



STEM Outside of School: a Meta-Analysis of the Effects of Informal Science Education on Students' Interests and Attitudes for STEM

Xin Xia¹ · Lillian R. Bentley² · Xitao Fan³ · Robert H. Tai⁴

Received: 12 February 2024 / Accepted: 6 September 2024
© The Author(s) 2024

Abstract

This meta-analysis explores the impact of informal science education experiences (such as after-school programs, enrichment activities, etc.) on students' attitudes towards, and interest in, STEM disciplines (Science, Technology, Engineering, and Mathematics). The research addresses two primary questions: (1) What is the overall effect size of informal science learning experiences on students' attitudes towards and interest in STEM? (2) How do various moderating factors (e.g., types of informal learning experience, student grade level, academic subjects, etc.) impact student attitudes and interests in STEM? The studies included in this analysis were conducted within the United States in K-12 educational settings, over a span of thirty years (1992–2022). The findings indicate a positive association between informal science education programs and student interest in STEM. Moreover, the variability in these effects is contingent upon several moderating factors, including the nature of the informal science program, student grade level, STEM subjects, publication type, and publication year. Summarized effects of informal science education on STEM interest are delineated, and the implications for research, pedagogy, and practice are discussed.

Keywords Informal science education · Meta-analysis · STEM · Student interests and attitudes

STEM careers and opportunities, encompassing Science, Technology, Engineering, and Mathematics, are fundamental drivers of economic stability within

✉ Xin Xia
baj5jy@virginia.edu

¹ University of Virginia, 405 Emmet St S, Charlottesville, VA, USA

² Georgia State University, 33 Gilmer Street SE, Atlanta, GA 30303, USA

³ The Chinese University of Hong Kong, Hong Kong SAR, China

⁴ Australian Catholic University, 25A Barker Road, Strathfield, NSW 2135, Australia

contemporary society (Xie & Killewald, 2012; Xie et al., 2015). The United States has demonstrated a strategic commitment to fostering successive generations of STEM-focused individuals to maintain global competitiveness within the modern market (National Research Council [NRC], 2007). An educational priority has thus been the cultivation of student interest in STEM disciplines. It is widely acknowledged that students exhibiting such interest possess heightened prospects for navigating the STEM pipeline toward attaining a career in the field (Beier & Rittmayer, 2008; Business-Higher Education Forum, 2011; Tai et al., 2006). Attainment of a STEM career not only bolsters individual economic stability but also amplifies the potential for national STEM-driven innovation (Xie et al., 2015).

Students gain STEM experiences through interacting in formal (such as that occurring in schools, colleges, and universities) and informal learning environments (that occur everywhere else; Krishnamurthi & Rennie, 2013). On average, children spend approximately 20% of their time learning in formal educational environments (Eshach, 2007; Falk & Dierking, 2010; Sacco et al., 2014). This suggests that informal learning experiences could contain enormous potential to strengthen and enrich school STEM experiences (Bevan et al., 2010; NRC, 2009; Phillipset al., 2007). The United States has a variety of informal learning institutions for people to engage in STEM learning opportunities, including public libraries, zoos, aquariums, and museums (Falk & Dierking, 2010). Nevertheless, there is limited understanding regarding the specific types of informal experiences that ignite and sustain children's STEM learning (National Academies of Sciences, Engineering, and Medicine [NASEM], 2016). Therefore, understanding how these opportunities support STEM interest is an important area of research.

Research into the effects of informal science programs highlights that situational factors can create conflicting results. On the one hand, informal science learning was shown to promote and strengthen student understanding of school science and interest in STEM (Bevan et al., 2010; Maiorca et al., 2021; Phillips et al., 2007). On the other hand, however, informal science experiences could have negative effects on students' attitudes toward STEM (Migliarese, 2011; Shields, 2010). The inconsistency of such research findings suggest that research attention is needed to understand how informal science learning may affect students' attitudes toward STEM.

The aim of this investigation was to conduct a current quantitative meta-analysis examining the impacts of informal science programs on students' attitudes and interests towards STEM disciplines. This meta-analysis offers fresh perspectives on potential determinants underlying the variation in effect sizes observed across diverse studies. These nuanced insights offer valuable guidance for educators, policymakers, and practitioners in their endeavor to create informal science learning initiatives aimed at cultivating and nurturing students' interest and attitudes towards STEM.

Literature Review

Formal and Informal Learning Experiences

Formal and informal learning environments can support STEM-focused learning experiences (Decoito, 2014; Frantz et al., 2011). Formal STEM learning is any experience or activity that happens within a normal class period and at a school. Informal science learning is any activity or experience that occurs outside of the regular school day, and it includes activities that are not developed as part of an ongoing school curriculum (Crane, 1994). Formal STEM learning experiences tend to be mandatory, whereas informal science learning experiences are mostly voluntary (Crane, 1994). For this meta-analysis, we include school-based field trips, student out-of-school projects (after-school camps), community-based science youth programs, visits to museums and zoos, after-school programs, summer camps, and weekend school camps as informal science activities (Hofstein & Rosenfeld, 1996; So et al., 2018).

Attitude and Interest Toward STEM

Various conceptualizations of attitudes and interests in STEM exist within scholarly discourse. Gauld and Hukins (1980) delineated attitudes towards scientists, scientific inquiry, science learning, science-related activities, science careers, and the embrace of scientific attitudes as affective behaviors indicative of interests in science. Alternatively, Osborne et al. (2003) posited that attitudes towards science encompass emotions, beliefs, and values pertaining to science and its societal impact. Additionally, Potvin and Hasni (2014) categorized attitudes into distinct dimensions such as motivation, enjoyment, interest, self-efficacy related to being able to complete STEM tasks and STEM career aspirations.

There are diverse definitions of attitude and interest in STEM. Gauld and Hukins (1980) proposed that attitudes toward scientists, scientific inquiry, science learning, science-related activities, science careers, and the adoption of scientific attitudes were all affective behaviors representing interests in science. Osborne et al. (2003) proposed that attitudes toward science are composed of feelings, beliefs, and values held about science, and the impact of science on society. Attitudes can also include the ideas of enjoyment, motivation, interest, self-efficacy, and career aspirations (Potvin & Hasni, 2014).

Drawing from a comprehensive examination of pertinent scholarly literature concerning attitudes toward science, and extending this framework to encompass STEM disciplines, we operationalized attitude toward STEM across four distinct dimensions: interest, self-efficacy, overall attitude toward STEM, and career interests (Wiebe et al., 2018). Interest in STEM pertains to the affective domain, encapsulating individuals' emotions and sentiments regarding the learning of scientific subjects (e.g., Zhang & Tang, 2017). Self-efficacy refers to students' perceptions

of their own capabilities to excel in STEM domains, as articulated by Bandura (1995). Attitude toward STEM denotes individuals' perspectives on the value, utility, and societal ramifications of science (e.g., Dowey, 2013). Lastly, STEM career interest is delineated as the extent to which students can furnish meaningful insights into their aspirations for future careers within STEM disciplines (Maltese & Tai, 2011).

This meta-analysis extends the scope of prior research conducted by Young et al. (2017). Young et al. completed a meta-analysis spanning 2009–2015 on the impact of out-of-school time learning on students' interests in STEM. The authors operationally defined out-of-school time as encompassing summer enrichment programs and after-school activities. Their findings indicated a favorable influence of out-of-school time on students' inclination towards STEM subjects. Furthermore, the variability in these effects was found to be moderated by factors such as the methodological rigor of the research design, the thematic focus of the programs, and the grade level of the participants.

Our analysis built on this foundation in several keyways. First, we extended the scope of time by including 30 years of empirical studies (1992–2022), compared to six from Young et al. (2017). We also broaden the types of programs evaluated by looking at all informal programs including field trips, outreach programs, and mobile labs. Additionally, we looked at the moderating effects of STEM as an entire discipline, then Science, Technology, Engineering, and Mathematics separately. We included the additional moderators of publication type, and year of study in the overall analysis. Finally, we extended the outcome variable by examining student interest, attitude, and self-efficacy toward STEM. The additional moderators and scope of the study will considerably increase the breadth and depth of the empirical knowledge base about the effects of informal science education experiences on students' interests and attitudes toward STEM.

Potential Moderators

The following literature review aims to elucidate salient findings concerning the different moderators used in this analysis. Specifically, the review focuses on the scholarship concerning different distinct categories of informal science programs, the influence of grade-level variations on students' STEM interests, the diverse typologies of STEM activities, as well as the impact of publication type and publication year on the overall effect of informal programming on students interests and attitudes toward STEM. Additionally, an examination of pertinent empirical conclusions concerning student interest, attitude, and self-efficacy regarding informal science programs will be summarized.

