Effectiveness of Evidence-based Instructional Practices on Students' Mathematics Achievement: A Meta-Analysis

Edrian Peter B. Villanueva
Maricar S. Prudente, PhD

Abstract

Examining trends in the literature regarding evidence-based instructional practices (EBIPs) enables mathematics educators to make well-informed decisions when selecting effective teaching strategies for their classrooms. The evidence-based pyramid suggests that systematic reviews, such as meta-analyses, are the highest level of evidence in a particular field. This study aimed to systematically analyze existing empirical studies on the impact of different evidence-based instructional approaches on students' mathematical achievement. Using a meta-analysis research design, 28 studies were examined. The findings indicated that, overall, EBIPs are successful in improving students' mathematical content knowledge and skills. Furthermore, the study revealed that among the EBIPs explored, teaching with cases was the most effective for elementary learners, while upside-down pedagogy yielded the best results for high school students. Additionally, the results showed that teaching with cases facilitates short-term comprehension of mathematical concepts, whereas upside-down pedagogy promotes long-term understanding.

Keywords: *meta-analysis, upside-down pedagogy, teaching with cases, POGIL, mathematics education, evidence-based practice*

Introduction

In recent years, pedagogical research has been focused on finding ways on how students' achievement may be enhanced. One of the trends in education research is the implementation of evidence-based teaching (EBT) (Borrego & Henderson, 2014). The evidence-based practice was originally implemented in the fields of clinical medicine and nursing wherein the available empirical evidence in the literature is integrated into clinical practice (Groccia & Buskist, 2011). In the field of education, one way to implement this evidence-based approach is by employing various evidence-based instructional practices (EBIPs). EBIPs are approaches to teaching that have been empirically shown to be effective in promoting and developing students' conceptual understanding (Burns & Ysseldyke, 2009; Sturtevant & Wheeler, 2019).

While there are a lot of teaching strategies available in the current literature that have been found to be effective in improving students' achievement, to our knowledge, there is no existing list of EBIPs in mathematics. In the existing literature, the only available list of EBIPs is for Science, Technology, Engineering and Mathematics (STEM) in general (Sturtevant & Wheeler, 2019). Such a list of EBIPs in mathematics is deemed to be of importance for teachers since it can serve as a guide on how and why students' mathematics proficiency is attained in a particular setting (Petty, 2009). The list is likewise important so teachers can easily choose and employ different teaching strategies that work.

Furthermore, within the realm of mathematics education, there exists a wide array of instructional approaches that can be classified as EBIPs. Given this extensive array, the present study narrows its focus to just three specific subsets of EBIPs. These subsets were identified as being the least commonly employed methods among mathematics educators, as indicated by our prior investigation (Villanueva & Prudente, 2022). The three underutilized EBIPs are Teaching with Cases (TWC), Process-Oriented Guided Inquiry Learning (POGIL), and Upside-down Pedagogy (UP). The rationale behind concentrating on these less-utilized EBIPs is to offer mathematics instructors insight into alternative and effective instructional strategies for incorporation into their classrooms. In this paper, when we refer to EBIP, we are specifically alluding to these three least-used practices. Presented in Table 1 are the definitions of TWC, POGIL, and UP.

TABLE 1

EBIP	Definition
Teaching with Cases	A teaching approach that uses a case from a book, article, story, simple question, or a real-life problem with sufficient details that allow the students to analyze and come up with a?
Process-Oriented Guided Inquiry Learning	A student-centered pedagogical approach that emphasizes small group collaboration, guided inquiry, and active learning to promote deeper understanding and critical thinking in STEM education.
Upside-down Pedagogy	Also known as inverted or flipped learning, is an instructional approach where traditional classroom activities such as lectures are moved outside of class, and homework-like activities such as problem-solving and discussion occur inside the classroom, allowing for more active and interactive learning.

As part of a much larger study, this systematic review aimed to collect substantial evidence which showcases the effectiveness of different pedagogical approaches in students' mathematics performance. Regarding this, the conduct of meta-analysis, which is considered to be the highest and the most common form of establishing evidence in the evidence-based pyramid (Murad et al., 2016), was employed in this study.

This current study will be the first to investigate the factors that influence the overall effect of the abovementioned EBIPs. This would be helpful to analyze the trend in the literature about each least-used evidence-based instructional practice. Accordingly, this study aimed to provide a systematic analysis of the existing empirical studies on the effect on students' mathematical achievement of different evidence-based instructional approaches. Specifically, this sought to answer the following questions:

1. What is the level of effectiveness of evidence-based instructional practices on students' mathematical content knowledge and math skills?
2. What is the relative effectiveness of each evidence-based instructional practice compared to other EBIPs in mathematics?
3. How do moderator variables influence the effects of EBIPs?

DESIGN & METHODS

Research Design

This study employed the meta-analysis research design. Meta-analysis is a technique of combining the empirical findings of previous research to create a synthesis of evidence (Basu, 2017). In the current study, the numerical findings from the empirical studies in the literature are pooled to arrive at an estimated effect of the EBIPs on students' mathematics performance.

Inclusion and Exclusion Criteria

The following inclusion criteria needed to be met by each study to be considered for inclusion in this current meta-analysis.

1. The publication date is from 2011 to 2020.
2. Samples are students in basic education or K-12 curriculum.

3. Research design is a quasi-experimental or experimental design where a group of participants that underwent an EBIP (treatment group) were compared against a control group (no treatment group).

4. Either mathematics achievement or mathematics skills (critical thinking, problem-solving, logical thinking) are assessed in the study.

5. Sufficient statistical data needed to compute for the effect size are present (mean, standard deviation, results of the tests of difference, effect size, and sample size).

