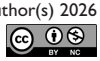
Check for updates

Manuscript

Review of Educational Research
1–35
© The Author(s) 2026

Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.3102/00346543261495960
journals.sagepub.com/home/rer


A Meta-Analysis of Outcomes for Steiner/Waldorf and Mainstream Education

Sebastian P. Suggate[1,2], Bernhard Schmalenbach[3], and Jürgen Peters[3]

Abstract

Despite being over 100 years old, empirical research on Waldorf (or Steiner) education has been scarce, with no published meta-analyses comparing outcomes to mainstream education. We identified 30 studies comparing outcomes for Waldorf ($N = 7{,}823$) and mainstream ($N = 187{,}015$) students, netting 40 treatment-control comparisons and 506 effect sizes, from 11 countries. Mixed-effects meta-regressions investigated the role of measurement, publication, sample features, and outcome (i.e., academic achievement, creativity, attitudes and emotions, school and classroom climate, and moral development). Overall, Waldorf students performed significantly better ($d = .216$, after trim-and-fill, $d = .369$). Waldorf students had significantly greater academic achievement and reported a significantly better school and classroom climate, with differences in aesthetic creativity, attitudes and emotions, and moral development not being statistically significant. Given the comparative designs employed, further research using quasi-experimental and longitudinal designs is needed, both on the domains investigated here and further domains.

Keywords

achievement, moral education/development, educational reform, classroom research, attitude, meta-analysis, quasi-experimental analysis, Waldorf education, Steiner education, academic skills, creativity, classroom climate, moral development, aesthetics, drawing

[1]University of Regensburg, Galgenberg, Germany
[2]University of Otago, Dunedin, New Zealand
[3]Alanus University, Alfter, Germany

Corresponding Author:
Sebastian P. Suggate, University of Regensburg, Universitätsstr. 31, 93040, Germany.
Email: sebastian.suggate@otago.ac.nz

Introduction

Since its inception in 1919, Waldorf or Rudolf Steiner (hereafter, Waldorf) education has spread across the globe to comprise 1,092 schools in 64 countries (Waldorf World List, 2025). The educational approach itself has been praised as visionary, innovative, and child-centred (Dahlin, 2017; Edwards, 2007; Lim, 2004), supporting social and aesthetic development (Schieren, 2023; Ullrich, 2024). In contrast, it has been argued that Waldorf has not been sufficiently subjected to critical, methodological, and empirical research; is for privileged families; or is based on outdated educational approaches (Buck & Hoffmann, 2024; Ullrich, 2024).

Until recent decades, there was little systematic empirical research on Waldorf education (Guerrero et al., 2024; Tyson, 2024), particularly for studies comparing children attending Waldorf and state schools. Yet, additional research would enable parents to make informed choices about where to send their children. Further, comparative and nuanced research on Waldorf education might promote greater cross-pollination of ideas and practices across different approaches, thereby enriching education for all children. Therefore, we conduct the first, to our knowledge, quantitative meta-analysis of this research, focusing on Waldorf versus mainstream student performance on measures of (a) academic skill, (b) creativity, (c) attitudes and behaviour, (d) school and classroom climate, and (e) moral development.

Waldorf Education

Waldorf education was founded in Germany in the social ruin resulting from the First World War as an attempt to provide a more well-rounded, co-educational, and cosmopolitan education (Rawson, 2021). Waldorf schools were opened in other European countries and the US, but the educational movement suffered during the Second World War, being banned by the Nazi, and later in East Germany, by the Communist governments. In the second half of the 20th century, schools were opened across the globe, with parents and teachers purportedly seeking a schooling that favoured a broader education over what was often seen as a narrow focus on academic skills in many state curriculum schools. Waldorf and mainstream schools differ in organisational, curriculum, assessment, social, and pedagogical and developmental conceptions, which will be briefly described.

In the early years (age birth to six or seven years), Waldorf education generally focuses on nurturing sensory development, imagination, and language (Waite & Rees, 2011). This is supposed to occur through free play, oracy (e.g., storytelling), music, practical activities (e.g., baking, weaving), and a focus on real-world aesthetic and sensory experiences to the exclusion of virtual worlds (e.g., electronic media). Explicit academic skills, such as learning the letters of the alphabet, reading, writing, or mathematical operations, are not deliberately taught (Cunningham & Carroll, 2011), with the approach instead on children laying sensorimotor and experiential foundations of cognition through play, language, and sensorimotor interactions (Grunelius, 1991; Nicol, 2015; Suggate & Suggate, 2019).

In the primary school years (ages 7 to 14), children are initially taught through methods that have a strong aesthetic focus (Lim, 2004), again activating the senses,

imagination, and preparing the ground for the unfolding of critical thinking through grounded conceptual development (Suggate & Suggate, 2019). Thus, the curriculum is delivered with the aid of music and song; poems, drama, role play, legends, myths, sagas, and stories; drawing, painting, and handicraft, along with the introduction of language, mathematics, and scientific content (Rawson & Bransby, 2025). Somewhat uniquely, foreign languages are often taught from the first class onwards (Rawson, 2022).

Academically, textbooks are avoided as children, under teacher guidance, produce their own workbooks. Accordingly, there is a strong focus on children being active creators of their learning experiences, although less through open instruction and self-directed learning typical in strongly constructivist approaches such as Montessori education. Socially, children usually remain with the same classes, retaining the same main lesson teacher during this entire seven- or eight-year period (i.e., looping; McCown & Sherman, 2002). Further, the main lessons are conducted in epochs (i.e., three weeks of maths, followed by language or science), which is intended to promote deeper learning and consolidation (see Becker & Wiehl, 2024). To increase intercultural awareness and understanding, second and third languages are often taught in this period. In contrast to most state schools, grades and high-stakes testing are avoided, where possible.

From age 13 or 14 years, the Waldorf curriculum focuses on developing autonomy and critical thinking (Sommer, 2010). Although many subjects available in state schools are now taught with a similar intensity, key methodological-didactic differences remain. Specifically, teaching aims to promote an immediate experience of learning, through phenomenological observation, experimentation, and description, leading in a final step to the theoretical derivation of laws and principles (Schumacher, 2020; Sigler, 2020; Sommer, 2020). This may be contrasted with more classical methods that begin with the theory, proceeding to hypothesis generation and testing (Sommer, 2020). Outside of the main subjects, there is still a strong focus on personal and social development through theatre, musical, and artistic performances, along with class trips and outdoor experiences (Rawson, 2021). Many schools also prepare students for university entrance examinations toward the end of secondary school.

Funding for Waldorf education varies internationally and nationally (Mepham, 2024), with schools in some countries being fully private (e.g., Austria, Switzerland, the United Kingdom), whereas others are largely state-funded (e.g., Germany, Australia, Czech Republic, New Zealand). Research indicates that parents are not typical "private" school clientele, but tended to be left-leaning and environmentally minded (Dahlin, 2007, p. 19), a little better educated without having greater incomes (Dahlin, 2007; Koolmann et al., 2018; Suggate et al., 2013), well-informed about educational options available to them (Dahlin, 2007), but often showing minimal demographic differences to their state-school counterparts (Koolmann et al., 2018; Suggate et al., 2013).

Mainstream (State) Education

To evaluate student outcomes for Waldorf students, a comparison group is required, which ideally constitutes the dominant educational curriculum experienced by most

students. The problem is that there is a large variance in what this dominant, mainstream, or conventional curriculum is called and what it contains. Further, the term *state school* is also inadequate because in many countries Waldorf schools also receive state funding and are bound in some ways to state curricula, usually retaining their special character to teach according to a Waldorf curriculum. Accordingly, we adopt the term *mainstream schools* in the current study to refer to those schools that are funded by the government and follow the national curriculum in that given country. Despite vast international differences in education, there are some common features for mainstream schools. Particularly in the English-speaking world, these typically have an early focus on academic skills (Marcon, 2002; Suggate, 2009) and regularly assign school grades and administer regular tests (OECD, 2023a). Further, many mainstream schools have a comparatively lesser emphasis on aesthetic and practical subjects (Rawson, 2021) and include electronic media in the early years and primary school (OECD, 2023b).

Possible Outcome Differences Between State and Waldorf Pupils

Our previous knowledge of the literature, combined with an initial scoping review, led us to focus initially on five key areas for which there appeared to be previous studies—namely: (a) academic and cognitive skills, (b) aesthetic creativity, (c) attitudes to learning and emotions, (d) school and classroom climate, and (e) moral development.

Academic and Cognitive Skills. We define academic skills as learning and performance in a subject that is typically instructed in school contexts, such as reading, mathematics, or science. As mentioned, Waldorf education shows marked difference in the timing and method of teaching academic skills and is characterised by minimal testing and formal assessment before grade nine (Peters & Randoll, 2018). Specifically, Waldorf curricula avoid explicit academic skill instruction (e.g., decoding, counting, numerical operations) in the early years. These explicit academic skills are gradually introduced in primary school via a soft approach, rich in music, art, and sensorimotor experiences (Lim, 2004; Suggate et al., 2013). In contrast to state curricula, scientific analytical reasoning in a number of subjects is taught with mainstream intensity in the secondary school, from around age 13 or 14 years. Although a meta-analysis is lacking, previous research has indicated the concept of equifinality, that both Waldorf and state curricula lead to similar academic outcomes despite large initial earlier differences at the beginning of primary school (Cunningham & Carroll, 2011; Suggate et al., 2013; Telfer-Radzat, 2024).

Aesthetic Creativity. Creativity is a complex multidimensional construct that has been defined and measured in a number of different ways (Kupers et al., 2019). We adopt a pragmatic task-oriented definition that views creativity as the ability to generate novel forms, sequences, or ideas, including artistic drawing, written compositions, or, as operationalised on tests of creativity, symbol sequences (Gao & Hall, 2024; Kirkham & Kidd, 2017; LeBoutillier & Marks, 2003). Thus, creativity can encompass artistic,

musical, literary, and problem-solving skills. The Waldorf curriculum is known for its focus on imagination, music, and art, from the early years through to secondary school (Dahlin, 2017; Lim, 2004). Instead of teaching from textbooks, children produce their own books under guidance from the teachers, which extends from simple forms in primary school to detailed anatomical drawings in secondary school (Becker & Wiehl, 2024). Additionally, there are regular class and school productions, reflecting a rich engagement in theatre and music (Rawson & Bransby, 2025). In contrast, it might be argued that Waldorf schools, particularly in primary school, place too much emphasis on formulaically reproducing artistic forms that the teachers draw or write on the blackboard, or through the predominance of teacher-directed instruction. Accordingly, an evaluation of aesthetic creativity outcomes for Waldorf versus state schools would appear insightful.

Student Attitudes and Emotions. Next, we turn to consider student attitudes and emotions that reflect positive beliefs, emotions, and engagement toward school; academic self-concept; and reduced stress and burnout. Research finds that academic attitudes (Mao et al., 2021), positive emotions (Camacho-Morles et al., 2021), and intrinsic motivation predict school achievement (Howard et al., 2021). Further, self-worth relates to perseverance in academic settings (Howard et al., 2021). Accordingly, there is good reason to hypothesise that aspects of educational approaches might relate to student attitudes and emotions.

Waldorf schools avoid awarding grades, preferring instead to provide detailed feedback on learning itself (Randoll & Peters, 2015; Rawson & Bransby, 2025). This lack of external academic pressure and evaluation is thought to heighten intrinsic motivation, improve attitudes to learning, and lead to more of a growth mindset. Additionally, attempting to educate through movement, aesthetic experiences, a rounded curriculum, and active engagement might work positively on attitudes and emotional well-being. Further, the familiarity with the teacher due to the generally smaller school sizes, looping, and the strong class community is thought to lead to more positive and emotionally supportive environments.

