



A Meta-Analysis to Gauge the Effectiveness of STEM Informal Project-Based Learning: Investigating the Potential Moderator Variables

Malavika Santhosh¹ · Hafsa Farooqi¹ · Mohammad Ammar · Nitha Siby · Jolly Bhadra^{1,2} · Noora J. Al-Thani¹ · Abdellatif Sellami³ · Noshin Fatima⁴ · Zubair Ahmad^{1,2}

Accepted: 23 June 2023 / Published online: 17 July 2023
© The Author(s) 2023

Abstract

In recent decades, informal project-based learning (PjBL) for science, technology, engineering, and mathematics (STEM) education has gained immense significance. Though some meta-analytical studies have reported the effectiveness of PjBL models, none of them has been explicit for informal PjBL models executed for school students in STEM education. Therefore, this meta-analytical study sought to investigate the effect of the informal PjBL model on students' learning gains when compared to the traditional classroom setting. The study also intends to explore the effect of potential moderator variables that might impact the effectiveness of the model. For this, a systematic review aligning to Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) methodology has been employed, by incorporating 26 peer-reviewed empirical studies, encompassing data from 3202 students from eight countries. The meta-analytical results revealed overall moderate effectiveness ($d=0.248$, $p<0.001$), indicating that informal PjBL has moderately significant effects on students' gains when compared to the traditional classroom setting. Correspondingly the moderator analysis revealed that the overall effectiveness of informal PjBL was influenced by the teaching model, assessment method, students' group size, and course duration. However, the participants' educational level and study location (country) did not emerge as being significant. This study contributes to the bulk of literature that can help academicians to design, develop and implement effective STEM-based informal PjBL models, with special consideration for influential variables (teaching method, assessment method, grade, location, course duration, group size, and subject area).

Keywords Meta-analysis · Informal · Informal PjBL models · STEM education · Moderator analysis · Effect sizes

Introduction

There is a constant urge to adapt to the demands of the changing world, where individuals are stimulated to use their knowledge to explore, create, and re-construct solutions for challenging problems. This could be achievable

by inculcating STEM (science, technology, engineering, mathematics) literacy, where the role of science education is substantially significant. Many researchers have found that informal education especially STEM-focused Project-Based Learning (PjBL) is becoming one of the prominent pedagogical methods employed to improve the quality of science education that enables students to become STEM-literate individuals (Balemen & Keskin, 2018; Chen & Yang, 2019; Ramdhayani et al., 2019; Sivia et al., 2019).

An informal PjBL is a widely employed model, that has proven to instill students with STEM skills (Habig & Gupta, 2021; Mateos-Nunez et al., 2020; Nadelson et al., 2021; Vela et al., 2020). An informal setting typically relates to an out-of-school experience such as in a camp, zoo, museum, summer program, laboratory-oriented setting, research environment, after-school program, maker space, etc. While a formal setting relates to standardized & structured classroom education. Regardless of whether the setting is formal or

✉ Zubair Ahmad
zubairtarar@qu.edu.qa

¹ Qatar University Young Scientists Center (QUYSC), Qatar University, 2713 Doha, Qatar

² Center for Advanced Materials (CAM), Qatar University, 2713 Doha, Qatar

³ Educational Research Center (ERC), Qatar University, 2713 Doha, Qatar

⁴ UCSI: UCSI University, Alan Puncak Menara Gading, Taman Connaught, Cheras, 56000 Federal Territory of Kuala Lumpur, Malaysia

informal, the main framework for PjBL remains the same i.e., it is a systematic pedagogical model that helps students experience and indulge in complex, real-world tasks, resulting in an idea, a product, or prototypes (Johnson et al., 2013). There are two critical parts of a PjBL model: a question that drives learning activities as per the objectives and the solution in the form of students' artifacts (Shpeizer, 2019). Some fundamental features of PjBL include inquiry, investigations, peer discussions, creation, revision, reflection, sharing of findings, design thinking, etc. (Larmer & Mergendoller, 2010). In this present study, PjBL is viewed as a student-centered learning model (Kokotsaki et al., 2016), with the projects being a critical component (Mergendoller & Thomas, 2005). An informal PjBL model has been opted for our research since diverse meta-analytical studies on PjBL had shown varied effectiveness [i.e., $d=0.71$ (Chen & Yang, 2019), $d=0.99$ (Ayaz & Söylemez, 2015) and $d=1.06$ (Balemen & Keskin, 2018)]. Moreover very little has been known about moderator variables that might affect learning gains associated with informal PjBL.

Therefore, despite these recent meta-analytical studies on the significance of PjBL when compared to the traditional classroom setting (Ayaz & Söylemez, 2015; Balemen & Keskin, 2018; Chen & Yang, 2019; Ramdhayani et al., 2019), there is a dearth of meta-analytical literature focusing on the field of informal PjBL. Moreover, most of these review studies have not specified the mode of PjBL, whether employed in a formal or informal setting. The novelty of this study lies in the fact that it encompasses literature (of 12 years) employing STEM-focused informal PjBL models for school-level students exclusively, along with the moderator analysis. Wherein the moderator analysis has illustrated the impact of diverse potential variables that are likely to influence the effectiveness of informal PjBL (i.e., subject area, teaching method, assessment method, grade, location, course duration, and group size).

Review of Literature and Conceptual Framework

Comprehending the impact of potential moderator variables is essential for tailoring the pedagogies in the best possible manner (Chen & Yang, 2019). Many individual studies have revealed the implication of informal PjBL in the teaching of diverse STEM subjects. More often, they have been employed in teaching various science disciplines such as physics (Awad, 2021; Mateos-Nunez et al., 2020), biology (Bokor et al., 2014; Covert et al., 2019), chemistry (Nadelson et al., 2021), mathematics (Calabrese & Capraro, 2021; Kwon et al., 2021), engineering (Hirsch et al., 2017; Innes et al., 2012), and technology (Smit et al., 2021). Many studies have also shown the success of integrated informal PjBL models, where these informal PjBL models are integrated with other pedagogies, such as 1) problem-based learning (Awad, 2021; Hirsch et al.,

2017; Kwon et al., 2021; Nugent et al., 2010), 2) inquiry-based learning (Covert et al., 2019), 3) game-based learning (Newton et al., 2020), and 4) forensic discovery-based learning models (Todd & Zvoch, 2019). However, learning is not only influenced by the subjects taught or the learning models employed, but also by many other moderator variables such as assessment method, students' grades, location, course duration, and students' group size, etc.

Indeed, it is widely acknowledged that assessment influences students' learning (Cilliers et al., 2012), and the relationship between assessment and student learning can be complex and unpredictable (Al-Kadri et al., 2012). Even though, assessment is typically used to evaluate a student's knowledge, skills, and abilities; the type of assessment employed can also have a direct influence on students' learning, including their reaction to the assessment, their involvement, and their motivation (Marriott & Lau, 2008). In this regard, the study by Gao et al. (2020) has proposed future directions for developing improved assessment methods for STEM education (Gao et al., 2020). Their findings indicate that despite many programs aiming to improve students' interdisciplinary skills, their assessments did not align with their objectives. Therefore their recommendations involve: assessments must be built to the set learning objectives, connections across disciplines must be operationalized and assessed to provide targeted student feedback, and development of practical assessment methods and guidelines must be prioritized. Typical assessment methods include direct and indirect assessments. Direct assessment methods are employed when students demonstrate their mastery of skills and knowledge through actual work. It is usually measured via professional licensure exams, standardized exams, knowledge-based pre-and post-tests, student posters, models, speeches, presentations, etc. Whereas, in the indirect assessment method, the students are assessed indirectly via focus group interviews, self-report surveys for perceived learning, exit survey, mentor/volunteers' feedback, etc. (Martell & Calderon, 2005). Nowadays, many educators prefer to evaluate student progress by using a cutting-edge method of course-embedded assessment (Gerretson & Golson, 2005). Embedded assessments, often known as "authentic assessments," include internal-mentor-created course-level assessment plans, including individual student design logs, worksheets, homework, assignments, class tests, etc. The strength of this approach lies in the fact that it is developed by mentors who are experts in the course field, which offer them fine-grained insights that help them to make instructional adjustments in the classroom as per students' needs (Kim et al., 2021). Correspondingly, a careful mix of embedded, direct, and indirect assessments in STEM education, a practice viewed as a triangulation of assessments that has a great significance (Ghrayeb et al., 2011), has rarely been discussed/ investigated in the literature.

Likewise, the time devoted to instruction has also been identified as a moderator variable of PjBL models (Chen & Yang, 2019). Different instructional times have been reported to be successful for informal PjBL. For example, many studies have productively employed and recommended informal PjBL instruction for a week (Bokor et al., 2014; Nugent et al., 2010; Smit et al., 2021), for 2 weeks (Biçer & Lee, 2019; Calabrese & Capraro, 2021; Todd & Zvoch, 2019; Todd & Zvoch, 2019), and more than a month (Awad, 2021; Covert et al., 2019; Stevens et al., 2016; Yin et al., 2020). Although some educators and parents believe that projects require a greater amount of instructional time, and thus blocks course content acquisition (Miller, 2018). Therefore, if we have a clearer understanding of the impact of instructional time on student learning in informal PjBL, it would demonstrate evidence supporting the amount of time that should be devoted to informal PjBL.

