

## On the use of flipped classroom across various disciplines: Insights from a second-order meta-analysis

**Khe Foon Hew, Shurui Bai, Weijiao Huang**

The University of Hong Kong

**Phillip Dawson**

Deakin University

**Jiahui Du, Guoyuhui Huang, Chengyuan Jia, Khongjan Thankrit**

The University of Hong Kong

Flipped classroom has become a popular buzzword in the post-secondary education setting, and it is one of the most visible trends in smart learning environments. Alongside this popularisation comes the view that the flipped classroom is something desirable. Yet, many educators remain divided over whether flipped classroom is really an improvement over traditional approaches. This paper is the first to synthesise all relevant meta-analytic information using a second-order meta-analysis approach on the effectiveness of the flipped classroom on student learning outcomes. By synthesising the findings of multiple primary meta-analyses instead of individual empirical studies, a second-order meta-analysis can more accurately account for publication bias and generate a more robust result. The present study synthesised and analysed the quality of 15 primary meta-analyses that involved 156,722 participants in flipped and non-flipped conditions to provide the most exhaustive test of the flipped classroom approach on its effect on student learning outcomes in higher education to date. The mean random effect size, after trim-and-fill adjustment, was 0.37,  $p < 0.001$  in support of flipped classrooms. To check the accuracy of the second-order meta-analysis results, we performed a study-level meta-analytic validation. We discuss possible contextual and methodological moderators.

### *Implications for practice or policy:*

- Educators should consider using the flipped classroom approach in their teaching because it increases learning performance compared with conventional non-flipped teaching.
- Educators could use the insights reported in this study to inform planning for future meta-analyses involving the flipped classroom approach.
- Educators could also use the insights reported in this study to inform planning for future empirical studies involving the flipped classroom approach.

*Keywords:* flipped classroom, flipped learning, inverted classroom, learning outcomes, meta-analysis

## Introduction

Unlike a conventional course, where instruction occurs mainly in the classroom, and students complete homework exercises outside class, a flipped classroom usually consists of online independent learning of basic concepts before class (pre-class), followed by face-to-face lessons (in-class) that focus on active learning activities such as collaborative group work (Abeysekera & Dawson, 2015). A recent search of Google Scholar (July 8, 2020) yielded about 94,000, 183,000, and 72,000 hits for the terms *flipped classroom*, *flipped learning*, and *inverted classroom* respectively. The flipped classroom approach has been growing fast, particularly in higher education settings (Lundin et al., 2018). Theoretically, the flipped classroom approach works better than non-flipped classrooms for two reasons: the flipped classroom's ability to better address student motivation and its ability to promote student active learning (Keengwe et al., 2014). In doing so, flipped classrooms aim to be effective, efficient and engaging, thus meeting the three general characteristics expected of smart learning environments (Spector, 2014).

The self-determination theory of motivation posits three basic universal human cognitive needs: the need for autonomy, relatedness and competence (Ryan & Deci, 2000). A flipped classroom may theoretically

satisfy students' need for autonomy better than a traditional classroom because flipped students can choose to skip viewing any pre-class content they have already understood, as well as choose to complete the pre-class work at their own pace and time (Sergis et al., 2018). Flipped classroom learners also have the autonomy to review the pre-class content as many times they wish.

Flipped classroom students are expected to apply what they have learned in pre-class sessions to solve problems or discuss issues in face-to-face class sessions. This can enhance their sense of competency and sense of relatedness with their peers particularly when group activities are employed (Abeysekera & Dawson, 2015). Flipped classroom teachers also have more class time to give feedback to help students understand the subject material more thoroughly and to enhance students' sense of competence (van Alten et al., 2019).

A flipped classroom also offers students more in-class time for active learning since the information transmission of a traditional lecture is now shifted out of class time (Abeysekera & Dawson, 2015; Berrett, 2012). Freeman et al. (2014) defined active learning as a method where students learn by participating in activities and/or discussion in class, as opposed to passively listening to an instructor. Methods of active learning include the use of tests and quizzes, writings, discussions, case studies and group problem-solving (Phillips, 2005). Active learning can help students construct better understandings of the subject material (Bransford et al., 1999). Recent research has found that active learning can significantly increase students' learning performance (Deslauriers et al., 2019; Freeman et al., 2014).

Nevertheless, not all institutions are sold on the idea of flipped classroom. A survey of 290 European universities (one response per institution, for which a senior institutional representative was asked to take responsibility, was collected) revealed that only 15% of institutions found the flipped classroom approach to be fully useful (Gaebel & Zhang, 2018). The traditional lecture format remains many instructors' preferred teaching method (Deslauriers et al., 2019; Stains et al., 2018). This is partly because designing flipped classroom materials can be costly in terms of time as it requires the instructor to develop quality video lectures and design appropriate face-to-face activities (Cheng et al., 2019). McLaughlin et al. (2014) estimated that a faculty professor has to spend 127% more time to develop and manage a flipped course and 57% more time to maintain a flipped course compared to a lecture course. The production of a mere 10-minute video may require approximately 2 to 3 hours (Altafi et al., 2017).

Although extensively studied over the years, there is still recent debate about the effectiveness of flipped classroom in improving learning outcomes (Strelan et al., 2020; Zainuddin et al., 2019). While some empirical studies reported significant improvements in student learning induced by flipped classrooms (e.g., Flynn, 2015; Lax et al., 2017), other studies reported no difference between non-flipped and flipped classrooms (e.g., F. Chen et al., 2017; DeSantis et al., 2015). Still other studies found that flipped classroom impaired student learning. For instance, a recent randomised controlled trial experiment published in a Massachusetts Institute of Technology discussion paper involving 1,328 students across 80 economics and mathematics classes found that flipped classroom exacerbated the achievement gaps between White and Black or Hispanic students when compared to the traditional lecture group (Setren et al., 2019).

### **Prior first-order meta-analyses**

Given the aforementioned concerns around the flipped classroom approach, more precise estimates of the overall impact of this approach should be a priority. To achieve this aim, many authors have begun to meta-analyse subsets of the existing empirical literature. The primary goal of these first-order meta-analyses (hereafter called *primary meta-analyses*) is to estimate an overall mean effect from *individual empirical studies* and to identify possible factors that moderate that effect (Gurevitch et al., 2018).

To the best of our knowledge, the earliest primary meta-analysis on flipped classroom was published in 2016 (Kang & Shin, 2016). Since then, the number of flipped classroom primary meta-analyses has skyrocketed. While primary meta-analyses do offer several advantages, they are usually hampered in their focus and scope (Causadias et al., 2018). Although each primary meta-analysis provides a useful piece of information, the result of any single primary meta-analysis cannot imply the overall effect of flipped classroom on student learning outcomes. Instead of doing (yet) another primary meta-analysis of flipped classroom, this study aimed to synthesise the quantitative results of all relevant primary meta-analyses to provide a general picture of what we currently know about the effects of flipped classroom.

## The present study

Specifically, this study employed an approach known as *second-order* meta-analysis to analyse comparable primary meta-analyses (Cooper & Koenka, 2012). This approach can be thought of as a meta-analysis of meta-analyses. Rather than synthesising results from individual empirical studies, a second-order meta-analysis combines data from primary meta-analyses. Second-order meta-analyses have become a widely accepted research method (Steenbergen-Hu et al., 2016). The original contributions of our second-order meta-analysis study are fivefold.

First, by summarising the findings of more than one primary meta-analysis using a second-order meta-analysis approach, we can generate a more robust generalisable result with a very large sample size (Busch & Friede, 2018). The present study synthesised 15 primary meta-analyses that covered 156,722 participants to provide the most exhaustive test of the flipped classroom approach on its effect on student learning to date. Such data are not easily available for analysis from an individual primary study or even in a primary meta-analysis.

Second, a second-order meta-analysis can more accurately account for publication bias by analysing both published and unpublished meta-analyses, on top of meta-analyses that examine unpublished empirical works (Causadias et al., 2018). Unpublished meta-analyses or empirical works include dissertations and theses (often collectively referred to as “gray literature”) (A. C. K. Cheung & Slavin, 2016, p. 288; Hartling et al., 2017). Gray literature is often used to refer to literature not formally published in journals (Lefebvre et al., 2008). Searching for gray literature such as dissertations or theses is a highly desirable practice recommended by the *Cochrane Handbook for Systematic Reviews of Interventions* in order to capture as many studies as possible and to minimise the risk of publication bias (Higgins et al., 2019). Dissertations are more likely to report negative or non-significant findings when compared to journal articles (Pigott et al., 2013). A second-order meta-analysis that includes gray literature meta-analyses is therefore more likely to report the whole story of the existing research into an intervention’s effectiveness.

Third, a second-order meta-analysis can take advantage of heterogeneity across primary meta-analyses to examine the role of moderators with regard to various contextual and methodological factors (Causadias et al., 2018). The present study examined nine moderating factors.

Fourth, this study assessed the quality of the primary meta-analyses using an empirically validated instrument, the Revised Assessment of Multiple Systematic Reviews (R-AMSTAR; Kung et al., 2010). R-AMSTAR has been found to be a readily applicable and validated tool to evaluate the quality of a meta-analysis (Sygouros & Acar, 2013). To date, we are unaware of any studies that have critically appraised the quality of flipped classroom primary meta-analyses. We examined whether the quality of flipped classroom primary meta-analyses may moderate the effect sizes. Hitherto, no other study has examined this issue to the best of our knowledge.

