bars indicate confidence intervals; shaded areas highlight uncertainty ranges for each estimate-sample-model combination. The dashed line at zero marks statistical significance threshold; right panel shows bootstrap estimates.

From an educational practice perspective, all estimated main effect sizes are small and arguably negligible after accounting for potential heterogeneity and publication bias. Having addressed these biases, we now turn to meta-regression analyses to explore whether intervention type influences these estimates.

Meta-regression

Table 8 presents effect size estimates for each intervention type using RE, PET, and PEESE meta-regression models. The RE model identifies three intervention types with statistically significant effect sizes: package interventions (*d* = -.308, 95% CI [-.510, -.105), school curriculum and learning environments (*d* = -.405, 95% CI [-.599, -.209) and support networks (*d* = -.250, 95% CI [-.432, -.068). These estimates are larger than the overall mean effect but have wide confidence intervals, limiting their precision.

Table 8

Meta-regression Results for Intervention Types using RE, PET and PEESE models

Intervention	**Model**	**ES**	**95% CI**	**BS ES**	**95% CI**
Aid programs	RE-MRA	-0.105	[-0.310, 0.101]	-0.088	[-0.202, 0.026]
	PET-MRA	0.004	[-0.003, 0.011]	0.007	[0.000, 0.032]
	PEESE-MRA	-0.005***	[-0.007, -0.003]	-0.005	[-0.015, 0.001]
Package	RE-MRA	-0.308*	[-0.510, -0.105]	-0.178***	[-0.272, -0.083]
	PET-MRA	0.046	[-0.001, 0.094]	0.044	[-0.004, 0.079]
	PEESE-MRA	0.003	[-0.049, 0.055]	-0.001	[-0.052, 0.029]
School Curriculum and Learning Environments	RE-MRA	-0.405**	[-0.599, -0.209]	-0.277*	[-0.512, -0.042]
	PET-MRA	0.009*	[0.001, 0.018]	0.009	[0.000, 0.018]
	PEESE-MRA	-0.004	[-0.015, 0.008]	-0.004	[-0.017, 0.003]
Support networks	RE-MRA	-0.250*	[-0.432, -0.068]	-0.137***	[-0.208, -0.067]
	PET-MRA	0.044	[-0.003, 0.092]	0.047	[0.006, 0.094]
	PEESE-MRA	-0.002	[-0.034, 0.031]	0.000	[-0.030, 0.030]

Note. ES = effect size Cohen’s *d*, BS = bootstrapped

$^*p < .05$; $^{**}p < .01$; $^{***}p < .001$.

Contradicting the RE-MRA results, the PEESE-MRA yields a small but precise—and thus statistically significant—effect size for aid programs (*d* = -.005, 95% CI -.007, -.003). However, PET-MRA estimates suggest a small but statistically significant effect in the opposite direction for school curriculum and learning environments programs (*d* =.009, 95% CI [.001, .013]). These inconsistencies highlight how sensitive results are to model assumptions; effects found with RE do not hold when explicitly controlling for potential sample bias using PET or PEESE.

To verify these findings, we applied WILS meta-regression (Table 9), which compares intervention types pairwise. WILS shows that package interventions *d* = -.003 (CI 95% [-.005, -

.002]), school curriculum and learning environments (*d* = -.005 (CI 95% [-.006, -.003]) and support networks (*d* = −.004, CI 95% [-.006, -002]) all still have statistically significant effects; yet they are very small in magnitude and become statistically insignificant after bootstrapping.

Table 9

WILS Pairwise Comparisons of Intervention Types

Intervention	ES (*d*)	95% CI	BS ES(*d*)	95% CI
Aid programs	0.001	[-0.001, 0.004]	-0.001	[-0.019, 0.008]
Package	-0.003***	[-0.005, -0.002]	-0.003	[-0.018, 0.007]
School Curriculum and Learning Environments	-0.005***	[-0.006, -0.003]	-0.005	[-0.023, 0.012]
Support networks	-0.004***	[-0.006, -0.002]	-0.004	[-0.022, 0.005]

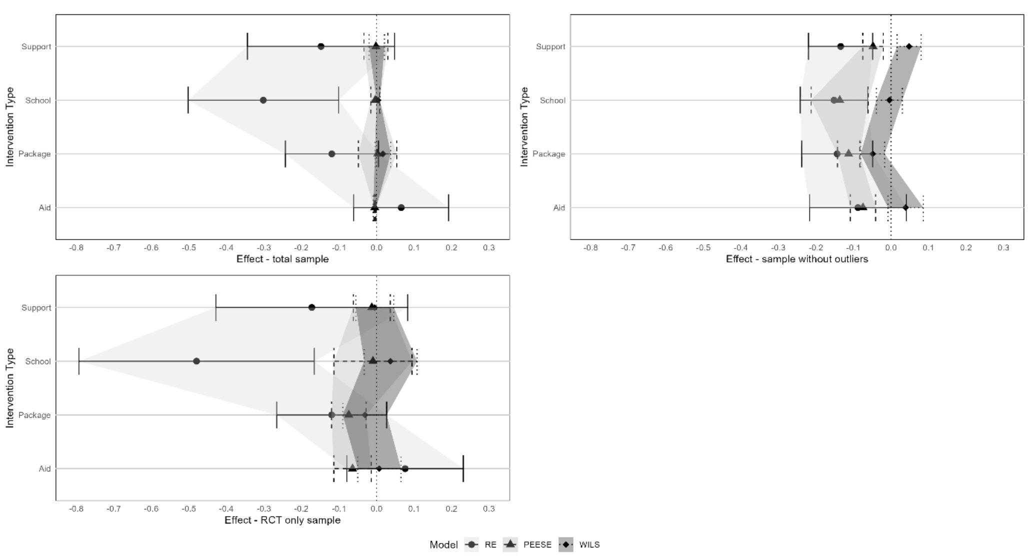
Note. ES = effect size Cohen's *d*, BS = bootstrapped.

*p < .05 ; **p < .01; ***p < .001.

Figure 4 compares effect sizes for intervention types across meta-analytic models. PEESE is preferred to be included over PET as it has narrower confidence intervals and better accuracy when correcting for publication bias when there is a lot of heterogeneity (see Figure S3, online only).

Figure 4

Comparative Effects of Intervention Types across Meta-analytic Models



Note. Points represent the estimated effects, error bars indicate confidence intervals, and shaded areas highlight the range of uncertainty for each model. The dashed line at zero marks the threshold confidence interval limits for statistical significance.

Upon examining effect sizes across meta-analytic models in Figure 4, differences between models appear minimal. The RE model consistently exhibits the widest confidence intervals across intervention types and samples. In the total sample, aid programs have small statistically significant effects according to both PEESE (*d* = -.005, 95% CI [-.007, -.003]) and WILS (*d* = -.005, 95% CI [-.009, -.001) while school curriculum and learning environment programs show statistical significance with RE (*d* = -.302, 95% CI [-.502, -.100]). Removing outliers generally leads to gains in model precision, resulting in more statistically significant effect sizes for different intervention types. All the point estimates of PEESE reach statistical significance, ranging from school curriculum and learning environment programs with *d* = -.137, 95% CI: [-.213, -.060] to support programs with *d* = -.048, 95% CI [-.075, -.020]. WILS only identifies package programs as on average effective with *d* = -.049, 95% CI [-.081, -.017]. RE point estimates for package and support programs reach statistical significance, with the results for school curriculum and learning environment programs—the only statistically significant intervention type based on RE models in the total sample—now being estimated to have an average effect size of *d* = -.152, 95% CI [-.242 -.162]).

In RCT-only samples, RE again finds statistically significant effects for school curriculum and learning environment programs, but even more sizable with *d* = -.480 (95% CI [-.793, -166]). PEESE identifies small but statistically significant effects for aid programs (*d* = -.064, 05% CI [-.114, -.014]), and package interventions (*d* = -.074, 95% CI [-.120, -028]). WILS identifies no statistical significant effects in this restricted sample.

Interpretation of meta-regression findings

The meta-regression analysis shows that each intervention type seems to have relatively minimal influence on absenteeism reduction when accounting explicitly for heterogeneity and publication bias through robust methods like PEESE or WILS regression models. While RE suggests moderate effect sizes for school curriculum and learning environments programs, these results are not corroborated with PEESE or WILS regression models. Removing outliers improves precision but does not fundamentally alter conclusions; all estimated effects remain small and close to zero after adjusting for biases like heterogeneity, sample, and publication selection bias.

Even with improved precision from PEESE or WILS iterative process and bootstrapping procedures, no consistent empirical evidence is observed for any of the specific intervention types in being effective in reducing absenteeism.

Discussion

Conclusion

This meta-analysis found no conclusive empirical evidence that school-based interventions significantly reduce absenteeism. While performing a traditional meta-analytic random effects model suggested an overall and moderate effect for certain intervention types, most notably school curriculum and learning environment programs, these effects all became negligible or inconsistent when more advanced meta-analytic models, such as PEESE and WILS, which explicitly address potential heterogeneity and publication bias were applied. Furthermore, in restricted samples focusing only on RCT studies or excluding outliers some differential patterns of statistically significant effect sizes were observed. However, again, for the more advanced meta-analytics models explicitly dealing with potential biases, these estimates were all relatively small and of limited practical relevance. Such inconsistencies highlight the complexity of analyzing interventions in educational contexts where heterogeneity is inherent and substantial.