Therefore, this review article is motivated by Mergendoller and Thomas's (2005) review of PjBL, which points to the demand for more evidence-based research to determine the effectiveness of PjBL, especially when compared with traditional classroom settings (Mergendoller & Thomas, 2005). According to them, future research must investigate the diverse potential moderator variables of PjBL, including subject areas, duration of the course, appropriate assessment tools, grades, location, etc., because designing, executing, and supervising PjBL are all related to students' learning gains. The conceptual framework of this review is established on the study by Chen & Yang, 2019, which included a 20-year meta-analytical study on PjBL for STEM & non-STEM disciplines. In their study, formerly the overall effectiveness of informal PjBL has been established, followed by moderator analysis (subject area, school location, hours of instruction, information technology support, educational stage, and, students' group size). Indeed, the significance of the various moderator variables impacting PjBL has been inconsistently reported in the literature (Ayaz & Söylemez, 2015; Balemén & Keskin, 2018; Chen & Yang, 2019), with no distinction on formal and informal PjBL. Hence, this current study strives to investigate the effectiveness of informal PjBL, since the informal (out-of-school) PjBL approach should be viewed as a process that must be tailored to assure the needs of diverse learners. In addition, the study also intends to perform a moderator analysis to compute the most effective sub-variables that might influence the effectiveness of informal PjBL. Therefore, the research questions guiding this meta-analytical article are as follows:

1. What is the measure of the effectiveness of informal PjBL in developing students' learning gains (compared to the traditional classroom setting)?
2. Does the effectiveness of informal PjBL influenced by moderator variables such as teaching model, educational level, group size, subject area, study location, assessment model, and, instruction time?

Methods

Search Strategy

The relevant research articles were gathered via common web search engines such as Scopus, Web of Science, and Journals for Educational Research Information Center (ERIC). Specific keywords and operators (AND, OR, *) were employed to locate the research papers; i.e., ("STEM" OR "science" OR "technology" OR "engineering" OR "mathematics") AND ("project" OR "research" OR "inquiry" OR "problem-based" OR "virtual" OR "game-based" OR "blended" OR "flip*") AND ("out of school" OR "informal" OR "outreach" OR "summer" OR "winter" OR "internship" OR "workshop" OR "camp" OR "zoo" OR "museum" OR "park"). A systematic literature review technique aligning with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines was adopted (Page et al., 2021) (refer to Figure). PRISMA is an evidence-based protocol for reporting in systematic reviews and meta-analyses (Page et al., 2021). A total of 755 papers were identified using the keyword search: i.e., Scopus ($n = 330$), ERIC ($n = 250$), and Web of Science ($n = 175$). Followingly, 326 and 221 articles were discarded due to duplication and exclusion criteria respectively (refer to Section 2.2: [Inclusion and Exclusion Criteria](#)).

Inclusion and Exclusion Criteria

The articles were primarily reviewed for title aptness, gist relevance, context-relatedness, abstract relevance, and retrieval (refer to Fig. 1). For this, articles were reviewed as per inclusion and exclusion criteria, established under the scope of the study. The inclusion criteria for article selection were: studies employing the informal PjBL model at the school level (e.g., primary, middle, and high school); informal PjBL model for STEM education; articles extracted during the last 12 years (2010–2022 March); peer-reviewed articles; articles written in English.

Correspondingly, the initial exclusion criteria were the education level pertaining to kindergarten, undergraduate, and postgraduate level; teaching model for other non-STEM subjects (e.g., commerce, business, social sciences, language, etc.); qualitative research, and review papers. Final exclusion criteria were based on study specifications i.e., non-empirical studies; no pre-/post-test design or control/experiment design; study data not aligning to CMA format (e.g., not including the mean, standard deviation, *Cohen's d*, *t*-value, *p*-value, etc.); insufficient data for calculating the effect size (ES) (i.e., p -*t*-value = 0). Finally, aligning with the inclusion & exclusion criteria, 26 peer-reviewed empirical articles were finalized for the meta-analysis.

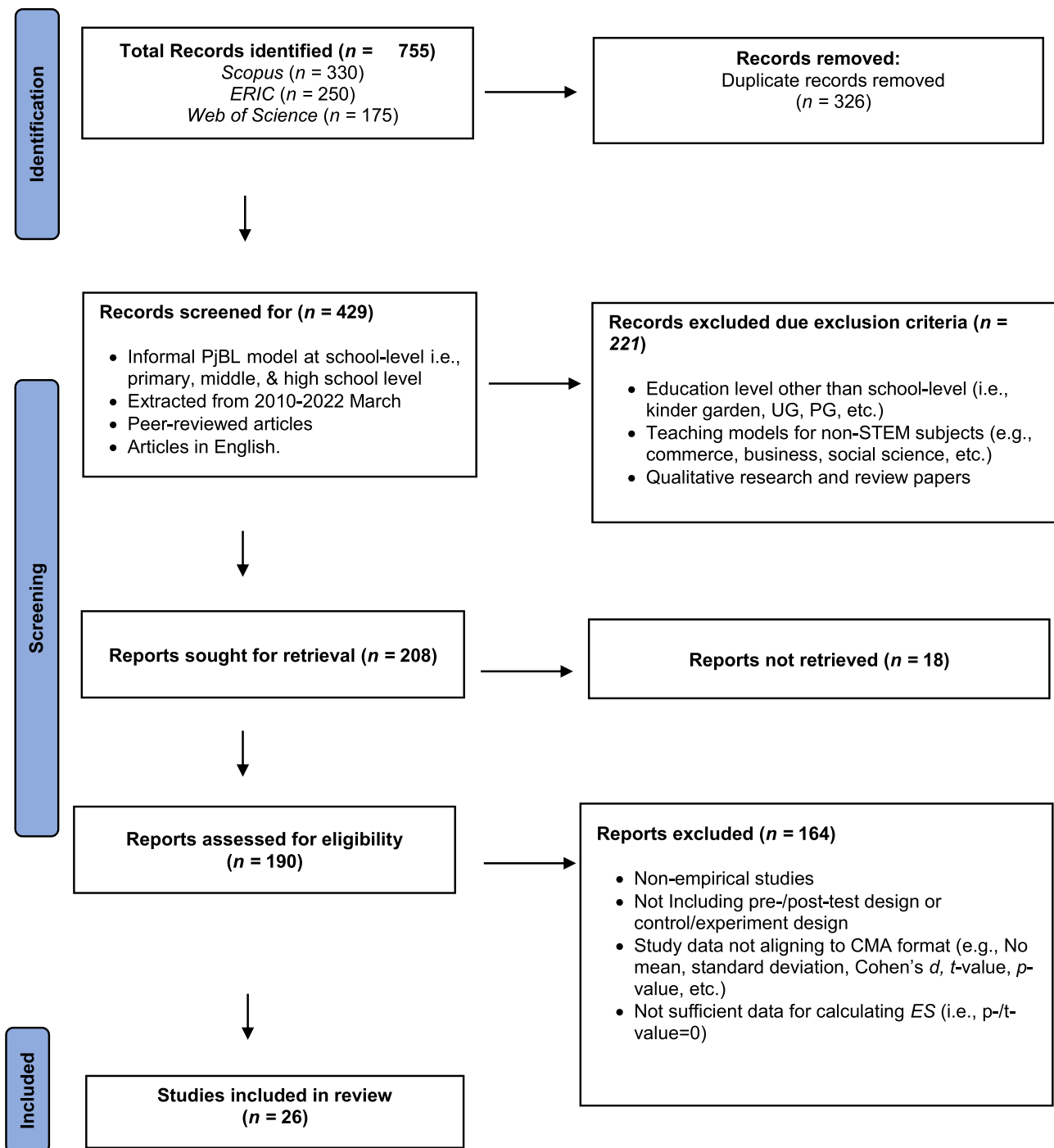


Fig. 1 PRISMA 2020 flow diagram revealing the inclusion–exclusion criteria for studies included in the meta-analysis (Page et al., 2021)

Conclusively, Table 1 shows the descriptive features of the 26 studies shortlisted for the meta-analysis. Figure 2 depicts the number of studies reviewed, by year of publication (2010 to May 2022). Results indicated that most of the studies were extracted from 2019 ($n = 7$), followed by 2021 ($n = 6$) and 2020 ($n = 4$). On examining the country of publication, most of the studies were from the U.S.A. ($n = 13$), followed by 1 study from Israel, Georgia, Spain, Taiwan,

Switzerland, Turkey, and Malaysia ($n = 1$; each). While analyzing the educational level, most of the studies focused on middle schools ($n = 11$), followed by high schools ($n = 6$) and primary schools ($n = 3$). Some studies involved a combination of school levels, such as middle and high schools ($n = 4$) and middle and primary schools ($n = 2$). After short-listing the articles, a meta-analysis has been performed and the results were interpreted (refer to the following sections).