6. Any journal article, thesis, or dissertation that is peer-reviewed or published in reputable journals (Web of Science, Scopus, Taylor and Francis, EBSCO, Publish and Perish).

Study Search Procedure

After setting the criteria for inclusion, the research started the article identification using Harzing's Publish or Perish 7. This initial article identification includes the databases of Google Scholar and SCOPUS. Each database was searched using 11 keywords (Table 2) which were paired with the terms "Math" or "Mathematics." Additionally, separate searches were conducted on EBSCO, ProQuest, and Taylor and Francis databases. These three databases were purposively chosen since the authors have legal access to these databases through De La Salle University. Furthermore, EBSCO and Taylor and Francis publish major journals in the Social Sciences, especially in the field of Education. Meanwhile, ProQuest publishes thesis and dissertations from various reputable institutions all over the world. The researchers used the Advanced Search options in identifying the records published on EBSCO, Taylor and Francis, and ProQuest. The terms per category in Table 1 were used interchangeably.

TABLE 2

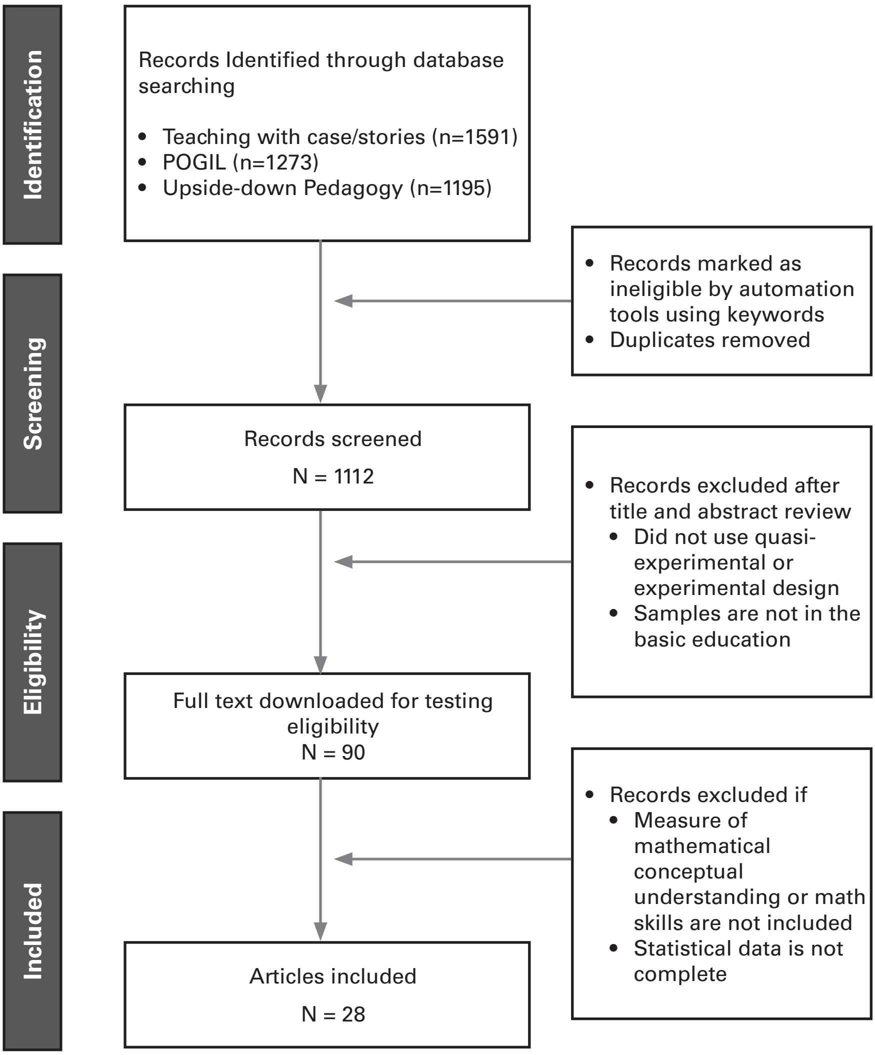
Keywords Used in the Study Search

Teaching with Cases/Stories	POGIL	Upside-down Pedagogy
Case-based	Process-oriented Guided Inquiry Learning	Upside-down
Teaching with cases	Guided Inquiry Learning	SCALE-UP
Story-based approach	POGIL	Flipped

Considering the inclusion criteria and study search procedure, 28 studies were exhausted from 4059 records which came from the initial article identification. Using Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram, Figure 1 summarizes the entire search procedure for this meta-analysis.

FIGURE 1

Flow Diagram of Different Articles at Different Phases



Coding Procedure

The 90 studies which passed the screening were narrowed down to 28 studies. This is because only the 28 studies contained sufficient information to calculate the effect sizes. The included studies were coded using the following identification and the corresponding scheme: study identification (first author's last name and year of publication), source of study (article, thesis, or dissertation), students' grade level (elementary, high school, or both), EBIP model used (conventional, or with innovation), assessment coverage (single topic or cumulative), type of test used (researcher-made or adopted).

Data Analysis

The researchers utilized Meta-essentials v.1.5 Workbook 3 (van Rhee et al., 2015). Hegde's g was used to represent the effect sizes of each study. There are studies in the present meta-analysis that compared means of pretest and post-test scores of the respondents. However, in terms of means, Workbook 3 of the meta-essentials only asks for means of two independent groups. Considering this, the Cohen's d for some studies were manually calculated using the following formula:

$$Cohen's\ d = \frac{M_t - M_c}{SD_{pooled}}$$

$$SD_{pooled} = \sqrt{\frac{(n_t - 1)SD_t + (n_c - 1)SD_c}{n_t + n_c - 2}}$$

Where:

M = mean
n = sample size
SD = standard deviation
t = treatment group
c = control group (non-treatment group)

After the necessary and sufficient data for each study was entered in the Meta-essentials, data analysis including the Forest plot, estimates of heterogeneity, and publication bias were conducted. The Forest plot was utilized to understand the effect size of each study relative to the overall effect size of the meta-analysis. Additionally, the estimates of heterogeneity used in this study were Q statistic, I2, and prediction interval. Q statistic is the widely used measure of heterogeneity in a meta-analysis. However, van Aert et al. (2019), posited the Q statistic is greatly influenced by the sample size. Thus, I2 was used to compensate for the limitation concerning the sample size. Yet, Borenstein et al. (2009), argued that while I2 is not greatly affected by the sample size, it is influenced by the accuracy of statistics.