Our findings also emphasize the importance of meta-analytic methodological innovations in educational research. By applying WILS, the results presented here demonstrate how meta-analyses can move beyond traditional assumptions, like those underlying RE models, to empirically better address potential heterogeneity and publication bias. The results revealed nuanced patterns of empirical evidence, opening new pathways for future research to build a solid empirical evidence base for addressing universal educational challenges, such as absenteeism.

Limitations

Our study has several limitations that may explain the lack of robust findings. First, our broad selection of publication years and countries may have compounded heterogeneity issues by introducing multifaceted biases related to differing standards over time and across regions. These biases could potentially violate key assumptions underlying our models. However, as demonstrated in the supplemental material (Table S6), our estimates remain similarly inconclusive when the sample is restricted to studies conducted in North America or limited to the decade of the 2010s. The results presented in this study indicate that even disaggregating studies by intervention type, a very likely source of variability in effectiveness, does not seem to alter the main conclusion of inconclusive evidence, and is therefore particularly relevant in light of this concern of compounded heterogeneity. Additionally, our decision to include only peer-reviewed studies may have introduced selection bias by excluding unpublished but potentially relevant studies. Future work could elaborate on whether expanding the evidence base to also include non-published evaluations can provide different insights.

Second, RCTs conducted decades apart may not be directly comparable due to shifts in educational practices and research standards over time. Acknowledging these dynamics in educational systems over time highlights the importance of periodically updating the evidence base for important educational challenges, such as reducing absenteeism, with state-of-the-art meta-analysis.

Third, while advanced methods like PET-PEESE and WILS allowed us to address heterogeneity and publication bias more effectively than traditional RE models, these approaches also have limitations. For example, WILS assumes that the probability of publication is based on retrospective power calculations derived from the sample's known distribution. This assumption may not hold when comparing studies across countries or decades with varying publication standards, as the distribution of effect sizes at the time of publication is unknown. As a result, WILS may inadvertently eliminate studies based on contextual factors rather than true sources of bias.

Reflection on meta-analytic techniques

This study has emphasized the relevance of not solely relying on applying traditional meta-analytic standards. While transparency and reporting guidelines are essential (Higgins et al., 2024; Page et al., 2021), they should not limit researchers to only apply traditional methods like forest plots, RE models, funnel plots, and meta-regression based on RE as default tools for meta-analysis. Instead, guidelines should encourage the application of diverse methods with fundamentally different assumptions to ensure robustness of the results across a multitude of approaches.

For example, requiring researchers to apply at least one "mirror meta-analytical method"—a secondary approach with contrasting assumptions—could enhance the reliability of findings by systematically addressing heterogeneity and publication bias. Transparency in search criteria and data extraction remains equally critical but can be paired with innovative approaches that challenge existing norms.

Meta-analyses in education should also prioritize optimizing the use of existing research rather than solely recommending the need for higher-quality new evaluation studies. While this remains an important goal for safeguarding the quality of the evidence available, there could be more emphasis on developing techniques to optimize the use of existing evaluation using more novel techniques such as synthetic control methods.

Future directions: heterogeneous ways to address heterogeneity

The lack of conclusive empirical evidence in this study may reflect either an absence of effective school-based interventions or limitations in our ability to detect them due to, for example, unreported moderators or insufficient studies of effective interventions. Future meta-analyses should combine multiple methods with different assumptions about heterogeneity and publication bias rather than relying on a single approach. For example, Maier et al.'s (2022) selection models offer promising directions, but still assume constant probabilities for publication across time and contexts—an assumption that may not hold in highly heterogeneous fields like education.

By systematically addressing heterogeneity through innovative methods and flexible frameworks for causal designs, meta-analyses can achieve greater policy relevance in educational contexts. This shift would require moving beyond descriptive reporting toward frameworks that define meta-analytic null results while providing clear guidance on how different methods should be applied based on research questions and sample characteristics.

In conclusion, while our findings were inconclusive regarding specific intervention effectiveness, they underscore the importance of methodological innovation in educational meta-analyses. By optimizing current research strategies and embracing diverse analytical approaches, we can better understand and address complex challenges in education such as absenteeism.

References

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

*Anderson, K. P., Egalite, A. J., & Mills, J. N. (2019). Discipline reform: The impact of a statewide ban on suspensions for truancy. *Journal of Education for Students Placed at Risk*, *24*(1), 68-91. https://doi.org/10.1080/10824669.2018.1537794

*Awaworyi Churchill, S., Iqbal, N., Nawaz, S., & Yew, S. L. (2021). Unconditional cash transfers, child labour and education: Theory and evidence. *Journal of Economic Behavior & Organization*, *186*, 437-457. https://doi.org/10.1016/j.jebo.2021.04.012

*Axford, N., Bjornstad, G., Clarkson, S., Ukoumunne, O. C., Wrigley, Z., Matthews, J., Berry, V., & Hutchings, J. (2020). The effectiveness of the kiva bullying prevention program in wales, uk: Results from a pragmatic cluster randomized controlled trial. *Prevention Science*. https://doi.org/10.1007/s11121-020-01103-9

Balfanz, R., & Byrnes, V. (2018). Using Data and the Human Touch: Evaluating the NYC Inter-Agency Campaign to Reduce Chronic Absenteeism. *Journal of Education for Students Placed at Risk*, *23*(1-2), 107-121. https://doi.org/10.1080/10824669.2018.1435283

Balkis, M., Arslan, G., & Duru, E. (2016). The school Absenteeism among high school students: Contributing Factors. *Educational Sciences: Theory & Practice*. https://doi.org/10.12738/estp.2016.6.0125

*Bannink, R., Broeren, S., Heydelberg, J., van ’t Klooster, E., van Baar, C., & Raat, H. (2014). Your health: An intervention at senior vocational schools to promote adolescents’ health and health behaviors. *Health Education Research*, *29*(5), 773-785. https://doi.org/10.1093/her/cyu038

*Bavarian, N., Lewis, K. M., DuBois, D. L., Acock, A., Vuchinich, S., Silverthorn, N., Snyder, F. J., Day, J., Ji, P., & Flay, B. R. (2013). Using social-Emotional and character development to improve academic outcomes: A matched-Pair, cluster-randomized controlled trial in low-Income, urban schools. *Journal of School Health*, *83*(11), 771-779. https://doi.org/10.1111/josh.12093

Belfield, C. R., & Levin, H. M. (Eds.). (2007). *The price we pay: Economic and social Consequences of Inadequate Education*. Brookings Institution Press. http://www.jstor.org/stable/10.7864/j.ctt126269

*Borgen, N. T., Olweus, D., Kirkebøen, L. J., Breivik, K., Solberg, M. E., Frønes, I., Cross, D., & Raaum, O. (2021). The potential of anti-bullying efforts to prevent academic failure and youth crime A case using the olweus bullying prevention program (obpp). *Prevention Science*. psyh. https://doi.org/10.1007/s11121-021-01254-3

*Borgen, N. T., Raaum, O., Kirkebøen, L. J., Sørlie, M.-A., Ogden, T., & Frønes, I. (2021). Heterogeneity in Short- and Long-Term Impacts of School-Wide Positive Behavior Support (SWPBS) on Academic Outcomes, Behavioral Outcomes, and Criminal Activity. *Journal of Research on Educational Effectiveness*, *14*(2), 379-409. https://doi.org/10.1080/19345747.2020.1862375

*Borman, G. D., Rozek, C. S., Pyne, J., & Hanselman, P. (2019). Reappraising Academic and Social Adversity Improves Middle-School Students' Academic Achievement, Behavior, and Well-Being. *Grantee Submission*. https://doi.org/10.1073/pnas.1820317116

*Brathwaite, R., Namuwonge, F., Magorokosho, N., Tutlam, N., Neilands, T. B., Namirembe, R., Ssentumbwe, V., & Ssewamala, F. M. (2024). Impact of economic and family intervention on adolescent girls' education performance, school absenteeism, and behavior in school: The Suubi4Her study. *Journal of Adolescent Health*, *74*(2), 340-349. https://doi.org/10.1016/j.jadohealth.2023.08.043

*Cabus, S. J., & De Witte, K. (2011). Does School Time Matter?—On the Impact of Compulsory Education Age on School Dropout. *Economics of Education Review*, *30*(6), 1384-1398. https://doi.org/10.1016/j.econedurev.2011.07.003

Chappell, S. L., O'Connor, P., Withington, C., & Stegelin, D. A. (2015). *A Meta-Analysis of Dropout Prevention Outcomes and Strategies*. Clemson, SC: National Dropout Prevention Center/Network at Clemson University. https://dropoutprevention.org/wp-content/uploads/2022/11/A-Meta-Analysis-of-DOP-Outcomes-and-Strategies-Chappell-et-al-2015.pdf

*Cho, H., Catherine Ryberg, R., Hwang, K., Pearce, L. D., & Iritani, B. J. (2017). A school support intervention and educational outcomes among orphaned adolescents: Results of a cluster randomized controlled trial in Kenya. *Prevention Science*, *18*(8), 943-954. https://doi.org/10.1007/s11121-017-0817-x

*Converse, N., & Lignugaris/Kraft, B. (2009). Evaluation of a School-Based Mentoring Program for At-Risk Middle School Youth. Remedial and Special Education, 30(1), 33-46. https://doi.org/10.1177/0741932507314023