Table 1 Descriptive features of the 26 shortlisted studies for the meta-analysis

S.No	Study	Grade	Pedagogical model	Assessment method	Group Size	Subject area	Study location	Course duration
1	Awad (2021)	Middle school	Project & problem-based	Triangulation	5	Science	Israel	15 weeks
2	Baran et al. (2019)	Middle school	Project & problem-based	Indirect	2	STEM	Turkey	3 weeks
3	Bicer and Lee (2019)	Middle & High school	Project-based	Indirect	-	STEM	Southern U.S.A	2 weeks
4	Bokor et al. (2014)	High school	Project-based	Triangulation	3–4	Science	Eastern U.S.A	1 week
5	Calabrese and Capraro (2021)	High school	Project-based	Indirect	3	Mathematics	Southwestern America	2 weeks
6	Collins et al. (2020)	High school	Project-based	Triangulation	5	Science	Western U.S.A	3 weeks
7	Covert et al. (2019)	High school (public)	Project & Inquiry-Based	Triangulation	-	Science	Eastern U.S.A	6 weeks
8	Habig and Gupta (2021)	Middle school & high school	Project-based	Triangulation	3–4	Science	Eastern U.S.A	3 weeks
9	Hirsch et al. (2017)	Middle school	Project & problem-based	Triangulation	4	Engineering	Eastern U.S.A	2 weeks
10	Innes et al. (2012)	Primary and middle school	Project-based	Direct/indirect	2	Engineering	The U.S.A (Southwestern)	1 day
11	Kwon et al. (2021)	Middle & high school	Project & problem-based	Indirect	-	Mathematics	Southern U.S.A	1 or 2 weeks
12	Lu et al. (2021)	Primary	Project-based	Triangulation	3–4	Science	Taiwan	2 weeks
13	Mateos-Nunez et al. (2020)	Primary	Project-based	Triangulation	3–4	Science	Spain	Within a week
14	Mohd Shahali et al (2019)	Middle school	Project-based	Direct & indirect	3	Science	Malaysia	1 week
15	Moreno et al. (2016)	Middle school	Project-based	Triangulation	Small groups	Science	Southern U.S.A	-
16	Nadelson et al. (2021)	High school	Project-based	Triangulation	9–12	Science	Southern U.S.A	4 weeks
17	Newton et al. (2020)	Primary (3–6)	Project & Game-based	Triangulation	3–5	STEM	Eastern U.S.A	10 weeks
18	Nugent et al. (2010)	Middle school	Project & problem-based	Direct & indirect	3–4	STEM	Western U.S.A	1 week
19	Smit et al. (2021)	Middle school	Project-based	Direct & indirect	4	Science & Technology	Switzerland	1 day -1 week
20	Stevens et al. (2016)	Primary & Middle school	Project-based	Indirect	Small groups	Engineering	Southern U.S.A	2 months
21	Tekbryik et al. (2022)	Middle school	Project & problem-based	Triangulation	4	STEM	Turkey	1 week
22	Todd and Zvoch (2017)	Middle school	Project & discovery-based	Direct & indirect	Small groups	Science & Technology	Western U.S.A	2 weeks
23	Todd and Zvoch (2019)	Middle school	Project & discovery-based	Triangulation	2	Science & technology	Western U.S.A	2 weeks

Table 1 (continued)

S.No	Study	Grade	Pedagogical model	Assessment method	Group Size	Subject area	Study location	Course duration
24	Vela et al. (2020)	Middle & High school	Project-based	Indirect	-	STEM	Southwestern U.S.A	1 or 2 weeks
25	Yin et al. (2020)	High school	Project-based	Triangulation	Small groups	Science & Technology	Western U.S.A	5 months
26	Zhou et al. (2017)	Middle school	Project-based	Triangulation	3–4	Engineering	Western U.S.A	2 weeks

Meta-Analysis and Interpretation

Comprehensive Meta-Analysis (CMA) software package (version 3.3.070) has been employed for meta-analysis. The Der Simonian and Laird methods were used to determine the individual and overall effectiveness [in terms of effect sizes (*ESs*)] with 95% confidence intervals (*CI*) of the studies ((DerSimonian & Laird, 2015). The raw empirical data were extracted from the shortlisted studies in the form of pre-/ post- or control/treatment means, *SDs*, *t*-values, *p*-values, etc., to calculate the *ESs* (in terms of *Cohen's d* index). *ESs* are exemplified in the form of forest plot diagrams illustrating the dispersal of the *ESs* of all the shortlisted studies. According to (Cohen, 2013), *Cohen's d* values ≤ 0.2 , ~ 0.5 , and ≥ 0.8 are regarded as low, moderate, and high effectiveness respectively. A random-effects model was employed to calculate the mean *ES*, as this model assumes that each study tends to have a different “true” effect (Borenstein et al., 2009). To test if a set of *ESs* is homogeneous, *Q* statistics and degree of freedom (*Df*) are analyzed. A statistically significant *Q_T* (overall variance) value suggests that the

difference between the studies is not due to sampling errors. Therefore, a further grouping of the studies is required by setting up potential moderator variables (grouping factors) to evaluate their influence on the informal PjBL.

Setting Up Moderator Variables

To determine whether the effectiveness of informal PjBL is impacted by influential variables, this study sets potential moderator variables. The seven moderator variables investigated in the study are as follows: teaching models, assessment methods, school level, instructional time, subject area, study location, and student group size. The study features were initially coded and categorized by two authors, experts in the field.

When screening the studies based on the “assessment models” employed, the categories generated were: (a) indirect assessment, (b) a combination of direct & indirect, and (c) triangulation methods. Surprisingly, no published study is known to have employed only direct assessment alone in an informal PjBL setting. To classify the studies in accordance with the assessment methods used, the assessment tools were coded as per the keywords devised. For direct assessment, the keywords employed were: pre-and post-tests, students’ posters, students’ presentations, working models, students’ designs, standard examinations, summative assessment, portfolio, rubrics, etc. For indirect assessment, the keywords screened were pre- and post-test (affective), surveys, focus group interviews, interviews, student self-report surveys, exit surveys, teachers/ mentors/ volunteers’ feedback, etc. For course-embedded assessment, the keywords were assignments, homework, worksheets, design logs, assay, course-embedded class tests, etc. For categorizing studies based on “teaching methods”, two categories were devised: (1) Informal PjBL and (2) Informal integrated PjBL. In the “Informal PjBL” category, studies exclusively employed the PjBL model, while for the “integrated PjBL” category, studies executing the PjBL model in integration with other teaching models (e.g., problem-based, inquiry-based, game-based, etc.) were included. Similarly, the articles’ categorization based on “educational level” were as per (1) high school, (2) middle & high school, (3) middle school, (4) primary and middle school, and (5) primary school students (Table 2). Likewise, studies were grouped based on “students’ group size” as per

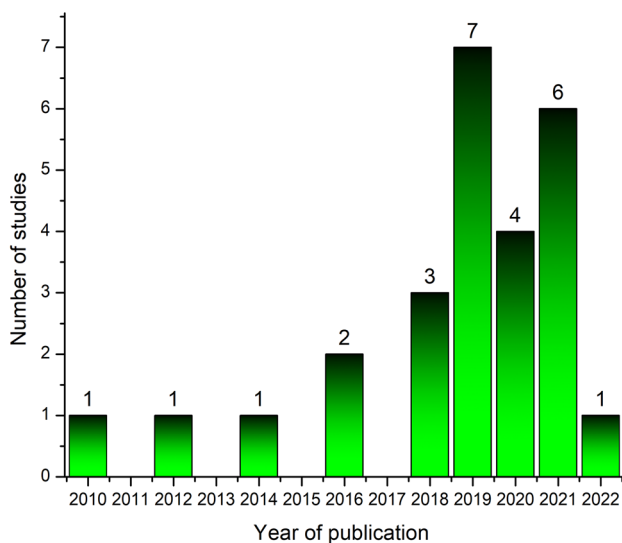


Fig. 2 Distribution of studies included in the meta-analysis by year of publication

Table 2 Results of the moderator analysis

Moderator variables	Heterogeneity between groups (Q_B)	K	p-value	Overall ES (Cohen's d)	95% CI		Heterogeneity within groups (Q_W)
					lower	Upper	
Teaching method							
Total	7.787*	26	0.005				
Integrated PjBL		10	0.087	0.132	0.045	0.219	15.149
PjBL		16	0.000	0.277	0.223	0.331	74.651*
Assessment method							
Total	9.086*	26	0.011				
Indirect		6	0.801	0.140	-0.092	0.327	2.335
Direct & indirect		5	0.340	0.178	0.097	0.258	4.521
Triangulation		15	0.000	0.303	0.240	0.367	81.646*
Educational level							
Total	7.829	26	0.098				
High school		6	0.000	0.286	0.202	0.369	25.305*
Middle & high school		4	0.832	0.181	0.050	0.312	0.910
Middle school		11	0.000	0.293	0.210	0.376	57.237*
Primary & middle school		2	0.491	0.143	0.035	0.251	0.474
Primary		3	0.054	0.155	0.007	0.303	5.832
Group Size							
Total	8.684*	26	0.034				
Less than 3		2	0.510	0.142	0.036	0.247	0.453
Within 3–4		13	0.000	0.237	0.194	0.352	38.412*
More than 5		5	0.776	0.169	0.070	0.267	0.506
Not specified		8	0.000	0.315	0.226	0.404	49.55*
Subject area							
Total	11.775	26	0.019				
Engineering		4	0.610	0.147	0.049	0.245	1.821
Mathematics		2	0.363	0.204	-0.008	0.145	0.827
Science		10	0.000	0.321	0.251	0.392	61.933*
Science & Technology		4	0.050	0.260	0.124	0.3960	7.794
STEM		6	0.020	0.149	0.047	0.252	13.437*
Study Location							
Total	5.887	26	0.315				
Asia		3	0.185	0.229	0.093	0.366	3.378
Europe		4	0.007	0.282	0.137	0.428	12.207*
Eastern U.S.A		5	0.000	0.219	0.099	0.339	25.977*
Southern U.S.A		5	0.000	0.291	0.205	0.378	42.260*
Southwestern U.S.A		3	0.793	0.138	0.038	0.239	0.463
Western U.S.A		6	0.192	0.263	0.149	0.378	7.414
Course Duration							
Total	49.679*	26	0.000				
≥ a month		6	0.029	0.175	0.095	0.255	12.477*
3 weeks		3	0.819	0.120	-0.068	0.308	0.399
Within 1–2 weeks		9	0.904	0.151	0.051	0.252	3.436
Within a week		7	0.000	0.284	0.209	0.359	31.596*
Not specified		1	1.000	1.071	0.817	1.324	0.000