Type of Program

There are many different types of informal science programs. In this analysis, we consider after-school programs, summer camps, outreach programs, weekend school camps, field trips, mobile labs, and a mixture of these different

out-of-school experiences to be included in informal education. Young et al. (2017) concluded that out-of-school programming (after school or summer school) did not produce significant effects on students' attitudes toward STEM. We are interested in re-examining the effects of after-school programs on students' attitudes towards STEM and will also include other types of informal science experiences in our analysis. Variability in the outcomes of studies examining the association between informal STEM programs and students' interest levels in STEM may be attributable to the diversity of program structures investigated. Consequently, there exists potential for enriching our comprehension of informal science programming through the exploration of alternative forms of informal science experiences.

After-School Programs After-school programs are any type of informal program that happens after-school and within or outside of school property. The programs can be administered by the school or by community-based organizations (Dryfoos, 1999). In this analysis, we define after-school programs as programs administered by the school.

Summer Enrichment Programs and Weekend School Camps Summer camps and weekend school camps represent extracurricular educational endeavors that occur beyond the confines of the regular academic calendar, particularly during weekends and summer (Young et al., 2017). In the context of our investigation, we defined summer enrichment programs and weekend school camps as initiatives with a distinct focus on Science, Technology, Engineering, and Mathematics (STEM). Summer enrichment programs have been associated with accelerated learning trajectories including early graduation (Berliner, 2009), while concurrently catering to the educational needs of culturally and linguistically diverse students, as well as those hailing from economically disadvantaged backgrounds (Keiler, 2011; Matthews & Mellom, 2012).

However, discrepant to these positive associations, Young et al. concluded a lack of significant moderation effect exerted by summer camps on student interest in STEM disciplines. Concurrently, empirical inquiry into the impact of weekend school camps remains notably scarce, thus creating a paucity in the research space to investigate how these activities support interest in and attitudes toward STEM.

Outreach Programs We define outreach programs as any STEM activity that is organized by an outside institution. There can be various formats to the program, from lecture-style outreach where a visiting professor will talk to a group of students about STEM, to a field trip to a university where STEM career pathways are highlighted (Tillinghast et al., 2020). Limited studies have measured the effects of these programs on students' interests and attitudes toward STEM, highlighting a need for research expansion, especially using meta-analysis techniques.

Field Trips and Virtual Museums Alternative forms of informal science program engagements include field trips and virtual museum experiences. We delineate

field trips and virtual museums as occasions wherein students embark on guided tours of STEM facilities, either physically or virtually, during school hours but outside the structured curriculum timeframe. These experiences have the potential to influence students' academic achievement and foster interest in STEM careers. For instance, secondary school students who interacted with scientists during field trips demonstrated heightened awareness regarding STEM career pathways (Jensen & Sjaastad, 2013). Furthermore, research indicates that students exhibited enhanced performance in mathematics subsequent to participating in a field trip to the New York Hall of Science (Alliance, 2011). Notably, despite these findings, there remains a paucity of systematic investigations utilizing meta-analytical approaches to scrutinize the collective impact of such field trip experiences.

Mobile Labs The last type of informal science experience that we analyzed was the use of mobile labs. Mobile labs are mobile vehicles and buses that transport STEM labs for students to experience hands-on science at their schools. They first became popular in the late-1990s and currently have 29 member programs in 17 different states (Jones & Stapleton, 2017).

Mixed Studies Some programs are a combination of informal experiences. When we could not categorize an informal experience as one separate group, we analyzed the effect of the program in a mixed category.

Grade

Recent empirical scholarship focused on the association of students' grade level and their interest and attitude towards STEM has produced conflicting conclusions. Scholars indicated a notable trend whereby students' attitudes and perceptions towards science and STEM disciplines diminish with increasing age as they progress through their educational schooling (George, 2000; Morrell & Lederman, 1998; Murphy & Beggs, 2003; Osbourne et al., 2003; Silver & Rushton, 2008). Additionally, these scholars underscored a discernible decline in student interest and enjoyment of science from intermediate to high school (George, 2000; Morrell & Lederman, 1998). Conversely, research conducted by Maltese and Tai (2011) revealed a positive association between eighth-grade students' perceptions of science's utility for their future and their likelihood of pursuing STEM degrees. Additionally, Sadler et al. (2012) established that students' career aspirations in STEM fields upon entering high school emerged as robust predictors of their vocational interests upon completing high school.

In summary, student attitudes towards STEM and their STEM career inclinations undergo a dynamic evolution throughout their elementary and secondary educational experiences. Once students reach early high school, research indicates that students' interests and attitudes towards STEM solidify and can become predictors

of their career choices later in college. Leveraging meta-analytic methodologies, our study aimed to augment the existing body of research by exploring the differential impacts of elementary and secondary education levels on students' overarching attitudes and interests in STEM disciplines as a result of their informal STEM experiences.

Subjects

Scholarship investigating how students' attitudes and interests might change when they study Science, Technology, Engineering, and Mathematics (STEM) as individual subjects and in conjunction, represents a research space that has limited conclusions. Looking at STEM experiences in conjunction, Wiebe et al. (2018) concluded that there was a reciprocal relationship between students' STEM experiences and the development of specific STEM career aspirations. Notably, success in mathematics (independent of STEM) was positively associated with the likelihood of students pursuing advanced education in STEM fields (Wang, 2012). Informal STEM learning activities have been shown to affect students' self-efficacy related to mathematics (Jiang et al., 2024). Likewise, heightened self-efficacy related to science among students augments the propensity for embarking upon a career trajectory within the STEM domain. Additionally, throughout high school, an elevated interest in STEM is associated with a student's sustained commitment to undertaking advanced coursework in both mathematics and science (Simpkins et al., 2006). Overall, when students have positive experiences in STEM both interdisciplinary and individually, these experiences have a positive association with their STEM career interests and higher education aspirations.

Most STEM research focuses on science and mathematics, neglecting technology and engineering's impact on student attitudes and interest. This opens up a paucity in this research space to investigate the moderating effect of STEM as an entire discipline, and science, technology, engineering, and mathematics as separate subjects as they are associated with students' interests and attitudes towards informal science experiences.

Learning Outcomes

As delineated earlier, we operationalized attitude in four distinct facets: interest in STEM, self-efficacy related to STEM, attitude towards STEM, and STEM career interests. Interest in STEM pertains to the emotions and sentiments pertaining to the process of learning STEM subjects (Zhang & Tang, 2017). Self-efficacy, on the other hand, revolves around students' convictions regarding their competencies to excel in STEM endeavors, as well as their perseverance in persisting within STEM domains (Bandura, 1995). Attitude towards STEM signifies individuals' perceptions regarding the inherent value and significance of STEM disciplines (Dowey, 2013). Lastly, STEM career interest denotes students' proclivity towards prospective careers within STEM domains (Maltese & Tai, 2011).

Publication Type

Within the realm of meta-analysis, publication type stands out as a prominent moderator variable. Our study notably incorporates both peer-reviewed journal articles and dissertations. It is well-established that studies reporting statistically significant findings are more likely to be published compared to those reporting non-significant results, a phenomenon commonly termed publication bias (Rosenthal, 1979). The presence of heterogeneous findings across studies may arise from differences in publication types rather than alternative moderator variables. Consequently, we systematically examined publication type as a potential moderator variable within this meta-analytic inquiry, classifying it into two categories: "journal article" and "thesis_dissertation," with the latter encompassing both theses and dissertations.

Publication Year

Given the dynamic nature of education policies spanning from 1992 to 2022, exemplified by the publication of the Next Generation Science Standards (NGSS), the formats and instructional methodologies employed in science education have undergone significant transformations over time (NRC, 2011). Consequently, these alterations can change long-term attitudes towards and interest in STEM. Considering the possible changes in STEM education over the past 30 years, we included publication year as a potential moderator.

Aim of the Meta-analysis

Informal science learning plays an important role in science learning and contains various formats. Nonetheless, the effect of informal science learning in general, and the possible difference among different informal science learning settings in particular, have not been fully examined. The purpose of this meta-analysis is to explore whether informal science education is effective in increasing students' learning interests and attitudes toward STEM. The following research questions were explored:

- (1) What is the overall effect size of informal science learning experiences on students' attitudes towards and interest in STEM?
- (2) How do various moderating factors, including the type of informal learning experience (such as museum visits, out-of-school programs, after-school activities), student grade level, academic subjects (science, technology, engineering, mathematics, or the broader STEM domain), type of publication (dissertation or peer-reviewed journal article), and publication year, impact student attitudes and interests in STEM?

Method

Literature Search for Primary Studies

In this study, we sought primary studies from Proquest, EBSCO, Web of Science, and ScienceDirect. We concentrated on empirical studies exploring the effect between informal science education and K-12 students' interest in science. This encompassed peer-reviewed journal articles as well as unpublished dissertations or conference papers. Keywords used were as follows: (“Interest” OR “attitude”) AND (“out-of-school” OR “informal” OR “after school”) AND (“science education”). In addition, we reviewed articles that were cited in a previous meta-analysis (Young et al., 2017) and examined each article for potential addition to our analysis using Google Scholar (scholar.google.com). Articles were filtered by time-frame (1992–2022), language (English), and location (United States). In this study, we were interested in the effects of informal science programs within the United States. We conducted our search using the keywords individually or in various combinations and did not restrict the publication status, hence we can have both gray literature and journal articles.