On the other hand, Waldorf schools can be seen as a "last resort" school by some parents, who enrol their children because they have had difficulties in other schools. In turn, often faced with low student numbers and a precarious financial position, along with the educational ethos to try to help every child, some teachers report an overrepresentation of challenging behaviours at Waldorf schools (Graudenz et al., 2013), reflecting worse attitudes and lower emotional well-being.

School and Classroom Climate. School climate has been referred to as the personality of the school, reflecting teacher support, leadership and governance, rules and discipline, participation, and educational goals (Berkowitz et al., 2017). By a positive school and classroom environment, we refer to teachers shaping positive learning experiences, encouraging active learning and participation, tailoring instruction to individual student needs, constructive teacher feedback, and prosocial helping behaviour from staff and students (Doll & Dooley, 2014). School climate can be conceived of as relating to

academic, community, safety, and institutional aspects, with consequences for academic and behavioural outcomes (Wang & Degol, 2016).

Many of the features of Waldorf education are designed to produce a positive school and classroom environment. Waldorf schools are typically much smaller than state schools, which can create a sense of community where each person can be more directly perceived and feel a sense of belonging. This is further heightened by the principle of looping (McCown & Sherman, 2002), whereby children in the primary school ideally remain in the same class with the same main lesson teacher for up to eight years. Children almost universally progress from one year to the next without being held back due to academic achievement. Further, schools often engage in learning activities that could be described as team building, including class excursions, cooperative games, theatre, projects, and role play.

Moral Development. Perceived moral behaviour, attitudes, and beliefs refer to ethical principles that benefit others, especially in the absence of immediate personal gain, self-interest, or negative impacts on other people (Killen & Malti, 2015; Molnar-Szakacs, 2011). State schools in many countries include subjects that specifically address ethical dilemmas through practising democratic education (Sant, 2019), anti-bullying programmes, project work, and prosocial games, either in the context of social studies, ethics education, or arts and literature curricula (OECD, 2021). In the lower classes in Waldorf schools, the approach to moral issues is more practised than discussed as a separate subject, through the class community and implicitly through fairy tales, myths, sagas, role play, and storytelling (Armon, 1997; Dahlin, 2010). Waldorf education has been criticised by former pupils for not including enough explicit teaching on political issues, which might reflect the implicit approach to moral education (Randoll & Peters, 2021). Thus, research is needed synthesising how Waldorf and mainstream students differ in terms of moral development.

Previous Studies

Although there have been many individual studies on Waldorf education, the vast majority of these employ qualitative or survey designs without a control group. Few studies have attempted a systematic review. In one report, Tyson (2024) surveyed research studies on Waldorf education; however, he did not provide a systematic quantitative analysis, focusing instead on scoping available literature. In another review, Guerrero et al. (2024) conducted a review of alternative schools compared to conventional education. They looked broadly at sociocognitive development, identifying 24 studies, with only eight looking at Waldorf education. Their findings were synthesised narratively and indicated mixed findings on measures of creativity, reading achievement, extrinsic motivation, and drawing. In short, reviews provide little insight into how the outcomes of Waldorf versus state-school children differ, being hampered by a lack of quantitative synthesis and a small number of identified studies. However, we had reason to believe based on our own knowledge of the literature that there were additional studies not identified by Guerrero et al. (2024) that would be amenable to a quantitative synthesis.

The Current Study

Waldorf education can be described as a small but significant international movement, with key differences in educational approaches, curriculum, and organisation, that could be expected to result in different student outcomes. Empirical research on Waldorf education is widely spread, with studies seeming to mostly come from European and Anglo-Saxon countries, using a diverse range of outcome variables and pursuing different research aims. However, the lack of a definitive meta-analysis makes it difficult to understand what the outcomes associated with Waldorf education are, making it difficult for parents, teachers, and educational policymakers to provide informed recommendations. Therefore, the purpose of the current meta-analysis is to quantitatively synthesise available evidence, thereby providing the first meta-analysis on student outcomes in Waldorf education. Importantly, although there are many descriptive and qualitative studies on Waldorf education, an investigation of student outcomes is ideally provided by studies with a mainstream comparison group.

Accordingly, we conducted a broad literature research and identified empirical studies comparing Waldorf and state-school children across five domains.

Due to the novel nature of the research question, we adopted an explorative approach, being unsure as to whether we would locate a sufficient number of studies satisfying inclusion criteria. If a sufficient number of studies could be found, we intended to conduct moderator analyses, investigating the role of age and outcome variable domain. In terms of academic domain, we had reason to believe that there were studies looking at academic skills, creativity, student attitudes and behaviour, school and classroom climate, and moral development. Additionally, we wanted to investigate the possible influence of methodological variables given an expected heterogeneity in study quality. Given the pioneering nature of this review, we adopted a mixture of research questions and, in a few instances, hypotheses, where we felt that the previous literature we were aware of was sufficiently clear to allow this.

Research Questions. Beginning with research questions, we first explored moderator effects for publication and sample features. Of particular interest was the role of publication type (i.e., whether peer-reviewed), measurement features, and the country where the study was conducted. In terms of peer review, it is common to find different effects in published peer-reviewed papers (Polanin et al., 2016). Further, given controversy around funding models in education, we also explored the overall effect sizes as a function of whether countries funded Waldorf education well versus minimally (less than 50%). Additionally, we explored whether age and school level (preschool, primary, secondary) predicted effect sizes.

Do Waldorf Students Show Better Attitudes and Emotions? On the one hand, we expected that attitudes and emotions would differ due to (a) attempts to engage children differently during instruction, (b) Waldorf pupils being more intrinsically motivated due to there being less extrinsic reward in the educational curriculum, and (c) their potentially enjoying school more and having a greater sense of belongingness.

On the other hand, it is well known that many parents send their children to Waldorf schools when they struggle in other systems (Graudenz et al., 2013). This might result in greater attitudinal and behavioural difficulties for some Waldorf students, resulting in non-significant or even negative attitudes to learning.

Do Waldorf Schools Have a Better School and Classroom Climate? In terms of school and classroom climate, our expectations were uncertain. On the one hand, school climate is the personality of the school (Berkowitz et al., 2017), so given the attempt in Waldorf education to create a special character, there should be differences. Specifically, Waldorf schools place emphasis on social cohesion through a stable class constellation (i.e., ideally remaining with the same students across the first seven or eight school years) and generally smaller schools with less competition, teacher support, and less academic pressure. On the other hand, informal reports suggest that Waldorf classrooms can suffer from disruptive behaviour or teaching quality (Graudenz et al., 2013) and mainstream education can be better resourced in many contexts, reflected in greater teacher support.

Hypotheses

Hypothesis 1: Waldorf Students Show Similar Academic Achievement We expected that Waldorf students would show similar academic achievement to state-school students (i.e., equifinality). This expectation arises from previous research showing that initial disadvantages at the beginning of primary school may wash out, or even be reversed (Suggate et al., 2013; Telfer-Radzat, 2024).

Hypothesis 2: Waldorf Students Are More Aesthetically Creative. Given previous research on creativity (Kirkham & Kidd, 2017; Rose et al., 2012) and the strong focus in the Waldorf curriculum on this, we expected there to be an advantage for Waldorf students on measures of aesthetic creativity.

Hypothesis 3: Waldorf Students Show Similar Moral Development. We expected the experience of being schooled in a stable classroom constellation, where the results of one's actions are more likely to be perceived and reflected in the community, to support Waldorf students' moral development. Negating this influence, mainstream schools often include more social sciences, along with ethics, and programmes to support democratic education. On the balance of these different approaches, we hypothesised similar levels of moral development in both samples.

Methods

Literature Search Strategy

We anticipated difficulty in finding a sufficient number of studies that contained a Waldorf sample compared to another educational sample, and adopted a correspondingly broad search strategy across six databanks. The first three databanks selected are

established scientific portals known to include education and/or psychological literature, namely ERIC, PsycINFO, and PubMed. We additionally conducted three further searches in databases that were specific to the German and Waldorf contexts, with which the third author was familiar and that were known to include studies on Waldorf education, also based in the country with the largest number of Waldorf schools worldwide. These databases were: (a) the DIPF–Fachportal, which is a prominent educational research institution, (b) the Karlsruhe Virtueller Katalog, which is a German educational server, and (c) the Bund der Freien Waldorfschulen, which is a compendium of studies conducted on Waldorf education, compiled by the Federation of Waldorf Schools. The searches were conducted in February of the year 2025.

Search Strings

Three strings of search terms were compiled relating to: (a) Waldorf or Steiner education, (b) research methodology, and (c) the sample. The Waldorf-Steiner education search strings were designed to capture the educational approach that is the focus of the meta-analysis. The methodology search strings were selected to exclude studies without an empirical focus, and the sample string focused the research on actors in the educational sector, further removing redundant hits. The search strings were combined with OR command and between groups with AND (or the equivalent commands in the respective databanks).

The first search was of three databanks, namely, using the following strings:

1. (Waldorf* OR Steiner OR Rudolf Steiner* OR anthroposoph*)
2. AND (education OR school* OR kindergarten OR early years OR training OR university OR college OR childhood)
3. AND (Empirical OR qualtitativ* OR quantitativ* OR research OR survey OR design OR data OR experiment OR intervention)
4. AND (parent* OR child* OR student* OR adolescent* OR pupil* OR teach* OR alumni*)

Conducted on 6 February 2025, the searches netted 780 hits in PsycINFO, 392 in ERIC, and 2,677 in PubMed. Given that PubMed resulted in an unexpectedly large number of hits—especially given that it has the smallest educational focus of the three databases—we conducted a screening of the results and found that many medical studies were included due to phenomena such as the author being called Steiner or conditions such as Wiedemann–Steiner syndrome. Restricting this PubMed search to titles and abstracts resulted in 55 hits. These hits were then combined, resulting in 1,695 potential studies for screening, which are available in the OSF materials linked to this paper.

Inclusion and Exclusion Criteria

Generally, to be included in this meta-analysis, studies needed to include (a) a sample that attended Waldorf education, (b) a sample that attended another educational

approach, (c) quantitative data sufficient for the estimation of effect sizes, and (d) measures in an educationally relevant domain, such as academic achievement, aesthetic creativity, or classroom climate.

Title and Abstract Screening Criteria. Due to the relatively young body of empirical research on Waldorf education and a tendency for research to have preferred qualitative and critical analyses in this educational system, we anticipated many non-quantitative studies resulting from our search terms. Accordingly, in a first step we screened titles and abstracts according to three criteria, using a knock-out exclusion approach, namely: (a) "clearly not educational"—for example, when titles or abstracts referred to mental health or commercial settings—(b) clearly not quantitative, such as when the abstract indicated that the methods were qualitative or discursive; and (c) clearly not Waldorf, such as when no mention was made of Waldorf or Steiner education. If studies could not be assigned as "clearly not" according to these criteria, they progressed to full article screening. In a second stage, abstracts were screened more closely again to remove further duplicates and studies using qualitative methods alone.

Full Article Screening Criteria. After abstract and title screening, screening of full documents was undertaken. Similar to the title and abstract screening phase, studies were excluded if they (a) did not include an investigation of Waldorf education, (b) did not contain a comparison group attending, or having attended, a different educational approach, and (c) did not include quantitative data allowing the calculation of effect sizes. Regarding inadequate data reporting, this included studies reporting only means or percentages without standard deviations or the results from statistical testing (e.g., t or F values).

Article and Effect Size Coding

Title and Abstract Screening. Prior to title and abstract screening, the three authors screened the titles and abstracts of 100 randomly selected studies, achieving an inter-rater reliability of $\alpha_{\text{Cronbach}} = .96$. Having achieved excellent agreement, the remaining titles and abstracts were screened. Exclusions were due to articles being clearly not educational ($n = 649$), empirical ($n = 296$), or Waldorf ($n = 1{,}019$), with many duplicates identified ($n = 263$). This left 304 studies whose full texts were selected to be screened.

