



OPEN ACCESS

EDITED BY

Gavin T. L. Brown,
The University of Auckland,
New Zealand

REVIEWED BY

Katrien Struyven,
University of Hasselt, Belgium
Chung Kwan Lo,
The Education University
of Hong Kong, Hong Kong SAR, China

*CORRESPONDENCE

Manu Kapur
manukapur@ethz.ch

SPECIALTY SECTION

This article was submitted to
Higher Education,
a section of the journal
Frontiers in Education

RECEIVED 30 May 2022

ACCEPTED 23 August 2022

PUBLISHED 26 September 2022

CITATION

Kapur M, Hattie J, Grossman I and
Sinha T (2022) Fail, flip, fix, and feed –
Rethinking flipped learning: A review
of meta-analyses and a subsequent
meta-analysis.
Front. Educ. 7:956416.
doi: 10.3389/educ.2022.956416

COPYRIGHT

© 2022 Kapur, Hattie, Grossman and
Sinha. This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s)
are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

Fail, flip, fix, and feed – Rethinking flipped learning: A review of meta-analyses and a subsequent meta-analysis

Manu Kapur^{1*}, John Hattie², Irina Grossman² and
Tanmay Sinha¹

¹Professorship for Learning Sciences and Higher Education, ETH Zürich, Zürich, Switzerland,
²Melbourne Graduate School of Education, University of Melbourne, Parkville, VIC, Australia

The current levels of enthusiasm for flipped learning are not commensurate with and far exceed the vast variability of scientific evidence in its favor. We examined 46 meta-analyses only to find remarkably different overall effects, raising the question about possible moderators and confounds, showing the need to control for the nature of the intervention. We then conducted a meta-analysis of 173 studies specifically coding the nature of the flipped implementation. In contrast to many claims, most in-class sessions are not modified based on the flipped implementation. Furthermore, it was flipping followed by a more traditional class and not active learning that was more effective. Drawing on related research, we proposed a more specific model for flipping, “Fail, Flip, Fix, and Feed” whereby students are asked to first engage in generating solutions to novel problems even if they fail to generate the correct solutions, before receiving instructions.

KEYWORDS

flipped learning, productive failure, meta-analysis, higher education, active learning

Introduction

Flipped learning is an instructional method that has gained substantial interest and traction among educators and policymakers worldwide. The Covid-19 pandemic will likely accelerate this trend. One of the first to use the term, [Bergmann and Sams \(2012\)](#) defined flipped learning as a teaching method in which “that which is traditionally done in-class is now done at home and that which is traditionally done as homework is now completed in-class” (p. 13). Conceived as a two-phase model, the first phase of flipped learning involves getting students to learn basic content online and prior to class. The second phase then allows teachers to make use of the freed-up in-class time to clarify

students' understandings of the concept and design learning strategies that will enable students to engage deeply with the targeted concept.

It is important to note that the traditional and flipped learning methods share the same two-phase pedagogical sequence: first instruction on the basic content, followed by problem-solving practice and elaboration. The underlying logic is, that when teaching a new concept, it is best to give students instruction on the content first followed by problem-solving, elaboration, and mastery of the content (Kirschner et al., 2006). In the traditional method, the instruction and practice phases are both carried out during in-class time, and homework is provided for after-class time.

Proponents of flipped learning argue that because the instruction phase is usually passive in nature and is not the best use of in-class time, which could be better used for active learning. By passive, they refer to activities such as watching and listening to a face-to-face or an online video lecture with few opportunities for deeper engagement. By active, they refer to activities that afford deeper engagement in the learning process, such as problem-solving, class discussion, dialog and debates, student presentations, collaboration, labs, games, and interactive and simulation-based learning under the guidance of a teacher (Chen, 2016; Tan et al., 2017; Hu et al., 2019; Karagöl and Esen, 2019).

Therefore, without changing the two-phase pedagogical sequence of instruction followed by practice, flipped learning moves typically from a passive, in-class, face-to-face lecture component to a pre-class, online lecture, thereby making more room for active learning during the in-class time. It is precisely because flipped learning allows for more active learning during the in-class time that proponents of flipped learning claim that it leads to better academic outcomes, such as higher grades and better test and examination scores, than the traditional method (Yarbro et al., 2014; Lag and Saele, 2019; Orhan, 2019). Examining this claim is precisely the aim of our paper.