Therefore, they also suggested the use of prediction interval which is neither influenced by sample size nor accuracy of the statistics in a study. Lastly, to determine whether publication bias exists in the current meta-analysis, Funnel plot coupled with Eger's test and Begg and Mazumdar's test were utilized.

Results

The 28 articles that met the criteria are presented in Table 3. A total of 6026 students were included in the study. It can be noticed that most of the included studies implemented Upside-down pedagogy. Meanwhile, POGIL and teaching with cases both have five empirical studies involved in the meta-analysis.

TABLE 3

Coding and the Effect Size of the Included Studies

First Author	Sample Size	Hedge's g	SE	Grade level	EBIP model	Assessment Coverage
POGIL						
Ucang (2013)	188	0.98	0.15	High School	Conventional	Single topic
Muhammad (2020)	60	1.25	0.28	High School	Conventional	Single topic
Kartono (2020)	64	0.50	0.25	High School	Innovative	Cumulative
Adiningsih (2013)	50	0.14	0.28	Elementary	Conventional	Single topic
Andriani (2019)	49	0.64	0.29	Elementary	Conventional	Single topic
UPSIDE-DOWN PEDAGOGY						
	206	-0.29	0.14	Elementary	Innovative	Cumulative
	1329	0.92	0.06	Mixed	Conventional	Cumulative
	2370	-0.17	0.14	Elementary	Conventional	Cumulative

First Author	Sample Size	Hedge's g	SE	Grade level	EBIP model	Assessment Coverage
	46	0.82	0.30	High School	Conventional	Single topic
	100	1.13	0.21	High School	Conventional	Single topic
	520	1.95	0.11	Both	Conventional	Cumulative
	117	3.40	0.29	High School	Conventional	Cumulative
	58	0.05	0.26	High School	Conventional	Cumulative
	82	2.06	0.27	High School	Conventional	Single topic
	90	-0.74	0.22	High School	Conventional	Single topic
	46	0.20	0.29	Elementary	Conventional	Single topic
	67	-0.09	0.24	High School	Conventional	Single topic
	24	-0.78	0.41	High School	Conventional	Single topic
	60	0.02	0.26	High School	Conventional	Cumulative
	44	1.13	0.32	Elementary	Innovative	Single topic
	79	0.98	0.24	High School	Conventional	Cumulative
	91	0.42	0.21	High School	Conventional	Cumulative
	49	0.82	0.29	Elementary	Innovative	Single topic
TEACHING WITH CASES						
Gunbas (2014)	85	0.65	0.22	Elementary	Innovative	Cumulative
Özpinar (2017)	58	0.50	0.26	High School	Innovative	Cumulative
Ahmed (2014)	27	0.98	0.41	High School	Conventional	Cumulative
Zankour (2017)	30	0.92	0.37	Elementary	Innovative	Single topic
Huang (2020)	37	0.84	0.34	Elementary	Conventional	Single topic

TABLE 4

Overall Effect Size

Model	k	ES	SE	95% CI		z	p	Heterogeneity					
				Lower	Upper			Q	df	p_q	I^2		PI
Fixed	28	0.69	0.03	0.62	0.76	20.06	0.0	459.39	27	0.0	94.04%	- 0.93	2.30
Random	28	0.69	0.16	0.26	1.02	4.24	0.0						

As shown in Table 4, the test of heterogeneity indicates that the effect sizes in the current meta-analysis are heterogeneous ($Q(27) = 453.39$, $p < 0.05$). Since there is heterogeneity, random effects model was considered. Furthermore, the overall effect size of 28 studies is 0.61 which implies EBIPs have a positive medium effect on students' mathematical content knowledge and math skills. Additionally, the overall prediction interval is from negative to positive values (-0.93 to 2.30). This indicates that future studies involving TWC, POGIL, and UP may have a negative or positive effect on students' mathematical content knowledge and math skills.

The forest plot in Figure 2 exhibits the effect sizes of each included studies relative to each other and the overall effect size. Six studies included in the meta-analysis were found to have negative effect sizes. Such negative values indicate that the