Fifth, our study identified three major limitations of previous primary meta-analyses, consisting of a lack of control to address possible student initial differences, a lack of explanation as to how independence of effect size was addressed and sole reliance on funnel plots to assess publication bias. More importantly, we provide useful suggestions to address these limitations.

To the best of our knowledge, this is the first second-order meta-analysis of flipped classroom approach that cumulates the findings from multiple prior primary meta-analyses. Student learning outcomes refer to domain-specific knowledge of a subject. Student learning outcomes are usually assessed using teacher-developed or standardised tests (Davis et al., 2003). To validate the results, we conducted a synthesis of all available study-level effect sizes reported in the primary meta-analyses included in our second-order meta-analysis (see the Study overlap and validation of the second-order meta-analysis section for details). Each individual empirical study was included only once in this validation stage, which contrasts with the second-order meta-analysis which may include some empirical studies multiple times if they were included in multiple primary meta-analyses. We then compared the findings from the present second-order meta-analysis with the findings from the validation study, where study overlap had been removed, to determine whether the average effect sizes were similar. This study addresses two specific research questions:

- What is the overall effect of the flipped classroom on students' learning outcomes as shown by synthesising findings of existing primary meta-analyses?
- What factors may moderate the effect of the flipped classroom on students' learning outcomes?

## Method

### Searching for eligible primary meta-analyses

We searched 11 major academic databases for eligible primary meta-analyses: ACM Digital Library; EBSCO (e.g., Academic Search Premier, ERIC); Emerald Insight; IEEE Xplore; ProQuest Dissertations & Theses A&I; Science Direct; Scopus; Springer; Web of Science; JSTOR; PubMed; and Google Scholar. We included primary meta-analyses that were published in journals or conference proceedings. We used the following search string (“review” OR “synthesis” OR “meta-analysis”) AND (“flip\*” OR “invert\*”) AND (“class\*” OR “learn\*”). “Flip\*” and “invert\*” would capture morphological variants such as flipping, flipped, inverting and inverted, while “class\*” and “learn\*” captures various expressions including class, classroom, learning, learner. The date of publication remained open for the initial search. The search was completed on May 22, 2020. The following inclusion criteria for selecting a primary meta-analysis were used:

- It compared the effects of flipped classrooms with non-flipped classrooms using between-group research designs.
- It measured students' learning outcomes using teacher-developed or standardised tests and exams.
- It reported an overall mean effect size and the standard error or confidence intervals of the mean effect size. The type of effect size metric used must also be clearly described.
- It provided a list of the primary empirical studies analysed.
- It was written in English.
- It was publicly available or through library database subscription.

The unit of analysis in our second-order meta-analysis was the individual primary meta-analysis on flipped classroom. All the primary meta-analysis articles that we examined compared the effects of flipped and traditional classroom on student learning. In a traditional classroom, the instructor typically introduces the course materials in the classroom using direct instruction such as lectures; students are then given a few in-class exercises, followed by homework problems (Dove & Dove, 2015; Hew & Lo, 2018; Jungic et al., 2015; Låg & Sæle, 2019; Orhan, 2019; van Alten et al., 2019). We used the term *non-flipped* to refer to a traditional classroom, following van Alten et al. (2019).

Figure 1 shows the preferred reporting of items for systematic reviews and meta-analyses flowchart (Moher et al., 2009), which illustrates the entire article screening process. The initial academic databases search resulted in 2,912 records. Eighteen additional records were identified by searching the Web. After removing duplicates, 2,527 remained. The titles and abstracts of the remaining 2,527 records were screened. Many records were excluded because they were irrelevant to the purpose of the present study (e.g., FLIP as a therapeutic target in cancer). Subsequently, 48 full-text records were assessed for eligibility. Of these 48 full-text records, 20 were excluded because they did not focus on student learning outcomes, did not provide a list of the primary empirical studies analysed or did not focus on the specific student-related challenges of flipped classroom implementation. Ultimately, 15 meta-analyses were included in the present study. A list summarising the pertinent information about the included 15 primary meta-analyses (e.g., effect sizes, standard errors) is presented in Table 1. We included effect sizes related to higher education only.

Table 1

*Summary of the 15 meta-analyses examining the effects of flipped classroom on student learning outcomes*

Primary meta-analysis: author	Total sample size	Effect type	size	Mean effect size <sup>a</sup> (Std error)	Lang. of primary studies	Lit. source	Subject	Publication bias
K. S. Chen et al. (2018)	9,354	SMD <sup>b</sup>		0.47 (0.082)	English	Journals + theses + conference	Combination	No bias
Cheng et al. (2018)	6,779	g <sup>c</sup>		0.193 (0.064)	English	Journals + theses + conference	Combination	No bias
Gillette et al. (2018)	1,465	g		0.366 <sup>d</sup> (0.156)	English	Journals	Health	Yes
Hew & Lo (2018)	4,715	SMD		0.33 (0.066)	English	Journals	Health	No bias
Hu et al. (2018)	1,484	SMD		1.06 (0.179)	Chinese	Journals	Health	No bias
Karagöl & Esen (2018)	2,640	g		0.594 (0.098)	English, Turkish	Journals + theses	Combination	No bias
Låg & Sæle (2019)	48,211	g		0.34 (0.026)	English	Journals + theses + conference	Combination	Yes
Lo & Hew (2019)	5,238	g		0.272 (0.064)	English	Journals + conference	Engineering	No bias
Lo et al. (2017)	2,919	g		0.30 (0.078)	English	Journals	Mathematics	No bias
Orhan (2019)	571	g		0.779 (0.096)	English, Turkish	Journals + theses	Combination	No bias
Shi et al. (2019)	6,947	SMD		0.53 (0.087)	English	Journals	Combination	No bias
Strelan et al. (2020)	31,650	g		0.48 (0.040)	English	Journals	Combination	No bias
Tan et al. (2017)	3,694	SMD		1.13 (0.184)	Chinese	Journals + theses	Health	No bias
van Alten et al. (2019)	19,446	g		0.353 (0.048)	English	Journals + theses + conference	Combination	No bias
Zhang (2018)	11,609	g		0.320 (0.064)	English	Journals + conference	Science	No bias

<sup>a</sup> We included effect sizes related to higher education only. The total sample size refers to higher education samples.

<sup>b</sup> Standard mean difference.

<sup>c</sup> Hedges's *g*.

<sup>d</sup> We included all 6 studies shown in Table 3 (Gillette et al., 2018) and calculated the mean effect size using Hedges's *g*.

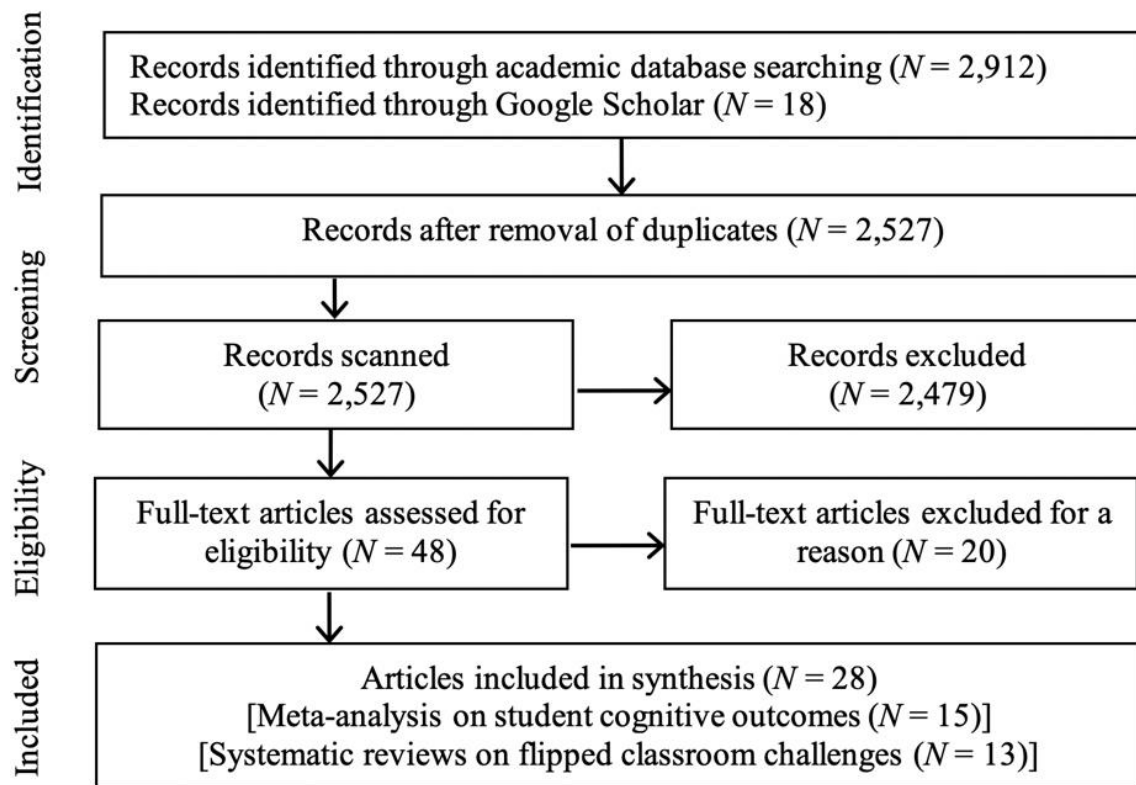


Figure 1. Flowchart of article selection

Primary meta-analyses were excluded if they focused solely on K-12 flipped classroom studies, they measured student learning outcomes using some form of self-reported data such as questionnaires and they focused only on student behavioural outcomes (e.g., student skill in performing a psychomotor task).