Inclusion and Exclusion Criteria

For inclusion in this meta-analysis, we listed the criteria that studies must satisfy below:

1. A study necessitates the exploration of an informal science learning setting wherein explicit documentation of informal activities is provided. The study should align with the established criteria for informal learning activities as mentioned previously. The criteria were based on those used by Hofstein and Rosenfeld (1996).
2. A study was a quasi-experimental study, with an experimental group or groups (e.g., groups of students involved in an informal science learning program or activities) and a control or comparison group (business as usual or not involved in an informal science program or activities), respectively. Studies that lacked a control/comparison group, wherein participants did not engage in informal science learning programs, were systematically excluded from the analysis. For instance, Knapp and Barrie (2001) compared two field trips to a science center, and Wilson et al. (2012) tested the effectiveness of two versions of a film for part of the science center planetarium demographic to compare children's learning and attitude changes in response to films. These studies lacked a control group, and were not included in this study.
3. A pretest–posttest design was included in the study, typically involving collecting data from participants at two distinct time points: before the implementation of an informal science learning intervention or treatment (pretest) and after the intervention or treatment had been administered (posttest). The studies involve

the same group of participants being measured on the same variables at both the pretest and posttest stages. Studies that did not collect the pretest data were excluded from this study.

4. A study had to include students' interests, attitudes, and/or self-efficacy as outcomes. In our meta-analysis, we define attitude towards STEM in four ways (as mentioned above): interest, self-efficacy, attitude toward STEM, and career interests, as mentioned above. Studies that investigated the impacts of informal science learning on achievement were excluded from our study.
5. To be eligible for inclusion, a study had to provide sufficiently detailed quantitative data that facilitated the calculation and extraction of the relevant relationship as an effect size. This criterion ensured that the selected studies provide the necessary information required for effect size estimation, thereby enabling a rigorous analysis of the relationship under investigation. If a study only contained a qualitative interview for analysis, we did not include this study in our analysis because we could not calculate effect sizes.
6. The present study exclusively focused on samples comprising students from grades K-12, excluding college students or other participant groups.
7. A study must be published or available in English.

Study Selection

As illustrated in a PRISMA flowchart (Fig. 1), our initial search yielded 1042 potentially relevant studies as previously used by Maltese and Tai (2011). We included studies from published journal articles, and gray literature which included conference papers, and theses/dissertations. On the first round of screening, we removed 291 duplicate articles, hence 751 studies remained. Secondary round screening was to review the title and abstract of each article and determine if the study focused on the impact of informal science education on students' interests and attitudes. At this stage, 192 studies were reminded for further full-text examination.

During the third round of screening, a comprehensive review of all articles was conducted independently by both the first and second authors to assess the eligibility of the 192 studies against the pre-established inclusion and exclusion criteria. Although an initial discrepancy emerged between the two reviewers, resolution was achieved through deliberation among the entire research team to determine the final inclusion status of each study.

As is common in many meta-analytic reviews, among the included primary studies, some, or even many, contain multiple effect sizes due to the following reasons: a) testing students' attitudes and interests in different subjects (e.g. science, mathematics, technology, or engineering, etc.); b) measuring students' attitudes in various dimensions (e.g. attitudes towards STEM content learning, attitudes towards STEM career). Subsequently, a total of 19 studies, comprising 68 effect sizes, were deemed to meet the predetermined inclusion criteria and were consequently incorporated into the meta-analysis. The cumulative sample size across these 19 studies amounted to 6160 participants.

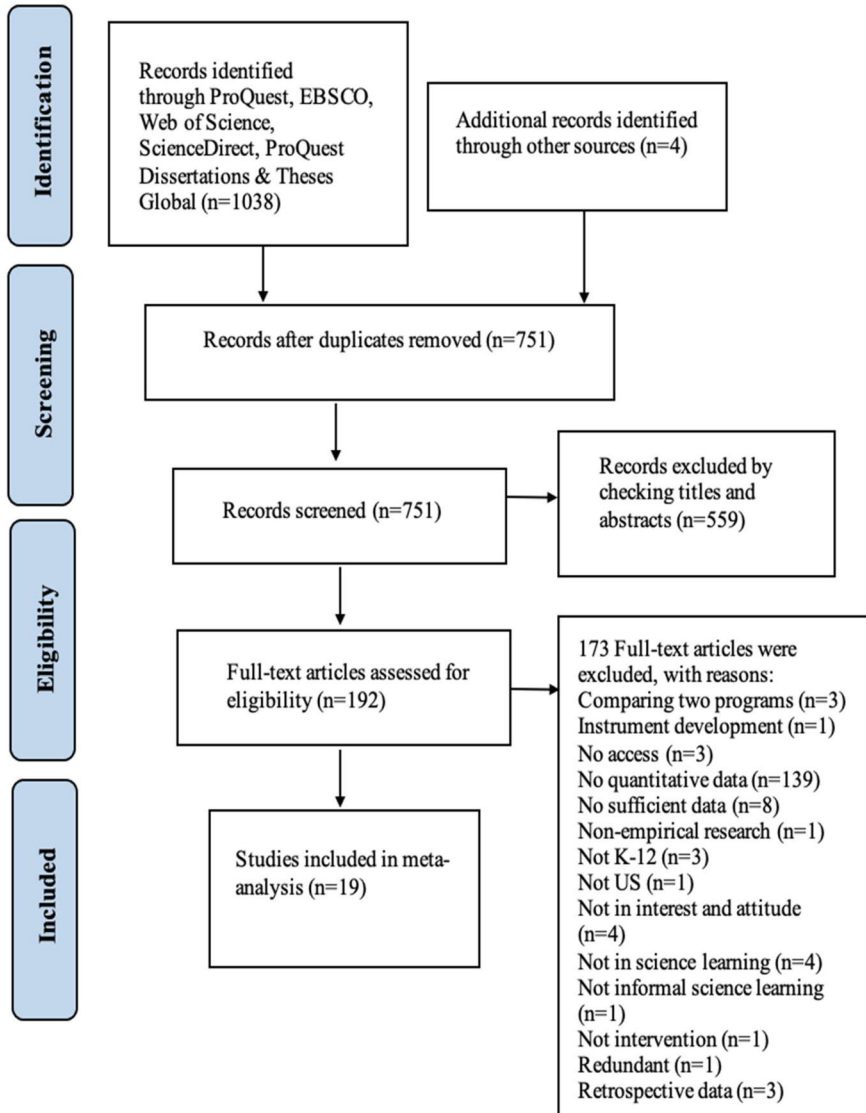
Flowchart of the inclusion and exclusion in the meta-analysis

Fig. 1 Flowchart of the inclusion and exclusion in the meta-analysis

Coding Process

Two authors of the article finished coding. Initially, the two coders collaborated in coding 25% of the primary studies together, which was the “trial coding” phase. The discordances encountered between the two coders were effectively addressed through a process of deliberation undertaken within the research team, whereby the

identified issues observed during the coding procedure were thoroughly examined and subject to comprehensive discussion. Following the initial trial coding phase, the research team proceeded to establish a definitive coding table. Subsequently, two coders independently conducted coding on the remaining primary studies. Upon completion of the final coding phase, no significant disparities were observed, although minor variances were addressed through collaborative discussion.

Coding of Study Characteristics as Variables

Information from each study was coded about informal science learning program characteristics, student sample, publication year, and publication type as detailed below.

Type of Program

Studies were conducted on informal science learning in different formats. We divided the type of programs into eight categories: after-school, summer camp, outreach program, weekend school camp, field trip or a virtual museum, mobile lab, and mixed program. While studies contained more than one category of informal activities, we coded them as mixed programs.

Grade

Four categories of grades were coded: elementary (grade K-5), middle (grade 6–8), high (grade 9–12), and mixed grade in which studies included participants across school levels.

Subjects

Research studies focusing on the examination of students' learning interests or attitudes in individual subjects, including science, mathematics, engineering and technology, and in STEM as a whole, were specifically considered in this analysis. Only those studies that encompassed all four domains of science, mathematics, engineering, and technology were coded and categorized as STEM. In this study, primary studies examined engineering and technology together, separate from science and mathematics. Therefore, we included an engineering and technology category.

Outcomes

Studies were categorized into four outcomes: self-efficacy, attitude, interest (i.e., in subject content), and career interest (i.e., in STEM careers).

Publication Type

Another frequently utilized moderator variable in meta-analysis is Publication type (Cai et al., 2017). In this study, we coded studies into two categories: journal articles or thesis/dissertations.

Publication Year

The last moderator in this study is the publication year that recorded the year of studies that were published.