Q_B Heterogeneity between the studies, Q_W Heterogeneity within the studies, Q_T Heterogeneity of total studies, CI Confidence interval, K Number of studies, ES Effect size in terms of *Cohen's d*

$Q_T=97.59, I^2=74.38$, *: significant p -value associated with Q_B & Q_W (i.e. p -value < 0.05)

i.e., (1) less than 3 students in a group, (2) within 3–4 students, (3) more than 5 students, and (4) not specified. For the “subject area”, the studies were classified based on the following categories: (1) science, (2) engineering, (3) mathematics, (4) science & technology, and (5) STEM. Grouping the studies as per the “study location” involved the following categories (1) Asia, (2) Europe, (3) Eastern U.S.A., (4) Southern U.S.A., (5) Southwestern U.S.A., and (6) Western U.S.A. Finally, the classification of articles based on “instruction time”, included the following categories i.e., (1) instructional delivery within a week, (2) within 1–2 weeks, (3) for 3 weeks, (4) equal to or more than a month, and (5) not specified.

Results

Prior to addressing the proposed research questions, the reliability of the meta-analysis has been verified by performing tests for heterogeneity, sensitivity, and publication bias. The findings have been briefed below:

Test for Heterogeneity

In order to assess the heterogeneity of the studies shortlisted for the meta-analysis, Cochran's Q statistic and I^2 statistic were employed (Cochran, 1954). A significant Q statistic demonstrates that ES s come from different populations (heterogeneity), signifying the use of the random-effects model. Typically, the Q statistic is a test for the null hypothesis, where the null hypothesis (which states that the true ES is the same for all studies) is rejected, if the Q value is not equal to degrees of freedom (Df). In this study, since the Q -value is 97.59 and the Df is 25, the null is rejected. Correspondingly, Cochran's Q statistic was employed to investigate if the overall effectiveness of informal PjBL is affected by the moderator variables. For

this, the between-class variance component Q_B was computed ($Q_B = Q_T - Q_W$, where Q_T and Q_W are overall and within-class variance components respectively). If the Q_B value is significant, then it relates to the fact that the moderator variables influence the overall effectiveness of the studies. Similarly, the I^2 statistic is the ratio of observed variance to true variance in ES s. I^2 values between < 20%, 20–50%, 50%–75%, and $\geq 75\%$ depict low, moderate, high, and very high heterogeneity respectively. Our study estimated an average I^2 value of 74.38%, with a significant p -value < 0.001, indicating the need for moderator analysis.

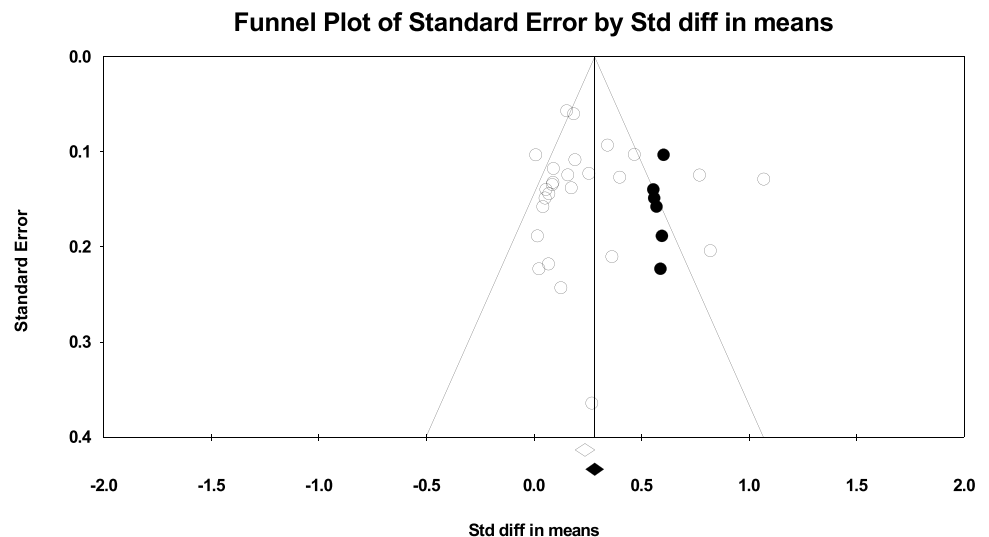
Test For Sensitivity

Sensitivity analysis was used to determine the existence of any ES /studies that unnecessarily influence the central tendency and variability of the data. The ES s of the studies were analyzed using the “one study removed procedure” in CMA software (Borenstein et al., 2009). The result showed that, for each study removed, the highest mean in the random model was $d = 0.260$, $n = 26$, $SE = 0.051$, and the lowest mean was $d = 0.213$, $n = 26$, $SE = 0.039$. These two new weighted average ES s lie within the confidence interval of the whole dataset [$n = 26$, $d = 0.248$, 95% CI {0.151; 0.346}, $p < 0.001$]. This suggested that no incongruities were observed to have an influential effect on the pooled average ES calculated. Finally, a test for publication bias was conducted to ensure the authenticity of the meta-analysis.

Test for Publication Bias

The existence of publication bias was estimated using a funnel plot Borenstein et al., 2009), a trim-and-fill model (Duval & Tweedie, 2000), and a classic fail-safe N (Rosenthal, 1979). The funnel plot demonstrates the relationship between ES s and standard errors (SE). An asymmetrical funnel plot (Fig. 3)

Fig. 3 Funnel plot for publication bias



revealed some amount of publication bias due to small-scale studies. To further estimate the influence of any significant publication bias, Rosenthal’s method of the classic fail-safe N was employed (Rosenthal, 1979). The publication bias is not significant if the number of missing studies in the classic fail-safe N is higher than the tolerance level of $5n + 10$, where n is the number of studies included in the study ($n = 26$). The classic fail-safe number (i.e., 609), is greater than the tolerance level value of 140 [i.e., $5(26) + 10$], demonstrating the fact that publication bias is insignificant to this study, and its results are reliable.

Meta-Analysis and Results

This section of the results addresses the proposed research questions (RQs). For RQ1 and RQ2, the measure of the effectiveness of studies employing informal PjBL was estimated using *Cohen’s d* value (Cohen, 2013). *Cohen’s d* value is the “standardized difference in means” that has been employed to compute the study’s effectiveness (effect size; *ES*). A positive *ES* would favor the informal PjBL model (treatment intervention) over the traditional classroom settings (control intervention), while a negative *ES* would favor the control over the treatment intervention. Furthermore, *Cohen’s d* values ≤ 0.2 , $0.2-0.5$, $0.5-0.7$, ≥ 0.8 are regarded as low, medium, high, and very high effects respectively.

Results Regarding Effectiveness

When compared to the control setting, the impact of informal PjBL shows moderate effectiveness, with an *ES* value of 0.248 [$n = 26$, $d = 0.248$, 95% *CI* {0.151;0.346}, $p < 0.001$]. These results illustrate the significance of informal PjBL in terms of students’ academic gains, compared to the traditional classroom setting. In this context, Fig. 4 depicts a forest plot revealing the distribution of *ES*s (of the individual studies) shortlisted in the paper. In the forest plot, the squares (on the right) demonstrate the *ES*s of the different studies while the diamond (at the bottom) presents the overall *ES*. The confidence interval is shown by the lines over the squares and diamonds. The *ES*s range from low ($d = 0.011$) to high ($d = 1.071$) effects. The distribution of true *ES*s was also calculated using CMA prediction interval software (Fig. 5). The true *ES*s in 95% of all comparable populations fall in the prediction interval range of -0.19 to 0.69.