FIGURE 2

Forest Plot of the Meta-analysis of 28 Studies

#	Study name	Hedges' g	CI Lower limit	CI Upper limit	Weight
1	Ucang (2013)	0.98	0.68	1.28	3.83%
2	Muhammad (2020)	1.25	0.71	1.83	3.52%
3	Shora (2020)	0.50	0.00	1.00	3.60%
4	Adiningsih (2013)	0.14	-0.41	0.71	3.52%
5	Andriani (2019)	0.64	0.07	1.23	3.49%
6	Flick (2019)	-0.29	-0.57	-0.02	3.86%
7	Martin (2015)	0.92	0.81	1.04	3.96%
8	Ripley (2015)	-0.17	-0.44	0.10	3.86%
9	Zeineddine (2018)	0.82	0.22	1.44	3.45%
10	Charles-Ogan (2015)	1.13	0.71	1.57	3.70%
11	Ramaglia (2015)	1.95	1.74	2.16	3.91%
12	Montgomery (2015)	3.40	2.85	3.99	3.50%
13	Saunders (2014)	0.05	-0.47	0.57	3.58%
14	Kumar (2015)	2.06	1.54	2.62	3.54%
15	Segumpan (2018)	-0.74	-1.18	-0.32	3.69%
16	Jackson (2019)	0.20	-0.38	0.79	3.49%
17	Tekin (2020)	-0.09	-0.57	0.40	3.63%
18	Casem (2016)	-0.78	-1.65	0.05	3.10%
19	Vang (2017)	0.02	-0.49	0.54	3.59%
20	Lai (2016)	1.13	0.51	1.80	3.40%
21	Jarrah (2019)	0.98	0.52	1.46	3.64%
22	Esperanza (2016)	0.42	0.01	0.84	3.71%
23	Ku (2019)	0.82	0.25	1.43	3.48%
24	Gunbas (2015)	0.65	0.22	1.10	3.68%
25	Ozpinar (2017)	0.50	-0.02	1.04	3.57%
26	Ahmed (2014)	0.98	0.17	1.85	3.11%
27	Zankour (2017)	0.92	0.18	1.71	3.22%
28	Huang (2020)	0.84	0.17	1.54	3.34%

result of the comparison of means favored the non-treatment group. Moreover, this result suggests that students who did not experience the EBIP treatment have performed better compared to those who underwent the treatment.