### Coding of the primary meta-analyses

We developed a codebook based on the work conducted by Steenbergen-Hu et al. (2016) and Tamim et al. (2011). The codebook contains four main sections: basic information (e.g., year of publication); study context (e.g., discipline); method (e.g., number of participants, research design); and results (e.g., effect size data). The first author performed the study coding. To test the reliability of the coding, the second author coded five randomly selected primary meta-analyses independently. There was perfect agreement between the two coders.

### Quality of the primary meta-analyses

To assess the quality of the primary meta-analyses, we employed an empirically validated instrument, R-AMSTAR (Kung et al., 2010). R-AMSTAR was developed based on the widely used AMSTAR instrument (Kung et al., 2010). The R-AMSTAR instrument rates each meta-analysis on 11 items. An example of an item is “Was a list of studies (included and excluded) provided?” (see Kung et al., 2010, for more details). Compared to the initial AMSTAR, R-AMSTAR has more detailed sub-criteria options for each of the 11 items, thus making R-AMSTAR a more detailed and sensitive quality assessment tool (Kohl et al., 2013). Scores for each of the 11 items range from 1 to 4 points, making the full marks in R-AMSTAR 44 points. The first and second authors assessed all 19 primary meta-analyses using R-AMSTAR. Discrepancies were resolved through mutual discussion. Following Sygouros and Acar (2013), the total scores were then categorised according to the percentile ranking of our sample.

### Study overlap and validation of the second-order meta-analysis

An important concern in second-order meta-analysis is the issue of study overlap across the various primary meta-analysis (Polanin et al., 2017). Study overlap occurs when the primary meta-analyses includes the

same empirical studies (Cooper & Koenka, 2012). Although several approaches to address study overlap have been proposed, it is not clear which approach is the most appropriate (Steenbergen-Hu, 2016). Some scholars suggest eliminating a primary meta-analysis with a high percentage of overlapping empirical studies (e.g., Young, 2017) or selecting only the most recent reviews (Cooper & Koenka, 2012). Others, however, disagree by suggesting that discarding overlapping meta-analyses may not be optimal because meta-analyses with a high degree of overlap may investigate different moderator variables (Cooper & Koenka, 2012).

In this study, we followed the example of Young (2017) and Tamim et al. (2011) to validate the results of our second-order meta-analysis. To do this, we extracted individual mean effect sizes and effect size standard errors from the available empirical studies reported in the primary meta-analyses. If the primary meta-analysis did not contain such information, we wrote to the corresponding author to request the data. If the author did not respond to our request, we excluded the individual empirical study data from our validation sample. We eliminated all empirical study overlap in the validation sample. Following Young (2017) and Tamim et al. (2011), if the mean effect size of the second-order meta-analysis was equivalent or closely similar to the mean effect size of the validation study, we can assume the result of second-order meta-analysis is valid in representing the aggregate effects of the included primary meta-analyses.

### **Analyses of mean effect size, publication bias and moderators of the primary meta-analyses**

We first extracted the mean effect size from every primary meta-analysis. All the primary meta-analyses used Hedges's *g* or Cohen's *d* to compute the overall mean effect size. We assume the differences between *g* and *d* are minimal, because both metrics represent the standardised mean difference, and the sample sizes were large in most of the primary meta-analyses (Young, 2017). We retrieved the standard error of each mean effect size directly from the article when available. Otherwise, we computed the standard error using 95% confidence intervals via the following formula (Higgins et al., 2019):

$$SE = (95\% CI_{upper\ limit} - 95\% CI_{lower\ limit}) / 3.92$$

We used the Comprehensive Meta-Analysis software package (Borenstein et al., 2009) to conduct our analyses of effect sizes, publication bias and moderators. Publication bias may lead to overestimates of an effect (Steenbergen-Hu et al., 2016). We conducted an evaluation of publication bias using the following tests: a classic fail-safe *N* test, an examination of the funnel plot, the Begg and Mazumdar rank correlation (Kendall's Tau with continuity correction) and Egger's regression. We employed the Duval and Tweedie's (2000) trim-and-fill method to adjust any possible publication bias.

We also examined different possible moderating factors. These moderators can be parsimoniously categorised into two main groups – contextual and methodological factors. We chose to examine these factors because they could influence the effects of a pedagogical innovation on student outcomes (Sailer & Homner, 2020).

The flipped classroom approach has been implemented in various contexts. This second-order meta-analysis includes the research context as a possible moderating factor for the effects of flipped classroom. Some of these contextual factors, as investigated by other meta-analyses of the flipped classroom (e.g., K. S. Chen et al., 2018), include subject discipline, language of primary studies, literature source and year of publication. We coded these factors as follows.

#### *Subject discipline*

Following Låg and Sæle (2019), we coded the subject discipline of the empirical studies included in the primary meta-analysis using the following codes: *HUM* = humanities; *STEM* = science, technology, engineering, mathematics; *MH* = medical and health science; *SS* = social sciences. If multiple subject disciplines were included in one meta-analysis, we coded it as *Combination*.

#### *Language of primary studies*

The language of primary studies refers to the language in which the studies were written. If a meta-analysis examined primary studies written in English, we coded the language of primary studies in the meta-analysis

as *English*. If a meta-analysis included primary studies written in both English and Turkish, we coded the language as *English+Turkish*.

#### *Literature sources*

Following Causadias et al. (2018), we coded this variable as *Published Sources* (e.g., journals, conference proceedings) or *Published and Unpublished Sources* (e.g., journals and theses) based on the inclusion criteria set by the primary meta-analyses. We did not find any primary meta-analysis that focused solely on unpublished sources (e.g., theses).

#### *Year of publication*

Some researchers (e.g., K. S. Chen et al., 2018) have suggested that there might be improvement in flipped classroom outcomes over time due to the increasing maturity of instructors' teaching skills and experience with flipped classroom as time progresses. We therefore examined whether the year of publication may cause any significant difference in effect sizes. We coded the year of publication of each primary meta-analysis (e.g., 2017, 2018, 2019) in the present study.

We also examined the following methodological rigor of the primary meta-analyses as a possible moderating factor.

#### *Instructor equivalence*

Establishing instructor equivalence is important in determining the pedagogical effectiveness of an instructional approach. Since different instructors have different teaching styles, it becomes unclear whether the flipped classroom approach might have caused the effect or if it is influenced by differences between teachers. For this reason, we included *Instructor Equivalence* as a potential moderating factor. We examined whether each primary meta-analysis specifically addressed the issue of instructor equivalence. Inspired by Causadias et al. (2018), we coded a primary meta-analysis as having *Identical Instructor* if it explicitly reported that the same instructor was employed in both the flipped and control groups in greater than 50% of its included empirical studies. We acknowledge that this is a compromised imputation as we could not identify any primary meta-analysis that solely examined empirical studies with identical instructors in both groups (see the Discussion section for a more in-depth treatment of this limitation). Otherwise, we coded the meta-analysis as having *Different Instructors/no data*. We are aware that not all primary meta-analyses explicitly stated whether the instructors were identical or different for the flipped and control groups. Although we can choose to ignore these studies, such a practice is not recommended since drawing conclusions only from studies that do report the items can be misleading (Lipsey & Wilson, 2001). Therefore, following Freeman et al. (2014), we coded the meta-analyses that did not explicitly report it as *Different Instructors/no data*. We then conducted moderator analysis to determine differences (if any) between studies with *Identical Instructor* and studies with *Different Instructors/no data*.

#### *Student initial equivalence*

Determining student initial difference is also important in evaluating pedagogical effectiveness. If students have different initial knowledge about the subject matter, it becomes unclear whether it is the flipped classroom approach that caused the effect or the student's initial knowledge that influenced the outcome. We examined whether a primary meta-analysis specifically addressed the issue of student initial equivalence. If a primary meta-analysis explicitly reported, with supporting statistical evidence, that students in both the flipped and control groups were initially equivalent in more than 50% of its included empirical studies, we coded it as *Equal Students*. We acknowledge that this is a compromised imputation as we could not identify any primary meta-analysis that only examined empirical studies with initial student equivalence in both groups. We also considered students who were randomly assigned to either flipped and control groups as *Equal Students*. Otherwise, we coded the meta-analysis as having *Unequal Students/Unsure*.

#### *Independence of data explicitly addressed*

Often a single primary meta-analysis may report multiple data from the same participants (e.g., Howell, 2013; Whitman Cobb, 2016). For instance, a study may report outcome measures from multiple tests (e.g., weekly tests, midterm test, final test) taken by the same participants (Freeman et al., 2014; Lo & Hew, 2019). The resulting effect sizes are dependent because the same participants were measured more than once (Scammacca et al., 2014). Borenstein et al. (2009) argued that we cannot treat multiple outcomes from the same participants as if they are independent as this would cause incorrect results. It is therefore



important for meta-analyses to explicitly report how they specifically address the issue of multiple outcomes within a same primary study. If a primary meta-analysis explicitly addressed the issue on independence, we coded it as *yes*; if not, we coded it as *no*.