Full Article Screening. Of the remaining 304 studies, when searching for the full texts and re-checking the abstracts, it became apparent that there were further duplicates ($n = 33$), often due to the same publications appearing as a translation or as part of an edited collection. A further 98 studies had qualitative designs only. This left 173 reports to assess for eligibility, plus an extra 20 that a reviewer suggested; further screening of these (see Figure 1) resulted in 110 candidates for full-text coding. Of these 110, a further 80 were discarded because they did not (a) investigate education ($n = 50$), were not Waldorf ($n = 75$), did not contain a comparison group ($n = 22$), were

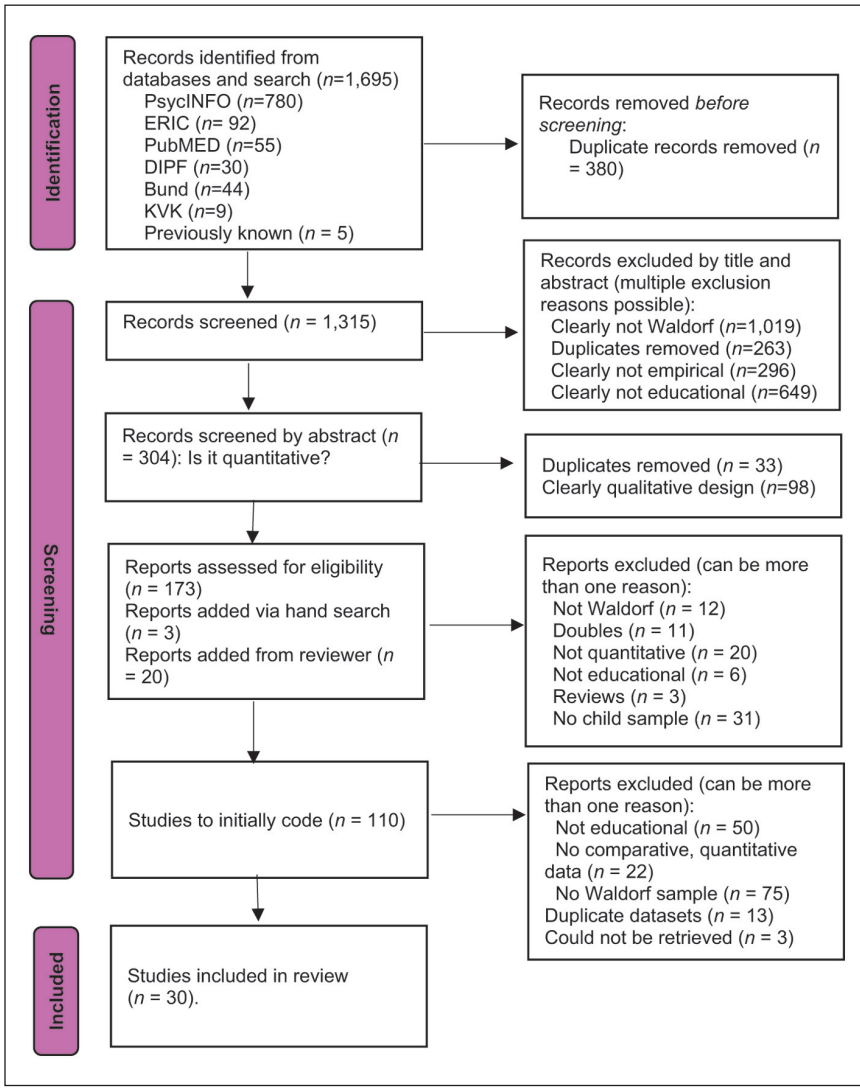


Figure 1. PRISMA flow diagram for article search and screening.

from duplicate datasets ($n = 13$), or could not be retrieved ($n = 3$). This resulted in a final sample of 30 studies. Interrater reliability across the three authors for exclusions based on these categories was acceptable, $\alpha_{Cronbach} = .513$, $\alpha_{Cronbach} = .617$, and $\alpha_{Cronbach} = .834$, with discrepancies being resolved by discussion.

Study features and potential moderator variables were coded, along with the outcome variables, as detailed below. Interrater reliability between the two raters (the first and second authors) was first calculated on 25% of the sample to test the coding categories;

these were then optimised and refined via mutual agreement. Then all studies were re-coded with the revised coding scheme, which indicated excellent agreement, ICC = .971. Any remaining disagreements were then resolved by mutual agreement.

Cohen's *d* effect sizes were calculated based on the means and pooled standard deviations derived from data presented in the papers. In many cases, chi-square tests, *F* values (main effects), and *t* values were used to calculate Cohen's *d* because the papers did not report individual standard deviations and means for Waldorf versus mainstream groups. Effect sizes were reverse coded, where appropriate, so that greater effect sizes always indicated that Waldorf students performed better or rated more positively on Likert scales. In cases where Waldorf was compared to state and Montessori children, only the state group was selected as the control condition. In one instance, the original report did not provide the standard deviations or another measure from which variance could be derived (e.g., *t*, *F*), so the author was contacted to provide these (Rohde, 2022).

Publication Features. Studies were coded for publication year, publication language, publication type (peer-review, chapter, dissertation, report, unpublished data, non-peer-review paper), along with title and author.

Sample Features. Sample features included the percentage of female participants, whether participants were low achieving/at-risk/low SES versus normally achieving (sample achievement status), the country in which the study was conducted, sample language(s) spoken, sample age, and grade. Where sample age was missing but grade was present, grade was imputed based on the typical age according to that country. To create comparable data across school systems, unless otherwise stated, for the school grade category, children up to age 6 were classified as preschool, from 6 to 12 as primary/elementary school, and secondary if older than 12 years.

Measure Type. Measurement type was coded according to whether it was a behavioural measure administered to the students versus a self-report scale.

Study Quality. To provide methodological quality indicators alongside publication type, studies were coded according to the National Institutes of Health's (2013) Study Quality Assessment Tools. The Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies was selected as it best matches the study designs targeted in this meta-analysis. However, two of the items were not relevant to the current study and were hence dropped (i.e., items 12 on blind assessors and 13 on loss to follow-up). This left 10 items pertaining to specification of (1) research objectives, (2) study sample description, (3) participation rate, (4) participant recruitment and inclusion criteria, (5) sample size justification, (6–8) time in respective education systems measured, sufficient, and controlled for in analyses, (9) evidence provided on the psychometric properties of outcome measures, and (10) whether analyses accounted for confounding variables. Up to two points for each item were awarded, except for items 4 to 8 for which one point each was awarded, giving a possible score range of 0–15 points. The

first author rated 28 studies, the second author rated 11 German-language studies, and the third author rated 11 English-language studies, achieving excellent interrater reliability with the first author, 85.71% and 95.15%, respectively.

Measure Domain

Academic Skills. Academic skills were defined as performance in a subject that is typically instructed in school contexts, such as reading, mathematics, or science (Steinmayr et al., 2014). In the current study, this included performance on achievement tests in mathematics, numeracy, reading fluency, decoding, writing, letter naming (Kaminski et al., 2008), spelling, reading comprehension (Woodcock, 1998), oral language (Reese et al., 2012), science including social science (OECD, 2007), geography, and general academic achievement (CTB McGraw-Hill). We initially intended to collect data on cognitive skills, including intelligence and executive functions. However, there were no studies measuring cognitive skills and only one measuring executive functions; therefore, the latter was included in this category. Executive functions included measures of inhibition, attention shifting, working memory, and regulation (MacBeath & Mortimore, 1994).

Aesthetic Creativity. Aesthetic creativity was defined as the ability to generate novel forms, sequences, or ideas, including artistic drawing, written compositions, or symbol sequences (Gao & Hall, 2024). However, at a practical level, most studies that were included in this meta-analysis assessing aesthetic creativity were studies of children's drawing, focusing on composition quality, content, and realism (Rose et al., 2012). Additional effect sizes were provided by measures of verbal creativity or symbol-drawing creativity (Lubart et al., 2011; Urban, 2005).

Attitudes and Emotions. Measures of attitudes and emotions were a broad category that reflected individual students' positive experiences, emotions, and beliefs toward school settings. This included constructs such as motivation (Vallerand et al., 1989), self-concept, self-esteem, stress and burnout, internalising problems, and school adjustment. Measures were exclusively self-report, including many researcher-developed items (e.g., "it is important to be good at Swedish," "I only participate in maths to get better grades") and published questionnaires (e.g., MacBeath & Mortimore, 1994; Maslach & Jackson, 1986; Midgley et al., 2000; Spinath et al., 2002). A list of the individual items is available in the open-source datafile.

School and Classroom Climate. Although school and classroom climate reflects the personality of the school (Berkowitz et al., 2017), the specific operationalisation often refers to features of schools or classrooms perceived as improving learning environments in terms of opportunities to learn, tailoring of instruction, learning support, constructive feedback, and promoting prosocial helping behaviour from staff and students (Doll & Dooley, 2014). A key differentiation between the category attitudes to learning and emotions was that the former represent evaluations of students, whereas school and classroom climate focused on institutional and instructional parameters,

not on individual students. Measures included in this meta-analysis were exclusively self-report scales, including a range of items (e.g., "teachers influence our sense of community," "the many rules in school limit me greatly," "school is a place where I learn useful things," "teachers positively influence my well-being"). A list of the individual items is available in the open-source datafile.

Moral Development. Moral development and behaviour were defined as perceived moral acts, attitudes, thoughts, and beliefs that refer to ethical principles that benefit others, especially in the absence of immediate personal gain, self-interest, or negative impacts on other people (Killen & Malti, 2015). Tasks included Kohlberg-type paradigms (e.g., Rest, 1980), self-report items (e.g., on feelings of moral responsibility for society, self-centredness, helping others, discussion of moral issues; Dahlin, 2010).

Transparency, Openness, and Positionality Statements

We adhered to the MARS guidelines for meta-analytic reporting (Appelbaum et al., 2018). All meta-analytic data, analysis code, and research materials (including our coding scheme) are available (https://osf.io/s6azr/?view_only=ce222427ba154dcc9b95e19cbd686e0c). Data were analysed using R, version 4.2.1 (R Core Team, 2026) and the package metafor, version 3.0-2 (Viechtbauer, 2010).

In accordance with open science practices and journal guidelines (Boveda et al., 2023), we offer positionality statements regarding previous experiences and motivations leading us to conduct the current research study. The authors have diverse backgrounds that have led them to be motivated to explore the current research question. Common to all authors is (a) an interest in educational philosophy and reform, (b) approaches that respect diversity as opposed to overlaying neoliberal criteria into the educational process, (c) critically appraising the effectiveness of education and teaching, and (d) experiences (both positive and negative) as parents of children in both Waldorf and mainstream education. The first author trains teachers for mainstream education and has conducted empirical research in both school systems, but is concerned that education is often too focused on narrow outcomes and performativity. The second author has taught in Waldorf schools in different countries (i.e., in China, Germany, Namibia, South Africa, the US) before moving into tertiary education and conducting empirical research into schools, including those in the reform movement. The third author teaches special and inclusive education at two different universities, with a specialisation in autism spectrum disorder, including 20 years of experience working in an inclusive community caring for people with disabilities.

Data Analysis

Initial analyses focused on determining the overall effect size, investigating publication bias via funnel plots, and exploring the influence of moderator variables. Given that we derived multiple effect sizes from single studies, we could not assume that

each effect size was statistically independent. Accordingly, we analysed the data using mixed-effects meta-regression models performed with the metafor package in R (Viechtbauer, 2010) and maximum likelihood estimation. Further, some studies reported multiple samples, necessitating a nesting of effect sizes within samples, the latter of which were in turn (initially) nested within studies. To test for publication bias, we conducted an Egger's mixed-effects test because this method appears to best accommodate dependent effect sizes inherent in nested data (Rodgers et al., 2020), along with Kendall rank and trim-and-fill tests.