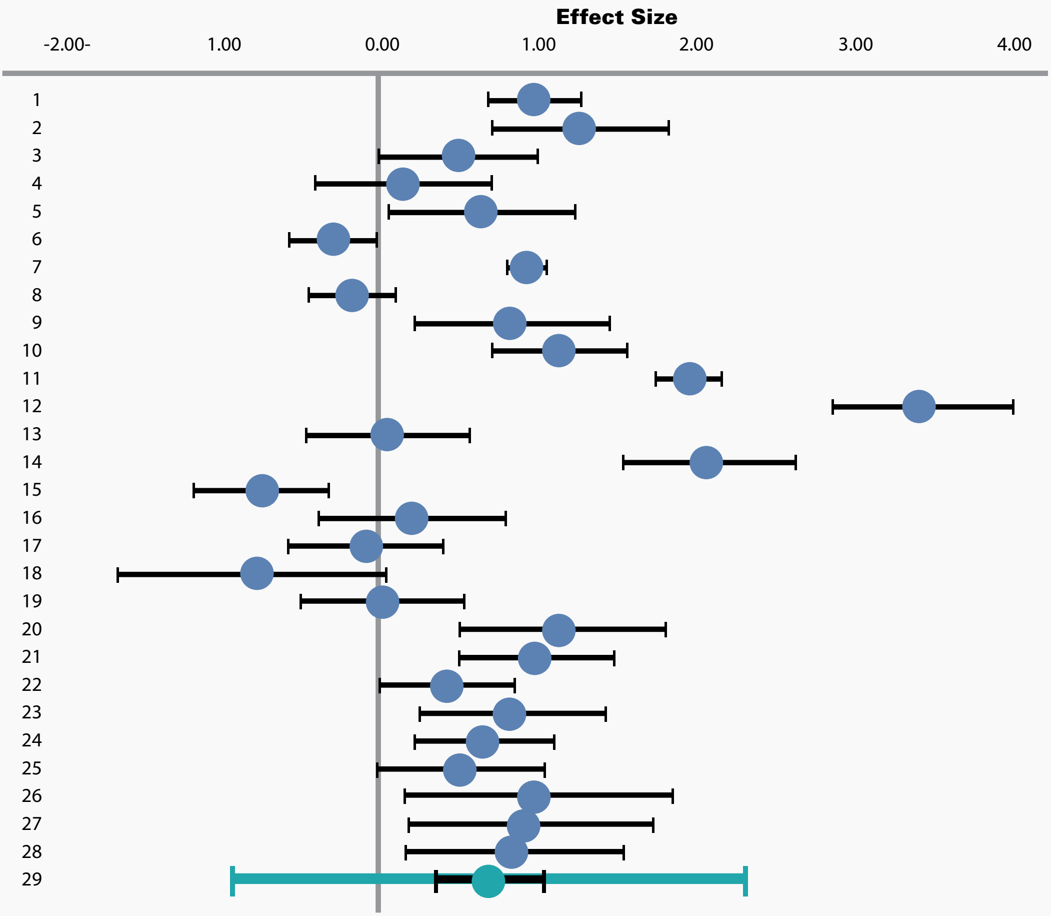


FIGURE 3

Funnel Plot of Standard Error by Effect Size

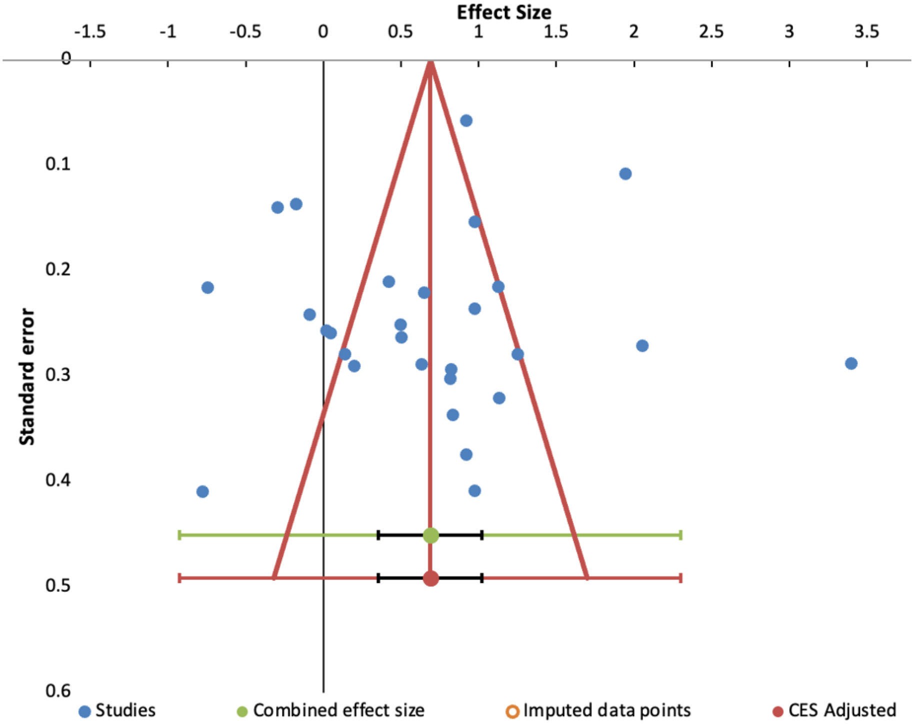


TABLE 5

Tests for Publication Bias

Egger's Test	
t test	0.03
p-value	0.978
Begg & Mazumdar's Test	
Δx-y	36.00
Kendall's Tau a	0.10
z	0.71
p	0.477
Rosenthal' failsafe-N test	
Overall Z-score	17.69
Failsafe-N	3211
Ad-hoc rule	FALSE

Figure 3 displays that there is a symmetry in the distribution of effect sizes. This symmetry indicates that there is no publication bias present in the meta-analysis. The estimate of the non-existence of the publication bias in the meta-analysis is supported by the Egger's test and Begg and Mazumdar's test in Table 5. The p-values (p = 0.978 and p = 0.477, respectively) from the two tests denote that there is no significant difference between the distribution of effect sizes in the meta-analysis and a symmetrical distribution. Moreover, Rosenthal's failsafe-N test indicates that 3211 studies averaging a z-value of zero are missing so that the combined effect size will become insignificant.

TABLE 6

Relative Effectiveness of Each Evidence-based Instructional Practice

Model	k	ES	SE	95% CI		Heterogeneity				
				Lower	Upper	Qw	p	I^2	PI Lower	PI Upper
Type of EBIP										
TWC	5	0.77	0.09	0.6	0.94	1.56	0.816	0.00%	-1.48	3.02
POGIL	5	0.70	0.37	0.33	1.03	11.20	0.024	64.29%	-1.60	3.01
UP	18	0.66	0.25	0.18	1.15	440.28	0.00	96.14%	-1.12	2.44
Students' Grade Level										
Elementary	10	0.47	0.16	0.17	0.78	46.44	0.00	80.62%	-1.48	2.43
High School	16	0.59	0.28	0.05	1.13	300.84	0.00	95.01	-1.31	2.49
Both	2	1.43	0.51	0.43	2.44	70.44	0.00	98.57%	-11.18	14.04
Assessment Coverage										
Single Topic	15	0.49	0.22	0.05	0.92	334.74	0.00	92.63%	-1.43	2.41
Cumulative	13	0.75	0.27	0.22	1.29	189.88	0.00	96.42%	-1.23	2.74
EBIP model used										
Conventional	21	0.59	0.22	0.25	0.94	518.99	0.00	96.15%	-1.61	2.8
Innovative	7	0.62	0.18	0.18	1.06	34.28	0.00	95.30%	-1.28	2.52

Table 6 displays the categorization of the 28 studies in the meta-analysis utilizing random effects model with respect to different codes. In terms of the type of EBIP used, it can be noticed that the number of studies ranged from 5 (TWC and POGIL) to 18 (UP) and the standard error of these is from 0.09 to 0.37. The overall effect size of studies involving TWC (Hedge's g = 0.77) is the largest while UP has the lowest overall effect size (Hedge's g = 0.66). Additionally, the result shows that POGIL ($Q = 11.20, p < 0.05$) and UP ($Q = 440.28, p < 0.05$) have statistically significant variances within groups which imply that the studies across POGIL and UP do not share a common effect size. On the other hand, the result reveals that there is not enough evidence to say that studies across TWC do not share common effect size ($Q = 1.56, p > 0.05$).

In terms of the students' grade level, it can be noted that the number of studies involving elementary, high school, and a mixture of both elementary and high school students ranged from 2 (both) to 16 (high school) and the standard error of these is from 0.05 to 0.43. The overall effect size of studies involving a combination of elementary and high school (Hedge's g = 1.43) is the largest while studies involving elementary students has the lowest overall effect size (Hedge's g = 0.47). Also, the result shows studies that involve elementary ($Q = 46.44$, $p < 0.05$), high school ($Q = 300.84$, $p < 0.05$), and a combination of both ($Q = 70.44$, $p < 0.05$) have statistically significant variances within groups which imply that the studies across these three categories do not share a common effect size.

Table 6 also exhibits the characteristics of the 28 included studies according to assessment coverage. The analysis of the effect sizes within the studies that utilized single topic coverage ($k = 15$) and cumulative coverage ($k = 13$) reveals that EBIPs have greater effect on studies that utilized comprehensive assessments (Hedge's g = 0.75). Furthermore, the difference between the effect sizes of studies within the single-topic assessment is statistically significant which means they do not share common effect sizes. Additionally, there is a significant difference across the studies that made use of cumulative assessment indicating that each study does not share a common effect size.

Lastly, the overall effect sizes of studies that utilized conventional and innovative models are 0.59 and 0.62, respectively. The analysis of variance within each subgroup reveals that studies which employed conventional models do not share a common effect size. Likewise, studies that made use of an innovative model during the implementation have no common effect size.