*Corcoran, S. P., Elbel, B., & Schwartz, A. E. (2016). The effect of breakfast in the classroom on obesity and academic performance: Evidence from New York City. *Journal of Policy Analysis and Management*, *35*(3), 509-532. https://doi.org/10.1002/pam.21909

*Dee, T. S. (2011). Conditional Cash Penalties in Education: Evidence from the Learnfare Experiment. Economics of Education Review, 30(5), 924-937. https://doi.org/10.1016/j.econedurev.2011.05.013

*DeSocio, J., VanCura, M., Nelson, L. A., Hewitt, G., Kitzman, H., & Cole, R. (2007). Engaging Truant Adolescents: Results from a Multifaceted Intervention Pilot. *Preventing School Failure*, *51*(3), 3-9. https://doi.org/10.3200/PSFL.51.3.3-11

*Durkin, K., Lipsey, M. W., Farran, D. C., & Wiesen, S. E. (2022). Effects of a statewide pre-kindergarten program on children's achievement and behavior through sixth grade. *Developmental Psychology*. pdh. https://doi.org/10.1037/dev0001301

Duval, S., & Tweedie, R. (2000). *Trim and fill: A simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis*. Biometrics, 56(2), 455-463. https://doi.org/10.1111/j.0006-341X.2000.00455.x

Egger, H. L., Costello, J. E., & Angold, A. (2003). School Refusal and Psychiatric Disorders: A Community Study. *Journal of the American Academy of Child & Adolescent Psychiatry*, *42*(7), 797-807. https://doi.org/10.1097/01.CHI.0000046865.56865.79

Eklund, K., Burns, M. K., Oyen, K., DeMarchena, S., & McCollom, E. M. (2020). Addressing Chronic Absenteeism in Schools: A Meta-Analysis of Evidence-Based Interventions. *School Psychology Review*, *51*(1), 95-111. pbh. https://doi.org/10.1080/2372966X.2020.1789436

*Felver, J. C., Clawson, A. J., Morton, M. L., Brier-Kennedy, E., Janack, P., & DiFlorio, R. A. (2019). School-based mindfulness intervention supports adolescent resiliency: A randomized controlled pilot study. *International Journal of School & Educational Psychology*, *7*, 111-122. pbh. https://doi.org/10.1080/21683603.2018.1461722

Fernandes, C.-S. F., Kannoth, S., Pendergrass Boomer, T. M., Hieftje, K. D., & Fiellin, L. E. (2024). Systematic Review of Interventions with Some School Involvement for School

Refusal in High School–Age Adolescents. *Children & Schools*, *46*(2), 85-95. https://doi.org/10.1093/cs/cdae003

*Flannery, K. B., Kato, M. M., Kittelman, A., McIntosh, K., & Triplett, D. (2020). A tier 1 intervention to increase ninth grade engagement and success: Results from a randomized controlled trial. *School Psychology*, *35*(1), 88-98. pdh. https://doi.org/10.1037/spq0000347

*Flannery, K. B., Kato, M. M., Kittelman, A., Sampson, N. K., & McIntosh, K. (2024). Check-In/Check-Out (CICO) in High School: Results from a small randomized controlled trial. *Behavioral Disorders*, *49*(3), 149-159. psyh. https://doi.org/10.1177/01987429231225461

Gershenson, S., Jacknowitz, A., & Brannegan, A. (2017). Are Student Absences Worth the Worry in U.S. Primary Schools? *Education Finance and Policy 2017; 12 (2): 137–165.* https://doi.org/10.1162/EDFP_a_00207

Goodman, J. (2014). *Flaking Out: Student Absences and Snow Days as Disruptions of Instructional Time* (No. w20221; p. w20221). National Bureau of Economic Research. https://doi.org/10.3386/w20221

*Gottfredson, D., Cross, A. B., Wilson, D., Rorie, M., & Connell, N. (2010). Effects of Participation in after-School Programs for Middle School Students: A Randomized Trial. *Journal of Research on Educational Effectiveness*, *3*(3), 282-313. https://doi.org/10.1080/19345741003686659

Gottfried, M. A. (2009). Excused versus Unexcused: How Student Absences in Elementary School Affect Academic Achievement. *Educational Evaluation and Policy Analysis*, *31*(4), 392-415. https://doi.org/10.3102/0162373709342467 \

Gottfried, M. A. (2014). Chronic absenteeism and its effects on students' academic and socioemotional outcomes. *Journal of Education for Students Placed at Risk*, *19*(2), 53-75. psyh. https://doi.org/10.1080/10824669.2014.962696

Gottfried, M. A. (2015). Prekindergarten and kindergarten center-based child care and students' early schooling outcomes. *Teachers College Record*, *117*(11), 1-28. https://doi.org/10.1177/016146811511701105

*Graeff-Martins, A. S., Oswald, S., Comassetto, J. O., Kieling, C., Gonçalves, R. R., & Rohde, L. A. (2006). A package of interventions to reduce school dropout in public schools in a developing country: A feasibility study. *European Child & Adolescent Psychiatry*, *15*(8), 442-449. https://doi.org/10.1007/s00787-006-0555-2

*Grant, A. A., Mac Iver, D. J., Byrnes, V., Clark, E., Balfanz, R., & Lofton, R. (2023). Combining restorative practices with diplomas now: Impacts on practices, problems, suspensions, and chronic absenteeism. *Journal of Education for Students Placed at Risk*. https://doi.org/10.1080/10824669.2023.2278047

Gubbels, J., Van Der Put, C. E., & Assink, M. (2019). Risk Factors for School Absenteeism and Dropout: A Meta-Analytic Review. *Journal of Youth and Adolescence*, *48*(9), 1637-1667. https://doi.org/10.1007/s10964-019-01072-5

*Guryan, J., Christenson, S., Cureton, A., Lai, I., Ludwig, J., Schwarz, C., Shirey, E., & Turner, M. C. (2021). The Effect of Mentoring on School Attendance and Academic Outcomes: A Randomized Evaluation of the Check & Connect Program. *Journal of Policy Analysis and Management*, *40*(3), 841-882. https://doi.org/10.1002/pam.22264

Hancock, K.J., Shepherd, C.C.J., Lawrence, D., Zubrick, S.R. (2013). *Student attendance and educational outcomes: every day counts.* Report for the Department of Education, Employment and Workspace Relations, Canberra. https://apo.org.au/node/35141

*Heppen, J. B., Zeiser, K., Holtzman, D. J., O'Cummings, M., Christenson, S., & Pohl, A. (2018). Efficacy of the Check & Connect Mentoring Program for At-Risk General Education High School Students. *Journal of Research on Educational Effectiveness*, *11*(1), 56-82. https://doi.org/10.1080/19345747.2017.1318990

Heyne, D., Strömbeck, J., Alanko, K., Bergström, M., & Ulriksen, R. (2020). A Scoping Review of Constructs Measured Following Intervention for School Refusal: Are We Measuring Up? *Frontiers in Psychology*, *11*, 1744. https://doi.org/10.3389/fpsyg.2020.01744

Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (Eds.). (2024). *Cochrane Handbook for Systematic Reviews of Interventions* (Version 6.5). Cochrane. https://doi.org/10.1002/9780470712184

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Thousand Oaks, CA: Sage Publications. https://doi.org/10.4135/9781412985031

*James, S. A., Erickson, D. J., Lammert, S., & Widome, R. (2023). School start time delays and high school educational outcomes: Evidence from the START/LEARN study. *Journal of Adolescence*, *95*(4), 751-763. https://doi.org/10.1002/jad.12151

*Johnsen, D. B., Lomholt, J. J., Heyne, D., Jensen, M. B., Jeppesen, P., Silverman, W. K., & Thastum, M. (2024). The effectiveness of modular transdiagnostic cognitive behavioral therapy versus treatment as usual for youths displaying school attendance problems: A randomized controlled trial. *Research on Child and Adolescent Psychopathology*. https://doi.org/10.1007/s10802-024-01196-8

Johansson, B. (2019). *DROPPING OUT OF SCHOOL : a systematic and integrative researchreview on risk factors an interventions.* Örebro University. http://www.diva-portal.org/smash/get/diva2:1369852/FULLTEXT01.pdf

*Jones, C. J., & Christian, M. (2021). The Results of a Randomized Control Trial Evaluation of the Spark Literacy Program: An Innovative Approach That Pairs One-on-One Tutoring with Family Engagement. *Journal of Education for Students Placed at Risk*, *26*(3), 185-209. https://doi.org/10.1080/10824669.2020.1809419

*Jones, L. P., Harris, R., & Finnegan, D. (2002). School Attendance Demonstration Project: An evaluation of a program to motivate public assistance teens to attend and complete school in an urban school district. *Research on Social Work Practice, 12*(2), 222-237. https://doi.org/10.1177/104973150201200202

Kaufman, P., Alt, M. N., & Chapman, C. D. (2004). *Dropout Rates in the United States: 2000*. (NCES 2002-114). U.S. Department of Education, National Center for Education Statistics. https://doi.org/10.1037/e672832007-008

Kearney, C. A. (2002). *Identifying the Function of School Refusal Behavior: A Revision of the School Refusal Assessment Scale -Revised (SRAS-R, SRAS, SRAS-C-R, SRAS-P-R, SRAS-C, SRAS-P)* [Database record]. APA PsycTests. https://doi.org/10.1037/t11398-000