For the main models, to account for small sample sizes, cluster-robust estimation (of effect sizes and confidence intervals, etc.) using the robust package was carried out (Wang et al., 2024). Restricted maximum likelihood estimation was used. Analyses modelled sample and individual effect sizes as random effects (Harrer et al., 2021). Level 1 predictors were measure type and domain. Level 2 predictors concerned sample (e.g., country, age, language) and publication features (year, publication language, peer-review).

Results

Description of the Reviewed Literature

Overall, there were 30 studies included in the meta-analysis, with 40 independent Waldorf versus state-school comparison samples because publications contained more than one independent sample, resulting in 506 effect sizes. Across the studies, there were 7,823 Waldorf students and 187,015 mainstream students, the latter being particularly large because four studies used broader national data for their comparison group. Eighteen studies were published in English, 11 in German, and one in Hungarian. The studies contained samples from 11 different countries, with eight from Germany, four from the US, six from the UK, and the rest from a combination of countries (e.g., Austria, Australia, Czech Republic, France, Hungary, New Zealand, Sweden, Turkey). Most studies were peer-reviewed ($j=17$), with the remainder being book chapters ($j=3$), dissertations ($j=3$), reports ($j=4$), or researcher data from which we extracted the effect sizes ($j=2$). The samples were described as normally achieving, except in one case where the students were described as explicitly coming from low socioeconomic status backgrounds.

An overview of the number of effect sizes available per moderator variable is presented in Table 1. As Table 1 shows, there is a large number of learning climate effect sizes that were not peer-reviewed, from secondary school students, and from Germany, which influences the cell frequencies. There is also a one-sidedness for the aesthetic creativity domain, with most effect sizes coming from English peer-reviewed papers. Behavioural measures of attitudes and behaviour and learning climate are also absent.

Overall Effect Size and Publication Bias

The overall mean effect size, unweighted and unnested, across all 506 effect sizes was small, $d=.221$, $SD=.452$, median $d=.189$. A null model using mixed-effects meta-analysis with three levels, effect size, sample, and study, was initially calculated.

Table 1. Number of Effect Sizes and Sample Comparisons (in Parentheses) as a Function of Domain and Publication, and Sample Features.

Domain	Type	Overall	Academic skills	Aesthetic creativity	Attitudes and emotions	Learning climate	Moral development
Publication type	Peer-reviewed	137 (21)	27 (7)	66 (7)	27 (4)	2 (1)	15 (2)
	Other	369 (19)	81 (9)	0 (0)	82 (5)	197 (3)	9 (2)
Publication language	English	178 (24)	31 (8)	66 (7)	53 (6)	16 (1)	12 (2)
	German	324 (15)	7 (8)	0 (0)	52 (2)	183 (3)	12 (2)
	Other	4 (1)	0 (0)	0 (0)	4 (1)	0 (0)	0 (0)
Sample country	English-speaking	97 (17)	26 (8)	64 (7)	4 (1)	2 (0)	1 (1)
	Germany	317 (13)	73 (6)	0 (0)	51 (2)	181 (3)	12 (2)
	Europe	92 (10)	9 (2)	2 (0)	54 (6)	16 (1)	11 (1)
School level	Preschool	33 (4)	13 (2)	16 (1)	4 (1)	0 (0)	0 (0)
	Primary	63 (11)	12 (4)	48 (6)	3 (1)	0 (0)	0 (0)
	Secondary	410 (25)	83 (8)	2 (1)	102 (7)	199 (3)	24 (4)

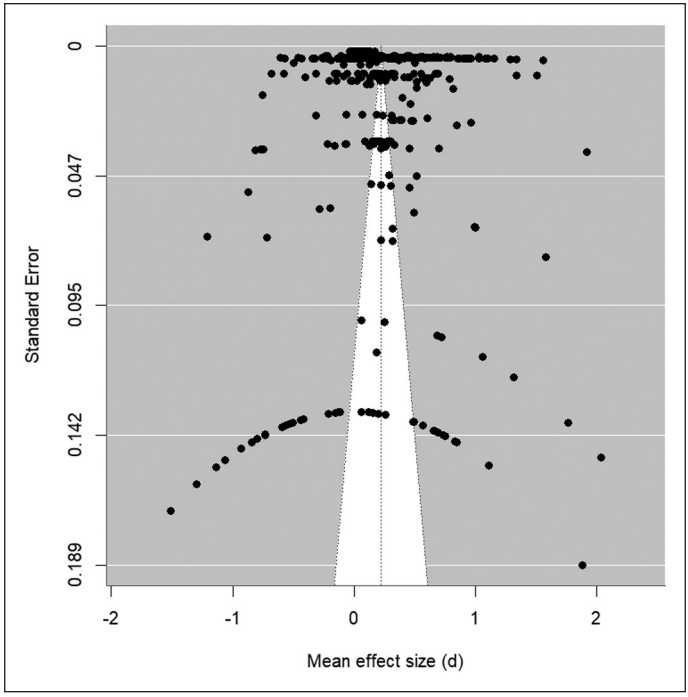


Figure 2. Funnel plot of effect sizes.

However, the ICC indicated that the study level did not explain much variance (i.e., 1.20%), so this was dropped. Subsequently, a two-level null model was re-run, with 52.99% of the variance being explained at the sample level, and 47.01% at the effect size level. The estimated mixed-effects mean effect size was $d = .216$, $Z = 3.725$, $p < .001$, indicating a small positive and statistically significant effect favouring Waldorf education across all measures.

To investigate for publication bias, a funnel plot was calculated and appears in Figure 2. The funnel plot appears to show some asymmetry, which was confirmed with a regression test, $z = -3.476$, $p < .001$, and a Kendall's tau = .101, $p < .001$. A trim-and-fill analysis was conducted, which on the 12th iteration estimated that 104 studies ($SE = 14.709$) were missing on the right side of the funnel plot, which when filled would give rise to an estimated mean small-to-medium effect size favouring the Waldorf students, $d = .369$, $z = 17.098$, $p < .001$. Accordingly, the following effect size estimates in subsequent analyses are likely underestimated.

Testing the Role of Study Quality

The study quality ratings are presented in Table 2. Table 2 shows that most studies provided detail on research aims, appropriately described their samples and sample

Table 2. Summary of Ratings of Study Quality.

	j	Min.	Max.	*M*	*SD*	Skew.	Kurtosis
Research aims (0/1/2)	40	0	2	1.512	.675	−1.070	−.005
Sample specification (0/1/2)	40	0	2	1.439	.634	−.687	−.445
Participation rate (0/1/2)	40	0	2	.707	.844	.614	−1.321
Recruitment criteria (0/1)	40	0	1	.610	.494	−.467	−1.876
Sample size justification (0/1)	40	0	1	.220	.419	1.407	−.023
Educational exposure measure (0/1)	40	0	1	.244	.435	1.238	−.493
Educational exposure sufficiency (0/1)	40	0	1	.220	.419	1.407	−.023
Educational exposure analyses (0/1)	40	0	0	.000	.000	—	—
Measure properties and development (0/1/2)	40	0	2	1.293	.782	−.578	−1.112
Control for confounding variables (0/1/2)	40	0	2	.366	.698	1.664	1.280
Study quality total (0–15)	40	2	12	6.610	3.073	.394	−.829

Note. Possible score ranges appear in parentheses.

recruitment and provided evidence for the psychometric properties of their measures. No study included time in Waldorf or comparison educational approaches as a covariate in analyses, although some studies accounted for this indirectly by excluding participants with mixed educational experiences. Reporting of participation rates, power analyses or sample size justifications, exposure to the respective educational systems, and controlling for confounding variables in multivariate analyses was less common. A correlation between effect size and the study quality indicators, while controlling for the standard error of the effect size (similar to a weighted regression), was conducted to investigate the role of the coded study quality features. This appears in Table 3.

As can be seen in Table 3, effect sizes correlated positively with reporting of research aims, specifying samples, reporting participation rates and time in the respective educational systems, and considering confounding variables. Negative effect sizes were found for links between reporting of study recruitment criteria and the psychometric properties of the instruments. Regarding the latter, a partial correlation was conducted between reporting on psychometric properties of measures and the effect sizes, while controlling for whether the measure was a self-report or behavioural measure. The reason for this additional analysis was that self-report measures were often author-designed and hence missing psychometric data. The resulting correlation found that the link between reporting psychometric properties of measures and effect size was not significant, $r(503) = -.014$, $p = .746$. Next, a mixed-effects regression model was conducted to test the contribution of study quality in explaining variance in effect sizes. Study quality was first z-transformed, so that the model intercept represented the sample effect size after accounting for study quality. This model did not indicate that the study quality measure explained significant variance in effect sizes, $d = .029$, $SE = .035$, $p = .415$.

Publication, Sample, and Measure Features

Next, we conducted a series of mixed-effects regressions to identify and test the effect of potential confounding variables on effect sizes. Because of the small sample sizes in some cells, analyses were conducted using cluster-robust estimation (Wang et al., 2024). Each of the analyses below was obtained by adding the moderator variable into the null model described in the previous section.

Publication Type: Are Effect Sizes From Peer-Reviewed Papers Different? In terms of independent Waldorf-state comparisons, these arose from 21 peer-reviewed papers, eight book chapters, three dissertations, six unpublished reports, and two unpublished datasets. The resulting model using cluster-robust estimates indicated that peer-review papers and book publications resulted in significant effect sizes, $d = .200$, $p = .043$, $d = .236$, $p = .038$, respectively, whereas other categories did not: unpublished researcher data, $d = .211$, $p = .080$, dissertations, $d = .399$, $p = .356$, and unpublished reports, $d = .173$, $p = .071$. Given the small number of studies in many categories, we collapsed these into peer-reviewed versus "other" and re-ran the model. Both peer-reviewed, $d = .237$, $p = .002$, and other publications, $d = .199$, $p = .043$, resulted in statistically significant effect sizes, which in an ANOVA post hoc test did not differ, $p = .747$.

Table 3. Partial Correlations Between Study Quality Measures and Effect Sizes, Controlling for Standard Errors.

	Dependent variable	1	2	3	4	5	6	7	8	9	10	11
1	Effect size (*d*)	—	.128*	.143*	.254*	−.143*	.004	.019	.266*	−.150*	.103*	.171*
2	Research aims		—	.545*	.268*	−.264*	−.373*	−.77	.191*	.098*	.101*	.531*
3	Sample specification			—	.409*	.005	−.252*	−.030	.159*	.036	.188*	.641*
4	Participation rate				—	.060	.061	.014	.730*	−.577*	−.019	.554*
5	Recruitment criteria					—	.374*	.565*	−.215*	−.266*	.017	.194*
6	Sample size justification						—	.428*	.240*	.111*	−.003	.276*
7	Educational exposure measure							—	.116*	−.082	.142*	.397*
8	Educational exposure sufficiency								—	−.322*	.061	.526*
9	Measure properties and development									—	.264*	.141*
10	Control for confounding variables										—	.433*
11	Study quality total											—

*$p < .05$.

Does Study Country and State Funding Moderate Effect Size? We tested whether effect sizes for Waldorf education were different across countries. To reflect the distribution of studies, we grouped these into Germany, English-speaking countries (i.e., the UK, US, Australia, and New Zealand), and Europe (including Turkey). Cluster-robust estimation indicated that effects were not significant in the English-speaking countries, $d = .214$, $p = .086$, but were in Europe, $d = .242$, $p = .041$, and Germany, $d = .199$, $p = .013$. However, the post hoc ANOVA found no significant differences between these three categories, $ps > .850$.