TABLE 7

Interaction of EBIP and Students' Grade Level

	PARTICIPANTS											
EBIP	Overall			Elementary			High School			Both		
	k	ES	SE	k	ES	SE	k	ES	SE	k	ES	SE
TWC	5	0.7	1	3	0.75	0.08	2	0.64	0.08	-	-	-
POGIL	5	0.37	1	2	0.39	0.25	3	0.24	0.66	-	-	-
UP	18	0.69	0.71	5	0.32	0.28	11	0.67	0.67	2	1.43	0.51

Considering that studies within elementary and high school subgroups do not share common effect size, the researchers analyzed the interaction between each EBIP and students' grade level (see Table 7). In elementary level, the result reveals that studies utilizing TWC have the largest effect size compared to POGIL and UP. On the other hand, UP has the largest effect size among the studies which involved high school students. Moreover, it can be noted that POGIL has the lowest effect on high school students. Furthermore, the most notable difference was the effect sizes for UP participated by elementary and high school students (difference of 0.35).

TABLE 8

Interaction of EBIP and Studies' Assessment Coverage

ASSESSMENT COVERAGE									
EBIP	Overall			Single Topic			Cumulative		
	k	ES	SE	k	ES	SE	k	ES	SE
TWC	5	0.78	1	2	0.87	0.04	3	0.65	0.11
POGIL	5	-	-	4	-	-	1	-	-
UP	18	0.62	1	9	0.51	0.31	9	0.8	0.4

Cognizant of the abovementioned results that studies within single topic and cumulative subgroups do not share common effect size, the researchers analyzed the interaction between each EBIP and studies' assessment coverage (see Table 8). POGIL studies with cumulative assessments are not well represented. The use of single-topic assessment in TWC has a larger effect size compared to the use of cumulative assessment. On the contrary, the use of single-topic assessment in UP has a smaller effect size compared to the use of cumulative assessment.

TABLE 9

Interaction of EBIP and Model

EBIP MODEL USED									
EBIP	Overall			Conventional			Innovative		
	k	ES	SE	k	ES	SE	k	ES	SE
TWC	5	0.78	1	2	0.89	0.07	3	0.65	0.1
POGIL	5	-	-	4	-	-	1	-	-
UP	18	0.64	1	14	0.69	0.29	3	0.53	0.44

Considering that studies within conventional and innovative subgroups do not share common effect size, the researchers analyzed the interaction between each EBIP and EBIP model used by each study (see Table 9). POGIL studies using traditional models are not well represented. Moreover, conventional models were found to have a generally larger effect compared to innovative models.

Discussion

The first research question aimed at determining the level of effectiveness of EBIPs on students' mathematical content knowledge and math skills. The result shows that, in general, EBIPs have a positively moderate effect on students' mathematics achievement. The reason for such a moderate result is because five studies (Casem, 2016; A. Flick, 2019; Ripley, 2015; Segumpan et al., 2018; Tekin & Emmioğlu-Sarıkaya, 2020) were found to have negative effect sizes. This means that these studies' control groups, or the groups which received no treatment, have performed better than the treatment group. This goes to show that while EBIPs improve mathematics content knowledge, there are cases wherein the non-treatment group would perform better. These negative effect sizes from the five studies might also be the reason why publication bias does not exist. This is because even though the results from the five studies failed to show that EBIPs are more effective than traditional teaching, they were still published in reputable journals or publications.

Furthermore, the second research question aimed to assess the comparative effectiveness of each Evidence-Based Instructional Practice (EBIP). The findings indicated that Teaching with Cases (TWC) demonstrated the highest impact, with each study yielding moderate to substantial improvements in students' mathematics achievement. Moreover, the analysis of TWC studies in this investigation revealed a consistent pattern of effect sizes, suggesting the strategy's effectiveness across diverse educational settings.

Interestingly, among the three EBIPs examined, Upside-down Pedagogy (UP) garnered the most substantial body of evidence over the past decade, as evidenced by numerous studies (Casem, 2016; Charles-Ogan & Williams, 2015; Esperanza & Toto, 2016; Flick, 2019; Jackson, 2019; Jarrah & Khaled Mohammed Abdel Baki Mohammed Diab, 2019; Ku et al., 2019; Kumar Bhagat, Chang, et al., 2016; Lai & Hwang, 2016; Martin, 2015; Ramaglia, 2015; Ripley, 2015; Saunders, 2014; Segrin et al., 2015; Segumpan & Tan, 2018; Tekin & Emmioğlu-Sarıkaya, 2020; Vang, 2017; Zeineddine, 2018). This predominance can be attributed to the fact that all the studies encompassing Upside-down Pedagogy employed a flipped classroom approach, which is one of the most extensively investigated pedagogical methods in contemporary mathematics education. The popularity of the flipped classroom model stems from its incorporation of technology into instruction, aligning it with the ongoing technological revolution.

Lastly, our third research inquiry aimed to uncover the influence of moderator variables on the effectiveness of Evidence-Based Instructional Practices (EBIPs). With respect to students' grade levels, it was observed that Teaching with Cases or Stories and Process-Oriented Guided Inquiry Learning (POGIL) demonstrated comparatively lower effect sizes when implemented with high school students as opposed to elementary students. Conversely, Upside-down Pedagogy (UP) exhibited a higher impact on high school students.

Furthermore, concerning the choice of EBIP models, we observed that POGIL studies employing traditional models were underrepresented. Notably, within the Teaching with Cases (TWC) framework, the utilization of traditional models yielded larger effect sizes compared to innovative models. Similarly, within the UP approach, the use of traditional models generated larger effect sizes in contrast to innovative models.

Regarding the coverage of moderator assessments, there was a shortage of POGIL studies incorporating cumulative assessments. Additionally, we found that the implementation of single-topic assessments in TWC was associated with larger effect sizes, whereas in UP, the use of cumulative assessments resulted in smaller effect sizes compared to single-topic assessments.