*Research design*

Some meta-analyses include both quasi-experimental and randomised controlled trial (RCT) studies. While participants in RCTs are randomly assigned to different groups to avoid selection bias in the experiment (Sailer & Homner, 2020), quasi-experimental studies cannot do so (Wouters et al., 2013). Therefore, randomisation is likely to be a potential moderator in terms of methodological rigor. If the primary meta-analyses included only empirical studies in which participants were randomly assigned to the experimental and control groups, it was coded as *randomised controlled trials (RCT)*. Following Causadias et al. (2018), if a meta-analysis included more than one research designs, we coded the most frequent research design in it (greater than 50%). If a primary meta-analysis did not explicitly indicate the type of research design, we coded it as *Not Reported*.

*Quality of meta-analyses*

We applied the R-AMSTAR checklist to categorise the quality of each meta-analysis. The quality of the meta-analyses ranged from 23 to 36 R-AMSTAR points. Following Sygouros and Acar (2013), we categorised the total R-AMSTAR points for each meta-analysis based on a percentile ranking of our sample. Two meta-analyses could be classified as high-quality studies ranking within the top 75th percentile; four meta-analyses were of good quality ranking between the 50th and 75th percentiles of our sample; three meta-analyses were of fair quality ranking between the 25th and 50th percentiles; and six meta-analyses were of low-quality ranking below the 25th percentile.

The first author coded all the aforementioned moderating factors. The second author coded 50% of the randomly selected primary meta-analyses. The overall percentage agreement of the coding was 93%. Discrepancies were resolved through mutual discussion.

**Results**

**Effect size synthesis**

We used the random-effects model to synthesise the effect sizes. The results revealed a significant positive effect in favour of the flipped classroom approach on students’ learning outcomes (Hedges’s  $g = 0.45$ , confidence interval (CI) = 0.37–0.53,  $p < 0.001$ ) (see Figure 2).

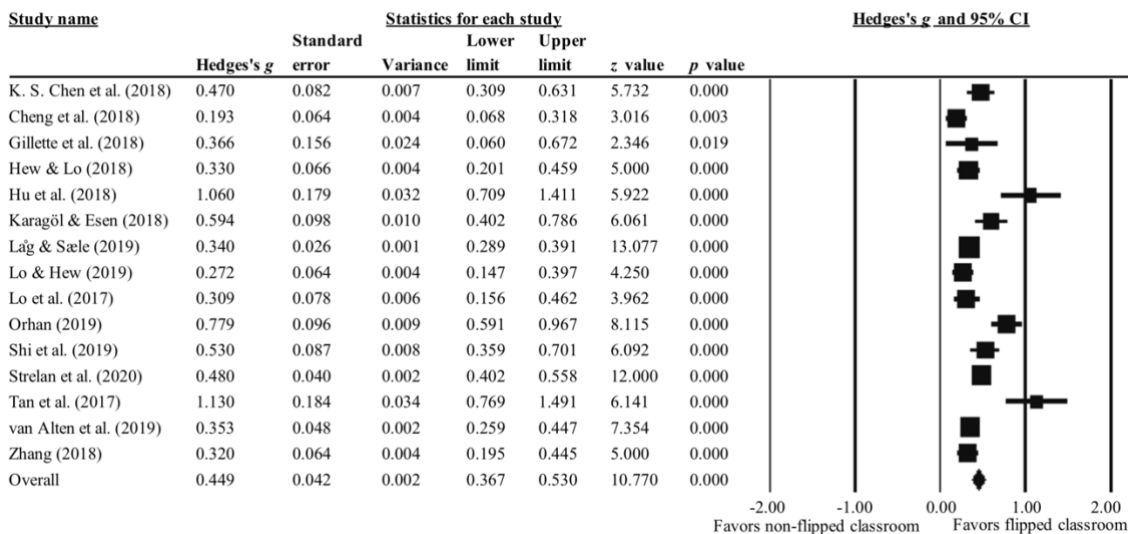


Figure 2. Forest plot of effect sizes

### Validation study

To validate the results of the second-order meta-analysis, we also undertook a meta-analytic validation at the study level. We performed data validation by extracting the raw individual effect sizes used in the 15 primary meta-analyses and using these to perform a regular meta-analysis. A total of 385 available individual effect sizes and standard errors were extracted. The overall random-effect size also showed a significant positive effect in favour of the flipped classroom approach (Hedges's  $g = 0.41$ ,  $CI = 0.36 - 0.45$ ,  $p < 0.001$ ). The mean effect sizes were closely similar in both the second-order meta-analysis and the validation study. The mean effect size of the second-order meta-analyses was 0.45 (random effects model), while the mean effect size of the validation study was 0.41 (random effects model). There was only a difference of 0.04; a magnitude which can be deemed trivial (Cohen, 1988). Therefore, the second-order meta-analysis in the study was considered valid.

### Analyses of publication bias

Next, we conducted an examination of publication bias, which refers to situations when authors selectively report positive and/or significant studies or when journal editors or reviewers preference such studies for publication. Figure 3 shows the funnel plot of the 15 primary meta-analyses. The classic fail-safe  $N$  was 2,289. According to Carson et al. (1990), "if the fail-safe  $N$  ( $X$ ) is relatively small in comparison to the number of studies in the meta-analysis ( $k$ ), then only tenuous conclusions should be drawn" (p. 239). They suggested using Rosenthal's (1979) guideline in which  $X$  should reach  $5k + 10$  to ensure  $X$  is large relative to  $k$ . Using this formula, the  $X$  in our meta-analysis should be larger than 85 (i.e.,  $5(15) + 10$ ). The result of our classic fail-safe  $N$  test showed that 2,289 additional missing studies with a mean effect size of 0 would be required to make the overall effect size statistically insignificant. There would have to be an unreasonably large number of undetected studies with zero effect to bring the reported effect sizes to values that may be statistically insignificant. However, Kendall's Tau was 0.50 (one-tailed  $p = 0.004$ ) and Egger's regression intercept was 2.51 (one-tailed  $p = 0.021$ ), which suggest evidence of publication bias. The trim and fill-method suggests that three studies are missing from the left side of the mean effect; after re-estimation and imputation to account for these missing studies, the mean effect size changed from 0.45 to 0.37 [ $CI = 0.28, 0.46$ ].

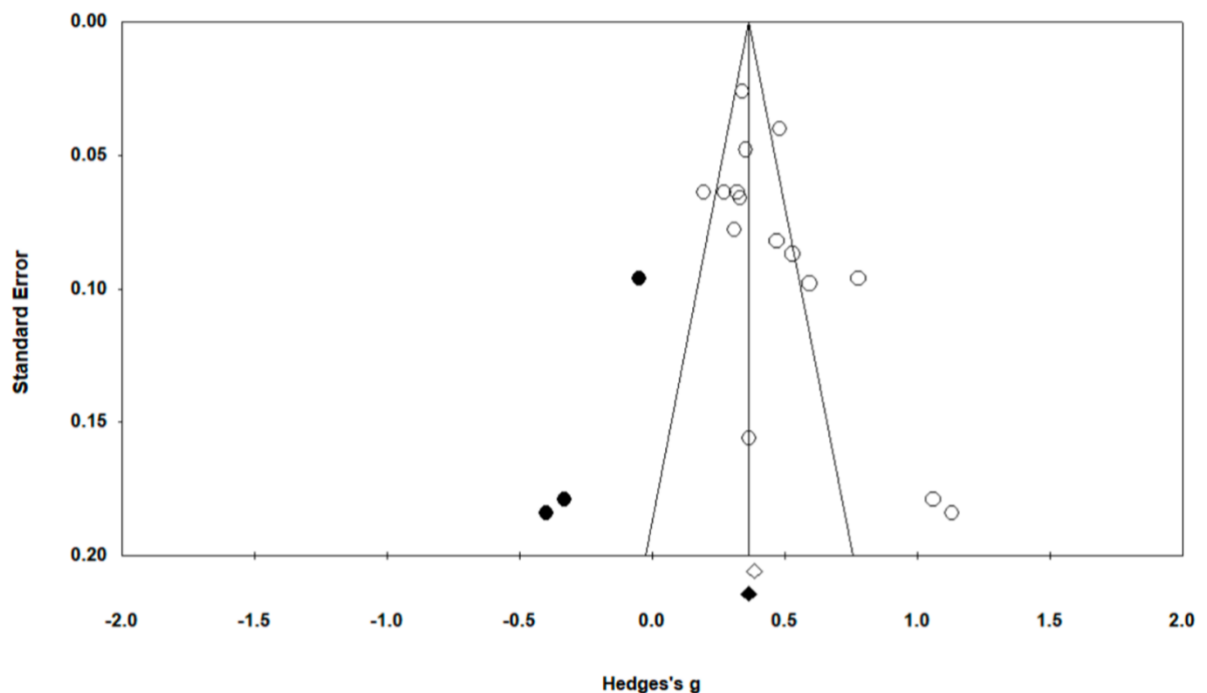


Figure 3. Funnel plot of standard error by Hedges's  $g$  (white data points are observed, black data points are imputed)

In summary, although the large fail-safe  $N = 2,289$  suggests no obvious publication bias, the funnel plot (Figure 3) and statistical tests (Kendall's Tau and Egger's regression) seem to suggest otherwise. Therefore, to be conservative, we conducted trim-and-fill adjustments, which yielded an effect size estimate of 0.37 for the 15 primary meta-analysis. While trim-and-fill analyses probably yield not very precise estimates (Peters et al., 2007), the adjusted effect size is around one third of its standard deviation when taking publication bias into consideration.

### Moderator analyses

To explore the variability between flipped and conventional approaches, we conducted moderator analyses on contextual and methodological features under the random-effects model. The results of the moderator analyses are summarised in Table 2.