In a subsequent analysis, we divided the countries into those that received substantial state funding (well over 50%) to those that received no or minimal state funding (well under 50%). Five countries were classified as receiving substantial state funding (i.e., Germany, New Zealand, Australia, Sweden, Denmark). Mixed-effects estimates indicated that both state-funded, $d = .237$, $p = .096$, and privately funded schools showed significant positive effect sizes, $d = .203$, $p < .001$, with the difference not being statistically significant, $p = .781$.

Measure Type: Are Behavioural Measures Associated With Smaller Effects? The next analyses tested whether behavioural measures resulted in different effect sizes to self-report measures. In total there were 161 behavioural and 345 self-report effect sizes. The mixed-effects model and cluster-robust estimates indicated a small difference between behavioural, $d = .245$, $p = .032$, and self-report measures, $d = .192$, $p = .099$, which was not statistically significantly different, $p = .636$.

Exploratory Analyses on Role of Publication Year. Studies were published between 1999 and 2024, $M = 2013$, $SD = 7.144$. Modelling publication year as a continuous variable in a mixed-effects model indicated that studies published more recently resulted in greater effect sizes, $d = .0001$, $t = 3.733$, $p < .001$.

Exploratory Analysis on School Level and Age. The sample was on average 13.466 years old ($SD = 4.947$). The cluster-robust estimate from a mixed-effects model with age added to it found that age was positively related to effect size, $d = .014$, $t = 4.392$, $p < .001$. Not surprisingly, entering the categorical variable of school level into the model found that effect sizes were largest, but not robustly significant in the early years, $d = .462$, $t = 2.627$, $p = .093$, or in primary school, $d = .118$, $t = .733$, $p = .479$, but was in secondary school, $d = .221$, $t = 4.046$, $p < .001$. None of the effect sizes for school level were significantly different to each other, $ps > .116$.

Effect Sizes as a Function of Domain. To test whether the effect sizes differ for Waldorf versus mainstream students, a mixed-effects model included domain as a moderator variable. To save numerator degrees of freedom and given the equivocal evidence for the role of the publication, sample, or measure variables, no additional moderators were added to the model. The main model had 40 samples, ICC = .434 at the effect size level, 506 effect sizes, and had unexplained heterogeneity, $Q(501) = 7{,}964.733$, $p < .001$. The domain moderator variable was statistically significant, $Q(5) = 25.131$, $p < .001$.

Table 4. Summary of Cluster-Robust Effect Sizes for Waldorf Versus Mainstream Students as a Function of Domain.

	d	*SE*	*t*	*df*	*p*	CI lower	CI upper
Academic skills	.198	.066	3.003	15.00	.009	.058	.339
Aesthetic creativity	.381	.241	1.579	6.50	.162	−.199	.961
Attitudes and emotions	.097	.094	1.024	14.31	.323	−.105	.298
School and classroom climate	.308	.107	2.878	9.58	.017	.068	.547
Moral development	.150	.098	1.540	7.03	.167	−.08	.381

Note. Positive effect sizes indicated greater performance for the Waldorf students. CI = confidence interval.

Not using cluster-robust modelling, there were significant effects for academic skills, $p = .033$, aesthetic creativity, $p = .004$, and school and classroom climate, $p = .001$. The effect sizes from cluster-robust modelling are presented in Table 4.

and will now be considered in terms of the domain-specific research questions and hypotheses.

Do Waldorf Students Show Better Emotions and Greater Attitudes to Learning? As Table 4 shows, the effect size for attitudes and emotions of the Waldorf school students compared to the state-school students was positive, but not statistically significantly greater.

Do Waldorf Schools Have a Better School and Classroom Climate? The effect size for school and classroom climate was small, but positive and significantly greater in the Waldorf schools.

Hypothesis 1: Waldorf Students Show Similar Academic Achievement. The cluster-robust estimates showed that the academic achievement in Waldorf schools was small and positive, and statistically significantly greater than in state schools.

Hypothesis 2: Waldorf Students Are More Creative. The mixed-effects model indicated that Waldorf students showed greater aesthetic creativity scores; however, after cluster-robust adjustment, this was no longer significantly different.

Hypothesis 3: Waldorf Students Show Similar Moral Development. The estimated effect size was small and positive but not significantly different. Further, adding study quality to the model did not change the findings as it was not a significant predictor.

Discussion

The current study is the first quantitative meta-analysis of a range of outcomes for students attending Waldorf compared to mainstream schools. Findings indicated overall positive and statistically significant—but small—differences favouring students attending Waldorf education. Specifically, the effect size was small in the overall

mixed-effects null model, $d = .216$; however, small-to-medium, $d = .369$, when corrected for publication bias using a trim-and-fill procedure. These effect sizes were derived from 30 studies containing 40 independent samples comparing nearly 8,000 Waldorf and 187,000 mainstream students. However impressive the sample sizes might seem, the comparison is still dependent on the methodology employed in the individual studies. Therefore, before considering the hypotheses regarding subdomains, we analyse the robustness of the findings.

In terms of study quality, only about 57% of papers were peer-reviewed, with the rest often comprising research reports evaluating Waldorf education. Viewed over time, there was a moderate-to-strong correlation between publication year and whether the paper was peer-reviewed ($r = .493$, $p = .006$), with the first peer-reviewed papers published in 2010—some 11 years after the first published report. This trend reflects a growing academic interest in studying Waldorf education quantitatively. Reflecting this trend, the study quality measures indicated that much research did not control for potential confounding variables (e.g., educational history, SES, IQ). However, there was no compelling evidence that publication, sample, or measurement features systematically biased the effect size estimates. Specifically, these were generally not significant predictors in the analyses. Nevertheless, the measure of study quality appeared to show face and criterion validity, generally correlating positively with effect size (i.e., greater quality, greater d), yet in the mixed-effects models, study quality did not change the pattern of findings. Thus, taking these findings together, effect sizes do not appear to be biased by study quality, even if there is evidence of publication bias downwardly estimating effect sizes for the Waldorf students.

Turning to the specific hypotheses, we first posited that Waldorf students would show similar academic achievement to state students. However, contrary to expectation, this hypothesis was rejected, with Waldorf students performing significantly better on academic measures. Importantly, the academic measures employed in the original studies were given to a range of different ages and contained a number of different subdomains, including reading, mathematics, and science. Measures included well-known tests, such as standardised reading tests (e.g., the Woodcock), PISA tasks, and university entrance qualifications (i.e., the German Abitur). Interestingly, a clear focus and claim made by Waldorf educators is that preschool and primary school should focus on foundational skills, such as sensorimotor, language, aesthetic, and social development, to create a basis for analytical thinking in secondary school. Given that this approach has been controversial (Lonigan & Phillips, 2012), the positive finding here may be of some interest to this discussion.

Our second hypothesis, that Waldorf students would be more creative, was not convincingly supported. Although there was a small-to-medium effect size ($d = .381$), this contained a large confidence interval and was hence not reliably different from zero, once cluster-robust estimation was used. One potential reason for this is due to the dependent variables in the aesthetic creativity domain. Most creativity effect sizes were derived from ratings of drawings according to specific features (e.g., use of space, colour, realism) often without a priori theory as to which specific dimensions indicate better performance. This is not a criticism of the studies but represents the difficulty with aesthetic creativity judgments—it is easy to imagine that in some form

of art realism represents high quality whereas in others the opposite might be the case. In conducting a meta-analysis we coded every effect size, without a priori theory as to which dimensions represent more creativity, which likely explains the heterogeneity obtained, and hence the positive but non-significant effect size.

Third, we hypothesised that Waldorf students would show similar moral development to state students. This hypothesis was supported, with there nevertheless being a small but non-significant effect size favouring the Waldorf students. One potential reason for this finding is that most schools now include ethics and discussions related to moral development in their curricula (OECD, 2021; Schuitema et al., 2008). It is also noteworthy that moral development itself appears to show only a moderate relation to actual moral behaviour (Villegas de Posada & Vargas-Trujillo, 2015). Given the relatively small number of studies included in the meta-analysis looking at moral development ($j = 5$), it remains to be seen as to whether features of Waldorf education (smaller schools, looping, stable classes, group activities) or mainstream education result in different moral reasoning and whether this extends to behaviour.

Moving to consider the exploratory research questions, both Waldorf and mainstream students showed similar attitudes and emotions. Again, there was a small positive but statistically non-significant effect size favouring Waldorf students. Interestingly, the confidence interval around attitudes appeared broad, which likely indicates that this category is affected by an unidentified moderator variable—perhaps having been operationalised too broadly in the current study. Given previous research indicating that Waldorf families tend to have greater educational capital (Koolmann et al., 2018), it may be that any small differences between the samples in terms of attitudes to learning are due to parental factors rather than to the educational approach.

There was a statistically significant difference in how the Waldorf students rated their classroom and school climates, showing a small but reliable effect. Thus, Waldorf students consistently rated their schools as being positive learning environments that were focused on their needs and well-being. Items more strongly endorsed by Waldorf than mainstream students included those relating to: feeling less learning pressure, feeling safe and happy at school, learning a lot, having teachers shaping the teaching to be interesting, being able to play an active role in learning, feeling that their problems were taken seriously, or being able to find friends easily. The current finding on climate appears plausible in light of key features of Waldorf education designed to improve classroom climate, such as reduced assessment pressure, more stable classes with less grade retention, a round curriculum with a focus on aesthetics, and looping. Additionally, given evidence that climate can positively influence academic achievement (Roorda et al., 2011; Thapa et al., 2013), it could be that features of the school and classroom climate influenced the greater academic achievement in Waldorf schools.

We found a positive correlation between age and the effects of Waldorf education. This finding was initially surprising because differences between Waldorf and mainstream schools are perhaps most pronounced in the early years and the first years of primary school (Suggate & Suggate, 2019). When the analysis was broken down categorically into school stage, the effects were more nuanced. The medium effect for the early years and elementary school was likely not significant due to the relatively high heterogeneity in domains, for instance, containing both academic and aesthetic creativity measures,

which are different foci in Waldorf education (e.g., aesthetic creativity should be higher, and academic skills lower, cancelling each other out and creating greater variance). However, in secondary school the effect sizes were small yet robust, perhaps because effects accumulate over time and the educational aims of Waldorf and mainstream schools converge on explicitly academic instruction.

Finally, there was an interesting finding regarding whether schools received minimal versus substantial state funding. Although not statistically significantly different, the raw effect sizes appeared larger for schools not receiving state funding. Such schools often have more freedom to implement their own curricula but can suffer from staff recruitment due to often being only able to pay teachers lower wages, depending on where the school is. Furthermore, non-funded schools are more dependent on parental and teacher assistance for day-to-day running. Hence, non-funded schools might be expected to have greater heterogeneity in school quality.

Limitations

The current meta-analysis was firstly limited by the still relatively small number of studies on Waldorf education outcomes. More research is needed in all of the domains assessed here. Additionally, there were only two studies looking at cognitive outcomes (i.e., executive functioning, reasoning), preventing us from examining this category separately. Second, due to the small number of studies, it was not possible to conduct a large-scale meta-regression analysis to test the independent contributions of the sample, measure, publication, age, and domain moderators. This prevented us from definitively teasing potential moderators apart. Third, the current analysis cannot be taken as causal evidence of the effects of Waldorf education as there was no random assignment to the schools and we do not know whether parental or family variables leading to the selection of Waldorf education are not the main drivers of the small effect sizes found here. Fourth, there were some gaps in the data, with there being a lack of behavioural measures of attitudes, emotions, school and classroom climate. Accordingly, study quality and measure domain were confounded, which may explain why study quality correlated with effect size, despite it not contributing to the model with domain as a predictor variable.

The analysis of study quality conducted here revealed several weaknesses in the available literature, which limit the claims that can be made from this meta-analysis. For instance, many studies did not measure the educational history of participants, that is, the amount of time that students had spent in different educational institutions. Further, reporting of participation rates, power analyses, and controlling for IQ or SES in analyses was uncommon. Together, this makes it difficult to be certain that educational differences lead to the observed differences here.