Kearney, C.A. (2008a). School absenteeism and school refusal behavior in youth: A contemporary review. *Clinical Psychology Review*, *28*(3), 451-471. https://doi.org/10.1016/j.cpr.2007.07.012

Kearney, C. A. (2008b). An Interdisciplinary Model of School Absenteeism in Youth to Inform Professional Practice and Public Policy. *Educational Psychology Review*, *20*(3), 257-282. https://doi.org/10.1007/s10648-008-9078-3

Kearney, C. A., Gonzálvez, C., Graczyk, P. A., & Fornander, M. J. (2019). Reconciling Contemporary Approaches to School Attendance and School Absenteeism: Toward Promotion and Nimble Response, Global Policy Review and Implementation, and Future Adaptability (Part 1). *Frontiers in Psychology*, *10*, 2222. https://doi.org/10.3389/fpsyg.2019.02222

Kearney, C. A., & Childs, J. (2023). Translating sophisticated data analytic strategies regarding school attendance and absenteeism into targeted educational policy. Improving Schools, 26(1), 5-22. https://doi-org.vu-nl.idm.oclc.org/10.1177/13654802231174986

*Kim, J., & Joo, M. (2011). Did PRWORA's mandatory school attendance policy increase attendance among targeted teenage girls? *Children and Youth Services Review*, *33*(9), 1616-1623. https://doi.org/10.1016/j.childyouth.2011.04.003

Kirksey, J. J. (2019). Academic Harms of Missing High School and the Accuracy of Current Policy Thresholds: Analysis of Preregistered Administrative Data from a California School District. *AERA Open*, *5*(3). https://doi.org/10.1177/2332858419867692

*Kirksey, J. J., & Gottfried, M. A. (2021). The Effect of Serving 'Breakfast After-the-Bell' Meals on School Absenteeism: Comparing Results from Regression Discontinuity Designs. *Educational Evaluation and Policy Analysis*, *43*(2), 305-328. https://doi.org/10.3102/0162373721991572

*Komisarow, S., & Gonzalez, R. (2023). Can Community Crime Monitoring Reduce Student Absenteeism? *Education Finance and Policy*, *18*(2), 319-350. https://doi.org/10.1162/edfp_a_00361

Kraft, M. A., & Novicoff, S. (2024). Time in School: A Conceptual Framework, Synthesis of the Causal Research, and Empirical Exploration. *American Educational Research Journal*, *61*(4), 724-766. https://doi.org/10.3102/00028312241251857

Kreitz-Sandberg, S., & Fredriksson, U. (2023). Comparative Perspectives on School Attendance, Absenteeism, and Preventive Measures in Europe and Beyond. *European Education*, *55*(3-4), 137-147. https://doi.org/10.1080/10564934.2023.2281394

Kremer, K. P., Maynard, B. R., Polanin, J. R., Vaughn, M. G., & Sarteschi, C. M. (2015). Effects of After-School Programs with At-Risk Youth on Attendance and Externalizing Behaviors: A Systematic Review and Meta-Analysis. *Journal of Youth and Adolescence*, *44*(3), 616-636. https://doi.org/10.1007/s10964-014-0226-4

*Lasky-Fink, J., Robinson, C. D., Chang, H. N.-L., & Rogers, T. (2021). Using Behavioral Insights to Improve School Administrative Communications: The Case of Truancy Notifications. *Educational Researcher*, *50*(7), 442-450. https://doi.org/10.3102/0013189X211000749

Leduc, K., Tougas, A.-M., Robert, V., & Boulanger, C. (2024). School Refusal in Youth: A Systematic Review of Ecological Factors. *Child Psychiatry & Human Development*, *55*(4), 1044-1062. https://doi.org/10.1007/s10578-022-01469-7

*Lee, W. F., McNeely, C. A., Rosenbaum, J. E., Alemu, B., & Renner, L. M. (2020). Can Court Diversion Improve School Attendance among Elementary Students? Evidence from Five

School Districts. *Journal of Research on Educational Effectiveness*, *13*(4), 625-651. https://doi.org/10.1080/19345747.2020.1760976

Lehr, C. A., Sinclair, M. F., & Christenson, S. L. (2004). Addressing Student Engagement and Truancy Prevention During the Elementary School Years: A Replication Study of the Check & Connect Model. *Journal of Education for Students Placed at Risk (JESPAR), 9*(3), 279-301. https://doi.org/10.1207/s15327671espr0903_4

*Lipsey, M. W., Farran, D. C., & Durkin, K. (2018). Effects of the Tennessee Prekindergarten Program on children's achievement and behavior through third grade. *Early Childhood Research Quarterly*, *45,* 155-176. https://doi.org/10.1016/j.ecresq.2018.03.005

*Mac Iver, M. A. (2011). The Challenge of Improving Urban High School Graduation

* Lotyczewski, B.S., Whittington, R.A., Smith, L., Sanyshyn, S., Duprey, E.B., Allan, M., McFall, J.P., & White, A.M. (2024). The Effect of the Primary Project Program on School Attendance in Early Childhood. *Journal of Applied School Psychology*, *40*(1), 25-37. https://doi.org/10.1080/15377903.2023.2236982

Outcomes: Findings from a Randomized Study of Dropout Prevention Efforts. *Journal of Education for Students Placed at Risk, 16*(3), 167-184. https://doi.org/10.1080/10824669.2011.584497

*Mac Iver, M. A., Stein, M. L., Davis, M. H., Balfanz, R. W., & Fox, J. H. (2019). An Efficacy Study of a Ninth-Grade Early Warning Indicator Intervention. *Journal of Research on Educational Effectiveness*, *12*(3), 363-390. https://doi.org/10.1080/19345747.2019.1615156

* Mac Iver, M. A., Wills, K., Cruz, A., & Mac Iver, D. J. (2022). The Impact of Nudge Letters on Improving Attendance in an Urban District. Education and Urban Society, 54(2), 164-185. https://doi.org/10.1177/00131245211004561

Maier, M., VanderWeele, T. J., & Mathur, M. B. (2022). Using selection models to assess sensitivity to publication bias: A tutorial and call for more routine use. Campbell systematic reviews, 18(3), e1256. https://doi.org/10.1002/cl2.1256

*Marvul, J. N. (2012). If you build it, they will come: A successful truancy intervention program in a small high school. *Urban Education*, *47*(1), 144-169. https://doi.org/10.1177/0042085911427738

*Maynard, B. R., Kjellstrand, E. K., & Thompson, A. M. (2014). Effects of Check and Connect on Attendance, Behavior, and Academics: A Randomized Effectiveness Trial. *Research on Social Work Practice*, *24*(3), 296-309. https://doi.org/10.1177/1049731513497804

Maynard, B. R., McCrea, K. T., Pigott, T. D., & Kelly, M. S. (2013). Indicated Truancy Interventions for Chronic Truant Students: A Campbell Systematic Review. *Research on Social Work Practice*, *23*(1), 5-21. https://doi.org/10.1177/1049731512457207

*Mazerolle, L., Antrobus, E., Bennett, S., & Eggins, E. (2017). Reducing truancy and fostering a willingness to attend school: Results from a randomized trial of a police-school partnership program. *Prevention Science*, *18*(4), 469-480. https://doi.org/10.1007/s11121-017-0771-7

*Mazrekaj, D., & De Witte, K. (2020). The effect of modular education on school dropout. *British Educational Research Journal*, *46*(1), 92-121. pbh. https://doi.org/10.1002/berj.3569

* McCarthy, A. E., Young, J. F., Benas, J. S., & Gallop, R. J. (2018). School-Related Outcomes From a Randomized Controlled Trial of Adolescent Depression Prevention Programs. Journal of Emotional and Behavioral Disorders, 26(3), 170-181. https://doi.org/10.1177/1063426617717739

*McEwan, P. J. (2013). The Impact of Chile's School Feeding Program on Education Outcomes. *Economics of Education Review*, *32*, 122-139. https://doi.org/10.1016/j.econedurev.2012.08.006

*McNeely, C. A., Lee, W. F., Rosenbaum, J. E., Alemu, B., & Renner, L. M. (2019). Long-term effects of truancy diversion on school attendance: A quasi-experimental study with linked administrative data. *Prevention Science*, *20*(7), 996-1008. https://doi.org/10.1007/s11121-019-01027-z

*Mhurchu, C. N., Gorton, D., Turley, M., Jiang, Y., Michie, J., Maddison, R., & Hattie, J. (2013). Effects of a free school breakfast programme on children's attendance, academic achievement and short-term hunger: Results from a stepped-wedge, cluster randomised controlled trial. *Journal of Epidemiology and Community Health*, *67*(3), 257-264. https://doi.org/10.1136/jech-2012-201540

*Modestino, A. S., & Paulsen, R. (2023). School's Out: How Summer Youth Employment Programs Impact Academic Outcomes. *Education Finance and Policy*, *18*(1), 97-126. https://doi.org/10.1162/edfp_a_00371

*Newton, N. C., Andrews, G., Champion, K. E., & Teesson, M. (2014). Universal internet-based prevention for alcohol and cannabis use reduces truancy, psychological distress and moral disengagement: A cluster randomised controlled trial. *Preventive Medicine: An*

International Journal Devoted to Practice and Theory, *65*, 109-115. https://doi.org/10.1016/j.ypmed.2014.05.003

Nik Jaafar, N. R., Tuti Iryani, M. D., Wan Salwina, W. I., Fairuz Nazri, A. R., Kamal, N. A., Prakash, R. J., & Shah, S. A. (2013). Externalizing and internalizing syndromes in relation to school truancy among adolescents in high-risk urban schools. *Asia-Pacific Psychiatry*, *5*(S1), 27-34. https://doi.org/10.1111/appy.12072

Olssen, M., Codd, J., & O'Neill, A.-M. (2024). *Education Policy: Globalization, Citizenship and Democracy*. SAGE Publications. https://doi.org/10.4135/9781446221501

OECD (2024). "Evaluating post-pandemic education policies and combatting student absenteeism beyond COVID-19", *OECD Education Policy Perspectives*, No. 101, OECD Publishing, Paris, https://doi.org/10.1787/a38f74b2-en.