Table 2  
Results of moderator analyses

Moderator variables	<i>N</i>	<i>g</i>	<i>SE</i>	95% CI		$Q_B$ ( <i>p</i> )
				LL	UL	
Subject discipline						5.746 (0.057)
Combination	8	0.453	0.056	0.343	0.562	
Health professions	4	0.623	0.099	0.428	0.818	
STEM	3	0.300	0.091	0.122	0.478	
Languages of primary studies						39.742*** (< 0.001)
Chinese	2	1.094	0.137	0.826	1.362	
English	11	0.357	0.028	0.302	0.412	
English+Turkish	2	0.688	0.083	0.525	0.851	
Independence addressed						22.896*** (< 0.001)
No	5	0.739	0.072	0.598	0.880	
Yes	10	0.357	0.035	0.289	0.425	
Literature sources						0.587 (0.444)
Published	8	0.421	0.062	0.300	0.542	
Published and Unpublished	7	0.489	0.065	0.362	0.617	
Year of publication						1.095 (0.778)
2017	2	0.598	0.147	0.311	0.886	
2018	7	0.434	0.073	0.291	0.577	
2019	5	0.440	0.079	0.284	0.596	
2020	1	0.480	0.169	0.148	0.812	
Quality of meta-analysis						1.346 (0.718)
High	2	0.599	0.149	0.307	0.890	
Good	4	0.492	0.100	0.296	0.687	
Fair	3	0.441	0.119	0.207	0.675	
Low	6	0.414	0.078	0.260	0.567	
Student equivalence						6.801* (0.009)
Different/unsure	10	0.531	0.052	0.429	0.632	
Equal/randomised	5	0.315	0.064	0.189	0.442	
Instructor equivalence						3.312 (0.069)
Different/unsure	12	0.495	0.049	0.398	0.592	
Identical	3	0.309	0.089	0.133	0.484	
Research design						20.444*** (<0.001)
Quasi-experiment	4	0.412	0.065	0.285	0.539	
RCT	2	1.094	0.150	0.801	1.388	
Not reported	9	0.393	0.044	0.307	0.479	

Note. *N* = number of primary meta-analysis. *g* = Hedges's *g*; *SE* = standard error; CI = confidence interval; LL = lower limit; UL = upper limit. STEM = science, technology, engineering, and mathematics.

\* $p < .05$  \*\*\* $p < .001$

#### *Subject discipline*

The moderator analysis concerning subject discipline suggested no significant effect size difference between different disciplines ( $Q_B = 5.746$ ,  $df = 2$ ,  $p = 0.057$ ).

#### *Language of primary studies*

There was a significant difference between Chinese, English, and English + Turkish empirical studies ( $Q_B = 39.742$ ,  $df = 2$ ,  $p < 0.001$ ). Primary meta-analyses examining Chinese-reported primary studies displayed a higher effect size ( $g = 1.094$ ) than meta-analyses investigating English primary studies ( $g = 0.357$ ) and meta-analyses examining both English and Turkish studies ( $g = 0.688$ ).

#### *Independence of effect sizes*

We found evidence of significant variation in effect sizes between meta-analyses that explicitly handled effect size independence and meta-analyses which did not ( $Q_B = 22.896$ ,  $df = 1$ ,  $p < 0.001$ ). Primary meta-analyses that did not explicitly handle effect size independence reported a higher effect size ( $g = 0.739$ ) than those that did ( $g = 0.357$ ).

#### *Literature included*

There was no significant difference in effect sizes between only published sources, and both published and unpublished sources ( $Q_B = 0.587$ ,  $df = 1$ ,  $p = 0.444$ ).

#### *Year of publication*

There was no significant difference among 2017, 2018, 2019 and 2020 ( $Q_B = 1.095$ ,  $df = 3$ ,  $p = 0.778$ ).

#### *Quality of primary meta-analysis*

There was no significant difference in effect sizes between the different groups of meta-analytic quality ( $Q_B = 1.346$ ,  $df = 3$ ,  $p = 0.718$ ), although the mean effect size of the high-quality meta-analyses was greater than the other groups.

#### *Student initial equivalence*

Moderator analysis showed a significant difference between primary meta-analyses that reported students in both the flipped and control groups were initially equivalent and meta-analyses that did not report student initial equivalence ( $Q_B = 6.801$ ,  $df = 1$ ,  $p = 0.009$ ). Meta-analyses that did not report student initial equivalence showed a higher mean effect size ( $g = 0.531$ ) than their counterparts ( $g = 0.315$ ).

#### *Instructor equivalence*

We found no evidence of significant variation in effect sizes between meta-analyses that reported the same instructor was employed in both the flipped and control groups and meta-analyses that did not report instructor equivalence ( $Q_B = 3.312$ ,  $df = 1$ ,  $p = 0.069$ ).

#### *Research design*

A significant difference in effect sizes was found pertaining to the research design used ( $Q_B = 20.444$ ,  $df = 3$ ,  $p < 0.001$ ). Results showed that studies employing RCTs had a higher effect size ( $g = 1.094$ ) than studies that predominantly used quasi-experiments ( $g = 0.412$ ).

## **Discussion**

In this study, we conducted the largest quantitative test to date on the effect of the flipped classroom approach on student learning performance, with data from 15 primary meta-analyses covering 156,722 total participants. We found a significant mean effect size of 0.45 (0.37 after trim-and-fill adjustment) supporting the flipped classroom approach. Compared to students in a non-flipped classroom, students in a flipped classroom have a greater opportunity to do self-paced learning due to the availability of pre-class activity. Students can choose to watch the video or read the course materials at any time and in whatever pace they desire. The flipped classroom also provided students with more than one exposure to the course materials, as they are first exposed to the course materials before class during the pre-class activity as well as in the classroom (Lo & Hew, 2019). Multiple exposure to course materials can help improve student understanding of the lesson (Lo & Hew, 2019; Yelamarthi & Drake, 2015; Yelamarthi et al., 2016).

The overall effect size of 0.37 may be considered small (Cohen, 1988). Therefore, educators might question whether it is worth the time and effort to flip a course. But the real question is: how big should a flipped classroom effect size be before it is considered worthwhile to be used? In terms of learning achievement in the education field, an increase of 0.1 in the effect size is argued to be a marked improvement if it is a result of a small and inexpensive change (Coe, 2002). Glass et al. (1981) pointed out that the practical importance of an effect depends entirely on its relative costs and benefits. One of the major costs involved in flipping a course is the considerable demand on an instructor's effort (Altaii et al., 2017; Cheng et al., 2019; McLaughlin et al., 2014). Yet, although a significant amount of start-up effort is required to create flipped classroom resources, these resources can be reused in subsequent semesters, which may make the flipped classroom less expensive in the long term.

In this study, we also examined if contextual and methodological moderators affect the disparities of effect sizes. We found no evidence of significant difference in effect sizes regarding meta-analytic quality, subject disciplines, literature sources, sample sizes, year of publication and instructor equivalence. Although there was no significant effect size difference regardless of meta-analytic quality, it is still important to pay attention to quality in meta-analyses. This is because quality can help minimise the problem of the mass production of ambiguous and misleading meta-analyses (Ioannidis, 2016).

Likewise, even though we found no significant difference in effect size between meta-analyses that reported the same instructor was employed in both groups and meta-analyses that did not report an identical instructor, it is still a matter of concern that many primary meta-analyses have failed to account for instructor equivalence. The use of different instructors in the treatment and control groups is, after all, a poorly controlled study. We suggest two simple methods that future meta-analysts can employ to deal with the issue of instructor equivalence. First, meta-analysts could choose to select and analyse only empirical studies that used identical instructors in both the treatment and control groups. Second, meta-analysts could create two categories to characterise the quality of the controls over instructor equivalence in the included empirical studies; for example, (a) no data or different instructors, and (b) identical instructor (Freeman et al., 2014). Later, a moderator analysis could be conducted to reveal the differences (if any) between the two categories of studies.

Some of the findings should be viewed with some caution. For instance, with regard to subject disciplines, eight of the primary meta-analyses examined a combination of disciplines. As a result, we could not report a more fine-grained result concerning the effect of specific subject disciplines on effect size heterogeneity. The four categories used here, from Låg and Sæle (2019), are broad and have the potential to conflate disciplines where the flipped approach may work quite differently, such as education and economics, which are both categorised as social sciences. Further work is necessary to identify the particular subject disciplines that are best suited to the flipped approach.

We found evidence of effect size differences with regard to the language of the empirical studies, type of research design, student initial equivalence and issue of effect size independence. Concerning the language of the empirical studies, we found that the largest mean effect size ( $g = 1.094$ ) came from studies by Hu et al. (2018) and Tan et al. (2017) conducted exclusively with Chinese nursing students reported in Chinese-medium publications. At this moment, we could not provide any conclusive reasons for this. However, we can speculate that one possible reason for this could be that both meta-analyses accessed Chinese-language databases (e.g., China National Knowledge Infrastructure, Wanfang, Chinese Scientific Journals Database), which were not accessed by the other meta-analyses. Another possible reason is that compared to other contexts, teachers in conventional Chinese classroom contexts are more authoritarian, and students are expected to be quiet and obedient and not question the information given by teachers during lectures (Sit, 2013). This strict expository teaching culture can greatly diminish students' learning enthusiasm (Hu et al., 2018) and makes learning less effective. This may explain why the flipped learning approach, which emphasises student active learning, may lead to even larger benefits when compared with expository teaching (Freeman et al., 2014).