Implications and Future Work

Bearing in mind that this meta-analysis is not conducive to causal claims, it would nevertheless appear incumbent upon us to offer a brief analysis of potential causal mechanisms. Despite some research indicating minimal differences between Waldorf and mainstream families (at least those who agree to take part in research), there are likely

key differences that may not have been captured. Specifically, parents who inform themselves about alternative educational systems are likely invested in the education of their children. This background educational motivation and capital, at least in some instances, may then drive both the choice of school and the subsequent educational trajectory.

Additionally, there are a range of specific didactical methods and curriculum differences that may interact cumulatively across the course of school to result in the differences obtained here, each of which could be subjected to its own empirical investigation. Examples include oracy and narrative approaches, early literacy, embodied mathematical instruction, form drawing, artistic and handicraft activities, early onset of foreign language learning, teaching in scheduled blocks/epochs, eurythmy, methods of teaching science and literature studies. Conceivably, some of these differences may work in complex dynamics, some positively, others negatively, or cumulatively, or in interaction with each other—disentanglement of the mechanisms would pose a supreme challenge for research.

However, a key implication of this meta-analysis is that educational outcomes for students attending Waldorf schools were not deleterious, and in contrast, were associated with a small-to-medium positive effect overall. Thus, although the precise causal mechanisms have yet to be established, the educational approach broadly receives support across a range of countries and contexts. This effect appeared to exist regardless of whether the schools received none or minimal state funding. Accordingly, students attending an educational approach that was founded over 100 years ago may receive an education not behind those of their peers—at least in the domains assessed here.

A further implication is of developmental importance. Despite appreciable differences in the timing of academic instruction, with this beginning much later and more implicitly in Waldorf schools, there was no evidence that academic achievement suffers. Accordingly, the current study provides insight into the potential benefits of fostering unconstrained skills (Paris, 2005) and soft skills that are foundational to later development, such as sensorimotor (Karle et al., 2026; Moeller et al., 2012; Schmidt et al., 2019), language (Dickinson et al., 2010), and social development (Gallagher, 1993; Guerrero et al., 2024; Hoff, 2006). Thus, early and explicit (Good et al., 1998; Riley, 1996) is not the only approach as a later and initially indirect approach appears to result in similar academic outcomes. However, more longitudinal research is needed to test this hypothesis.

An important task for future research is to expand the number of studies on Waldorf education. This work could extend the number of educational and psychological outcomes investigated, with work missing in key domains such as numerical cognition, science performance, social competence, and long-term effects. Further, this work needs to be conducted longitudinally and control for educational history and SES, probably using quasi-experimental designs due to the difficulty in using random assignment—unless a waiting-list control design can be used, as has been done in a study looking at Montessori education (Lillard & Else-Quest, 2006). Such work could also compare outcomes in other alternative education approaches, as this may better account for educational capital in the family.

Additionally, research is needed that explores educational processes, with this review in the context of other research identifying school and classroom climate as an important aspect for further research (Thapa et al., 2013). In the context of

Waldorf education, there are many unique features that have undergone little quantitative testing—such as the effect of form drawing on sensorimotor development and cognition (Kutzli, 2007), teaching in epochs, the effect of looping on classroom climate and social identity, effects on mental health and well-being, and approaches to heterogeneity and inclusion. Further, there are no doubt practices in Waldorf education that would benefit from empirical investigation with a view to optimising educational processes and learning. Anecdotally, Waldorf schools have not always been open to participating in quantitative research—hopefully this meta-analysis will help encourage them to do so.

At the outset, we stated that research is needed to help parents make an informed decision as to what outcomes might be expected by different educational approaches. Taking research findings from a fairly large body of research and making recommendations for individual students is fraught with fallacy and danger. For instance, there is a large heterogeneity in many outcomes, suggesting that despite there being positive effect sizes for those in Waldorf education overall, there are also likely to be a number of students who might be better served in a different approach. Accordingly, parents and students are encouraged to consider each school—and possibly each class within that school if the children are to spend years together in the same class—individually in making their choice. That said, the current study suggests that attending a Waldorf school is associated with slightly more positive overall outcomes: Children can perform well and thrive in each system.

Summary

The Waldorf educational approach has spread across the globe and, despite being small, has captured attention for some key curricula and methodological differences, resulting in some anecdotally impressive outcomes—despite a lack of empirical research. We conducted the first systematic quantitative meta-analysis comparing educational outcomes across a range of domains for children attending Waldorf and mainstream education. Waldorf schools were associated with overall better outcomes, particularly with regard to school and classroom environment and academic achievement. However, more research across all domains is needed, particularly employing longitudinal designs controlling for background variables.

ORCID iD

Sebastian P. Suggate https://orcid.org/0000-0002-7755-068X

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Funding provided by Software AG. GK-P 16038

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

*Articles included in the meta-analysis are denoted with an asterisk

Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist*, *73*(1), 3–25. https://doi.org/10.1037/amp0000191

Armon, J. (1997). *The Waldorf curriculum as a framework for moral education: One dimension of a fourfold system*. https://eric.ed.gov/?id=ed420605

*Arslan, A., & Kartal, S. (2022). The effect of the Waldorf approach applied with the collaborative method on the early assessment skills of 48-60 months-old children. *International Online Journal of Education and Teaching*, *9*(3), 1096–1112.

*Barz, H., & Randoll, D. (2012). Unpublished data.

Becker, C., & Wiehl, A. (2024). *Epochenhefte: Theorie und Praxis eines Bildungsmediums*. Beltz Juventa. https://directory.doabooks.org/handle/20.500.12854/140521

Berkowitz, R., Moore, H., Astor, R. A., & Benbenishty, R. (2017). A research synthesis of the associations between socioeconomic background, inequality, school climate, and academic achievement. *Review of Educational Research*, *87*(2), 425–469. https://doi.org/10.3102/0034654316669821

*Besançon, M., Fenouillet, F., & Shankland, R. (2015). Influence of school environment on adolescents' creative potential, motivation and well-being. *Learning and Individual Differences*, *43*, 178–184. https://doi.org/10.1016/j.lindif.2015.08.029

*Bieg, S., Steuer, G., Rosentritt-Brunn, G., & Dresel, M. (2019). Motivationales Klima in Reform- und Regelschulklassen: Unterschiede, Gemeinsamkeiten und Effekte auf Schülermotivation [Motivational climate in progressive educational and traditional school environments: Differences, commonalities and effects on students' motivation]. *Unterrichtswissenschaft*, *47*(4), 405–426. https://doi.org/10.1007/s42010-019-00044-9

Boveda, M., Ford, K. S., Frankenberg, E., & López, F. (2023). Editorial vision 2022–2025. *Review of Educational Research*, *93*(5), 635–640. https://doi.org/10.3102/00346543231170179

Buck, M. R., & Hoffmann, A.-K. (2024). Waldorf education as an (inter)national phenomenon and subject of discourse. In A.-K. Hoffmann & M. F. Buck (Eds.), *Critically assessing the reputation of Waldorf education in academia and the public: Early endeavours of expansion, 1919–1955* (pp. 1–21). Routledge. https://doi.org/10.4324/9781003403609

Camacho-Morles, J., Slemp, G. R., Pekrun, R., Loderer, K., Hou, H., & Oades, L. G. (2021). Activity achievement emotions and academic performance: A meta-analysis. *Educational Psychology Review*, *33*(3), 1051–1095. https://doi.org/10.1007/s10648-020-09585-3

*Cox, M. V., & Rowlands, A. (2000). The effect of three different educational approaches on children's drawing ability: Steiner, Montessori and traditional. *British Journal of Educational Psychology*, *70*(4), 485–503. https://doi.org/10.1348/000709900158263

*Cunningham, A. J., & Carroll, J. M. (2011). The development of early literacy in Steiner- and standard-educated children. *British Journal of Educational Psychology*, *83*, 475–490. https://doi.org/10.1348/000709910X522474

*Dahlin, B. (2007). *The Waldorf school – cultivating humanity? A report from an evaluation of Waldorf schools in Sweden*. Education. Karlstad University. http://ecswe.net/wren/documents/Waldorf_School_Evaluation.pdf

*Dahlin, B. (2010). A state-independent education for citizenship? Comparing beliefs and values related to civic and moral issues among students in Swedish mainstream and Steiner Waldorf schools. *Journal of Beliefs & Values*, *31*(2), 165–180. https://doi.org/10.1080/13617672.2010.503629

Dahlin, B. (2017). *Rudolf Steiner: The relevance of Waldorf education*. SpringerBriefs in Education Ser. Springer International Publishing. https://doi.org/10.1007/978-3-319-58907-7

Dickinson, D. K., Golinkoff, R. M., & Hirsh-Pasek, K. (2010). Speaking out for language: Why language is central to reading development. *Educational Researcher*, *39*, 305–310. https://doi.org/10.3102/0013189x10370204

Doll, B., & Dooley, K. (2014). Classroom climate. In A. C. Michalos (Ed.), *Encyclopedia of quality of life and well-being research* (pp. 944–949). Springer. https://doi.org/10.1007/978-94-007-0753-5_402

Edwards, C. P. (2007). Three approaches from Europe: Waldorf, Montessori, and Reggio Emilia. *Early Childhood Research & Practice*, *4*(1), 1–14. https://ecrp.illinois.edu/v4n1/edwards.html

*Erasmus, H. (2015). *Topographische Kenntnisse und Fähigkeiten im Geographieunterricht der Waldorfschule. Eine Untersuchung auf der Grundlage zweier empirischer Erhebungen [Topographical knowledge and skills in geography instruction at the Waldorf school: Findings across two empirical investigations]*. Universitätsbibliothek. http://doi.org/10.22029/jlupub-9689

*Erasmus, H., & Obermeier, G. (2010). Sind Waldorfschüler interessierter als Schüler von Regelschulen? Eine empirische Untersuchung der Schülerinteressen am Beispiel Hessen [Are Waldorf students more interested than students from mainstream schools? An empirical investigation of student interest from Hessen]. In I. Hemmer & M. Hemmer (Eds.), *Schülerinteresse an Themen, Regionen und Arbeitsweisen des Geographieunterrichts. Ergebnisse der empirischen Forschung und deren Konsequenzen für die Unterrichtspraxis*, Geographiedidaktische Forschungen. 46 (pp. 165–184). Selbstverlag des Hochschulverbandes für Geographie und ihre Didaktik e.V.

*Föller-Mancini, A., Heusser, P., & Büssing, A. (2010). Self-centeredness in adolescents: An empirical study of students of Steiner schools, Christian academic high schools, and public schools. *Research on Steiner Education*, *1*(2), 148–164.