Furthermore, in addition to analyzing the effectiveness of informal PjBL on students’ gains, this study also investigates possible moderator variables that might impact the informal PjBL ($Q_T = 97.587$, $df = 25$, $p < 0.001$) Therefore, the study investigates the impact of the following moderator variables: Teaching method, assessment method, grade, location, course duration, group size, and subject area. Table 2 depicts the findings of moderator analysis, illustrating that the teaching

Effectiveness of Informal PjBL

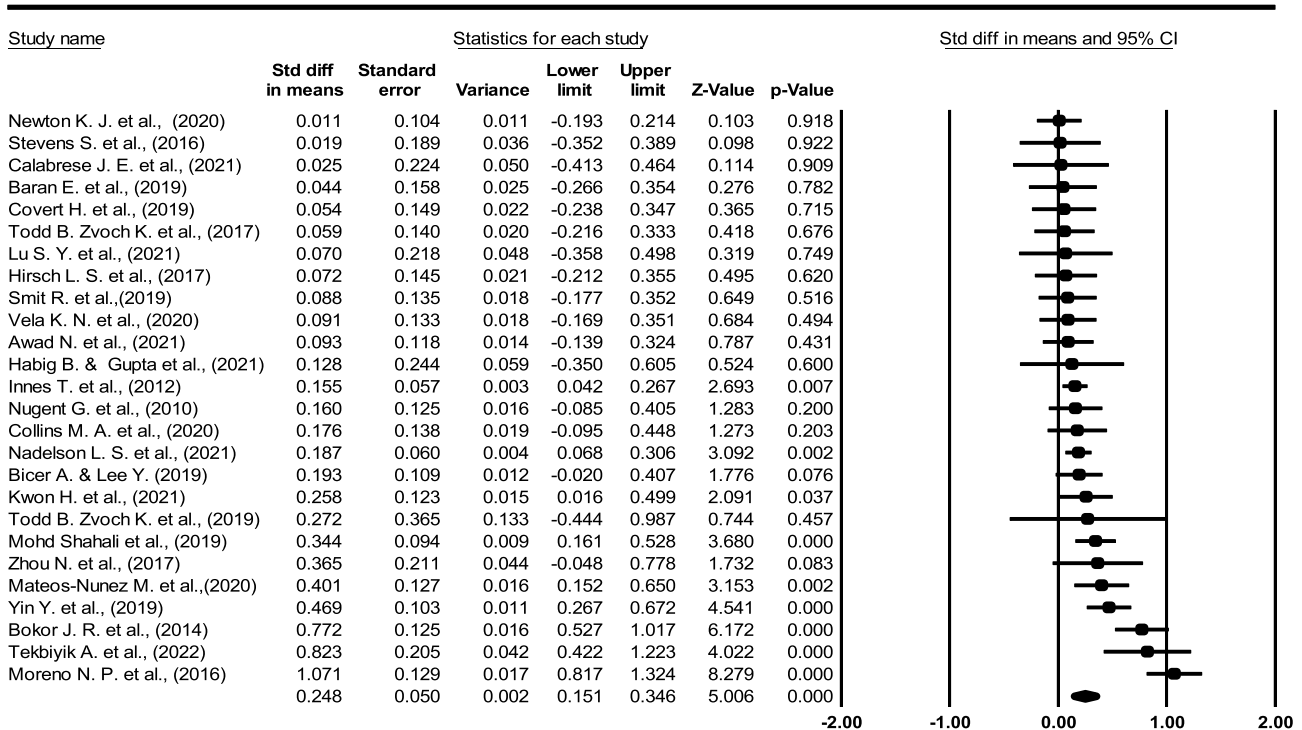
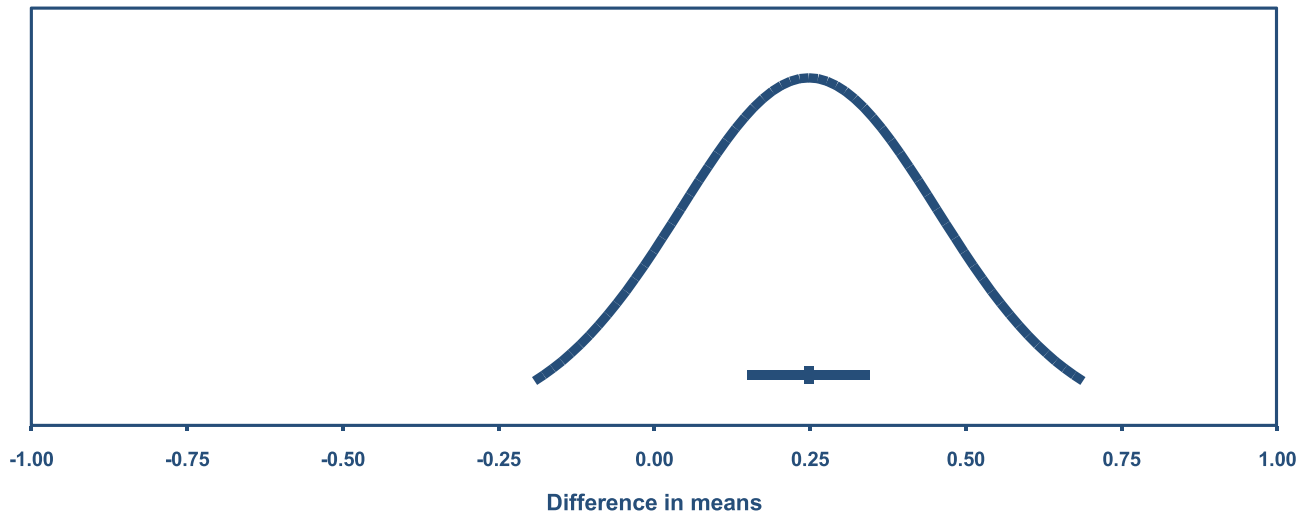


Fig. 4 Forest plot depicting the effectiveness of informal PjBL on students’ learning gains

Distribution of True Effects



The mean effect size is 0.25 with a 95% confidence interval of 0.15 to 0.35
 The true effect size in 95% of all comparable populations falls in the interval -0.19 to 0.69

Fig. 5 The distribution of true effects

method, assessment model, student group size, subject area, and instructional time do influence the effectiveness of informal PjBL. Meanwhile, educational level and study location were shown to have no impact on the effectiveness of informal PjBL. A comprehensive dissection of the individual moderator results has been described followingly.

Results Regarding Teaching Models

Q_B value was significant (p -value < 0.05 ; refer to Table 2) for the “teaching models”, depicting that utilizing informal PjBL is more effective than integrated PjBL in improving students’ gains (for STEM education). Indeed, the p -value is not significant for the integrated PjBL model, and more studies are required to explore the impact of integrated PjBL on students’ learning. The overall ES was significantly greater for the informal PjBL model ($n = 16$, $d = 0.277$, 95% CI {0.223;0.331}, $p < 0.000$) than for informal integrated PjBL models [$n = 10$, $d = 0.132$, 95% CI {0.045;0.219}, $p = 0.087$] (refer to Fig. S1).

Results Regarding Assessment Methods

Q_B statistics was significant (p -value < 0.05) for “assessment methods” and, resultantly, a significant difference existed between the individual studies (Table 2). In other words, findings indicate that informal PjBL had a significantly greater impact when employing the triangulation assessment method [$n = 15$, $d = 0.334$, 95% CI {0.240;0.367}, $p < 0.001$] than indirect assessment exclusively [$n = 6$, $d = 0.140$, 95%

CI {-0.092;0.327}, $p = 0.272$] or a combination of direct and indirect assessment methods [$n = 5$, $d = 0.178$, 95% CI {0.097;0.258}, $p = 0.119$] (refer to Fig. S2). As the p -value is not significant for the indirect and direct/indirect assessment models, it is not possible to conclude their impact on students’ gain during informal PjBL for STEM education. Therefore, more studies employing various assessment models need to be conducted to verify the impact of informal PjBL in a more accurate manner.

Results Regarding the Educational Level and Study Location

Q_B value was not significant for the “educational stage” and “study location” (as the associated p -value is more than the critical value), depicting that there are no significant differences between the individual studies (Table 2). In other words, the effects of PjBL conducted on primary, secondary, or high school students, hailing from different locations (countries) were not different and do not influence the effectiveness of informal PjBL (refer to Figs. S3 and S4).

Results Regarding Student Group Size

Q_B value was significant for group size, indicating that it does influence informal PjBL (p -value < 0.05) (Table 2). The results indicate that PjBL had a significantly greater

impact when students were grouped in small groups of 3–4 members [$n=13$, $d=0.237$, 95% CI {0.194;0.352}, $p<0.001$], rather than “less than 3” [$n=2$, $d=0.142$, 95% CI {0.036;0.247}, $p=0.510$] or “more than 5” [$n=5$, $d=0.169$, 95% CI {0.070;0.267}, $p=0.776$] (refer to Fig. S5). As the p -value is not significant for the “less than 3” and “more than 5”, it is not possible to conclude their influence on students’ gain during informal PjBL for STEM education.

Results Regarding the Subject Area

Q_B statistics have been significant (p -value <0.05) for the subject area, indicating a significant difference between the individual studies (Table 2). The findings illustrate that PjBL had a significantly greater impact when students were taught the general science subjects i.e., physics, chemistry, biology [$n=10$, $d=0.321$, 95% CI {0.251;0.392}, $p<0.001$], rather than engineering [$n=4$, $d=0.147$, 95% CI {0.049;0.254}, $p=0.003$] mathematics [$n=2$, $d=0.204$, 95% CI {0.008;0.415}, $p=0.059$]; science & technology [$n=4$, $d=0.260$, 95% CI {0.124;0.396}, $p<0.001$], STEM [$n=6$, $d=0.149$, 95% CI {0.047;0.252}, $p=0.004$] (refer to Fig. S6).

Results Regarding Instruction Time

Table 2 also showcases that the Q_B was significant (p -value <0.05) for the instruction time, indicating a significant difference between the individual studies. The findings illustrate that PjBL had a significantly greater impact when the course duration was within a week [$n=7$, $d=0.284$, 95% CI {0.209;0.359}, $p<0.001$] (refer to Fig. S7). Though for the “equal to or more than a month” category, the p -value was statistically significant, the overall ES showed low impacts ($d=0.175$) on students’ gains. For the other categories (2 weeks and 3 weeks), the p -value associated with ES is statistically insignificant. Therefore, more studies employing 2 and 3 weeks of instruction need to be conducted to verify the results.