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

*Reid, W. J., & Bailey-Dempsey, C. (1995). The effects of monetary incentives on school performance. *Families in Society*, *76*(6), 331-340. psyh.

*Robinson, C. D., Lee, M. G., Dearing, E., & Rogers, T. (2018). Reducing Student Absenteeism in the Early Grades by Targeting Parental Beliefs. *American Educational Research Journal*, *55*(6), 1163-1192. https://doi.org/10.3102/0002831218772274

*Rosário, P., Núñez, J. C., Vallejo, G., Azevedo, R., Pereira, R., Moreira, T., Fuentes, S., & Valle, A. (2017). Promoting Gypsy children's behavioural engagement and school success: Evidence from a four-wave longitudinal study. *British Educational Research Journal*, *43*(3), 554-571. https://doi.org/10.1002/berj.3271

*Rose, R. A., & Bowen, N. K. (2021). The effect on high school drop-out of a middle school relevance intervention. *Journal of Educational Research*, *114*(6), 526-536. https://doi.org/10.1080/00220671.2021.1993123

Sánchez-Meca, J., Marín-Martínez, F., & Chacón-Moscoso, S. (2003). Effect-size indices for dichotomized outcomes in meta-analysis. *Psychological Methods, 8*(4), 448–467. https://doi.org/10.1037/1082-989X.8.4.448

Simon, O., Nylund-Gibson, K., Gottfried, M., & Mireles-Rios, R. (2020). Elementary absenteeism over time: A latent class growth analysis predicting fifth and eighth grade outcomes. *Learning and Individual Differences*, *78*, 101822. https://doi.org/10.1016/j.lindif.2020.101822

*Sinclair, M. F., & Christenson, S. L. (1998). Dropout Prevention for Youth with Disabilities: Efficacy of a Sustained School Engagement Procedure. *Exceptional Children, 65*(1), 7-21. https://doi.org/10.1177/001440299806500101

*Sinclair, M. F., Christenson, S. L., & Thurlow, M. L. (2005). Promoting School Completion of Urban Secondary Youth With Emotional or Behavioral Disabilities. *Exceptional Children*, *71*(4), 465-482. https://doi.org/10.1177/001440290507100405

Slavin, R., & Smith, D. (2009). The Relationship Between Sample Sizes and Effect Sizes in Systematic Reviews in Education. Educational Evaluation and Policy Analysis, 31(4), 500-506. https://doi.org/10.3102/0162373709352369

*Smolkowski, K., Seeley, J. R., Gau, J. M., Dishion, T. J., Stormshak, E. A., Moore, K. J., Falkenstein, C. A., Fosco, G. M., & Garbacz, S. A. (2017). Effectiveness evaluation of the Positive Family Support intervention: A three-tiered public health delivery model for middle schools. *Journal of School Psychology*, *62*, 103-125. https://doi.org/10.1016/j.jsp.2017.03.004

*Snyder, F., Flay, B., Vuchinich, S., Acock, A., Washburn, I., Beets, M., & Li, K.-K. (2010). Impact of a Social-Emotional and Character Development Program on School-Level Indicators of Academic Achievement, Absenteeism, and Disciplinary Outcomes: A Matched-Pair, Cluster-Randomized, Controlled Trial. *Journal of Research on Educational Effectiveness*, *3*(1), 26-55. https://doi.org/10.1080/19345740903353436

*Sommer, T. E., Sabol, T. J., Chase-Lansdale, P. L., Small, M., Wilde, H., Brown, S., & Huang, Z. Y. (2017). Promoting Parents' Social Capital to Increase Children's Attendance in Head Start: Evidence from an Experimental Intervention. *Journal of Research on Educational Effectiveness*, *10*(4), 732-766. https://doi.org/10.1080/19345747.2016.1258099

Stanley, T. D. (2017). Limitations of PET-PEESE and Other Meta-Analysis Methods. *Social Psychological and Personality Science*, *8*(5), 581-591. https://doi.org/10.1177/1948550617693062

Stanley, T. D., & Doucouliagos, H. (2010). Picture this: a simple graph that reveals much ado about research. *Journal of Economic Surveys*, *24*(1), 170-191. https://doi.org/10.1111/j.1467-6419.2009.00593.x

Stanley, T. D., & Doucouliagos, H. (2017). Neither fixed nor random: Weighted least squares meta-regression. *Research Synthesis Methods*, *8*(1), 19-42. https://doi.org/10.1002/jrsm.1211

Stanley, T. D., & Doucouliagos, H. (2024). Harnessing the power of excess statistical significance: Weighted and iterative least squares. *Psychological Methods*, *29*(2), 407-420. https://doi.org/10.1037/met0000502

Stanley, T. D., Doucouliagos, H., & Ioannidis, J. P. A. (2022). Beyond Random Effects: When Small-Study Findings Are More Heterogeneous. *Advances in Methods and Practices in Psychological Science*, *5*(4), 251524592211204. https://doi.org/10.1177/25152459221120427

*Stormshak, E. A., Connell, A., & Dishion, T. J. (2009). An adaptive approach to family-centered intervention in schools: Linking intervention engagement to academic outcomes in middle and high school. *Prevention Science*, *10*(3), 221-235. https://doi.org/10.1007/s11121-009-0131-3

*Sun, M., Penner, E. K., & Loeb, S. (2017). Resource- and Approach-Driven Multidimensional Change: Three-Year Effects of School Improvement Grants. *Grantee Submission*, *54*(4), 607-643. https://doi.org/10.3102/0002831217695790

Sutton, A. J., & Higgins, J. P. (2008). Recent developments in meta-analysis. *Statistics in medicine*, *27*(5), 625-650. https://doi.org/10.1002/sim.2934

Tanner-Smith, E. E., & Wilson, S. J. (2013). A meta-analysis of the effects of dropout prevention programs on school absenteeism. *Prevention Science*, *14*(5), 468-478. https://doi.org/10.1007/s11121-012-0330-1

*Tobias, A. K., & Myrick, R. D. (1999). A Peer Facilitator-Led Intervention with Middle School Problem-behavior Students. *Professional School Counseling*, *3*(1), 27-33. ERIC Document Number: EJ626040.

Viechtbauer, W., & Cheung, M.W. L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods, 1*(2), 112-125. https://doi.org/10.1002/jrsm.11

*Villares, E., Bowers, H., Brigman, G., & Bottini, C. (2024). The effects of Student Success Skills on attendance and emotion regulation. *Journal of Counseling & Development*, *102*(2), 163-174. https://doi.org/10.1002/jcad.12503

*Wallace, L. B., Hai, A. H., & Franklin, C. (2020). An evaluation of working on what works (WOWW): A solution-focused intervention for schools. *Journal of Marital and Family Therapy*, *46*(4), 687-700. https://doi.org/10.1111/jmft.12424

Wang, Q., Hsiao, Y.-Y., Hushman, C., & Armstrong, J. (2024). The effectiveness of dropout intervention programs among k-12 students: A meta-analysis. *Journal of Education for Students Placed at Risk*. https://doi.org/10.1080/10824669.2024.2342779

*Weinstein, J., Villares, E., & Brigman, G. (2021). The effect of the student success skills small group intervention on factors associated with dropout potential. *Journal for Specialists in Group Work*, *46*(3), 256-271. https://doi.org/10.1080/01933922.2021.1945175

Wilson, S. J., Tanner-Smith, E. E., Lipsey, M. W., Steinka-Fry, K., & Morrison, J. (2011). Dropout Prevention and Intervention Programs: Effects on School Completion and Dropout among School-aged Children and Youth. *Campbell Systematic Reviews*, *7*(1), 1-61. https://doi.org/10.4073/csr.2011.8

Table of contents

1. Search strategy 66
2. Study, intervention, and sample characteristics overview 68
3. Initial Coding, Categorization, and Descriptive Statistics Overview 74
4. Categorization level and duration 75
5. Descriptives of level and exogenous variation in quasi-experimental studies 76
6. Robustness checks WILS 77
7. Robustness checks mean samples restricted: decade (2010), country (US) and risk of bias 78
8. Flowchart 79
9. Correlation matrix 80
10. PET-PEESE 81

References 82

Search strategy

Theoretical variables:

1. Theoretical stimulus: Exposure to treatment - school based interventions to reduce absenteeism of pupils and students
2. Theoretical response: The (change in) school attendance of pupils and students

Operationalized variables:

1. Independent variable(s): the school based interventions or practices to prevent or reduce absenteeism. The interventions can vary from a distinct program or procedure to an individual approach and is school based: operated in a school building, by school staff or under school or system auspices. In the treatment we will distinguish different types of interventions..
2. Dependent variable(s): School absenteeism outcomes. The school attendance of pupils and students, measured by attendance, nonattendance, withdrawal from school, and eventually school dropout.