We found meta-analyses that employed RCTs reported a significantly higher mean effect size than meta-analyses that predominantly employed quasi-experiments. This finding appears to contradict those of previous reviews which reported either non-significant differences in effect sizes between randomised and quasi-experiments (de Boer et al., 2014) or a significantly higher mean effect size in quasi-experimental studies (A. C. K. Cheung & Slavin, 2016). The significantly larger mean effect size in the randomised

studies found in the present review may be accounted for in part by the large number of randomised studies conducted exclusively with Chinese nursing students reported in Chinese-medium publications. We were unable to retrieve the actual empirical studies for further verification. Hence, we could provide only some possible reasons, as mentioned in the preceding paragraph.

We will discuss the significant differences in effect sizes regarding the student initial equivalence and effect size independence moderators in the following subsection.

### **Limitations of previous primary meta-analyses**

We have identified three major limitations of previous primary meta-analyses: a lack of control to address possible student initial differences, a lack of explanation as to how independence of effect size was addressed and sole reliance on funnel plots to assess publication bias.

The results of a moderator analysis showed a significant difference between primary meta-analyses that reported students in both the flipped and control groups were initially equivalent and meta-analyses that did not report student initial equivalence. More specifically, meta-analyses that did not explicitly report student initial equivalence showed a higher mean effect size. The large number of primary meta-analyses that failed to account for student initial equivalence is a matter of concern. If students have different initial knowledge about the subject matter, it becomes unclear whether it is the flipped classroom approach that caused the effect or the student's initial knowledge that influenced the outcome.

We therefore suggest two possible methods for future meta-analysts to address the potential problem of student initial differences. These two methods are similar to what we may use to address the problem of instructor difference. The first and easier method is to simply impose a stricter set of inclusion criteria to accept only empirical studies that explicitly reported student initial equivalence based on some form of pre-tests. Such a claim should be supported by relevant statistical data as we have, based on our experience, found empirical studies that merely claimed that the students were equal in terms of their initial knowledge of the subject matter, but the statistical evidence was missing. The first method, however, may yield too few primary studies to be included in a meta-analysis. The second method is to adapt the procedure employed by other scholars (e.g., Freeman et al., 2014) by categorising the empirical studies into one of several groups in terms of how well the empirical studies controlled for student equivalence. For example, an analyst may decide to categorise the empirical studies into studies with no data on student initial equivalence, studies with data showing student initial equivalence and studies with data showing students were unequal. The analyst can then conduct between-level moderator analysis to identify the differences in effect size (if any) among the different groups of studies.

One of the key assumptions in a primary meta-analysis is that the effect sizes are independent (M. W.-L. Cheung, 2019). For example, if more than one effect sizes are computed involving the same sample of participants, this will make such effect sizes correlated (Nakagawa et al., 2017). Consequently, the non-independent effect sizes can lead to incorrect conclusions (Nakagawa & Santos, 2012). In this study, close to half of the primary meta-analyses did not explicitly state they corrected for non-independence of effect sizes. We found that meta-analyses that did not explicitly state they corrected for non-independence had a significantly higher mean effect size than those that explicitly corrected for non-independence. To correct for non-independence among effect sizes in a single empirical study, the meta-analyst could use methods such as averaging the non-independent effect sizes (Cheng et al., 2019; Nakagawa et al., 2017) or selecting one among several non-independent effect sizes (Hew & Lo, 2018).

Publication bias was assessed in all 15 primary meta-analyses. Of these, four relied solely on visual inspection of the funnel plots (Gillette et al., 2018; Hu et al., 2018; Shi et al., 2019; Tan et al., 2017). Although funnel plots can provide an easy way to visualise the presence of asymmetry (which indicates the presence of publication bias) (Egger et al., 1997), researchers can be misled by its shape (Terrin et al., 2005). Further tests, such as the classic fail-safe  $N$ , as well as statistical tests (e.g., Begg and Mazumdar rank correlation) which can “quantify the amount of bias” (Borenstein, 2005, p. 195) should be conducted. Although there are no perfect solutions for correcting for publication bias, existing techniques should still be used (Nakagawa et al., 2017). An example of these existing techniques is the trim-and-fill approach. Based on the funnel plot of the data set, this approach trims off the asymmetric outlying part of the funnel after estimating the number of studies that lie within the asymmetric part (Duval & Tweedie, 2000).

Possible studies would be imputed in the trim-and-fill analysis, which will lead to an adjustment of the mean effect size. We have demonstrated that the use of these techniques leads to a slightly smaller overall effect size, which suggests that some researchers are not publishing unsuccessful flipped classroom interventions.

## Conclusion

The increasing popularity of flipped classroom has spawned many empirical studies. Along with this growing number of empirical studies, there has also been a corresponding increase of meta-analytic studies of the flipped classroom approach. Rather than conducting yet another primary meta-analysis, this study employed a second-order meta-analysis method to systematically synthesise the findings of 15 primary meta-analyses. We conclude by presenting the main limitation of the present second-order meta-analysis and several implications for future flipped classroom research.

The main limitation of our second-order meta-analysis is that we could not examine other possible moderators due to the lack of available data. Flipped classroom approaches may differ from one another. For example, some flipped classrooms may include the use of quiz and video lectures before class, while others do not. However, since we conducted a second-order meta-analysis which is a meta-analysis of primary meta-analyses (Young, 2017), we can only analyse what was reported in the primary meta-analyses. Although all the primary meta-analyses that we examined hold the common view that in a flipped approach, students study instructional material before class and apply the learning material during class, not every primary meta-analysis reported the details of the flipped classroom approaches used. For instance, we could not examine the role of the use of quizzes across all the primary meta-analyses because most of the included primary meta-analyses did not report it. Likewise, we could not extract data on the flipped classroom intervention duration because many primary meta-analyses did not report or did not clearly specify the durations. We encourage future meta-analysts to think carefully about the potential moderators that may be of practical or theoretical significance, and we similarly encourage future empirical study authors to report greater details to enable extraction of these moderators. This would enable us to move beyond the simplistic question of if the flipped classroom “works” and towards asking more nuanced questions such as “for whom?” and “in what circumstances?” (Pawson, 2006).

Despite this limitation, this second-order meta-analysis is the first study to provide evidence of this magnitude on the effect of the flipped classroom approach on student learning outcomes. The results showed that the flipped classroom approach, on the whole, does increase learning performance compared with conventional, non-flipped teaching. Future research should therefore move beyond doing yet another empirical study comparing the effectiveness of flipped learning versus conventional teaching. It should instead investigate the impact of the design of pre-class as well as in-class learning activities on student learning performance. Since flipped learning poses heavier demands on learners’ self-regulation (e.g., the requirement to watch videos outside of class), it would be profitable for future research to examine what self-regulation strategies can best promote students’ pre-class learning (Cheng et al., 2019), what strategies can best motivate students to complete the pre-class work as well as the long-term effects of flipped classroom approaches on students’ learning performance.

## Acknowledgements

The research was supported by a grant from the Research Grants Council of Hong Kong (Project reference no: 17610919).

## References

- Abeyssekera, L., & Dawson, P. (2015). Motivation and cognitive load in the flipped classroom: definition, rationale and a call for research. *Higher Education Research & Development*, 34(1), 1–14. <https://doi.org/10.1080/07294360.2014.934336>
- Altaii, K., Reagle, C. J., & Handley, M. K. (2017). Flipping an engineering thermodynamics course to improve student self-efficacy. In *Proceedings of the 124th American Society for Engineering Education Annual Conference & Exposition* (pp. 23529–23545). American Society for Engineering Education. <https://doi.org/10.18260/1-2--28368>