Discussion

Undoubtedly, the informal PjBL model requires students and teachers to reflect on, evaluate, and update themselves and their approaches continuously as it is a learning process that cannot be entirely predetermined (Chounta et al., 2017). Hence, this review incorporates a meta-analysis of 12 years of evidence-based empirical research, comparing the effects of informal PjBL (treatment intervention) and traditional classroom settings (control intervention) on students’ gains. The review also investigated the effectiveness of informal PjBL by examining several moderator variables that might influence its implementation. The findings reported moderate

effectiveness of the informal PjBL model in STEM education ($d=0.248$), revealing a positive effect on students’ learning gains (when compared to the traditional classroom setting). All the studies showed a positive ES value, favoring the informal PjBL. Of the 26 studies, 10 studies showed a statistically significant p -value (<0.05) [Kwon et al., 2021 (Bokor et al., 2014; Innes et al., 2012; Mateos-Nunez et al., 2020; Moreno et al., 2016; Nadelson et al., 2021; Tekbıyık et al., 2022; Yin et al., 2020; Zhou et al., 2017; Mohd Shahali et al., 2019). The distribution of ES value ranged from low [$d=0.011$; (Newton et al., 2020)] to high [$d=1.071$; (Moreno et al., 2016)]. The measure of effectiveness shown by this study ($d=0.248$) is lower than the previous meta-analytical studies showcasing the effectiveness of PjBL models [i.e., $d=0.71$ (Chen & Yang, 2019), $d=0.99$ (Ayaz & Söylemez, 2015) and $d=1.06$ (Balemen & Keskin, 2018)]. This disparity might be because this study is exclusively based on informal PjBL models, while the prior studies have not specified any distinctions in this regard (whether formal or informal PjBL). In addition, the aforementioned studies have included articles of primary to tertiary educational level, while this study has been limited to the school level (grades 1–12). Moreover, this study is specific to STEM education, while other studies (Balemen & Keskin, 2018; Chen & Yang, 2019) included articles with other subjects (such as social sciences). In addition, the findings of this meta-analysis (after investigating the Q statistics) suggested that students’ gains in informal PjBL for STEM education are significantly influenced by many moderator variables $Q_B=9.08$, $df=25$, $p<0.001$. Wherein the moderator analysis revealed that the mean weighted ES of informal PjBL was influenced by the teaching model, assessment method, students’ group size, and course duration, but not by the educational level of the participants and the study location. The particular results of the individual moderator analysis have been discussed followingly.

Correspondingly, the moderator analysis of the “teaching model” has revealed moderately high effectiveness for the informal PjBL model ($d=0.277$) in improving student gains, compared to the informal integrated PjBL models ($d=0.132$). A potential reason behind this could be the fact that employing a combination of pedagogies might make it more challenging for students to acquire the contents. Adopting a particular teaching model as per the student’s requirements and learning objectives has always been recommended (Bielefeldt, 2013).

Likewise, the moderator analysis of the “assessment method” has been performed since developing a valid assessment model for informal STEM learning has been quite challenging (Gao et al., 2020). The findings indicate that informal PjBL had a significantly greater impact on students’ gains (compared to a traditional classroom setting) when employing the triangulation assessment method ($d=0.303$) rather than indirect assessment alone ($d=0.140$)

or direct and indirect assessment methods ($d=0.78$). An adequate combination of the assessment tools, considering the triangulation method must be deemed since it is crucial for instructors to carefully interlink learning objectives, pedagogical models, and assessment methods (Tuunila & Pulkkinen, 2015). Wherein such a careful alignment of assessment methods with learning objectives increases the possibility for instructors to provide students with opportunities to learn, attain knowledge, and practice skills in a more efficient manner (Ghrayeb et al., 2011).

While PjBL has been practiced in Asia (e.g., Awad, 2021; Mohd Shahali et al., 2019; Lu et al., 2021), there has been a dearth of literature examining its effectiveness when compared to Western contexts. Thus, more cross-national research is required to support researchers and educators to better comprehend the effectiveness of PjBL in diverse cultural contexts. Thus the moderator analysis has been conducted and the Q_B statistics were found to be not statistically significant for “study location”, meaning they do not influence the effectiveness of informal PjBL. Similarly, it has been widely researched that effective implementation of informal PjBL should pertain to which educational level (Ayaz & Söylemez, 2015; Balemen & Keskin, 2018; Chen & Yang, 2019). PjBL has been widely employed in high school, and more work is required to showcase its effectiveness at the primary level (Han et al., 2015). Thus we employed moderator analysis and Q_B statistics were found to be not statistically significant for “educational level”, meaning they do not influence the effectiveness of informal PjBL. These findings (regarding study location & educational level) are in partial alignment with the meta-analytical review by Chen and Yang (2019). It’s also noteworthy that the context differs i.e., in terms of educational levels targeted (i.e., their study included articles of primary to tertiary level, while this study incorporated only school-level education). In addition, their study has categorized articles based on location in Europe, North America, and western and eastern Asia, whereas our study included categories of Asia, Europe, southern U.S.A., southwestern U.S.A, and western U.S.A.

While, the Q_B value was significant for “group size”, indicating that it does influence the informal PjBL model (p -value < 0.05). The results indicate that informal PjBL had a significantly greater impact when students are grouped in small groups of 3 to 4 members ($d=0.237$) rather than “less than 3” ($d=0.142$) or “more than 5” ($d=0.169$). Some studies have shown that a small group size of 3–4 is suitable for a class size of 18–32 students (Calabrese & Capraro, 2021; Tekbiyik et al., 2022; Todd & Zvoch, 2019), while others didn’t specify the batch/class size (Bokor et al., 2014; Mateos-Nunez et al., 2020; Newton et al., 2020). These findings are in accordance with the study by Bertucci et al. (2010) and Kooloos et al., 2011, which recommends

grouping students in groups of 3–4 and less than 5 members respectively for better learning gains, students’ participation, and satisfaction in PjBL (Bertucci et al., 2010; Kooloos et al., 2011). A potential justification could be that in a very small group, there might be less peer discussion, and in very large groups there might be a conflict of interest due to the diversity of ideas and perceptions. Interestingly, the findings from Apedoe et al. (2012), are also noteworthy and partially align with this study’s findings (Apedoe et al., 2012). Though their study reported the effectiveness of pairing and grouping in 3–4, they contended these findings to be seen in conjugation with the class type (e.g., basic vs. advanced) and the type of knowledge (basic vs. advanced). For the basic knowledge (i.e., requiring no transfer of gains) the size of the group is probably not significant. However, for advanced knowledge (i.e., requiring close transfer of gains) the students in basic classrooms benefit most from groups of 3–4, while students in advanced classrooms are served better by working in pairs (Apedoe et al., 2012). Despite our study results being fascinating, none of the shortlisted articles has incorporated the concept/significance of “role assignment in groups” (Schellens et al., 2007).

Followingly the moderator analysis of the “subject area” computed a significant Q_B (p -value < 0.05). The results indicate that informal PjBL had a significantly greater impact when students were taught general science subjects ($d=0.321$), rather than engineering, mathematics & technology. A possible reason for this finding may be that PjBL incorporates investigations, experiments, modeling, and interpreting and it is best employed in general science disciplines. These findings are in pact with the study by Ayaz and Söylemez (2015), which reported the highest effectiveness for “general science”, compared to biology, physics, and chemistry disciplines (Ayaz & Söylemez, 2015). Dissecting the various science disciplines, the study findings by Balemen and Keskin (2018) have been noteworthy (Balemen & Keskin, 2018). They reported the highest impact of PjBL models when employed for biology subjects ($d=1.147$), compared to physics and chemistry. However, they didn’t compute the effectiveness of PjBL when employed for technology and engineering subjects.

Finally, the moderator analysis of “instructional time”, has been performed to comprehend the time that should be devoted to informal PjBL for the effective acquisition of students’ gains. Findings depicted a significant Q_B value (p -value < 0.05), suggesting that informal PjBL has a higher impact when the course duration was within a week ($d=0.284$), rather than 1–2 weeks, 3 weeks, and \geq a month. These findings are partially in accordance with the meta-analytical findings of Ayaz & Söylemez (2015), which recommended $1 \leq h \leq 20$, to be the most effective course duration for PjBL models (Ayaz & Söylemez, 2015). Though much of the informal education span for a

day, none of the articles shortlisted for this meta-analysis employed a 1-day workshop or 1 session program. Conclusively, all the seven-moderator analysis performed in the study has proven critical in understanding the impact of various moderator variables on the effectiveness of informal PjBL for STEM education (with respect to the traditional classroom setting).

The study findings should be seen in light of some limitations. This study has solely focused on peer-reviewed articles and discarded data from theses, books, conference proceedings, etc. The review incorporated only empirical literature, aligning with the CMA format. Nonetheless, our study raises important opportunities for future research. For example, more study is needed to look at longitudinal informal PjBL interventions because some student variables might take time to be expressed and assessed. Further research could also include a meta-analysis incorporating undergraduate student participants and further investigate informal PjBL for non-STEM subjects (humanities, language) could also be explored.