In addition to distinguishing different types and goals of intervention, we distinguish between different assumed mechanisms at work, related to the malleable causes of student absenteeism. Third, we will distinguish between different types of absenteeism, varying from skipping classes to chronic absenteeism and dropouts. .

Search strategy

Databases

We will perform a keyword search in the following electronic bibliographic databases: APA PsycInfo, APA PsycArticles, ERIC, Psychology and Behavioral Sciences Collection with search platform EBSCOHost

Keywords (title, abstract and keywords)

Our search includes a combination of the following keywords (i) (“absentee*” OR “absence” OR “drop*out" OR “truan*” OR “attendance” OR “school refusal” OR “school phobia”) AND (ii) (“intervention*” OR “prevention” OR “nudg*” OR “program*” OR (“policy” OR “policies”)) AND (iii) (“school” OR “education”) AND (iv) (“effect*” OR “impact*” OR “random*”)

Time range: 01-02-1990 –01-07-2024

Screening - Eligibility Criteria

To be considered eligible, studies have to meet the following screening criteria:

- Intervention
    - Type: it has to be a school based intervention: operated in a school building, by school staff or under school or system auspices. This excludes interventions solely by medical specialists, such as individual therapy. In our analysis, we will differentiate types of interventions based on their presumed mechanisms.

- Causes of absenteeism: the working mechanism of the intervention addresses a malleable cause of student absenteeism which is not exclusively related to a medical reason, like gastrointestinal and respiratory illnesses.
- Educational context: we consider all compulsory education settings: primary education, secondary education and vocational education.

- Study design:
  - Experimental
    - Interventional (RCT)
    - Quasi-experimental (RD, dif-in-dif, IV)
- Outcomes, order of preference:
  - Most precise absence/attendance rate (in days/hours) preferred over dropout data or meeting a threshold (chronic absenteeism)
  - Order of preference: objective measures over self reported measures
  - Measured at student level

Study, intervention, and sample characteristics overview

Table S1

Overview study, intervention and sample characteristics clustered by study

Study characteristics					**Intervention characteristics**							**Outcomes**				**Sample**
Author	Publication	Country	Design	Ran. level	Type	Length year		School	Population	Age from	To age	*n*	Type	*d*	SE(*d*)	Sample *N*
Anderson et al.	2019	USA	DiD	student	Teaching time	≤ 1	secondary school	No risk	School attendance history	12	18	1	Absent rate (days)	-0.001	0.005	178064
Awaworyi Churchill et al.	2021	Pakistan	RDD	county or household	Conditional cash	≤ 1	elementary and secondary school	No risk	Family socio economic background	5	14	2*	Dropout	0.035	0.028	5835
Axford et al.	2020	Wales	RCT	school	Schoolwide program	≤ 1	elementary school	No risk	No risk	7	11	1	Absenteeism rate per 100 half days	-0.013	0.099	3214
Bannink et al.	2014	The Netherlands	RCT	school	Mentor	≤ 1	secondary school	Risk	School attendance history	15	17	1	Hours non permitted absenteeism	0.285	0.100	418
Bavarian et al.	2013	USA	RCT	school	Schoolwide program	>1	elementary school	Risk	School attendance history	4	11	1	School absenteeism rate	-0.780	0.056	1170
Bohdan et al.	2023	USA	RCT	student	Coach	≤ 1	elementary school	Risk	School attendance history	5	7	1	Absence rate	-0.337	0.129	249
Borgen et al.	2021	Norway	DiD	school	Schoolwide program	≤ 1	elementary school	No risk	No risk	6	12	4*	Did not drop out before 21 year old	0.004	0.006	131750
Borgen et al.	2021	Norway	DiD	school	Schoolwide program	>1	elementary school	No risk	Emotional or behavioral problems	9	12	1	Drop out before 17 year old	0.002	0.016	15403
Borman et al.	2019	USA	RCT	student	Mentor	≤ 1	secondary school	No risk	No risk	11	13	1	number of absences	-0.130	0.055	1304
Brathwaite et al.	2023	Sub Sahara Afrika	RCT	school	Conditional cash	>1	secondary school	Risk	Family socio economic background	14	17	4*	Days absent	-0.087	0.071	790
Cabus and De Witte	2011	The Netherlands	DiD	student	Teaching time	≤ 1	secondary school	No risk	No risk	16	17	1	Drop out share cohort	-0.041	0.018	12784

Study characteristics					**Intervention characteristics**								**Outcomes**			**Sample**
Author	Publication	Country	Design	Ran. level	Type	Length year		School	Population	Age from	To age	*n*	Type	*d*	SE(*d*)	Sample *N*
Cho et al.	2017	Kenya	RCT	school	Conditional cash	>1	secondary school	Risk	Family socio economic background	11	20	1	Stayed in school	-0.522	0.103	767
Converse et al.	2009	USA	RCT	student	Mentor	≤ 1	secondary school	Risk	Emotional or behavioral problems	13	15	1	Absent days	-0.233	0.354	31
Corcoran	2016	USA	DiD	school	Breakfast	≤ 1	elementary school	Risk	Family socio economic background	5	10	2*	Attendance rate	-0.005	0.004	310423
Dee	2011	USA	RCT	county or household	Conditional cash	>1	secondary school	No risk	Family socio economic background	13	19	1	Unexcused absences rate	-0.061	0.029	4697
Desocio et al.	2007	USA	RCT	student	Mentor	≤ 1	secondary school	Risk	School attendance history	13	16	1	Most days absent in any class ( )	-0.274	0.217	92
Durkin et al.	2022	USA	RCT	student	Teaching time	≤ 1	elementary school	No risk	Family socio economic background	4	5	6*	School attendance rate	0.061	0.030	4570
Felver et al.	2018	USA	RCT	classroom	Schoolwide program	≤ 1	secondary school	Risk	School attendance history	12	17	1	Attendance	-0.390	0.427	23
Flannery et al.	2020	USA	RCT	school	Package	≤ 1	secondary school	Risk	No risk	14	14	1	Days Attendance prevalence	-0.130	0.056	1319
Flannery et al.	2024	USA	RCT	student	Package	≤ 1	secondary school	Risk	Family socio economic background	14	15	1	Absences	-0.163	0.529	14
Gottfredson et al.	2010	USA	RCT	student	Teaching time	≤ 1	secondary school	Risk	Family socio economic background	11	13	1	School attendance rate	0.050	0.098	410
Graeff Martins et al	2006	Brazil	RCT	school	Package	≤ 1	elementary and	Risk	No risk	6	14	2*	Days absent	-0.640	0.214	80

Study characteristics					Intervention characteristics								Outcomes			Sample
Author	Publication	Country	Design	Ran. level	Type	Length year		School	Population	Age from	To age	*n*	Type	*d*	SE(*d*)	Sample *N*
							secondary school									
Grant et al.	2023	USA	RCT	school	Schoolwide program	>1	secondary school	Risk	School attendance history	12	16	1	Missing 10% or more	-0.123	0.027	1653
Guryan et al.	2021	USA	RCT	school	Package	>1	elementary school	Risk	School attendance history	6	10	4*	Days absent	-0.086	0.044	2372
Heppen et al.	2018	USA	RCT	student	Mentor	>1	secondary school	No risk	School attendance history	14	18	1	Dropped out before 14year old	0.119	0.128	552
James et al.	2023	USA	DiD	student	Teaching time	≤ 1	secondary school	No risk	No risk	14	15	2*	Absences	-0.341	0.043	2153
Johnsen et al.	2024	Denmark	RCT	student	Family intervention	≤ 1	elementary and secondary school	Risk	Emotional or behavioral problems	6	16	2*	Attendance rate	-0.187	0.168	144
Jones and Christian	2021	USA	RCT	student	Package	>1	elementary school	Risk	Family socio economic background	3	4	2*	Absences (day)	-0.271	0.175	128
Jones et al.	2002	USA	RCT	student	Package	>1	secondary school	No risk	Family socio economic background	16	18	1	Meet attendance requirement	-0.156	0.043	2486
Kim and Joo	2011	USA	DiD	student	Conditional cash	≤ 1	secondary school	No risk	Family socio economic background	15	18	2*	Percentage school attendance	-0.011	0.011	36951
Kirksey and Gottfried	2021	USA	RDD	school	Breakfast	≤ 1	elementary and secondary school	Risk	No risk	4	18	1	Chronic absenteeism rate	-0.249	0.125	258
Komisarow et al.	2023	USA	DiD	school	Coach	≤ 1	elementary school	Risk	Family socio economic background	5	12	1	Absence rate	-0.039	0.030	4421
Lasy Fink et al.	2021	USA	RCT	student	Family intervention	≤ 1	elementary and secondary school	Risk	School attendance history	7	16	1	Absences (day)	-0.152	0.055	82163