Procedures of Meta-analytic

Calculation of Effect Size

As described previously, the studies included in this meta-analytic review were based on the inclusion criteria mentioned above, and these studies had pre- and post-measures of interest/attitude of STEM subjects, thus providing interest/attitude change scores between pre- and post-measures from both experimental and control groups. Such a design was described as pretest–posttest-control (PPC) in the literature (Morris, 2008), which could be either experimental design with randomized assignment, or quasi-experimental design without randomized assignment (Morris, 2008). Becker (1988) presented an effect size measure for such PPC design, which was essentially the difference of standardized mean change scores between the two groups (treatment vs. control). Built upon Becker’s work on the effect size measure for the PPC design, Morris (2008) provided an in-depth discussion and empirical assessment about several alternative forms of effect size measures in such a pretest–posttest-control (PPC) design. The empirical findings in Morris (2008) showed that one effect size measure (labeled d_{ppc2} in Morris, 2008, p. 369), has better and more robust performance than other alternatives. Guided by the empirical findings and conclusions of Morris (2008), in the current meta-analytical review, we used d_{ppc2} as the effect size measure between the students in an informal science learning program vs. those with no informal science learning. d_{ppc2} is based on the difference of standardized change scores (i.e., posttest score – pretest score) of the two groups, and it is conceptually equivalent to Hedges’ g (Hedges & Olkin, 1985). The technical details of d_{ppc2} and other alternative forms are available in Morris (2008). For the sake of simplicity for our readers, in the following, we will use the well-known g , instead of d_{ppc2} , in our presentation and discussion.

If a study did not directly provide the components needed for calculating effect size as described above, but provided sufficient other statistics (e.g., t-statistic, F-statistic, odds ratio, etc.) that allowed us to obtain the effect size measure based on available conversion formula in the literature (e.g., Hedges & Olkin, 1985), the effect size was obtained. In this context, a positive value of g was interpreted as an

indication of improved performance by the treatment group of students who participated in an informal science learning program over the control group of students.

Random Effect Meta-Analysis Model

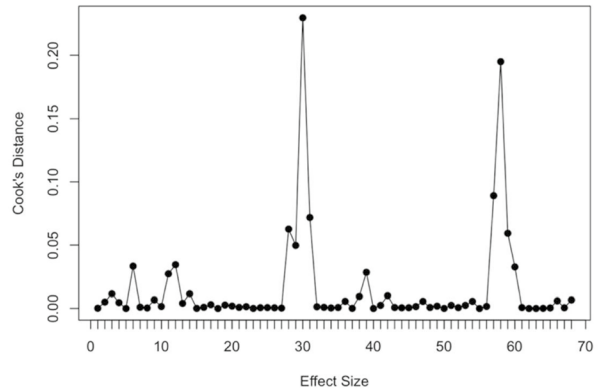
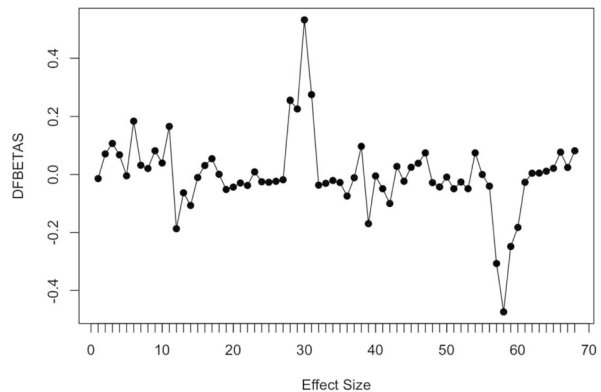
Some primary studies included in this meta-analysis presented multiple effect sizes within one study, thus, we used a random-effects model for analyzing the effect sizes. A random-effects model assumes that the effects of the variables are random and can vary across studies, depending on different study conditions. It is suitable when there is unobserved heterogeneity between the units or individuals. The weighting scheme in the random-effects model incorporates within-study variance alongside a constant value (T^2), representing between-study variance, reducing the relative differences among the weights. Consequently, a random-effects model promotes a more balanced distribution of relative weights across studies compared to a fixed-effect model (Borenstein et al., 2010). This study reported and interpreted the weighted average effect sizes, confidence intervals (lower and upper limits), and z-test results. We used the R statistical platform (R Core Team, 2019) with Viechtbauer's (2010) *metafor* package for analysis. The random effect coefficients were estimated using the maximum likelihood estimation method. Q statistics addressed the homogeneity of effect sizes.

Moderator Analysis

When Q statistics showed statistical heterogeneity across the studies, moderator analyses were conducted to test the potential study features that could have contributed to the inconsistency among the study's effect sizes. This study extracted four categorical variables (i.e., type of program, grade, disciplines, and type of publication) and one continuous variable (i.e., publication year). Significant moderation effects were assessed through the utilization of two distinct approaches. One is the omnibus test for categorical moderators which was employed to determine the presence of statistically significant moderation effects when dealing with categorical moderators, while the other is the slope analysis for the continuous moderator, which was utilized to identify significant moderation effects (Cai et al., 2022).

Publication Bias

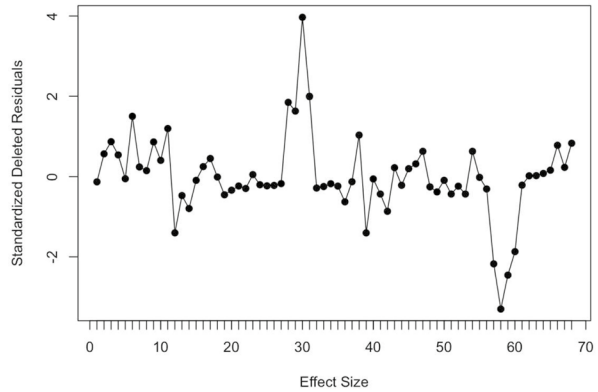
Publication bias poses a significant challenge to the validity of meta-analytic findings. Research demonstrating large effect sizes or statistically significant findings may have a higher likelihood of being published and consequently included in a meta-analysis compared to studies with small effect sizes or statistically nonsignificant results (Mao et al., 2021; Rosenthal, 1979). As publication bias has the capacity to distort estimations of the true effect being investigated, it poses a pervasive challenge when conducting a meta-analysis (Thornton & Lee, 2000). To address the publication bias, this study used the funnel plot and Egger's regression to assess potential publication bias. As Rothstein et al. (2005) discussed, the absence of bias can be inferred when the funnel plot demonstrates a symmetrical distribution. Egger

Fig. 2 Cook's distances of the effect sizes*Cook's Distances of The Effect Sizes***Fig. 3** EFBETA values of effect sizes*EFBETA Values of Effect Sizes*

et al. (1997) proposed a linear regression method for evaluating publication bias, which a p-value greater than 0.05 in this test suggests the absence of publication bias.

Results

Viechtbauer and Cheung (2010) demonstrated that meta-analyses need to include influential case diagnostics to identify outliers or extreme effect sizes and separate them from the rest of the data. We used Cook's distances (Fig. 2), DFBETAS (Fig. 3), and standardized deleted residuals (Fig. 4) to detect the potential outliers. As Figs. 2, 3, and 4 showed below, two effect sizes (i.e., the 30th and 58th effect sizes) were identified as influential outliers. After excluding the influential outlier, the revised pooled effect size was determined to be 0.21 (95% CI [0.12, 0.30], $p < 0.001$), demonstrating

Fig. 4 Standardized deleted residuals of effect sizes*Standardized Deleted Residuals of Effect Sizes*

close similarity to the previous pooled effect size of 0.21 (95% CI [0.10, 0.32], $p < 0.001$). Sensitivity analysis showed that the overall effect size remained virtually unchanged even after removing influential outliers. Therefore, we can conclude that including influential outliers did not change the main results of our meta-analysis.

We include independent studies in the analysis that spanned from 1997 to 2022. A total of 19 studies were incorporated, yielding a comprehensive set of 68 effect sizes. Data sheet can be provided by contacting the corresponding author. These effect sizes exhibited a diverse range, extending from -1.15 to 1.59, with a median value of 0.17. The magnitude of the effect size and the accuracy of its estimation differ. The majority of effect sizes were positive (81%), while 10 effect sizes were negative and three were close to zero.

The Q test was statistically significant (As shown in Table 1), indicating significant heterogeneity across the effect sizes (Q ($df=67$) = 1663.24, $p < 0.001$), with a total heterogeneity (I^2) of 91.63%. This means that the variation in effect sizes across the studies cannot be explained solely by sampling error (Borenstein et al., 2017). We can reject the null hypothesis that the true effect sizes are homogeneous. The true effectiveness of informal science programs appears to differ across the studies.