- Berrett, D. (2012, February 19). How 'flipping' the classroom can improve the traditional lecture. *The Chronicle of Higher Education*, 12(19), 1–3. <https://www.chronicle.com/article/how-flipping-the-classroom-can-improve-the-traditional-lecture/>
- Borenstein, M. (2005). Software for publication bias. In A. J. S. H.R. Rothstein & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 193–220). John Wiley & Sons, Ltd. <https://doi.org/10.1002/0470870168.ch11>
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Wiley. <https://doi.org/10.1002/9780470743386>
- Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (1999). *How people learn: Brain, mind, experience, and school*. National Academy Press. <https://doi.org/10.17226/9853>
- Busch, T., & Friede, G. (2018). The robustness of the corporate social and financial performance relation: A second-order meta-analysis. *Corporate Social Responsibility and Environmental Management*, 25(4), 583–608. <https://doi.org/10.1002/csr.1480>
- Carson, K. P., Schriesheim, C. A., & Kinicki, A. J. (1990). The usefulness of the “fail-safe” statistic in meta-analysis. *Educational and Psychological Measurement*, 50(2), 233–243. <https://doi.org/10.1177/0013164490502001>
- Causadias, J. M., Korous, K. M., & Cahill, K. M. (2018). Are Whites and minorities more similar than different? Testing the cultural similarities hypothesis on psychopathology with a second-order meta-analysis. *Development and Psychopathology*, 30(5), 2009–2027. <https://doi.org/10.1017/S0954579418000895>
- Chen, F., Lui, A. M., & Martinelli, S. M. (2017). A systematic review of the effectiveness of flipped classrooms in medical education. *Medical Education*, 51(6), 585–597. <https://doi.org/10.1111/medu.13272>
- Chen, K. S., Monrouxe, L., Lu, Y. H., Jenq, C. C., Chang, Y. J., Chang, Y. C., & Chai, P. Y. C. (2018). Academic outcomes of flipped classroom learning: a meta-analysis. *Medical Education*, 52(9), 910–924. <https://doi.org/10.1111/medu.13616>
- Cheng, L., Ritzhaupt, A. D., & Antonenko, P. (2019). Effects of the flipped classroom instructional strategy on students' learning outcomes: a meta-analysis. *Educational Technology Research and Development*, 67(4), 793–824. <https://doi.org/10.1007/s11423-018-9633-7>
- Cheung, A. C. K., & Slavin, R. E. (2016). How methodological features affect effect sizes in education. *Educational Researcher*, 45(5), 283–292. <https://doi.org/10.3102/0013189X16656615>
- Cheung, M. W.-L. (2019). A guide to conducting a meta-analysis with non-independent effect sizes. *Neuropsychology Review*, 29, 387–396. <https://doi.org/10.1007/s11065-019-09415-6>
- Coe, R. (2002, September 12–14). *It's the effect size stupid: What effect size is and why it is important* [Paper presentation]. Annual Conference of the British Educational Research Association, Exeter, England. <https://www.leeds.ac.uk/educol/documents/00002182.htm>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates. <https://doi.org/10.4324/9780203771587>
- Cooper, H., & Koenka, A. C. (2012). The overview of reviews: Unique challenges and opportunities when research syntheses are the principal elements of new integrative scholarship. *American Psychologist*, 67(6), 446–462. <https://doi.org/10.1037/a0027119>
- Davis, M. A., Curtis, M. B., & Tschetter, J. D. (2003). Evaluating cognitive training outcomes: Validity and utility of structural knowledge assessment. *Journal of Business and Psychology*, 18(2), 191–206. <https://doi.org/10.1023/A:1027397031207>
- de Boer, H., Donker, A. S., & van der Werf, M. P. C. (2014). Effects of the attributes of educational interventions on students' academic performance: A meta-analysis. *Review of Educational Research*, 84(4), 509–545. <https://doi.org/10.3102/0034654314540006>
- DeSantis, J., Van Curen, R., Putsch, J., & Metzger, J. (2015). Do students learn more from a flip? An exploration of the efficacy of flipped and traditional lessons. *Journal of Interactive Learning Research*, 26(1), 39–63. <https://www.learntechlib.org/primary/p/130133/>
- Deslauriers, L., McCarthy, L. S., Miller, K., Callaghan, K., & Kestin, G. (2019). Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. *PNAS*, 116(39), 19251–19257. <https://doi.org/10.1073/pnas.1821936116>
- Dove, A., & Dove, E. (2015). Examining the influence of a flipped mathematics course on preservice elementary teachers' mathematics anxiety and achievement. *Electronic Journal of Mathematics & Technology*, 9(2), 166–179. [https://php.radford.edu/~ejmt/deliverAbstract.php?paperID=eJMT\\_v9n2n2](https://php.radford.edu/~ejmt/deliverAbstract.php?paperID=eJMT_v9n2n2)



- Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56(2), 455-463. <https://doi.org/10.1111/j.0006-341X.2000.00455.x>
- Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, 315(7109), 629-634. <https://doi.org/10.1136/bmj.315.7109.629>
- Flynn, A. B. (2015). Structure and evaluation of flipped chemistry courses: organic & spectroscopy, large and small, first to third year, English and French. *Chemistry Education Research and Practice*, 16(2), 198-211. <https://doi.org/10.1039/c4rp00224e>
- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *PNAS*, 111(13), 8410-8415. <https://doi.org/10.1073/pnas.1319030111>
- Gaebel, M., & Zhang, T. (2018). *Trends 2018: Learning and teaching in the European higher education area*. European University Association. <https://eua.eu/downloads/publications/trends-2018-learning-and-teaching-in-the-european-higher-education-area.pdf>
- Gillette, C., Rudolph, M., Kimble, C., Rockich-Winston, N., Smith, L., & Broedel-Zaugg, K. (2018). A meta-analysis of outcomes comparing flipped classroom and lecture. *American Journal of Pharmaceutical Education*, 82(5), 433-440. <https://doi.org/10.5688/ajpe6898>
- Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Sage.
- Gurevitch, J., Koricheva, J., Nakagawa, S., & Stewart, G. (2018). Meta-analysis and the science of research synthesis. *Nature*, 555(7695), 175. <https://doi.org/10.1038/nature25753>
- Hartling, L., Featherstone, R., Nuspl, M., Shave, K., Dryden, D. M., & Vandermeer, B. (2017). Grey literature in systematic reviews: A cross-sectional study of the contribution of non-English reports, unpublished studies and dissertations to the results of meta-analyses in child-relevant reviews. *BMC Medical Research Methodology*, 17(64). <https://doi.org/10.1186/s12874-017-0347-z>
- Hew, K. F., & Lo, C. K. (2018). Flipped classroom improves student learning in health professions education. *BMC Medical Education*, 18, Article 38. <https://doi.org/10.1186/s12909-018-1144-z>
- Higgins, J. P. T., Li, T., & Deeks, J. (2019). Chapter 6: Choosing effect measures and computing estimates of effect. In J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, & V. A. Welch (Eds.), *Cochrane handbook for systematic reviews of interventions* (2nd ed., pp. 143-176). John Wiley & Sons. <https://doi.org/10.1002/9781119536604.ch6>
- Howell, D. (2013). *Effects of an inverted instructional delivery model on achievement of ninth-grade physical science honors students* [Doctoral dissertation, Gardner-Webb University]. Digital Commons @ Gardner Webb University. [https://digitalcommons.gardner-webb.edu/education\\_etd/35](https://digitalcommons.gardner-webb.edu/education_etd/35)
- Hu, R., Gao, H., Ye, Y., Ni, Z., Jiang, N., & Jiang, X. (2018). Effectiveness of flipped classrooms in Chinese baccalaureate nursing education: A meta-analysis of randomized controlled trials. *International Journal of Nursing Studies*, 79, 94-103. <https://doi.org/10.1016/j.ijnurstu.2017.11.012>
- Ioannidis, J. P. A. (2016). The mass production of redundant, misleading, and conflicted systematic reviews and meta-analyses. *The Milbank Quarterly*, 94(3), 485-514. <https://doi.org/10.1111/1468-0009.12210>
- Jungic, V., Kaur, H., Mulholland, J., & Xin, C. (2015). On flipping the classroom in large first year calculus courses. *International Journal of Mathematical Education in Science and Technology*, 46(4), 508-520. <https://doi.org/10.1080/0020739X.2014.990529>
- Kang, S., & Shin, I.-S. (2016). *The effect of flipped learning in Korea: Meta-analysis*. <https://www.semanticscholar.org/paper/The-Effect-of-Flipped-Learning-in-Korea-%3A-Kang-Shin/6d9d49af1e74a7b6cf6c6f27c24274e6f645d5d8>
- Karagöl, İ., & Esen, E. (2019). The effect of flipped learning approach on academic achievement: A meta-analysis study. *Hacettepe University Journal of Education*, 34(3), 708-727. <https://doi.org/10.16986/HUJE.2018046755>
- Keengwe, S. J., Onchwari, G., & Oigara, J. (2014). *Promoting active learning through the flipped classroom model*. IGI Global. <https://doi.org/10.4018/978-1-4666-4987-3>
- Kohl, L. F. M., Crutzen, R., & de Vries, N. K. (2013). Online prevention aimed at lifestyle behaviors: A systematic review of reviews. *Journal of Medical Internet Research*, 15(7), 1-14. <https://doi.org/10.2196/jmir.2665>
- Kung, J., Chiappelli, F., Cajulis, O. O., Avezova, R., Kossan, G., Chew, L., & Maida, C. A. (2010). From systematic reviews to clinical recommendations for evidence-based health care: Validation of revised assessment of multiple systematic reviews (R-AMSTAR) for grading of clinical relevance. *The Open Dentistry Journal*, 4(1), 84-91. <https://doi.org/10.2174/1874210601004020084>