Study characteristics					Intervention characteristics							Outcomes				Sample
Author	Publication	Country	Design	Ran. level	Type	Length year		School	Population	Age from	To age	*n*	Type	*d*	SE(*d*)	Sample *N*
Lee et al.	2020	USA	DiD	student	Family intervention	≤ 1	elementary school	No risk	School attendance history	7	8	4*	Attendance rate	0.015	0.026	6097
Lipsey et al.	2018	USA	RCT	student	Teaching time	≤ 1	elementary school	No risk	Family socio economic background	4	5	3*	School attendance rate	-0.001	0.038	2990
Mac Iver	2011	USA	RCT	school	Mentor	>1	secondary school	Risk	School attendance history	14	15	1	Dropout	-0.263	0.166	225
Mac Iver et al.	2019	USA	RCT	school	Coach	>1	secondary school	No risk	No risk	14	15	2*	Attendance rate	-0.014	0.022	7985
Mac Iver et al.	2022	USA	RDD	student	Family intervention	≤ 1	elementary school	No risk	School attendance history	5	10	3*	Attendance rate	-0.058	0.056	1404
Marvul	2022	USA	RCT	student	Schoolwide program	≤ 1	secondary school	Risk	Family socio economic background	14	18	1	Absences	-3.850	0.228	40
Maynard et al.	2014	USA	RCT	student	Package	≤ 1	secondary school	Risk	Emotional or behavioral problems	11	15	1	Number of days missed	-0.010	0.146	189
Mazerolle et al.	2017	Australia	RCT	student	Package	≤ 1	elementary and secondary school	Risk	School attendance history	10	16	2*	Absence rate	-0.041	0.199	100
Mazrekaj and De Witte	2019	Belgium	DiD	school	Teaching time	>1	secondary school	No risk	No risk	12	17	1	Did not dropout before 17 year old	-0.023	0.006	95850
McCarthy et al.	2018	USA	RCT	student	Family intervention	≤ 1	secondary school	No risk	Emotional or behavioral problems	12	15	5*	Absences	-0.076	0.146	186
McEwan	2013	Chile	RDD	school	Breakfast	>1	elementary and secondary school	Risk	Family socio economic background	0	19	1	Attendance rate	0.016	0.044	2024
Mcneely et al.	2019	USA	DiD	student	Package	≤ 1	secondary school	No risk	School attendance history	12	13	4*	Daily attendance rate	0.033	0.021	8874
Mhurchu et al.	2013	New Zeeland	RCT	school	Breakfast	≤ 1	elementary and	No risk	Family socio	5	13	1	Attendance rate <95%	-0.122	0.068	1478

Study characteristics					Intervention characteristics								Outcomes			Sample
Author	Publication	Country	Design	Ran. level	Type	Length year		School	Population	Age from	To age	*n*	Type	*d*	SE(*d*)	Sample *N*
							secondary school		economic background							
Modestina et al.	2023	USA	RDD	student	Teaching time	≤ 1	secondary school	Risk	Family socio economic background	13	17	4	Attendance rate	-0.094	0.039	2645
Newton et al.	2014	Australia	RCT	school	Teaching time	≤ 1	secondary school	No risk	Emotional or behavioral problems	13	14	3	Truancy scores/rates	-0.161	0.085	560
Powers et al.	2017	USA	RCT	student	Mentor	>1	elementary school	Risk	School attendance history	11	12	1	Days absent	-0.671	0.287	45
Reid and Bailey Dempsey	1995	USA	RCT	student	Conditional cash	≤ 1	secondary school	No risk	School attendance history	11	14	2	Absences	-0.131	0.230	75
Robinson et al.	2018	USA	RCT	student	Family intervention	≤ 1	elementary school	No risk	School attendance history	3	5	1	number of absences	-0.070	0.019	10966
Rose and Bowen	2021	USA	RCT	school	Teaching time	>1	elementary school	No risk	School attendance history	11	12	1	Dropout before 17 year old	-0.204	0.094	3413
Rosário et al.	2017	Portugal	RCT	student	Family intervention	>1	elementary school	Risk	Family socio economic background	6	11	4	non attendance	-0.723	0.349	30
Sinclair and Christenson	1998	USA	RCT	student	Mentor	>1	secondary school	No risk	Emotional or behavioral problems	12	15	1	Never dropped out	-0.703	0.309	94
Sinclair et al.	2005	USA	RCT	student	Package	>1	secondary school	Risk	Emotional or behavioral problems	14	15	1	dropped out	-0.444	0.206	144
Smolkowski et al.	2017	USA	RCT	school	Family intervention	>1	secondary school	No risk	No risk	12	15	1	Days absent	0.031	0.024	6827
Snyder et al.	2009	Hawaii	RCT	school	Schoolwide program	>1	elementary school	Risk	School attendance history	5	11	2	Average days absent	-0.602	0.304	40
Sommer et al.	2017	USA	RCT	student	Family intervention	≤ 1	elementary school	No risk	No risk	3	4	4	attendance rate (season)	-0.022	0.202	204

Study characteristics					**Intervention characteristics**								**Outcomes**			**Sample**
Author	Publication	Country	Design	Ran. level	Type	Length year		School	Population	Age from	To age	n	Type	d	SE(d)	Sample N
Stormshak et al.	2009	USA	RCT	student	Family intervention	>1	secondary school	No risk	No risk	11	17	5	Number of absences	0.026	0.106	514
Sun et al.	2017	USA	DiD	school	Schoolwide program	>1	elementary and secondary school	Risk	School attendance history	5	16	3	Average unexcused absences	-0.127	0.064	41571
Tobias and Myrick	1999	USA	RCT	student	Peer	≤ 1	elementary school	No risk	Emotional or behavioral problems	11	12	1	Number of absences	-0.080	0.162	150
Villares et al.	2023	USA	RCT	school	Schoolwide program	≤ 1	elementary school	Risk	Emotional or behavioral problems	10	11	2	Absences	-0.214	0.079	639
Vitaro et al.	1999	Canada	RCT	student	Package	>1	elementary school	No risk	Emotional or behavioral problems	7	9	1	Drop out before 17 year old	-0.465	0.104	149
Wallace et al.	2020	USA	RCT	classroom	Coach	≤ 1	elementary school	No risk	No risk	9	11	1	Days absent	-0.130	0.098	413
Weinstein et al.	2021	USA	RCT	student	Coach	≤ 1	secondary school	No risk	School attendance history	14	15	2	Days absent	-0.248	0.146	183

Note. Ran. level = randomization level; d = Cohen's d. RCT = randomized controlled trial; DiD = difference-in-differences; RDD = regression discontinuity design. "In our main analyses, DiD and RDD were categorized under Quasi Experimental (QE). Table S4 shows specific characteristics of the QE studies. The notation n indicates the number of outcomes per study, which may result from measuring outcomes at different time points. In this table, we averaged the statistics and characteristics where applicable. In the study, we used clustered standard errors at the study level to account for correlations between effect sizes within a single study. Some categories in this table are logical constructs derived from several coded categories, such as Type, length (year), and outcome type. Table S2 and S3 offer further details about how these constructs were developed.

Initial Coding, Categorization, and Descriptive Statistics Overview

Table S2

Categorization type of intervention based on program type

Initial coding		**Categorization**	
Program type	Frequency (%)	Intervention type	Frequency (%)
Breakfast	5 (4.2%)	Aid programs	14 (11.76%)
Conditional cash	9 (7.56%)		
Package	22 (18.49%)	Package	22 (18.49%)
Schoolwide program	17 (14.29%)	School Curriculum and Learning Environments	38 (31.93%)
Teaching time	21 (17.65%)		
Coach	6 (5.04%)	Support network	45 (37.82%)
Mentor	8 (6.72%)		
Peer	1 (0.84%)		
Family intervention	30 (25.21%)		

Categorization level and duration

Table S3

Categorization level and duration category

	Initial coding level		Categorization		Initial coding duration	Categorization	
Intervention type	Level	Frequency (%)	Categorization	Frequency (%)	Months *M*(*SD*)	Duration category (years)	Frequency (%)
Aid programs	student	14 (73.68%)	student	9 (64.29%)	28.0 (19.5)	≤ 1	9(64.3%)
	teacher	0 (0%)	two levels	5 (35.71%)		>1	5(35.7%)
	parents	3 (15.79%)					
	classroom	2 (10.53%)					
Package	student	22 (39.29%)	student	1 (4.55%)	21.5 (14.2)	≤ 1	11(50.0%)
	teacher	11 (19.64%)	two levels	10 (45.45%)		>1	11(50.0%)
	parents	21 (37.5%)	three or four levels	11 (50%)			
	classroom	2 (3.57%)					
School Curriculum and					21.5 (14.2)		
Learning Environments	student	37 (42.05%)	student	11 (28.95%)		≤ 1	28(73.7%)
	teacher	22 (25%)	other single levels	1 (2.63%)		>1	10(26.3%)
	parents	4 (4.55%)	two levels	5 (13.16%)			
	classroom	25 (28.41%)	three or four levels	21 (55.26%)			
					18.1 (14.5)		
Support networks	student	34 (44.74%)	student	11 (24.44%)		≤ 1	30(66.7%)
	teacher	7 (9.21%)	other single levels	10 (22.22%)		>1	15(33.3%)
	parents	30 (39.47%)	two levels	17 (37.78%)			
	classroom	5 (6.58%)	three or four levels	7 (15.56%)			

Note. *M* = mean, *SD* = standard deviation. Percentages following the frequencies represent the proportion of the (constructed) category within each intervention type.