Conclusion

The findings of the study demonstrated a positive impact of informal PjBL on students' learning gains ($d=0.248$) when compared to traditional classroom settings. Although previous meta-analytical studies have confirmed the effectiveness of the PjBL models (Ayaz & Söylemez, 2015; Balemen & Keskin, 2018; Chen & Yang, 2019), none of them have been explicit for informal PjBL model executed for school students in STEM education. Since in an informal setting, designing, and executing the PjBL model, considering all the influential factors, might be challenging, this study has investigated the effect of potential moderator variables that might impact informal PjBL models' effectiveness. The moderator analysis revealed that informal PjBL was influenced by the teaching model, assessment method, students' group size, and course duration, but not by the educational level of the participants and the study location. Thereby, the paper concludes that the “*informal PjBL*” model could be more effective for “*general science*” instruction when executed within “*a week*”, by dispersing students in small groups of “*3–4 members*”, and by employing “*triangulation assessment*”. Thus, we believe that this research would help academicians and researchers to comprehend the moderator variables affecting informal PjBL and assist them in designing and implementing effective informal PjBL strategies at the school level for STEM education.

Abbreviations CI: Confidence interval; CMA: Comprehensive meta-analysis; d : Cohen's d ; Df: Degree of freedom; ERIC: Educational Research Information Center; ES: Effect size; ESs: Effect sizes; n : Number of studies; PjBL: Project-based learning;

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analysis; Q_B : Between class variance component; Q_T : Total variance component; Q_W : Within class variance component; SD: Standard deviations; SE: Standard error; STEM: Science, technology, engineering, and mathematics

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10956-023-10063-y>.

Acknowledgements The authors thank Qatar University Young Scientists Center (QUYSC) or its support during this work.

Authors' Contributions Conception or design of the work: NS, ZA; Data collection: HF, AM, MS; Data analysis and interpretation: MS, HF, AM, NS, ZA, AS, NF; Drafting the article: MS; Critical revision of the article: ZA, JB, AS, NF; Final approval of the version to be published: MS, HF, MA, NS, JB, NA, AS, NF, ZA.

Funding Open Access funding provided by the Qatar National Library.

Availability of Data and Materials The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethical Statement Not applicable.

Consent Statement Not applicable.

Competing Interests The authors declare no conflict of interest.

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

- Al-Kadri, H. M., Al-Moamary, M. S., Roberts, C., & Van der Vleuten, C. P. (2012). Exploring assessment factors contributing to students' study strategies: literature review. *Medical teacher*, *34*(sup1), S42–S50. <https://doi.org/10.3109/0142159X.2012.656756>
- Apedoe, X. S., Ellefson, M. R., & Schunn, C. D. (2012). Learning together while designing: Does group size make a difference? *Journal of Science Education and Technology*, *21*(1), 83–94. <https://doi.org/10.1007/s10956-011-9284-5>
- Awad, N. (2021). Exploring STEM integration: assessing the effectiveness of an interdisciplinary informal program in fostering students' performance and inspiration. *Research in Science & Technological Education*, 1–25. <https://doi.org/10.1080/02635143.2021.1931832>
- Ayaz, M. F., & Söylemez, M. (2015). The effect of the project-based learning approach on the academic achievements of the students in

- science classes in Turkey: A meta-analysis study. *Egitim ve Bilim*, 40(178). <https://doi.org/10.15390/EB.2015.4000>
- Baran, E., Canbazoglu Bilici, S., Mesutoglu, C., & Ocak, C. (2019). The impact of an out-of-school STEM education program on students' attitudes toward STEM and STEM careers. *School Science and Mathematics*, 119(4), 223–235. <https://doi.org/10.1111/ssm.12330>
- Balemen, N., & Keskin, M. Ö. (2018). The effectiveness of Project-Based Learning on science education: A meta-analysis search. *International Online Journal of Education and Teaching*, 5(4), 849–865. Retrieved April 15, 2022, from <https://eric.ed.gov/?id=EJ1250564>
- Bertucci, A., Conte, S., Johnson, D. W., & Johnson, R. T. (2010). The impact of size of cooperative group on achievement, social support, and self-esteem. *The Journal of General Psychology: Experimental, Psychological, and Comparative Psychology*, 137(3), 256–272. <https://doi.org/10.1080/00221309.2010.484448>
- Bicer, A., & Lee, Y. (2019). Effect of STEM PBL embedded informal learning on student interest in STEM majors and careers. *Journal of Mathematics Education*, 12(1), 57–73. <https://doi.org/10.26711/007577152790038>
- Bielefeldt, A. R. (2013). Pedagogies to achieve sustainability learning outcomes in civil and environmental engineering students. *Sustainability*, 5(10), 4479–4501. <https://doi.org/10.3390/su5104479>
- Bokor, J. R., Landis, J. B., & Crippen, K. J. (2014). High school students' learning and perceptions of phylogenetics of flowering plants. *CBE—Life Sciences Education*, 13(4), 653–665. <https://doi.org/10.1187/cbe.14-04-0074>
- Borenstein, M., Hedges, L. v., Higgins, J. P. T., & Rothstein, H. R. (2009). Introduction to meta-analysis. *Introduction to Meta-Analysis*, 1–421. <https://doi.org/10.1002/9780470743386>
- Calabrese, J. E., & Capraro, R. M. (2021). The autonomy of informal STEM and benefits of andragogy with gifted children. *Journal of Research in Innovative Teaching & Learning*. <https://doi.org/10.1108/JRIT-08-2021-0060>
- Chen, C. H., & Yang, Y. C. (2019). Revisiting the effects of project-based learning on students' academic achievement: A meta-analysis investigating moderators. *Educational Research Review*, 26, 71–81. <https://doi.org/10.1016/j.edurev.2018.11.001>
- Chounta, I. A., Manske, S., & Hoppe, H. U. (2017). "From Making to Learning": Introducing Dev Camps as an educational paradigm for Re-inventing Problem-based Learning. *International Journal of Educational Technology in Higher Education*, 14(1), 1–15. <https://doi.org/10.1186/s41239-017-0061-2>
- Cilliers, F. J., Schuwirth, L. W., Herman, N., Adendorff, H. J., & van der Vleuten, C. P. (2012). A model of the pre-assessment learning effects of summative assessment in medical education. *Advances in Health Sciences Education*, 17, 39–53. <https://doi.org/10.1007/s10459-011-9292-5>
- Cochran, W. G. (1954). Some methods for strengthening the common χ^2 tests. *Biometrics*, 10(4), 417–451. <https://doi.org/10.2307/3001616>
- Cohen, J. (2013). Statistical power analysis for the behavioral sciences. *Routledge*. <https://doi.org/10.4324/9780203771587>
- Collins, M. A., Totino, J., Hartry, A., Romero, V. F., Pedrosa, R., & Nava, R. (2020). Service-learning as a lever to support STEM engagement for underrepresented youth. *Journal of Experiential Education*, 43(1), 55–70. <https://doi.org/10.1177/1053825919887407>
- Covert, H., Ilunga Tshiswaka, D., Ramkissoon, I., Sisskin, E., Lichtveld, M., & Wickliffe, J. (2019). Assessing science motivation among high school students participating in a supplemental science programme: The emerging scholars environmental health sciences academy. *International Journal of Science Education*, 41(17), 2508–2523. <https://doi.org/10.1080/09500693.2019.1689308>
- DerSimonian, R., & Laird, N. (2015). Meta-analysis in clinical trials revisited. *Contemporary Clinical Trials*, 45, 139–145. <https://doi.org/10.1016/j.cct.2015.09.002>
- Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56(2), 455–463. <https://doi.org/10.1111/J.0006-341X.2000.00455.X>
- Gao, X., Li, P., Shen, J., & Sun, H. (2020). Reviewing assessment of student learning in interdisciplinary STEM education. *International Journal of STEM Education*, 7(1), 1–14. <https://doi.org/10.1186/s40594-020-00225-4>
- Gerretson, H., & Golson, E. (2005). Synopsis of the use of course-embedded assessment in a medium sized public university's general education program. *The Journal of General Education*, 54(2), 139–149. <https://doi.org/10.2307/27798013>
- Ghrayeb, O., Damodaran, P., & Vohra, P. (2011). Art of triangulation: An effective assessment validation strategy. *Global Journal of Engineering Education*, 13(3), 96–101. Retrieved April 15, 2022, from <http://www.wiete.com.au/journals/GJEE/Publish/vol13no3/07-Vohra-P.pdf>
- Han, S., Capraro, R., & Capraro, M. M. (2015). How science, technology, engineering, and mathematics (STEM) project-based learning (PBL) affects high, middle, and low achievers differently: The impact of student factors on achievement. *International Journal of Science and Mathematics Education*, 13(5), 1089–1113. <https://doi.org/10.1007/s10763-014-9526-0>
- Habig, B., & Gupta, P. (2021). Authentic STEM research, practices of science, and interest development in an informal science education program. *International Journal of STEM Education*, 8(1), 1–18. <https://doi.org/10.1186/s40594-021-00314>
- Hirsch, L. S., Berliner-Heyman, S., & Cusack, J. L. (2017). Introducing middle school students to engineering principles and the engineering design process through an academic summer program. *International Journal of Engineering Education*, 33(1), 398–407. Retrieved April 15, 2022, from https://www.njit.edu/precollege/sites/njit.edu/precollege/files/International_Journal_of_Engineering_Education_Special_Edition_Hirsch.pdf
- Innes, T., Johnson, A. M., Bishop, K. L., Harvey, J., & Reisslein, M. (2012). The Arizona Science Lab (ASL): Fieldtrip based STEM outreach with a full engineering design, build, and test cycle. *Global Journal of Engineering Education*, 14(3), 225–232. Retrieved April 15, 2022, from <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.667.7339&rep=rep1&type=pdf>
- Johnson, D. R., Renzulli, L., Bunch, J., & Paino, M. (2013). Everyday observations: Developing a sociological perspective through a portfolio term project. *Teaching Sociology*, 41(3), 314–321. <https://doi.org/10.1177/0092055X13480642>
- Kim, Y. J., Murai, Y., & Chang, S. (2021). Implementation of embedded assessment in maker classrooms: Challenges and opportunities. *Information and Learning Sciences*, 122(3/4), 292–314. <https://doi.org/10.1108/ILS-08-2020-0188>
- Kokotsaki, D., Menzies, V., & Wiggins, A. (2016). Project-based learning: A review of the literature. *Improving Schools*, 19(3), 267–277. <https://doi.org/10.1177/1365480216659733>
- Kwon, H., Capraro, R. M., & Capraro, M. M. (2021). When I believe, I can: Success STEMs from my perceptions. *Canadian Journal of Science, Mathematics and Technology Education*, 21(1), 67–85. <https://doi.org/10.1007/s42330-020-00132-4>
- Kooloos, J. G. M., Klaassen, T., Vereijken, M., van Kuppeveld, S., Bolhuis, S., & Vorstenbosch, M. (2011). Collaborative group work: Effects of group size and assignment structure on learning gain, student satisfaction and perceived participation. *Medical Teacher*, 33(12), 983–988. <https://doi.org/10.3109/0142159X.2011.588733>