- Låg, T., & Sæle, R. G. (2019). Does the flipped classroom improve student learning and satisfaction? A systematic review and meta-analysis. *AERA Open*, 5(3), 1–17. <https://doi.org/10.2174/1874210601004020084>
- Lax, N., Morris, J., & Kolber, B. J. (2017). A partial flip classroom exercise in a large introductory general biology course increases performance at multiple levels. *Journal of Biological Education*, 51(4), 412–426. <https://doi.org/10.1080/00219266.2016.1257503>
- Lefebvre, C., Manheimer, E., & Glanville, J. (2008). Searching for studies. In J. P. T. Higgins, & S. Green (Eds.), *Cochrane handbook for systematic reviews of interventions: Cochrane Book Series* (pp. 95–150). John Wiley & Sons. <https://doi.org/10.1002/9780470712184.ch6>
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis. Vol. 49. Applied social research methods series*: Sage Publications, Inc.
- Lo, C. K., & Hew, K. F. (2019). The impact of flipped classrooms on student achievement in engineering education: A meta-analysis of 10 years of research. *Journal of Engineering Education*, 108(4), 523–546. <https://doi.org/10.1002/jee.20293>
- Lo, C. K., Hew, K. F., & Chen, G. (2017). Toward a set of design principles for mathematics flipped classrooms: A synthesis of research in mathematics education. *Educational Research Review*, 22, 50–73. <https://doi.org/10.1016/j.edurev.2017.08.002>
- Lundin, M., Bergviken Rensfeldt, A., Hillman, T., Lantz-Andersson, A., & Peterson, L. (2018). Higher education dominance and siloed knowledge: a systematic review of flipped classroom research. *International Journal of Educational Technology in Higher Education*, 15(1), Article 20. <https://doi.org/10.1186/s41239-018-0101-6>
- McLaughlin, J. E., Roth, M. T., Glatt, D. M., Gharkholonarehe, N., Davidson, C. A., Griffin, L. M., Esserman, D. A., & Mumper, R. J. (2014). The flipped classroom: A course redesign to foster learning and engagement in a health professions school. *Academic Medicine*, 89(2), 236–243. <https://doi.org/10.1097/ACM.0000000000000086>
- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA Statement. *Journal of Clinical Epidemiology*, 62(10), 1006–1012. <https://doi.org/10.1016/j.jclinepi.2009.06.005>
- Nakagawa, S., & Santos, E. S. A. (2012). Methodological issues and advances in biological meta-analysis. *Evolutionary Ecology*, 26(5), 1253–1274. <https://doi.org/10.1007/s10682-012-9555-5>
- Nakagawa, S., Noble, D. W. A., Senior, A. M., & Lagisz, M. (2017). Meta-evaluation of meta-analysis: ten appraisal questions for biologists. *BMC Biology*, 15(1), 1–14. <https://doi.org/10.1186/s12915-017-0357-7>
- Orhan, A. (2019). The effect of flipped learning on students' academic achievement: A meta-analysis study. *Çukurova University. Faculty of Education Journal*, 48(1), 368–396. <https://doi.org/10.14812/cufej.400919>
- Pawson, R. (2006). *Evidence-based policy: A realist perspective*. Sage.
- Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2007). Performance of the trim and fill method in the presence of publication bias and between-study heterogeneity. *Statistics in Medicine*, 26(25), 4544–4562. <https://doi.org/10.1002/sim.2889>
- Phillips, J. M. (2005). Strategies for active learning in online continuing education. *Journal of Continuing Education in Nursing*, 36(2), 77–83. <https://doi.org/10.3928/0022-0124-20050301-08>
- Pigott, T. D., Valentine, J. C., Polanin, J. R., Williams, R. T., & Canada, D. D. (2013). Outcome-reporting bias in education research. *Educational Researcher*, 42(8), 424–432. <https://doi.org/10.3102/0013189X13507104>
- Polanin, J. R., Maynard, B. R., & Dell, N. A. (2017). Overviews in education research: A systematic review and analysis. *Review of Educational Research*, 87(1), 172–203. <https://doi.org/10.3102/0034654316631117>
- 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>
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1), 68–78. <https://doi.org/10.1037/0003-066X.55.1.68>
- Sailer, M., & Homner, L. (2020). The gamification of learning: A meta-analysis. *Educational Psychology Review*, 32, 77–112. <https://doi.org/10.1007/s10648-019-09498-w>
- Scammacca, N., Roberts, G., & Stuebing, K. K. (2014). Meta-analysis with complex research designs: Dealing with dependence from multiple measures and multiple group comparisons. *Review of Educational Research*, 84(3), 328–364. <https://doi.org/10.3102/0034654313500826>

- Sergis, S., Sampson, D. G., & Pelliccione, L. (2018). Investigating the impact of flipped classroom on students' learning experiences: A self-determination theory approach. *Computers in Human Behavior*, 78, 368–378. <https://doi.org/10.3102/0034654313500826>
- Setren, E., Greenberg, K., Moore, O., & Yankovich, M. (2019). *Effects of the flipped classroom: Evidence from a randomized trial* (EdWorkingPaper: 19-113). Annenberg Institute at Brown University. <http://www.edworkingpapers.com/ai19-113>
- Shi, Y., Ma, Y., MacLeod, J., & Yang, H. H. (2019). College students' cognitive learning outcomes in flipped classroom instruction: A meta-analysis of the empirical literature. *Journal of Computers in Education*, 7(1), 79–103. <https://doi.org/10.1007/s40692-019-00142-8>
- Sit, H. H. W. (2013). Characteristics of Chinese students' learning styles. *International Proceedings of Economics Development and Research*, 62, 36–39. <http://www.ipedr.com/vol62/008-ICLMC2013-M10004.pdf>
- Spector, J. M. (2014). Conceptualizing the emerging field of smart learning environments. *Smart Learning Environments*, 1(1), Article 2. <https://doi.org/10.1186/s40561-014-0002-7>
- Stains, M., Harshman, J., Barker, M. K., Chasteen, S. V., Cole, R., DeChenne-Peters, S. E., Eagan, M. K., Esson, J. M., Knight, J. K., Laski, F. A., Levis-Fitzgerald, M., Lee, C. J., Lo, S. M., McDonnell, L. M., McKay, T. A., Michelotti, N., Musgrove, A., Palmer, M. S., Plank, K. M., ... Young, A. M. (2018). Anatomy of STEM teaching in North American universities. *Science*, 359(6383), 1468–1470. <https://doi.org/10.1126/science.aap8892>
- Steenbergen-Hu, S., Makel, M. C., & Olszewski-Kubilius, P. (2016). What one hundred years of research says about the effects of ability grouping and acceleration on K–12 students' academic achievement: Findings of two second-order meta-analyses. *Review of Educational Research*, 86(4), 849–899. <https://doi.org/10.3102/0034654316675417>
- Strelan, P., Osborn, A., & Palmer, E. (2020). The flipped classroom: A meta-analysis of effects on student performance across disciplines and education levels. *Educational Research Review*, 30, 100314. <https://doi.org/10.1016/j.edurev.2020.100314>
- Sygyouros, A., & Acar, A. (2013). Evidence-based orthodontics: Appraisal of the methodologies of systematic reviews and meta-analyses in controversial areas of orthodontics. *Journal of the World Federation of Orthodontists*, 2(3), 117–122. <https://doi.org/10.1016/j.ejwf.2013.05.004>
- Tamim, R. M., Bernard, R. M., Borokhovski, E., Abrami, P. C., & Schmid, R. F. (2011). What forty years of research says about the impact of technology on learning: A second-order meta-analysis and validation study. *Review of Educational Research*, 81(1), 4–28. <https://doi.org/10.3102/0034654310393361>
- Tan, C., Yue, W.-G., & Fu, Y. (2017). Effectiveness of flipped classrooms in nursing education: Systematic review and meta-analysis. *Chinese Nursing Research*, 4(4), 192–200. <https://doi.org/10.1016/j.cnre.2017.10.006>
- Terrin, N., Schmid, C. H., & Lau, J. (2005). In an empirical evaluation of the funnel plot, researchers could not visually identify publication bias. *Journal of Clinical Epidemiology*, 58(9), 894–901. <https://doi.org/10.1016/j.jclinepi.2005.01.006>
- van Alten, D. C. D., Phielix, C., Janssen, J., & Kester, L. (2019). Effects of flipping the classroom on learning outcomes and satisfaction: A meta-analysis. *Educational Research Review*, 28, Article 100281. <https://doi.org/10.1016/j.edurev.2019.05.003>
- Whitman Cobb, W. N. (2016). Turning the classroom upside down: Experimenting with the flipped classroom in American government. *Journal of Political Science Education*, 12(1), 1–14. <https://doi.org/10.1080/15512169.2015.1063437>
- Wouters, P., van Nimwegen, C., van Oostendorp, H., & van der Spek, E. D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology*, 105(2), 249–265. <https://doi.org/10.1037/a0031311>
- Yelamarthi, K., & Drake, E. (2015). A flipped first-year digital circuits course for engineering and technology students. *IEEE Transactions on Education*, 58(3), 179–186. <https://doi.org/10.1109/TE.2014.2356174>
- Yelamarthi, K., Drake, E., & Prewett, M. (2016). An instructional design framework to improve student learning in a first-year engineering class. *Journal of Information Technology Education: Innovations in Practice*, 15, 195–222. <https://doi.org/10.28945/3617>
- Young, J. (2017). Technology-enhanced mathematics instruction: A second-order meta-analysis of 30 years of research. *Educational Research Review*, 22, 19–33. <https://doi.org/10.1016/j.edurev.2017.07.001>

- Zainuddin, Z., Haruna, H., Li, X., Zhang, Y., & Chu, S. K. W. (2019). A systematic review of flipped classroom empirical evidence from different fields: what are the gaps and future trends? *On the Horizon*, 27(2), 72–86. <https://doi.org/10.1108/OTH-09-2018-0027>
- Zhang, S. (2018). A systematic review and meta-analysis on flipped learning in science education [Master's thesis, The University of Hong Kong]. The HKU Scholars Hub. <http://hub.hku.hk/handle/10722/265828>
- 

**Corresponding author:** Khe Foon Hew, [kfhew@hku.hk](mailto:kfhew@hku.hk)

**Copyright:** Articles published in the *Australasian Journal of Educational Technology* (AJET) are available under Creative Commons Attribution Non-Commercial No Derivatives Licence (CC BY-NC-ND 4.0). Authors retain copyright in their work and grant AJET right of first publication under CC BY-NC-ND 4.0.

**Please cite as:** Hew, K. F., Bai, S., Huang, W., Dawson, P., Du, J., Huang, G., Jia, C., & Thankrit, K. (2021). On the use of flipped classroom across various disciplines: Insights from a second-order meta-analysis. *Australasian Journal of Educational Technology*, 37(2), 132–151. <https://doi.org/10.14742/ajet.6475>