Conclusion

This meta-analysis was conducted with the primary goal of assessing the overall effectiveness of three of the least utilized EBIPs within the domain of mathematics education, while also endeavoring to make comparisons among these EBIPs. Our findings have led to several significant conclusions.

Firstly, in terms of the general effectiveness of EBIPs, the range of effect sizes observed, spanning from 0.26 to 0.97, can serve as valuable benchmarks for evaluating the effectiveness of other EBIPs that were not specifically examined in this study. These effect size ranges provide a reference point for educators and researchers seeking to gauge the potential impact of various instructional approaches within the field of mathematics education.

Secondly, our analysis has identified Teaching with Cases as the most promising EBIP for elementary-level learners. On the other hand, Upside-down Pedagogy has emerged as particularly effective for high school-level students. This distinction underscores the importance of tailoring instructional approaches to the specific developmental and educational needs of different student populations.

Thirdly, the choice of EBIP models has been shown to influence outcomes. Traditional models of EBIPs demonstrated a greater likelihood of success when compared to EBIP models characterized by innovation. This finding suggests that, in many cases, sticking to established and well-understood approaches may yield more consistent positive results.

Lastly, our analysis revealed that Teaching with Cases is especially valuable for facilitating short-term comprehension of mathematical concepts. In contrast, Upside-down Pedagogy appears to excel in promoting long-term understanding.

This distinction emphasizes the importance of considering the intended learning outcomes and the temporal dimension when selecting an appropriate EBIP for a given educational context.

In summary, this meta-analysis contributes valuable insights into the field of mathematics education by assessing the effectiveness of less commonly used EBIPs and providing guidance on their application. Educators, policymakers, and researchers can draw upon these findings to make informed decisions about instructional practices, considering the grade level of students, the chosen EBIP model, and the desired learning outcomes. This research ultimately supports the ongoing enhancement of mathematics education strategies and practices.

References

Adiningsih, S. (2013). *Mathematics learning experiment on guided inquiry method and open-ended method viewed from students' critical thinking ability in Grade VII of SMP N 1 Surakarta. Muhammadiyah University of Surakarta.*

Ahmed, N. (2014). *Impact of storytelling on students' learning of. California State University.*

Andriani, S., Nurlaelah, E., & Yulianti, K. (2019). *The effect of process oriented guided inquiry learning (POGIL) model toward students' logical thinking ability in mathematics. 42108. https://doi.org/10.1088/1742-6596/1157/4/042108*

Basu, A. (2017). *How to conduct meta-analysis: A basic tutorial. PeerJ Preprints, 1–15. https://peerj.com/preprints/2978.pdf*

Borenstein, M., Hedges, L. V, Higgins, J. P. T., & Rothstein, H. R. (2009). *Subgroup Analyses. In Introduction to Meta-Analysis. John Wiley & Sons, Ltd. https://www.meta-analysis.com/downloads/Meta-analysis%20Subgroups%20analysis.pdf*

Borrego, M., & Henderson, C. (2014). *Increasing the use of evidence-based teaching in STEM higher education: A comparison of eight change strategies. Journal of Engineering Education, 103(2), 220–252.*

Burns, M. K., & Ysseldyke, J. E. (2009). *Reported prevalence of evidence-based instructional practices in special education. The Journal of Special Education, 43(1), 3–11. https://doi.org/10.1177/0022466908315563*

Casem, R. Q. (2016). *Effects of flipped instruction on the performance and attitude of high school students in mathematics. European Journal of STEM Education, 1(2), 37–44. https://doi.org/10.20897/lectito.201620*

Charles-Ogan, G., & Williams, C. (2015). *Flipped classroom versus a conventional classroom in the learning of Mathematics. South Africa International Conference on Educational Technologies, 96–102. www.aa-rf.org*

Esperanza, P., & Toto, C. (2016). *Flipped classroom model: Effects on performance, attitudes and perceptions in High School Algebra. Lecture Notes in Computer Science, 9891, 85–97. https://doi.org/10.1007/978-3-319-45153-4_7*

Flick, A. (2019). *The Effects of flipped learning in the sixth-grade Mathematics classroom. Missouri Baptist University.*

Groccia, J. E., & Buskist, W. (2011). *Need for evidence-based teaching. New Directions for Teaching and Learning, 128. https://doi.org/10.1002/tl.463*

Gunbas, N. (2014). *Students' mathematics word problem-solving achievement in a computer-based story. https://doi.org/10.1111/jcal.12067*

Huang, C. W., Chen, H. R., Yu, S. C., Su, Y. L., & Chen, C. C. (2020). *Mathematic learning efficiency analysis of story-based situated learning in low-achieving elementary school students. In Innovative Technologies and Learning: Vol. 12555 LNCS (pp. 33–39). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/978-3-030-63885-6_4*

Jackson, N. (2019). *Beneficial or not: Flipped learning in an elementary Mathematics classroom [California State University]. https://digitalcommons.csumb.edu/caps_thes_all/641*

Jarrah, A., & Khaled Mohammed Abdel Baki Mohammed Diab. (2019). *The effect of flipped classroom model on students' achievement in the new 2016 scholastic assessment test mathematics skills. The Journal of Social Sciences Research, 5(3), 769–777. https://doi.org/10.32861/jssr.53.769.777*

Kartono, & Shora, R. Y. (2020). *Effectiveness of Process Oriented Guided Inquiry Learning with peer feedback on achieving students' mathematical reasoning capabilities. International Journal of Instruction, 13(3), 555–570. https://doi.org/10.29333/iji.2020.13338a*