- Larmer, J., & Mergendoller, J. R. (2010). Seven essentials for project-based learning. *Educational leadership*, 68(1), 34–37. Retrieved April 15, 2022, from https://d1wqtxts1xzle7.cloudfront.net/64271329/Seven_Essentials_for_Project_Based_Learn-with-cover-page-v2.pdf?Expires=1663505928&Signature=aaHCdBmj9VLXYn5SKVMqVvwoNID16-0h4BmuNqAbL3RzLq1VBjKIpBr4u3WmnByrCmcybTrqN14ZiujdJy2LzkQdrI8auJ6UkFKunwtHU-36XYc9a3~kHpUpxB43W6l3BzWrWmD4jBpXEHpBxQZEn8QTbp3gyBBXUPe7d8XWqumeOsOIRCGGqMmXBx~s-WTMW-LW1~hP-XQvc9-1K1hgCCi69901rAtOIR9pvyv2-9fN0WnApumsCcVqE64XoDDFYJbDhBd6E5KDSvYDrVhvrjDko3QVvYZXP6oD9YPP8cNi-9oabe6CTq4I58fx7TxfnoANOT2kz2O8TqnL65vA__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA
- Lu, S. Y., Lo, C. C., & Syu, J. Y. (2021). Project-based learning oriented STEAM: The case of micro-bit paper-cutting lamp. *International Journal of Technology and Design Education*, 1–23. <https://doi.org/10.1007/s10798-021-09714-1>
- Marriott, P., & Lau, A. (2008). The use of on-line summative assessment in an undergraduate financial accounting course. *Journal of Accounting Education*, 26(2), 73–90. <https://doi.org/10.1016/j.jaccedu.2008.02.001>
- Martell, K., & Calderon, T. (2005). Assessment in business schools: What it is, here we are, and where we need to go now. In K. Martell & T. Calderon (Eds.), *Assessment of student learning in business schools: best practices each step of the way* (pp. 1–22). Tallahassee FL: Association for Institutional Research.
- Mateos-Nunez, M., Martinez-Borreguero, G., & Naranjo-Correa, F. L. (2020). Learning science in primary education with STEM workshops: Analysis of teaching effectiveness from a cognitive and emotional perspective. *Sustainability*, 12(8), 3095. <https://doi.org/10.3390/SU12083095>
- Mergendoller, J. R., & Thomas, J. W. (2005). Managing project based learning: Principles from the field. Retrieved April 15, 2022, from http://dr-hatfield.com/science_rules/articles/Managing%20Project%20Based%20Learning.pdf
- Miller, A. (2018, April 17). Time management with PBL. Retrieved April 15, 2022, from <https://www.edutopia.org/article/time-management-pbl>
- Mohd Shahali, E. H., Halim, L., Rasul, M. S., Osman, K., & Mohamad Arsad, N. (2019). Students' interest towards STEM: A longitudinal study. *Research in Science & Technological Education*, 37(1), 71–89. <https://doi.org/10.1080/02635143.2018.1489789>
- Moreno, N. P., Tharp, B. Z., Vogt, G., Newell, A. D., & Burnett, C. A. (2016). Preparing students for middle school through after-school STEM activities. *Journal of Science Education and Technology*, 25(6), 889–897. <https://doi.org/10.1007/s10956-016-9643-3>
- Nadelson, L. S., Jemison, R. C., Soto, E., & Warner, D. L. (2021). Cultivating a New “SEED”: From an on-Ground to Online Chemistry Summer Camp. *Journal of Chemical Education*, 99(1), 129–139. <https://doi.org/10.1021/acs.jchemed.1c00280>
- Newton, K. J., Leonard, J., Buss, A., Wright, C. G., & Barnes-Johnson, J. (2020). Informal STEM: Learning with robotics and game design in an urban context. *Journal of Research on Technology in Education*, 52(2), 129–147. <https://doi.org/10.1080/15391523.2020.1713263>
- Nugent, G., Barker, B., Grandgenett, N., & Adamchuk, V. I. (2010). Impact of robotics and geospatial technology interventions on youth STEM learning and attitudes. *Journal of Research on Technology in Education*, 42(4), 391–408. <https://doi.org/10.1080/15391523.2010.10782557>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Systematic Reviews*, 10(1), 1–11. <https://doi.org/10.1136/BMJ.N71>
- Ramdhayani, H. G., Purwoko, A. A., & Muntari, M. (2019, October). Meta analysis: the effect of applying project-based learning models to students' science process skills. In *Journal of Physics: Conference Series*, 1321(3), 032089. IOP Publishing. <https://doi.org/10.1088/1742-6596/1321/3/032089>
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, 86(3), 638–641. <https://doi.org/10.1037/0033-2909.86.3.638>
- Schellens, T., van Keer, H., de Wever, B., & Valcke, M. (2007). Scripting by assigning roles: Does it improve knowledge construction in asynchronous discussion groups? *International Journal of Computer-Supported Collaborative Learning*, 2(2–3). <https://doi.org/10.1007/s11412-007-9016-2>
- Shpeizer, R. (2019). Towards a successful integration of project-based learning in higher education: Challenges, technologies and methods of implementation. *Universal Journal of Educational Research*, 7(8), 1765–1771. <https://doi.org/10.13189/ujer.2019.070815>
- Sivia, A., MacMath, S., Novakowski, C., & Britton, V. (2019). Examining student engagement during a project-based unit in secondary science. *Canadian Journal of Science, Mathematics and Technology Education*, 19(3), 254–269. <https://doi.org/10.1007/S42330-019-00053-X/TABLES/8>
- Smit, R., Robin, N., De Toffol, C., & Atanasova, S. (2021). Industry-school projects as an aim to foster secondary school students' interest in technology and engineering careers. *International Journal of Technology and Design Education*, 31(1), 61–79. <https://doi.org/10.1007/s10798-019-09538-0>
- Stevens, S., Andrade, R., & Page, M. (2016). Motivating young Native American students to pursue STEM learning through a culturally relevant science program. *Journal of Science Education and Technology*, 25(6), 947–960. <https://doi.org/10.1007/s10956-016-9629-1>
- Tekbıyık, A., Bulut, D. B., & Sandalcı, Y. (2022). Effects of a summer robotics camp on students' STEM career interest and knowledge structure. *Journal of Pedagogical Research*, 6(2), 91–109. <https://doi.org/10.33902/JPR.202212606>
- Todd, B. L., & Zvoch, K. (2019). The effect of an informal science intervention on middle school girls' science affinities. *International Journal of Science Education*, 41(1), 102–122. <https://doi.org/10.1080/09500693.2018.1534022>
- Todd, B., & Zvoch, K. (2017). Exploring girls' science affinities through an informal science education program. *Research in Science Education*, 49(6), 1647–1676. <https://doi.org/10.1007/s11165-017-9670-y>
- Tuunila, R., & Pulkkinen, M. (2015). Effect of continuous assessment on learning outcomes on two chemical engineering courses: Case study. *European Journal of Engineering Education*, 40(6), 671–682. <https://doi.org/10.1080/03043797.2014.1001819>
- Vela, K. N., Pedersen, R. M., & Baucum, M. N. (2020). Improving perceptions of STEM careers through informal learning environments. *Journal of Research in Innovative Teaching & Learning*, 13(1), 103–113. <https://doi.org/10.1108/jrit-12-2019-0078>
- Yin, Y., Hadad, R., Tang, X., & Lin, Q. (2020). Improving and assessing computational thinking in maker activities: The integration with physics and engineering learning. *Journal of Science Education and Technology*, 29(2), 189–214. <https://doi.org/10.1007/s10956-019-09794-8>
- Zhou, N., Pereira, N. L., George, T. T., Alperovich, J., Booth, J., Chandrasegaran, S., & Ramani, K. (2017). The influence of toy design activities on middle school students' understanding of the engineering design processes. *Journal of Science Education and Technology*, 26(5), 481–493. <https://doi.org/10.1007/s10956-017-9693-1>