As shown in Table 1, the estimated effect size of 0.21 with a 95% confidence interval ranging from 0.10 to 0.32 suggests a statistically significant effect. Kraft (2020) put forth a comprehensive framework aimed at elucidating effect sizes within the context of educational interventions targeting the academic achievement of pre-K-12 students. Specifically, the author delineated the categorization of effect sizes based on their magnitude and argued that effect sizes below 0.05 donate a small effect size,

Table 1 The overall analysis

Outcome classification	Studies	Effect sizes	Mean g (SE)	95% CI	z-score(sig)	Q_{total}	I^2
Overall analysis	19	68	0.21(0.06)	(0.10,0.32)	3.74 ***	1663.2439	91.63%

while effect sizes ranging from 0.05 to 0.2 represent a medium effect size, and effect sizes surpassing the threshold of 0.2 were identified as indicative of a large effect size. Therefore, interpreted within this framework, 0.21 in our study was characterized as a statistically significant medium to large effect size in terms of the overall effect of informal science learning programs on students' interests or attitudes in STEM.

"Forest plot" is a graphic tool that presents the effect sizes of each study and the overall effect size derived from random-effect modeling in one graph. Figure 5 presents all 68 effect sizes. The position of the square dot along the horizontal axis represents the estimated effect size for a given study and the size of the dot indicates the weight or precision of the study's estimate. Confidence intervals (CI) were presented by the horizontal bars extending from the square dot with two ends where the bar's length shows the confidence interval's width. In addition, the overall effect size is located below all individual effect sizes in a diamond shape, and its width represents the 95% confidence interval around the summary estimate. The vertical dot line is the "null" effect, meaning that there is zero effect by informal science learning activities on students' learning interests and attitudes. Since the "null" effect line did not cross the diamond shape, we concluded that the overall "effect" was statistically significant.

Specifically, in Fig. 5, the predominant trend manifests as positive outcomes, with a notable presence of 55 positive effect sizes positioned to the right of the "null"

Forest Plot

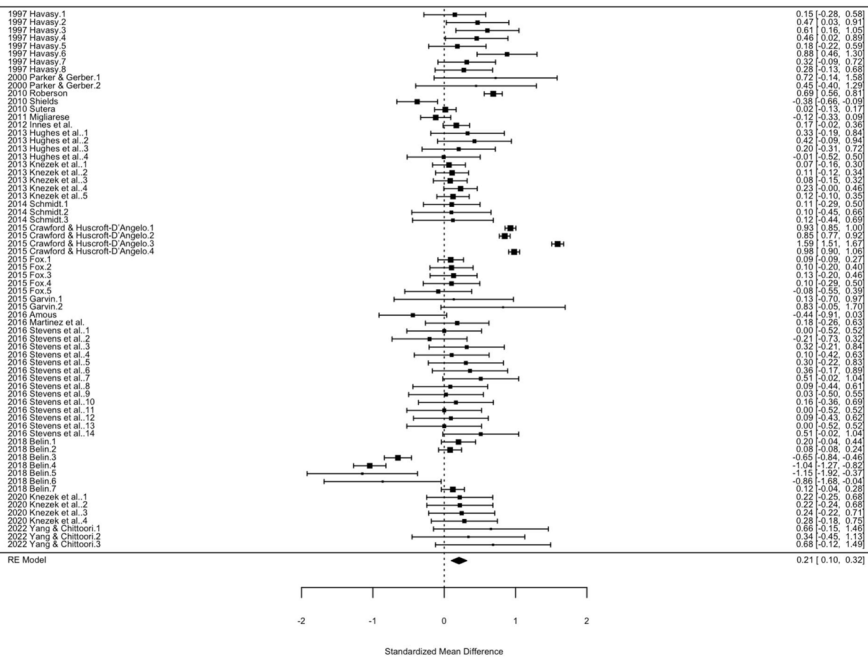


Fig. 5 Forest plot

effect line. Conversely, 10 effect sizes exhibited negative trends, while three were proximate to zero. It is noteworthy that certain effect sizes demonstrated exceptionally broad confidence intervals, as evidenced by Garvin (2015) and Parker and Gerber (2000), while others exhibited narrower confidence intervals, exemplified by Crawford and Huscroft-D'Angelo (2015) and Roberson (2010). The sample size associated with an effect size affected the confidence interval width, with smaller sample sizes associated with wider confidence intervals. Consequently, in order to mitigate against the potential influence stemming from variations in confidence intervals on the overall outcomes of the study, it was necessary to assign appropriate weights to the effect size within a meta-analysis. The accumulated overall effect size is based on weighted individual effect sizes across the studies, with effect sizes from larger samples weighted more than the effect sizes from smaller samples.

Moderators Analysis

We conducted random effects analysis to explore how moderators influence the effects of informal science learning programs. Table 2 summarizes the results, and the following discussion will address each moderator separately below.

Type of Program

The effect of informal science learning was found to be moderated by the type of program. As the omnibus test show, different types of the program explained the effect-size heterogeneity was statistically significant ($Q_{moderators} (df=7)=31.80, p<0.0001$), indicating that the effect sizes from the studies based on different programs could differ statistically. Under this moderator of *Type of Program*, Outreach Program stood out as showing the largest average effect size of 0.72, which is also statistically significant (Mean $g=0.72$; 95% CI [0.43, 1.02]; $p=0.0043$). Studies under other types of programs showed smaller effect sizes in general, although all positive.

Grade

Similar to the results for program type above, the grade was also a statistically significant moderator ($Q_{moderators} (df=4)=30.48, p<0.0001$). The studies involving middle school students (Mean $g=0.43$; 95% CI [0.12, 0.59]; $p=0.0033$) and high school students (Mean $g=0.42$; 95% CI [0.15, 0.69]; $p=0.0024$) showed larger effect sizes than those involving elementary school students (Mean $g=-0.24$; 95% CI [-0.77, 0.29]; $p=0.3657$) and mixed grade students (Mean $g=0.08$; 95% CI [-0.06, 0.21]; $p=0.2490$).

Subjects

We found a statistically significant moderating effect of subjects ($Q_{moderators} (df=4)=19.53, p=0.0006$). This suggests that the studies that examined interests

Table 2 Moderator analysis of moderators

Categorical moderators	No. studies	No.ES	$Q_{moderators}$ (<i>p</i> value)	Mean <i>g</i>	<i>z</i>	<i>p</i> value	95% CI
Type of program	19	68	31.80 (<0.0001****)				
After school	3	11		0.19	1.49	0.1363	(-0.06, 0.45)
Summer camp	3	7		0.18	1.02	0.3055	(-0.16, 0.52)
Outreach program	3	6		0.72	4.80	<0.0001****	(0.43, 1.02)
Weekend school camp	1	4		0.24	1.11	0.2654	(-0.18, 0.67)
Field Trip_ Virtual museum	4	17		0.02	0.18	0.8572	(-0.18, 0.21)
Mobile lab	2	6		0.19	1.19	0.2347	(-0.12, 0.51)
Mixed program	3	17		0.19	1.68	0.0933	(-0.03, 0.41)
Grade	19	68	30.48 (<0.0001****)				
Elementary	2	2		-0.24	-0.90	0.3657	(-0.77, 0.29)
Middle	4	16		0.43	4.37	<0.0001****	(0.24, 0.63)
High	2	10		0.42	3.03	0.0024**	(0.15, 0.69)
Mixed grade	11	40		0.08	1.15	0.2490	(-0.06, 0.21)
Subject *	19	68	19.53 (0.0006****)				
Science	17	51		0.14	2.15	0.0318*	(0.01, 0.26)
Math	4	5		0.44	2.18	0.0290*	(0.05, 0.84)
Engineer & technology	3	4		0.48	2.22	0.0263*	(0.06, 0.90)
STEM	5	8		0.36	2.29	0.0223*	(0.05, 0.68)
Outcomes	19	68	22.08 (0.0002****)				
Self_efficacy	1	2		0.27	0.81	0.4207	(-0.38, 0.91)
Attitude	13	38		0.32	4.43	<0.0001****	(0.18, 0.46)
Interest	8	15		-0.04	-0.34	0.7353	(-0.26, 0.18)
Career interest	4	13		0.17	1.31	0.1896	(-0.08, 0.42)
Publication type	19	68	22.50 (<0.0001****)				
Journal	9	38		0.34	4.69	<0.0001****	(0.20, 0.48)

Table 2 (continued)

Thesis_Dissertation	10	30	0.05	0.69	0.4913	(-0.10, 0.21)
Continues moderators	No. studies	No.ES	$Q_{moderators}$ (p-value)	β	p value	95% CI
Year	19	68	2.80 (0.0941)	-0.01	0.0941	(-0.03, 0.00)

“No. Studies” means the number of studies; “No. ES” means the number of effect sizes; $Q_{moderators}$, the omnibus tests for heterogeneity; Mean g means the mean effect size (g); CI means confidence intervals; p , $p^* < 0.05$; $** < 0.01$; $*** < 0.001$. For each categorical moderator, the last category is the reference group. For the continuous moderator of publication year (Year), a meta-regression was conducted, and β is the regression coefficient for publication year (Year)

*: The total number of the subject subgroups does not match 19 because some studies included multiple subject areas

or attitudes toward science (Mean $g=0.14$; 95% CI [0.01, 0.26]; $p=0.0318$), mathematics (Mean $g=0.44$; 95% CI [0.04, 0.84]; $p=0.03$), technology and engineering (Mean $g=0.48$; 95% CI [0.06, 0.90]; $p=0.03$), or full STEM (Mean $g=0.36$; 95% CI [0.05, 0.68]; $p=0.02$), although all having statistically significant effect sizes positively related to informal science activities, had statistical variations in their respective effect sizes related to the subject areas, with those studies focusing on interests or attitudes toward science showing smaller effect sizes.