*Friedlaender, D., Beckham, K., Zheng, X., & Darling-Hammond, L. (2015). *Growing a Waldorf-inspired approach in a public school district*. Stanford Center for Opportunity Policy in Education. https://www.proquest.com/reports/growing-waldorf-inspired-approach-public-school/docview/1941340745/se-2?accountid=13478

Gallagher, T. M. (1993). Language skill and the development of social competence in school-age children. *Language, Speech, and Hearing Services in Schools*, *24*, 199–205. https://doi.org/10.1044/0161-1461.2404.199

Gao, Q., & Hall, A. (2024). Early childhood teachers' beliefs about creativity and practices for fostering creativity: A review of recent literature. *Early Childhood Education Journal*. Advance online publication. https://doi.org/10.1007/s10643-024-01816-4

Good, R. H., Simmons, D. C., & Smith, S. B. (1998). Effective academic interventions in the United States: Evaluating and enhancing the acquisition of early reading skills. *School Psychology Review*, *27*, 45–56. https://doi.org/10.1080/02796015.1998.12085896

Graudenz, I., Peters, J., & Randoll, D. (2013). What it's like being a Waldorf teacher: Results of an empirical survey. *RoSE – Research on Steiner Education*, *4*(2), 104–114. https://rosejourn.com/index.php/rose/article/view/161

Grunelius, E. M. (1991). *Early childhood education and the Waldorf School plan*. Waldorf Curriculum Series. Rudolf Steiner College Publications. (Original work published 1950)

Guerrero, S., Valenciano-Valcárcel, J., & Rodríguez, A. (2024). Unveiling alternative schools: A systematic review of cognitive and social-emotional development in different educational approaches. *Children and Youth Services Review*, *158*, 107480. https://doi.org/10.1016/j.childyouth.2024.107480

Harrer, M., Cuijpers, P., Furukawa, T. A., & Ebert, D. D. (2021). *Doing meta-analysis with R: A hands-on guide*. Chapmann & Hall/CRC Press. ISBN 978-0-367-61007-4

*Hether, C. A. (2001). *The moral reasoning of high school seniors from diverse educational settings* [ProQuest dissertations & theses]. 3044032.

Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review*, *26*(1), 55–88. https://doi.org/10.1016/j.dr.2005.11.002

Howard, J. L., Bureau, J., Guay, F., Chong, J. X. Y., & Ryan, R. M. (2021). Student motivation and associated outcomes: A meta-analysis from self-determination theory. *Perspectives on Psychological Science*, *16*(6), 1300–1323. https://doi.org/10.1177/1745691620966789

*Jelinek, D., & Sun, L. (2003). *Does Waldorf offer a viable form of science education? A research monograph*. California State University College of Education. California State University Press.

Kaminski, R. A., Cummings, K. D., Powell-Smith, K. A., & Good, R. H. (2008). Best practices in using dynamic indicators of basic early literacy skills for formative assessment and evaluation. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (Vol. 5, pp. 1181–1204). National Association of School Psychologists.

Karle, V. L., Suggate, S. P., Winter, R., & Stoeger, H. (2026). When fine is large and gross is small: A meta-analysis of links between fine and gross motor skills with academic-cognitive skills. *Educational Psychology Review*, *38*(1), 73. https://doi.org/10.1007/s10648-026-10161-4

Killen, M., & Malti, T. (2015). Moral judgments and emotions in contexts of peer exclusion and victimization. *Advances in Child Development and Behavior*, *48*, 249–276. https://doi.org/10.1016/bs.acdb.2014.11.007

*Kirkham, J. A., & Kidd, E. (2017). The effect of Steiner, Montessori, and national curriculum education upon children's pretence and creativity. *The Journal of Creative Behavior*, *51*(1), 20–34. https://doi.org/10.1002/jocb.83

*Kitti, K., Balázs, J., Éva, S., Nóra, K., & Márta, F. (2023). Diák kiégés hagyományos és alternatív pedagógiai programú oktatási intézményekben [Student burnout in schools using traditional and alternative educational programs]. *Magyar Pszichológiai Szemle*, *78*(3), 329–344. https://doi.org/10.1556/0016.2023.00043

Koolmann, S., Petersen, L., & Ehrler, P. (2018). *Waldorf-Eltern in Deutschland: Status, Motive, Einstellungen, Zukunftsideen* (1. Auflage). Beltz Juventa. http://www.beltz.de/de/nc/verlagsgruppe-beltz/gesamtprogramm.html?isbn=978-3-7799-3761-6

Kupers, E., Lehmann-Wermser, A., McPherson, G., & van Geert, P. (2019). Children's creativity: A theoretical framework and systematic review. *Review of Educational Research*, *89*(1), 93–124. https://doi.org/10.3102/0034654318815707

Kutzli, R. (2007). *Creative form drawing* (Illustrated ed.). Learning Resources: Rudolf Steiner Education Series. Hawthorn Press; [distributor] BookSource.

*Kvintova, J., Kremenkova, L., Cuberek, R., Petrova, J., Stuchlikova, I., Dobesova-Cakirpaloglu, S., Pugnerova, M., Balatova, K., Lemrova, S., Viteckova, M., & Plevova, I. (2022). Preschoolers' attitudes, school motivation, and executive functions in the context of various types of kindergarten. *Frontiers in Psychology*, *13*, 823980. https://doi.org/10.3389/fpsyg.2022.823980

LeBoutillier, N., & Marks, D. F. (2003). Mental imagery and creativity: A meta-analytic review study. *British Journal of Psychology*, *94*(1), 29–44. https://doi.org/10.1348/000712603762842084

Lillard, A., & Else-Quest, N. (2006). Evaluating Montessori education. *Science*, *313*, 1893–1894. https://doi.org/10.1126/science.1132362

Lim, B. (2004). Aesthetic discourses in early childhood settings: Dewey, Steiner, and Vygotsky. *Early Childhood Development and Care*, *174*, 473–486. https://doi.org/10.1080/0300443032000153633

Lonigan, C. J., & Phillips, B. M. (2012). Understanding the contributions of early academic skills to children's success in school. In S. P. Suggate & E. Reese (Eds.), *Contemporary debates in child development and education* (pp. 145–158). Routledge, Taylor & Francis.

Lubart, T. I., Besançon, M., & Barbot, B. (2011). *Evaluation du Potentiel Créatif. (EPoC)*. Hogrefe.

MacBeath, J., & Mortimore, P. (1994). *Improving school effectiveness: A Scottish approach*. University of London: International School Effectiveness & Improvement Centre.

Mao, P., Cai, Z., He, J., Chen, X., & Fan, X. (2021). The relationship between attitude toward science and academic achievement in science: A three-level meta-analysis. *Frontiers in Psychology*, *12*, 784068. https://doi.org/10.3389/fpsyg.2021.784068

Marcon, R. A. (2002). Moving up the grades: Relationship between preschool model and later school success. *Early Childhood Research & Practice*, *4*(1), 24.

Maslach, C., & Jackson, S. E. (1986). *Maslach burnout inventory manual* (2nd ed.). Consulting Psychologists Press.

*Mayer, M. C. (2007). *Soziomoralischer Kompetenzvergleich von Regel- und Waldorfschülern nach Grundlagen des Kohlberg-Modells*. Universität Passau.

McCown, C., & Sherman, S. (2002). Looping for better performance in the middle grades. *Middle School Journal*, *33*(4), 17–21. https://doi.org/10.1080/00940771.2002.11494679

Mepham, T. M. (2024). Publicly funded Steiner education in England–beautiful anomaly? Missed opportunity? Or both? *Frontiers in Education*, *9*, 1364978. https://doi.org/10.3389/feduc.2024.1364978

Midgley, C., Maehr, M. L., Hruda, L. Z., Anderman, E., Anderman, L., Freeman, K. E., Gheen, M., Kaplan, A., Kumar, R., Middleton, M. J., Nelson, J., Roeser, R., & Urdan, T. (2000). *Manual for the patterns of adaptive learning scales (PALS)*. University of Michigan.

Moeller, K., Fischer, U., Link, T., Wasner, M., Huber, S., Cress, U., & Nuerk, H.-C. (2012). Learning and development of embodied numerosity. *Cognitive Processing*, *13*(S1), 271–274. https://doi.org/10.1007/s10339-012-0457-9

Molnar-Szakacs, I. (2011). From actions to empathy and morality – A neural perspective. *Journal of Economic Behavior & Organization*, *77*(1), 76–85. https://doi.org/10.1016/j.jebo.2010.02.019

National Institute for Health. (2013). *Quality assessment tool for observational cohort and cross-sectional studies*. Retrieved January 13, 2026, from https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools

Nicol, J. (2015). *Bringing the Steiner Waldorf approach to your early years practice* (3rd ed.). Routledge.

OECD. (2007). *PISA 2006 science competencies for tomorrow's world volume 1: Analysis*. OECD Publishing.

OECD. (2021). *Embedding values and attitudes in curriculum*. Organization for Economic Cooperation & Development. https://doi.org/10.1787/aee2adcd-en

OECD. (2023a). *Education at a glance 2023: OECD indicators*. OECD Publishing. https://doi.org/10.1787/e13bef63-en

OECD. (2023b). *OECD digital education outlook 2023* (1st ed.). OECD Digital Education Outlook Series. Organization for Economic Cooperation & Development. https://doi.org/10.1787/c74f03de-en

Paris, S. G. (2005). Reinterpreting the development of reading skills. *Reading Research Quarterly*, *40*, 184–202. https://doi.org/10.1598/rrq.40.2.3

*Peters, J., & Randoll, D. (2013). Unpublished data.
Peters, J., & Randoll, D. (2018). Leistungsprinzip und Leistungsverständnis an Waldorfschulen. Pädagogische Intentionen und empirische Befunde [Educational achievement in Waldorf schools: Educational intentions and empirical findings]. In L. Weiss & C. Willmann (Eds.), *Sinnorientiert lernen – zieloffen gestalten, Zum Leistungsverständnis der Waldorfpädagogik* (pp. 93–110). LIT Verlag.
Polanin, J. R., Tanner-Smith, E. E., & Hennessy, E. A. (2016). Estimating the difference between published and unpublished effect sizes. *Review of Educational Research*, *86*(1), 207–236. https://doi.org/10.3102/0034654315582067
*Randoll, D. (1999). *Waldorfpädagogik auf dem Prüfstand. Auch eine Herausforderung an das öffentliche Schulwesen? Mit einer vergleichenden Untersuchung zur Wahrnehmung von Schule durch Abiturienten aus Freien Waldorfschulen und aus staatlichen Gymnasien*. Verl. für Wiss. u. Bildung.
Randoll, D., & Peters, J. (2015). Empirical research on Waldorf education. *Educar Em Revista*, *56*, 33–47. https://doi.org/10.1590/0104-4060.41416
Randoll, D., & Peters, J. (2021). *Wir waren auf der Waldorfschule: Ehemalige als Experten in eigener Sache [We went to the Waldorf school: Former students report]*. Beltz. http://nbn-resolving.org/urn:nbn:de:bsz:31-epflicht-1838358
Rawson, M. (2021). *Steiner Waldorf pedagogy in schools: A critical introduction*. Routledge.
Rawson, M. (2022). L2 teaching and learning in Waldorf schools – why performative? *Scenario: A Journal of Performative Teaching, Learning, Research*, *16*, 1–21. https://doi.org/10.33178/scenario.16.1.1
Rawson, M., & Bransby, K. (2025). *Crafting a curriculum of coherence: Lessons from Steiner Waldorf education*. Routledge.
R Core Team (2026). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. https://doi.org/10.32614/R.manuals
Reese, E., Sparks, A., & Suggate, S. (2012). Assessing children's narratives. In E. Hoff (Ed.), *Research methods in child language: A practical guide* (pp. 133–148). Wiley-Blackwell.
Rest, J. (1980). Development psychology and value education. In B. Munsey (Ed.), *Moral development, moral education and Kohlberg* (pp. 101–129). Religious Education Press.
Riley, R. (1996). Improving the reading and writing skills of America's students. *Learning Disability Quarterly*, *19*, 67–69. https://doi.org/10.2307/1511248
Rodgers, M. A., & Pustejovsky, J. E. (2020). Evaluating meta-analytic methods to detect selective reporting in the presence of dependent effect sizes. *Psychological Methods*, *26*(2), 141–160. https://doi.org/10.1037/met0000300
*Rohde, D. (2022). *Waldorfschulen und das Landesabitur. Eine vergleichende Studie am Beispiel des Leistungsfaches Biologie in Hessen [Waldorf schools and the state-wide Abitur: A comparative study from the subject biology in Hessen]*. Beltz Verlag.
Roorda, D. L., Koomen, H. M. Y., Spilt, J. L., & Oort, F. J. (2011). The influence of affective teacher–student relationships on students' school engagement and achievement. *Review of Educational Research*, *81*(4), 493–529. https://doi.org/10.3102/0034654311421793
*Rose, S. E., & Jolley, R. P. (2016). Drawing development in mainstream and Waldorf Steiner schools revisited. *Psychology of Aesthetics, Creativity, and the Arts*, *10*(4), 447–457. https://doi.org/10.1037/aca0000070
*Rose, S. E., & Jolley, R. P. (2020). Children's creative intentions: Where do the ideas for their drawings come from? *Journal of Creative Behavior*, *54*(3), 712–724. https://doi.org/10.1002/jocb.405
*Rose, S. E., Jolley, R. P., & Charman, A. (2012). An investigation of the expressive and representational drawing development in National Curriculum, Steiner, and Montessori schools.