Descriptives of level and exogenous variation in quasi-experimental studies

Table S4

Descriptives of level and exogenous variation in quasi-experimental studies

	Difference in Differences	Randomized Controlled Trial	Regression discontinuity design	Overall
Level				
Student	4 (15.4%)	26 (31%)	2 (22.2%)	32 (26.9%)
Other Single Levels	5 (19.2%)	6 (7.1%)	0 (0%)	11 (9.2%)
Two Levels	5 (19.2%)	27 (32.1%)	5 (55.6%)	37 (31.1%)
Three Or Four Levels	12 (46.2%)	25 (29.8%)	2 (22.2%)	39 (32.8%)
Exogenous variation				
Implementation Timing	11 (42.3%)			11 (31.4%)
Policy Change	2 (7.7%)		1 (11.1%)	3 (8.6%)
Policy Implementation Differences	8 (30.8%)			8 (22.9%)
Policy Reform	5 (19.2%)			5 (14.3%)
Eligibility Criteria			6 (66.7%)	6 (17.1%)
Lottery			2 (22.2%)	2 (5.7%)

F

Robustness checks WILS

Table S5

Robustness checks WILS procedure

WILS	**WILS procedure [BS]**			**Estimates**		**Bootstrap estimates**	
Sample	it	*k*	*n*	ES (SE)	95% CI	ES (SE)	95% CI
Original procedure							
Total	4 [4]	46 [46]	85 [85]	-0.000 (0.003)	[-0.006, 0.006]	-0.000 (0.002)	[-0.005, 0.004]
No outliers	8 [8]	35 [35]	57 [57]	-0.090*** (0.009)	[-0.136, -0.094]	-0.093*** (0.015)	[-0.122, -0.064]
RCT only	6 [6]	35 [35]	58 [57]	-0.007 (0.023)	[-0.053, 0.039]	-0.007 (0.020)	[-0.051, 0.029]
Sensitivity check							
Total	3[3]	43 [43]	81 [81]	0.001 (0.003)	[-0.004, 0.006]	0.001 (0.002)	[-0.003, 0.006]
No outliers	7 [7]	29 [29]	44 [44]	-0.056*** (0.009)	[-0.073, -0.039]	-0.057*** (0.009)	[-0.074, -0.041]
RCT only	3 [3]	29 [29]	55 [55]	0.018 (0.019)	[-0.019, 0.054]	0.017 (0.011)	[-0.005, 0.039]

Note. ES = effect size Cohen's *d*, *it* = the number of iterations, *k* = the number of studies, *n* = the number of outcomes. The numbers in square brackets next to each statistic indicate the frequency of bootstrapping for that specific statistic. The original procedure follows the procedure as described by Stanley and Doucouliagos (2024), removing studies based on the relative largest residual (first iteration) and ESS in the subsequent iterations. In the sensitivity check, negative residuals (first iteration) or negative ESS values (subsequent iterations) are set to zero before proceeding with further calculations. Standard errors (SE) are reported in parentheses.

*p <.05; **p <.01; ***p < .001

Robustness checks mean samples restricted: decade (2010), country (North America) and risk of bias

Table S6

Robustness Checks Mean Effect Restricted samples

Model	ES	95% CI
Sample restricted to North America, k = 47, n = 86		
RE	-0.214	[-0.364, -0.064]
UWLS	-0.008	[-0.008, -0.000]
PET	0.005	[-0.004, 0.014]
PEESE	-0.006	[-0.012, 0.001]
Sample restricted to decade 2010s, k = 29, n = 51		
RE	-0.101*	[-0.182, -0.019]
UWLS	-0.001*	[-0.014, -0.001]
PET	0.000	[-0.005, 0.006]
PEESE	-0.007*	[-0.013, -0.013]
Sample - excluded risk of bias, k = 62, n = 112		
RE	-0.207***	[-0.322, -0.092]
UWLS	-0.005	[-0.013, 0.002]
PET	0.006	[-0.002, 0.015]
PEESE	-0.003	0.010, -.008

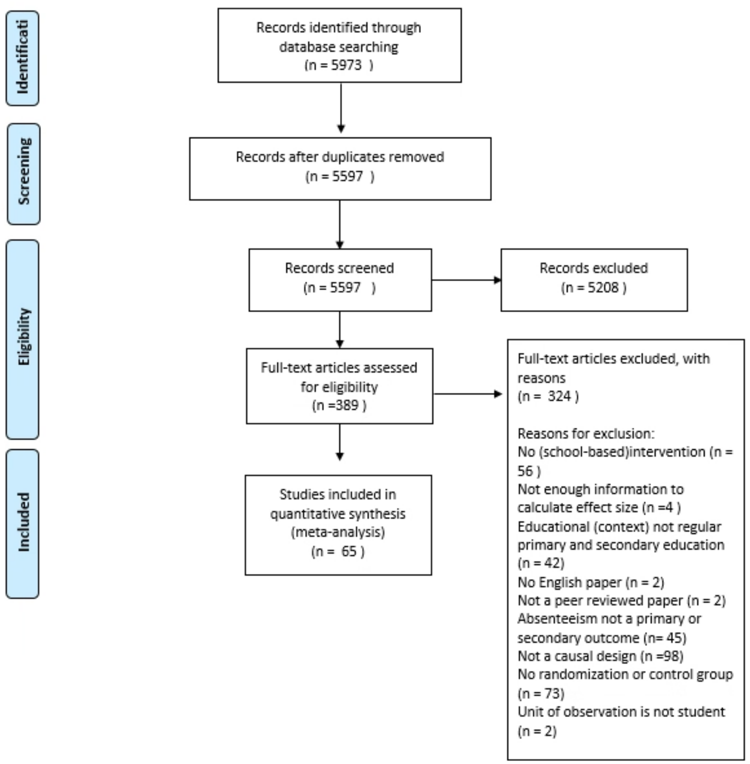
Note: ES = effect size Cohen's *d*, BS = bootstrapped, *k* = number of studies *n* = number of outcomes. Outcomes are clustered at the study level to account for potential correlations between multiple outcomes from the same study.

**p* <.05; ***p <.01;* ****p* < .001

Flowchart

Figure S1

Flowchart of Proces Selecting Studies in the Meta-analysis



Correlation matrix

Figure S2
Correlation Matrix of Characteristics

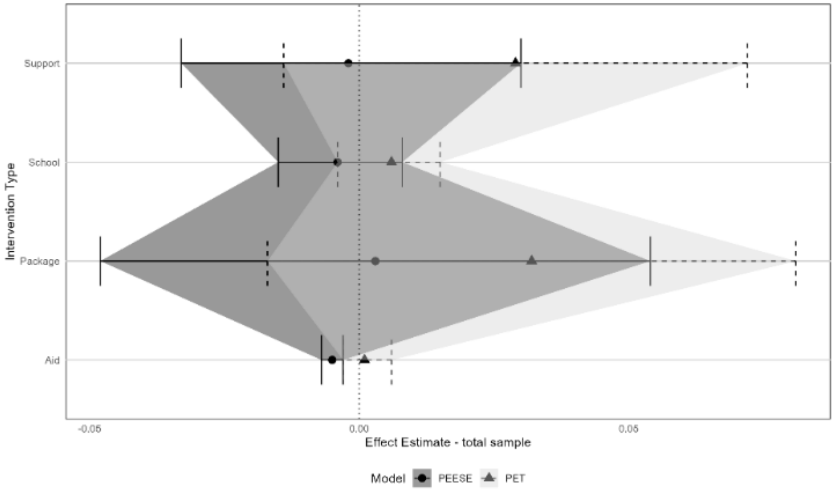
	Outcome	Duration	Intervention type	Level	Target schools	Target population	Context	Units of randomization
Outcome	**1.00**							
Duration	0.33	**1.00**						
Intervention type	0.16	0.08	**1.00**					
Level	0.09	0.08	0.36	**1.00**				
Target schools	0.19	0.27	0.28	0.28	**1.00**			
Target population	0.17	0.16	0.31	0.29	0.34	**1.00**		
Context	0.11	0.14	0.26	0.13	0.27	0.12	**1.00**	
Units of randomization	0.06	-0.01	0.17	0.15	0.08	0.19	0.15	**1.00**

Note. We used various measures appropriate for different variable types (Pearson, 1896; Cramér, 1946; Cohen, 1988)

PET-PEESE

Figure S3

Visual Comparison of PET and PEESE



Note. Points represent the estimated effects, error bars indicate confidence intervals, and shaded areas highlight the range of uncertainty for each model. The dashed line at zero marks the threshold for statistical significance.

References

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

Cramér, H. (1946). Mathematical Methods of Statistics. Princeton University Press. https://doi.org/10.1515/9781400883868

Pearson, K. (1896). Mathematical contributions to the theory of evolution. III. Regression, heredity, and panmixia. Philosophical Transactions of the Royal Society of London. Series A, 187, 253-318. https://doi.org/10.1098/rsta.1896.0007

Stanley, T. D., & Doucouliagos, H. (2024). Harnessing the power of excess statistical significance: Weighted and iterative least squares. *Psychological Methods*, *29*(2), 407-420. https://doi.org/10.1037/met0000502.