Outcomes

Similarly, we found a significant moderating effect of outcome types ($Q_{moderators} (df=4)=22.08, p=0.0002$). More specifically, the studies that examined self-efficacy and attitude showed larger effect sizes than those that examined interest or career interest. However, in regard to self-efficacy, there was only one article with two effect sizes that addressed this outcome. For this reason, the reliability of the finding is questionable because of the small sample size, and caution is warranted for the interpretation of this sub-group finding. In general, the finding suggests that the informal science learning activities showed a significant positive effect on students' attitudes toward STEM areas.

Publication Type

The dataset used in this meta-analysis consisted of 38 effect sizes from journal articles, and 30 effect sizes from dissertations or theses. The results revealed a statistically significant effect of the publication type moderator ($Q_{moderators} (df=2)=22.50, p<0.0001$), indicating publication type explains heterogeneity in the observed effect sizes. Specifically, the effect sizes derived from journal articles exhibited a larger magnitude (Mean $g=0.34$; 95% CI [0.205, 0.48]; $p<0.0001$) compared to those obtained from dissertations or theses (Mean $g=0.05$; 95% CI [-0.10, 0.21]; $p=0.4913$).

Publication Year

No statistically significant difference was revealed in terms of the effect size heterogeneity of publication year ($Q_{moderators} (df=1)=2.80, p=0.09$). The slope of publication year was statistically non-significant ($\beta=-0.01$; 95% CI [-0.03, 0.00]; $p=0.0941$), suggesting that the variable of publication year did not have a significant impact on the effect sizes.

Publication Bias

To assess the potential publication bias, we used both the funnel plot and Egger's regression method (Egger et al., 1997) in this study. As Fig. 6 shows, the funnel plot had an approximately symmetrical distribution, which indicates a general absence of publication bias. In addition, Egger's regression test ($p=0.28$) is not statistically significant, indicating a lack of evidence for publication bias.

Funnel Plot for The Effect Sizes

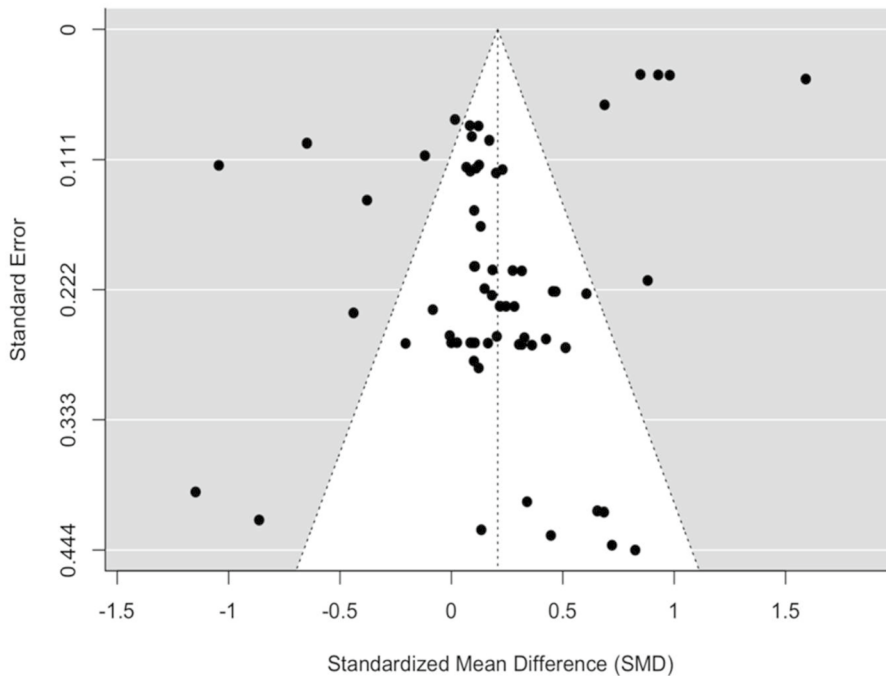


Fig. 6 Funnel plot for the effect sizes

Discussion

This meta-analysis served as a comprehensive summary of quantitative studies carried out within the context of informal science learning in the United States. Building upon previous meta-analytic research, we aimed to conduct a systematic meta-analysis to examine the impact of informal science learning by focusing on changes in interests or attitudes toward STEM areas before and after students participated in informal science activities. More specifically, we compared the interest or attitude change scores across treatment groups (i.e., with informal science learning activities) and control groups (i.e., without informal science learning activities). By addressing these aspects, our meta-analysis provides valuable insights into the role of informal science learning and its effects on students' interests and attitudes toward STEM areas.

Specifically, the random effect modeling revealed the statistically significant overall mean effect size (Mean $g=0.21$), indicating that informal science learning opportunities had a positive effect on students' STEM interests, and this positive effect, as discussed in Kraft (2020), could be characterized as a medium to large

effect. The result was consistent with previous funding by Young et al. (2017) and consistent with empirical findings on informal science learning toward STEM interests (e.g., Crawford & Huscroft-D'Angelo, 2015; Havasy, 1997; Yang & Chittoori, 2022).

The statistically significant outcomes of the heterogeneity test conducted among the effect sizes revealed notable statistical diversity in the impacts of informal science learning opportunities on students' attitudes and interests in STEM across the encompassed studies. Consequently, an examination of potential moderators was initiated to address the second research question. The significant moderators include the type of informal science program, grade level, publication type, and sample size. However, due to our small sample size in this meta-analysis, we highlight that the overall effect of informal STEM programming was large (0.21) and had a significant impact on students' interests and attitudes toward STEM, rather than focusing on the differences between moderators. The heterogeneity could be due to the program's focus. Laurer et al. (2006) concluded that informal program focus was a significant moderator of mathematics achievement among at-risk youth. Since STEM interest has been associated with achievement (Maltese & Tai, 2011), it is important to determine how program focus affects student interest in STEM (Young et al., 2017). Shaby et al. (2024) pointed out the importance of pedagogical design of Laboratory Group Activity in a Science Museum for students' interaction and learning outcomes. The various pedagogical designs of the same type of program might result in different effects. Future research in program focus might uncover why there is a difference between different types of informal science experiences.

Grade level emerged as a significant moderator on student interest in and attitudes toward STEM. However, the available data are insufficient to draw a statistical conclusion regarding the disparity between grade levels. Specifically, only a very limited number of studies conducted at the elementary school level were available for this meta-analytic review, with only 2 out of 19 studies involving elementary students. Nevertheless, there is accumulating evidence suggesting that early engagement in STEM learning, particularly during elementary school, may yield long-term benefits for sustained interest in STEM (Curran & Kitchin, 2019; Morgan et al., 2016). The scant representation of elementary-level studies in the research literature underscores a notable gap in the research landscape, emphasizing the necessity for further investigation to elucidate the impact of informal STEM experiences on elementary school students.

Although there was a lack of evidence for publication bias for studies included in this meta-analysis, the type of publications (i.e., journal articles vs. thesis/dissertations) showed different magnitude of effect sizes. Publication bias is influenced by various factors such as language bias, time lag bias, and selective reporting. The publication bias test as implemented in this study may not be sufficiently sensitive to capture all aspects of bias. Therefore, it is possible for the publication bias test to indicate no bias while the moderator test reveals an impact of publication type.

Conclusions and Future Directions

Students acquire STEM-related knowledge through formal and informal education experiences (Goldstein, 2015). Formal STEM experiences are part of the national curriculum provided to the students in their schools. Researchers can gauge the effectiveness of formal STEM programs by tracking students' progress in classrooms and with national testing data. However, it is more difficult to assess the impact of informal science learning experiences, since they happen outside of school. Based on this meta-analysis, informal STEM experiences have a positive effect on students' interest in and attitudes toward STEM, and should be incorporated into students' educational experiences.

Despite our exhaustive efforts in locating relevant studies conducted in a span of thirty years (1992–2022), we were only able to find a very limited number of studies that quantitatively assessed the effects of informal science learning on students' interest and attitude for science and STEM subjects. The majority of studies in this area were not appropriate for our quantitative synthesis because of different reasons: lack of sufficient information for effect size calculations for meta-analysis; studies of qualitative approaches that did not have quantifiable data on STEM interest changes; quantitative studies with only one-time data collection, but could not be used to assess the effect of informal science learning on students' interests/attitude for science/STEM. Future research may consider these and other similar issues for designing methodologically rigorous quantitative studies that examine how informal science learning experiences contribute to students' interest/attitude for STEM.