Psychology of Aesthetics, Creativity, and the Arts, *6*(1), 83–95. https://doi.org/10.1037/a0024460

Sant, E. (2019). Democratic education: A theoretical review (2006–2017). *Review of Educational Research*, *89*(5), 655–696. https://doi.org/10.3102/0034654319862493

*Schieffer, J., & Busse, R. T. (2001). Low SES minority fourth-graders' achievement from an urban public Waldorf and comparison school. *Waldorf Education Research Institute Research Bulletin*, *6*, 18–24.

Schieren, J. (2023). *Handbook of research on Waldorf education*. Routledge Taylor and Francis. https://doi.org/10.4324/9781003187431

Schmidt, M., Benzing, V., Wallman-Jones, A., Mavilidi, M.-F., Lubans, D. R., & Paas, F. (2019). Embodied learning in the classroom: Effects on primary school children's attention and foreign language vocabulary learning. *Psychology of Sport and Exercise*, *43*, 45–54. https://doi.org/10.1016/j.psychsport.2018.12.017

Schuitema, J., Dam, G. T., & Veugelers, W. (2008). Teaching strategies for moral education: A review. *Journal of Curriculum Studies*, *40*(1), 69–89. https://doi.org/10.1080/00220270701294210

Schumacher, R. (2020). *Literature in Waldorf schools: Understanding literature as a basis for teaching literature* (C. Gunn, Trans; 1. Auflage). Creation + discovery Literature. Edition Waldorf.

*Seebacher, S. (2016). *Freude an der Schule. Eine empirische Untersuchung der Schultypen Neue Mittelschule und Waldorfschule [Joy at school: An empirical investigation of new middle and Waldorf schools]*. GRIN Verlag.

Sigler, S. (2020). *Mathematics in Waldorf schools: Teaching mathematics in the upper grades* (C. Gunn, Trans; 1. Auflage). Creation + discovery Mathematics. Edition Waldorf.

Sommer, W. (2010). Oberstufenunterricht an der Waldorfschule: Kognitive Herausforderungen für das verkörperte Selbst [Upper-school teaching at Steiner Waldorf schools: Cognitive challenges for the embodied self]. *Research on Steiner Education*, *1*(1), 19–32.

Sommer, W. (2020). *Physics in Waldorf schools: Teaching physics phenomenologically* (J. Mcalice, Trans; 1. Auflage). Creation + discovery Physics. Edition Waldorf.

Spinath, B., Stiensmeier-Pelster, J., Schöne, C., & Dickhäuser, O. (2002). *Skalen zur Erfassung der Lern- und Leistungsmotivation (SELLMO) [Tests of learning and achievement motivation]*. Hogrefe.

Steinmayr, R., Meißner, A., Weidinger, A. F., & Wirthwein, L. (2014). Academic achievement. In R. Steinmayr, A. Meißner, A. F. Weidinger, & L. Wirthwein (Eds.), *Education*. Oxford University Press. https://doi.org/10.1093/obo/9780199756810-0108

Suggate, S. P. (2009). School entry age and reading achievement in the 2006 Programme for International Student Assessment (PISA). *International Journal of Educational Research*, *48*, 151–161. https://doi.org/10.1016/j.ijer.2009.05.001

*Suggate, S. P., Schaughency, E. A., & Reese, E. (2011). The contribution of age and formal schooling to oral narrative and pre-reading skills. *First Language*, *31*, 379–404. https://doi.org/10.1177/0142723710395165

*Suggate, S. P., Schaughency, E. A., & Reese, E. (2013). Children who learn to read later catch up to children who learn to read early. *Early Childhood Research Quarterly*, *23*, 33–48. https://doi.org/10.1016/j.ecresq.2012.04.004

Suggate, S. P., & Suggate, T. (2019). *Reclaim early childhood: The philosophy, psychology and practice of Steiner-Waldorf early years education*. Hawthorn Press.

*Telfer-Radzat, K. (2024). Exploring alternative education: A comparison on 3 levels—Waldorf charters, non-Waldorf charters, and local public schools, as measured by the California

Assessment of Student Performance and Progress. *Frontiers in Education*, *9*, 1323535. https://doi.org/10.3389/feduc.2024.1323535

Thapa, A., Cohen, J., Guffey, S., & Higgins-D'Alessandro, A. (2013). A review of school climate research. *Review of Educational Research*, *83*(3), 357–385. https://doi.org/10.3102/0034654313483907

Tyson, R. (2024). Waldorf education – a survey of empirical research. *Research on Steiner Education*, *15*(1), 1–24.

Ullrich, H. (2024). The educational discourse on Waldorf education in Germany. In A.-K. Hoffmann & M. F. Buck (Eds.), *Critically assessing the reputation of Waldorf education in academia and the public: Early endeavours of expansion, 1919–1955* (pp. 22–46). Routledge. https://doi.org/10.4324/9781003403609-2

Urban, K. K. (2005). Assessing creativity: The test for creative thinking-drawing production (TCT-DP). *International Education Journal*, *6*, 272–280.

Vallerand, R. J., Blais, M. R., Brière, N. M., & Pelletier, L. G. (1989). Construction et validation de l'échelle de motivation en éducation (EME) [Construction and validation of the Motivation toward Education Scale]. *Canadian Journal of Behavioural Science / Revue Canadienne Des Sciences Du Comportement*, *21*(3), 323–349. https://doi.org/10.1037/h0079855

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03

Villegas de Posada, C., & Vargas-Trujillo, E. (2015). Moral reasoning and personal behavior: A meta-analytical review. *Review of General Psychology*, *19*(4), 408–424. https://doi.org/10.1037/gpr0000053

Waite, S., & Rees, S. (2011). Imagination in Steiner kindergartens: Practices and potential purposes. In R. Parker-Rees (Ed.), *Meeting the child in Steiner kindergartens: An exploration of beliefs, values, and practices* (pp. 51–67). Routledge.

Waldorf World List. (2025). *Directory of Steiner-Waldorf schools, kindergartens and teacher education centers worldwide*. https://www.freunde-waldorf.de/en/waldorf-worldwide/waldorf-education/waldorf-world-list/

*Wallner-Paschon, C. (2018). Kompetenzen, Motivation und Selbstwahrnehmung der Waldorfschüler/innen im Vergleich [Competences, motivation, and self-perception of Waldorf students compared to others]. *R&E-Source 9*, 1–13. https://journal.ph-noe.ac.at/index.php/resource/article/view/548/543

Wang, M.-T., & Degol, J. L. (2016). School climate: A review of the construct, measurement, and impact on student outcomes. *Educational Psychology Review*, *28*(2), 315–352. https://doi.org/10.1007/s10648-015-9319-1

Wang, J., Zamar, R, Marazzi, A., Yohai, V., Salibian-Barrera, M., Maronna, R., Zivot, E., Rocke, D., Martin, D., Maechler, M., Konis, K., & Todorov, V. (2024). Robust: Port of the S+ "Robust Library". https://doi.org/10.32614/CRAN.package.robust

Woodcock, R. W. (1998). *Woodcock reading mastery tests–revised normative update forms G and H examiner's manual*. Pearson Assessments.

Author Biographies

Sebastian P. Suggate is an associate professor in psychology at the University of Otago and an adjunct professor in education at the University of Regensburg; email: sebastian.suggate@otago.ac.nz. His research focuses on reading and language, fine motor skills, mental imagery, and media.

Bernhard Schmalenbach is a professor of special education at Alanus University and lectures in inclusive education at the University of Bonn; email: bernhard.schmalenbach@alanus.edu. His research focuses on special needs in education, autism, and biographies of those with Down syndrome.

Jürgen Peters is a lecturer in education and research methods at Alanus University; email: juergen.peters@alanus.edu. His research involves the empirical evaluation of reform education schools.

Appendix. Overview of Studies Included in Meta-Analysis.

Authorship	Year	$N_{Waldorf}$	N_{state}	*k*	Domains	Publication	Age
Arslan & Kartal	2022	15	15	1	Academic	Peer-review	4.6
Barz & Randoll	2012	379	1051	22	Climate	Book, dataset	14
Besançon et al.	2015	41	90	9	Attitudes, Creativity	Peer-review	12.74
Bieg et al.	2019	117	770	10	Attitudes	Peer-review	13.1
Cox & Rowlands	2000	20	20	5	Creativity	Peer-review	6.94
Cunningham & Carroll	2011	30	31	2	Academic	Peer-review	6.92
Dahlin	2007	271	477	3	Attitudes, Climate, Morality	Report	26
Dahlin	2007	196	5941	47	Attitudes, Climate, Morality	Report	15
Dahlin	2007	160	203	9	Attitudes, Climate, Morality	Report	18
Dahlin	2010	196	5941	4	Attitudes, Climate, Morality	Peer-review	15.5
Dahlin	2010	196	5941	4	Attitudes, Climate, Morality	Peer-review	18.5
Erasmus	2015	616	357	68	Academic	Dissertation	16
Erasmus & Obermeier	2010	1050	1050	34	Attitudes	Book	15
Föller-Mancini et al.	2010	186	139	11	Morality	Peer-review	16.6
Friedlaender	2015	416	30028	3	Attitudes	Paper	9
Hether	2001	25	14	1	Morality	Dissertation	17
Jelinek	2003	40	40	1	Academic	Report	12
Kirkham & Kidd	2017	20	20	2	Attitudes	Peer-review	8
Kitti et al.	2023	72	219	4	Attitudes	Peer-review	16
Kvintova et al.	2022	30	731	9	Attitudes	Peer-review	6
Mayer	2007	120	120	1	Morality	Book	17
Peters & Randoll	2013	1564	107	18	Climate	Dataset	19
Randoll	1999	371	761	130	Attitudes, Climate	Book	19
Rohde	2022	224	21062	1	Academic	Book	18
Rohde	2022	225	22329	1	Academic	Book	18
Rohde	2022	220	22949	1	Academic	Book	18
Rohde	2022	230	24566	1	Academic	Book	18
Rohde	2022	244	31435	1	Academic	Book	18
Rose & Jolley	2016	80	80	8	Creativity	Peer-review	11.8
Rose & Jolley	2020	27	30	2	Creativity	Peer-review	11.8
Rose et al.	2012	15	15	16	Creativity	Peer-review	5.42
Rose et al.	2012	15	15	16	Creativity	Peer-review	9.21
Rose et al.	2012	15	15	16	Creativity	Peer-review	7.44
Schieffer & Busse	2001	60	51	5	Academic	Report	9
Seebacher	2016	69	83	4	Academic, Attitudes	Dissertation	13.4
Suggate et al.	2011	33	35	7	Academic	Peer-review	6
Suggate et al.	2013	32	67	1	Academic	Peer-review	11.5
Suggate et al.	2013	30	63	1	Academic	Peer-review	7.7
Suggate et al.	2013	30	75	1	Academic	Peer-review	9.6
Suggate et al.	2013	50	33	4	Academic	Peer-review	12.25
Telfer-Radzat	2024	73	12805	2	Academic	Peer-review	13.8
Wallner-Paschon	2018	149	8858	3	Academic	Peer-review	15