Ku, W.-P., Yang, K.-H., & Chang, W.-L. (2019). *The design and evaluation of interactive video-based flipped classroom on mathematics learning. 8th International Congress on Advanced Applied Informatics, 1041–1042. https://doi.org/10.1109/IIAI-AAI.2019.00217*

Kumar Bhagat, K., Chang, C.-N., & Chang, C.-Y. (2016). *The impact of the flipped classroom on mathematics concept learning in High school. Educational Technology & Society, 19(3), 124–132. https://www.researchgate.net/publication/286047548*

Lai, C.-L., & Hwang, G.-J. (2016). *A self-regulated flipped classroom approach to improving students' learning performance in a mathematics course. Technology, 43. https://doi.org/10.1016/j.compedu.2016.05.006*

Martin, A. G. (2015). *The impact of flipped instruction on middle school mathematics achievement. Texas A&M University-Commerce.*

Montgomery, J. (2015). *The effects of flipped learning on middle school students' achievement with Common Core Mathematics. California State University.*

Muhammad, M., & Purwanto, J. (2020). *The effect of Process Oriented Guided Inquiry Learning (POGIL) on mathematical problem-solving abilities. J. Phys. https://doi.org/10.1088/1742-6596/1469/1/012171*

Murad, M. H., Asi, N., Alsawas, M., & Alahdab, F. (2016). *New evidence pyramid. BMJ Evidence-Based Medicine, 21(4), 125–127. https://doi.org/10.1136/EBMED-2016-110401*

Özpinar I. Gökçe E. S., & Aydogan Yenmez, A. (2017). *Effects of digital storytelling in mathematics instruction on academic achievement and examination of teacher-student opinions on the process. Journal of Education and Training Studies, 5(10). https://doi.org/10.11114/jets.v5i10.2595*

Petty, G. (2009). *Evidence-based teaching: A practical approach: Vol. 2nd Ed. Nelson Thornes.*

Ramaglia, H. (2015). *The Flipped mathematics classroom: A mixed methods study examining achievement, active learning, and perception. In ProQuest. Kansas State University.*

Ripley, D. G. (2015). *An examination of flipped instructional method on sixth graders' mathematics learning: Utilizing propensity score matching. University of Nevada.*

Saunders, J. M. (2014). *The flipped classroom: Its effect on student academic achievement and critical thinking skills in high school mathematics. Liberty University.*

Segumpan, L., & Tan, D. A. (2018). *Mathematics performance and anxiety of junior high school students in a flipped classroom. European Journal of Education Studies, 4(12). https://doi.org/10.5281/zenodo.1325918*

Sturtevant, H., & Wheeler, L. (2019). *The STEM Faculty Instructional Barriers and Identity Survey (FIBIS): Development and exploratory results. International Journal of Science Education, 6(35), 22. https://doi.org/10.1186/s40594-019-0185-0*

Tekin, O., & Emmio lu-Sarıkaya, E. (2020). *Flipped classroom model in high school mathematics. Bartın University Journal of Faculty of Education, 9(2), 301–314. https://doi.org/10.14686/buefad.559990*

Ucang, J., & Tan, D. (2013). *Students' belief and mathematics performance in a Process-Oriented Guide Inquiry Learning (POGIL) environment. Central Mindanao University Journal of Science, October, 141–157.*

van Aert, R. C. M., van Assen, M. A. L. M., & Viechtbauer, W. (2019). *Statistical properties of methods based on the Q statistic for constructing a confidence interval for the between study variance in meta analysis. Research Synthesis Methods, 10(2), 225. https://doi.org/10.1002/JRSM.1336*

van Rhee, H., Suurmond, R., & Hak, T. (2015). *User Manual for Meta-Essentials: Workbooks for Meta-Analysis. SSRN Electronic Journal. https://doi.org/10.2139/SSRN.3241355*

Vang, Y. V. (2017). *The impact of the flipped classroom on high school mathematics students' academic performance and self-efficacy. California State University.*

Villanueva, E. P. B., & Prudente, M. (2022). *Mathematics teachers' knowledge, attitudes, and practices in using evidence-based instructional practices. ACM International Conference Proceeding Series. https://doi.org/10.1145/3514262.3514266*

Zankour, M. M. S. (2017). *The digital story in mathematics for developing mathematical imagination and numerical intelligence among primary school (Impulsive/Reflective) students. Assiut University.*

Zeineddine, D. (2018). *Investigate the effects of flipped learning on understanding of mathematics for secondary students. Journal of Mathematics Education , 11(1), 62–80. https://doi.org/10.26711/007577152790021*

ABOUT THE AUTHORS:

Edrian Peter Villanueva is an educator and researcher currently pursuing a Doctor of Philosophy in Science Education with a specialization in Mathematics at De La Salle University – Manila. His research interests encompass evidence-based teaching in mathematics education, mathematical discourse strategies, and the implementation of HyFlex learning setups. Alongside his doctoral studies, he is an Assistant Professor in the General Education Program at NU-Asia Pacific College (Corresponding author: edrianpeterv@apc.edu.ph).

Maricar S. Prudente, PhD, is a Full Professor of the Science Education Department of De La Salle University-Manila. As an educator, Dr. Prudente has served as administrator in various capacities at De La Salle University- Manila – as Academic Chairperson of the Science Education Department and as Research Director of the Lasallian Institute for Development and Educational Research (LIDER). Her involvement in research was recognized in 2015 when she was awarded the Lasallian Pillar of Excellence in Research by De La Salle University-Manila. Dr. Prudente was also the recipient of the 2015 Lifetime National Achievement Award of the National Research Council of the Philippines (NRCP). She was recently recognized as the 2018 Outstanding Filipino JSPS Fellow in the field of Education by the Department of Science and Technology (Email: maricar.prudente@dlsu.edu.ph).