Limitations

Several limitations were identified during the evaluation and analysis process of this meta-analysis. During the data collection process, several articles were not available for the researchers to review, due to the internal restrictions of our database access and limitations from our university library. Another limitation is the conflicting results of our publication type analysis and publication bias assessment. Publication type was used as a moderator in our analysis, and we concluded that journal articles had a statistically significant effect size when compared to dissertation articles. This contradicted our publication bias assessment results, which suggested a lack of sufficient statistical evidence for publication bias, and this could be a limitation of this analysis. But the overall very limited number of studies ($N=19$) available for this meta-analysis made it difficult to explore this issue in a more meaningful way. Because of these considerations, caution is warranted in the interpretation of the findings related to this issue.

Small sample size (in terms of both the number of studies and number of effect sizes) is a general limitation in this meta-analysis, especially for some moderator analysis. Obviously, this meta-analysis, like other synthesis studies, is at the mercy of what is available in the research literature. Because of the small sample size issue, we need to exercise caution when interpreting the findings, especially the findings in

the moderator analysis as shown in Table 2. As discussed above, future research may further examine some of the issues revealed in the study.

We did not include some other possible variables, such as dosage, duration, gender, race, school type, or sampling method. Although these could be potential moderators, the lack of relevant information in the primary studies made it impossible for us to consider these in the meta-analysis. Furthermore, the interaction effects of moderator variables on the effect of learning interest and attitudes were not addressed in this study, because the limited sample size made it statistically impractical to conduct interaction analysis.

In addition, we were not able to explore what components of different types of programs actually had an effect on students' interests and attitudes. Previous studies revealed that hands-on activities and challenge assessments could enhance students' interest and motivation (Hamari et al., 2016; Parsons & Taylor, 2011; Poudel et al., 2005). It is possible that the outreach program included in this meta-analytic study contained more challenging activities, which might explain that such studies' larger effect size than some other studies. But lack of relevant information in the primary studies included in our meta-analysis made it impossible for us to explore such potentially relevant issues.

Authors' Contributions All authors contributed to the study conception and design. Material preparation, data collection were performed by Xin Xia, Lillian Bentley; and analysis by Xin Xia. The first draft of the manuscript was written by Xin Xia and Lillian Bentley, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Parts of this work have been supported by the US National Science Foundation (NSF DRL 1811265). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the US National Science Foundation.

Data Availability The datasets used in this meta-analysis are derived from publicly available sources (Proquest, EBSCO, Web of Science, and ScienceDirect) and previously published studies. The references and citations for all included studies are provided in the reference list with the asterisk accompanying this publication.

Declarations

Ethical Approval Not applicable.

Competing Interests We declare that there is no competing interests relevant to this meta-analysis. No financial or personal relationships with individuals or organizations have influenced the study design, data collection, analysis, interpretation, or the decision to publish the findings. The research was conducted with impartiality and objectivity to ensure the integrity of the meta-analysis process and the credibility of its outcomes.

Conflict of Interest The authors have no conflicts of interest to declare. All co-authors have read and agree with the contents of the manuscript. All co-authors have no financial interest to report.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended

use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

*: A primary study with effect size(s) included in this meta-analysis.

- Alliance, A. (2011). *STEM learning in afterschool: An analysis of impact and outcomes*. Afterschool Alliance.
- Bandura, A. (1995). *Exercise of personal and collective efficacy in changing societies* (A. Bandura, Ed.). Cambridge University.
- Becker, B. J. (1988). Synthesizing standardized mean-change measures. *British Journal of Mathematical and Statistical Psychology*, *41*, 257–278.
- Beier, M., & Rittmayer, A. (2008). *Literature overview: Motivational factors in STEM: Interest and self-concept*. Assessing Women and Men in Engineering.
- Berliner, D. C. (2009). *Poverty and potential: Out-of-school-factors and school success*. Education and the Public Interest Center, University of Colorado/Education Policy Research Unit, Arizona State University. Retrieved October 26, 2015, from <http://epicpolicy.org/publication/poverty-and-potential>
- Bevan, B., Dillon, J., Hein, G. E., Macdonald, M., Michalchik, V., Miller, D., Root, D., Rudder-Kilkenny, L., Xanthoudaki, M., & Yoon, S. (2010). *Making science matter: Collaborations between informal science education organizations and schools*. Center for Advancement of Informal Science Education.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*, *1*(2), 97–111. <https://doi.org/10.1002/jrsm.12>
- Borenstein, M., Higgins, J. P., Hedges, L. V., & Rothstein, H. R. (2017). Basics of meta-analysis: I² is not an absolute measure of heterogeneity. *Research Synthesis Methods*, *8*(1), 5–18.
- Business-Higher Education Forum (2011). *Creating the workforce of the future: The STEM interest and proficiency challenge*. BHEF Research Brief. Business-Higher Education Forum.
- Cai, Z., Fan, X., & Du, J. (2017). Gender and attitudes toward technology use: A meta-analysis. *Computers & Education*, *105*, 1–13.
- Cai, Z., Mao, P., Wang, D., He, J., Chen, X., & Fan, X. (2022). Effects of scaffolding in digital game-based learning on student's achievement: A three-level meta-analysis. *Educational Psychology Review*, *34*(2), 537–574.
- Crane, V. (1994). *Informal science learning: What the research says about television, science museums, and community-based projects*. Research Communications, Limited.
- *Crawford, L., & Huscroft-D'Angelo, J. (2015). Mission to space: Evaluating one type of informal science education. *The Electronic Journal for Research in Science & Mathematics Education*, *19*(1). <https://ejrsmc.icrsmc.com/article/view/13753>
- Curran, F. C., & Kitchin, J. (2019). Early elementary science instruction: Does more time on science or science topics/skills predict science achievement in the early grades? *AERA Open*, *5*, 1–18. <https://doi.org/10.1177/2332858419861081>
- DeCoito, I. (2014). Focusing on science, technology, engineering, and mathematics (STEM) in the 21st century. *Ontario Professional Surveyor*, *57*(1), 34–36.
- Dowey, A. L. (2013). *Attitudes, Interests, and Perceived Self-Efficacy toward Science of Middle School Minority Female Students: Considerations for Their Low Achievement and Participation in STEM Disciplines* (Unpublished doctoral dissertation). University of California.
- Dryfoos, J. G. (1999). The role of the school in children's out-of-school time. *The Future of Children*, *9*(2), 117–134.
- Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, *315*(7109), 629–634. <https://doi.org/10.1136/bmj.315.7109.629>

- Eshach, H. (2007). Bridging in-school and out-of-school learning: Formal, non-formal, and informal education. *Journal of Science Education and Technology*, 16, 171–190.
- Falk, J. H., & Dierking, L. D. (2010). The 95 percent solution. *American Scientist*, 98(6), 486–493.
- Frantz, T. D., Miranda, M., & Siller, T. (2011). Knowing what engineering and technology teachers need to know: An analysis of pre-service teachers' engineering design problems. *International Journal of Technology Design Education*, 21, 307–320.
- *Garvin, B. A. (2015). *An investigation of a culturally responsive approach to science education in a summer program for marginalized youth* (Doctoral dissertation). University of South Carolina. Retrieved from ProQuest Dissertation Theses Global.
- Gauld, C. F., & Hukins, A. A. (1980). Scientific attitudes: A review. *Studies in Science Education*, 7(1), 129–161.
- George, R. (2000). Measuring change in students' attitudes toward science over time: An application of latent variable growth modeling. *Journal of Science Education and Technology*, 9(3), 213–225.
- Goldstein, D. (2015). *The teacher wars: A history of America's most embattled profession*. Anchor.
- Hamari, J., Shernoff, D. J., Rowe, E., Coller, B., Asbell-Clarke, J., & Edwards, T. (2016). Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning. *Computers in Human Behavior*, 54, 170–179.
- *Havasy, R. A. D. P. (1997). *The effect of informal science experiences on science achievement and attitude of high school biology students* (Ed.D.). Teachers College, Columbia University. Retrieved from ProQuest Dissertation Theses Global.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Academic press.
- Hofstein, A., & Rosenfeld, S. (1996). Bridging the gap between formal and informal science learning. *Studies in Science Education*, 28, 87–112.
- Jensen, F., & Sjaastad, J. (2013). A Norwegian out-of-school mathematics project's influence on secondary students' STEM motivation. *International Journal of Science and Mathematics Education*, 11(6), 1437–1461. <https://doi.org/10.1007/s10763-013-9401-4>
- Jiang, H., Chugh, R., Turnbull, D., Wang, X., & Chen, S. (2024). Exploring the effects of technology-related informal mathematics learning activities: A structural equation modeling analysis. *International Journal of Science and Mathematics Education*, 1–21. <https://doi.org/10.1007/s10763-024-10456-4>
- Jones, A. L., & Stapleton, M. K. (2017). 1.2 million kids and counting—Mobile science laboratories drive student interest in STEM. *PLoS biology*, 15(5), e2001692.
- Keiler, L. S. (2011). An effective urban summer school: Students' perspectives on their success. *The Urban Review*, 43(3), 358–378.
- Knapp, D., & Barrie, E. (2001). Content evaluation of an environmental science field trip. *Journal of Science Education and Technology*, 10(4), 351–357. <https://doi.org/10.1023/A:1012247203157>
- Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, 49(4), 241–253. <https://doi.org/10.3102/0013189X20912798>
- Krishnamurthi, A., & Rennie, L. J. (2013). *Informal science learning and education: Definition and goals*. Afterschool Alliance.
- Laurer, P. A., Akiba, M., Wilkerson, S. B., Apthorp, H. S., Snow, D., & Martin-Glen, M. L. (2006). Out-of-school-time programs: A meta-analysis of effects for at-risk students. *Review of Educational Research*, 76(2), 275–313.
- Maiorca, C., Roberts, T., Jackson, C., Bush, S., Delaney, A., Mohr-Schroeder, M. J., & Soledad, S. Y. (2021). Informal learning environments and impact on interest in STEM careers. *International Journal of Science and Mathematics Education*, 19, 45–64.
- Maltese, A. V., & Tai, R. H. (2011). Pipeline persistence: Examining the association of educational experiences with earned degrees in STEM among U.S. students. *Science Education*, 95(5), 877–907.
- 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.
- Matthews, P. H., & Mellom, P. J. (2012). Shaping aspirations, awareness, academics, and action outcomes of summer enrichment programs for English-learning secondary students. *Journal of Advanced Academics*, 23(2), 105–124.
- *Migliarese, N. L. (2011). *"That's my kind of animal!" Designing and assessing an outdoor science education program with children's megafaunaphilia in mind*. (Doctoral dissertation). University of California, Berkeley. Retrieved from ProQuest Dissertation Theses Global.

- Morgan, P. L., Farkas, G., Hillemeier, M. M., & Maczuga, S. (2016). Science achievement gaps begin very early, persist, and are largely explained by modifiable factors. *Educational Researcher*, 45(1), 18–35.
- Morrell, P. D., & Lederman, N. G. (1998). Student's attitudes toward school and classroom science: Are they independent phenomena? *School Science and Mathematics*, 98(2), 76–83.
- Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods*, 11(2), 364–386.
- Murphy, C., & Beggs, J. (2003). Children's perceptions of school science. *School Science Review*, 84, 109–116.
- National Research Council. (2007). *Taking science to school: Learning and teaching science in grades K-8*. National Academies Press.
- National Research Council. (2009). *Learning science in informal environments: People, places, and pursuits*. National Academies Press.
- National Research Council, Division of Behavioral, Board on Testing, Assessment, Board on Science Education, & Committee on Highly Successful Schools or Programs for K-12 STEM Education. (2011). *Successful K-12 STEM education: Identifying effective approaches in science, technology, engineering, and mathematics*. National Academies Press.
- National Academies of Sciences, Engineering, and Medicine. (2016). *Parenting matters: Supporting parents of children ages 0–8*. The National Academies Press.
- Osborne, J., Simon, S., & Collins, S. (2003). Attitudes towards science: A review of the literature and its implications. *International Journal of Science Education*, 25(9), 1049–1079.
- *Parker, V., & Gerber, B. (2000). Effects of a science intervention program on middle-grade student achievement and attitudes. *School Science and Mathematics*, 100(5), 236–242.
- Parsons, J., & Taylor, L. (2011). Improving student engagement. *Current Issues in Education*, 14(1).
- Phillips, M., Finkelstein, D., & Wever-Frerichs, S. (2007). School site to museum floor: How informal science institutions work with schools. *International Journal of Science Education*, 29(12), 1489–1507.
- Potvin, P., & Hasni, A. (2014). Interest, motivation and attitude towards science and technology at K-12 levels: A systematic review of 12 years of educational research. *Studies in Science Education*, 50(1), 85–129.
- Poudel, D. D., Vincent, L. M., Anzalone, C., Huner, J., Wollard, D., Clement, T., DeRamus, A., & Blake-wood, G. (2005). Hands-on activities and challenge tests in agricultural and environmental education. *The Journal of Environmental Education*, 36(4), 10–22.
- R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved March 10, 2020 from <https://www.R-project.org/>
- *Roberson, S. V. (2010). *Science skills on wheels: The exploration of a mobile science lab's influence on teacher and student attitudes and beliefs about science*. (Ed.D dissertation). University of Pennsylvania. Retrieved from ProQuest Dissertation Theses Global.
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, 86, 638–641. <https://doi.org/10.1037/0033-2909.86.3.638>
- Rothstein, H., Sutton, A. J., & Borenstein, M. (Eds.). (2005). *Publication bias in meta-analysis: Prevention, assessment and adjustments*. Wiley.
- Sadler, P. M., Sonnert, G., Hazari, Z., & Tai, R. (2012). Stability and volatility of STEM career interest in high school: A gender study. *Science Education*, 96(3), 411–427
- Sacco, K., Falk, J. H., & Bell, J. (2014). Informal science education: Lifelong, life-wide, life-deep. *PLoS Biology*, 12(11), e1001986.
- Shaby, N., Assaraf, O. B. Z., & Koch, N. P. (2024). Students' interactions during laboratory group activity in a science museum. *International Journal of Science and Mathematics Education*, 22(4), 703–720.
- *Shields, N. C. (2010). *Elementary students' knowledge and interests related to active learning in a summer camp at a zoo* (Doctoral dissertation). Purdue University. Retrieved from ProQuest Dissertation Theses Global.
- Silver, A., & Rushton, B. (2008). Primary-school children's attitudes towards science, engineering and technology and their images of scientists and engineers. *Education 3–13*, 36(1), 51–67.
- Simpkins, S. D., Davis-Kean, P. E., & Eccles, J. S. (2006). Math and science motivation: A longitudinal examination of the links between choices and beliefs. *Developmental Psychology*, 42(1), 70–83.

- So, W. W. M., Zhan, Y., Chow, S. C. F., & Leung, C. F. (2018). Analysis of STEM activities in primary students' science projects in an informal learning environment. *International Journal of Science and Mathematics Education, 16*, 1003–1023.
- Tai, R. H., Liu, C. Q., Maltese, A. V., & Fan, X. (2006). Career choice: Planning early for careers in science. *Science, 312*(5777), 1143–1144.
- Thornton, A., & Lee, P. (2000). Publication bias in meta-analysis: Its causes and consequences. *Journal of Clinical Epidemiology, 53*(2), 207–216. [https://doi.org/10.1016/S0895-4356\(99\)00161-4](https://doi.org/10.1016/S0895-4356(99)00161-4)
- Tillinghast, R. C., Appel, D. C., Winsor, C., & Mansouri, M. (2020, August). *STEM outreach: A literature review and definition*. Paper presented at 2020 IEEE Integrated STEM Education Conference (ISEC).
- 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>
- Viechtbauer, W., & Cheung, M. W. L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods, 1*(2), 112–125. <https://doi.org/10.1002/jrsm.11>
- Wang, M.-T. (2012). Educational and career interests in math: A longitudinal examination of the links between classroom environment, motivational beliefs, and interests. *Developmental Psychology, 48*(6), 1643–1657.
- Wiebe, E., Unfried, A., & Faber, M. (2018). The relationship of STEM attitudes and career interest. *EUR-ASIA Journal of Mathematics, Science and Technology Education, 14*(10), 1–17.
- Wilson, A. C., Gonzalez, L. L., & Pollock, J. A. (2012). Evaluating learning and attitudes on tissue engineering: A study of children viewing animated digital dome shows detailing the biomedicine of tissue engineering. *Tissue Engineering Part A, 18*(5–6), 576–586. <https://doi.org/10.1089/ten.tea.2011.0242>
- Xie, Y., Fang, M., & Shauman, K. (2015). STEM education. *Annual Review of Sociology, 41*, 331–357.
- Xie, Y., & Killewald, A. A. (2012). *Is American science in decline?* Harvard University Press.
- *Yang, D., & Chittoori, B. (2022). Investigating Title I school student STEM attitudes and experience in an after-school problem-based bridge building project. *Journal of STEM Education: Innovations and Research, 23*(1), 17–24.
- Young, J., Ortiz, N., & Young, J. (2017). STEMulating interest: A meta-analysis of the effects of out-of-school time on student STEM interest. *International Journal of Education in Mathematics, Science and Technology, 5*(1), 62–74.
- Zhang, D., & Tang, X. (2017). The influence of extracurricular activities on middle school students' science learning in China. *International Journal of Science Education, 39*, 1381–1402. <https://doi.org/10.1080/09500693.2017.1332797>