We will show that the basis of this claim is weak. We start with a review of existing reviews and meta-analyses, of which there are many. As this initial review will show, there is a large variance in the effects, in part because of the nature of pre-class and in-class activities (as defined earlier). Such variability necessitates closer attention to the nature of passive and active learning activities in flipped versus traditional learning methods. Therefore, we use this initial review to identify a set of moderators and confounds that may help explain the large variance. We then report findings from our meta-analyses that code for the identified moderators and confounds to explain the large variance in the effects of flipped learning on learning outcomes. Finally, we use the findings from our meta-analysis to derive an alternative model for flipped learning that helps (a) students acquire an understanding of what they do and do not know before in-class interaction, (b) identify foci from students in these pre-class experiences to tailor the in-class

teaching, and thus (c) teach in a way to enhance learning outcomes for students.

Review of existing quantitative meta-analyses

To date, we located at least 46 meta-analyses based on up to 2,476 studies. Across these meta-analyses, we found 765 references (three did not provide a list of references) and 471 (62%) were unique [see also Hew et al. (2021) for key country-specific moderated effects]. Not all meta-analyses reported the total sample size. Of the 19 that did, the total sample was 178,848. Using the average of these 19 ($N = 9,413$), we estimate about 451,827 students were involved overall (but there are approximately 40% overlapping articles, so the best estimate of sample size is between 100,000 and 410,674 students).

The average effect of 0.69 (range 0.19–2.29) and the standard error of 0.12 is substantial (Table 1). This large variance suggests that any average effect is of little value and efforts should be made to identify moderators and confounds that may help explain this large variance.

Furthermore, these meta-analyses focused on synthesizing effects across (a) different student populations in flipped classrooms (i.e., 3 in Elementary, 2 in High, 2 in K-12, 19 in College, and 23 across all levels of schooling), (b) used different meta-analytic methods (e.g., 40 used random-effect models for pooling with a higher average effect, $g = 0.72$, while the remaining eight used fixed-effect models with an average of $g = 0.50$), and (c) did not specifically code for the nature of the flipped learning intervention (pre and within the class) to account for moderators and confounds (especially the nature of the pre- and in-class activities and dosage). These factors make causal attribution to flipped learning difficult, limiting the value of conclusions based on these previous meta-analyses.

Identification of moderators and confounds

A review of the 46 meta-analyses (see Table 1) indicated the presence of major moderators and confounds that needed to be controlled. We reviewed these meta-analyses to identify (a) moderators between studies that may contribute to the high variance, and (b) confounds within each study, pointing to a more fundamental problem in interpreting the effects within a study. We distilled seven moderators and three confounds.

Moderators

Instructional domain

The majority of the studies were conducted in STEM and Medicine related areas ($g = 0.38$), and the

remainder were from the Humanities and Social Sciences ($g = 0.57$).

than in elementary schools ($g = 0.40$), high schools ($g = 0.63$), or K-12 ($g = 0.55$).

Students' education level

The majority of the studies were conducted at the university level. The effects were highest at the university level ($g = 0.93$)

Culture and educational system

The largest effect sizes all came from studies in non-Western countries (Asian, $g = 0.75$, Turkey and Iran $g = 0.79$), and the

TABLE 1 Summary information from 46 meta-analyses on flipped learning.

References	No. studies	No. people	No. effects	<i>d</i>	SE	Model	Level	Domain	1st author country
Algarni, 2018	34	8,598	36	0.27	0.02	Random	All	Math	United Kingdom
Aydin et al., 2020	25		25	0.71	0.06	Random	All	All	Turkey
Bong-Seok, 2018	29		29	0.56	0.38*	Random	Elem	All	Korea
Bredow et al., 2021	282	51,437	282	0.39	0.02	Random	College	All	United States
Chen et al., 2018	46	9,026	46	0.54	0.14	Random	All	All	Taiwan
Cheng et al., 2019	55	7,912	115	0.19	0.04	Fixed	All	All	United States
Cho and Lee, 2018	95		95	0.54	0.21*	Random	All	All	Korea
Deshen and Yu, 2021	28		28	0.49	0.38*	Random	College	All	Japan
Doğan et al., 2021	30		30	0.73	0.38*	Random	All	Science	Turkey
Farmus et al., 2020	10		10	0.43	0.10	Random	College	Statistics	Canada
Ge et al., 2020	19	2,114	19	1.86	0.50	Random	College	Radiology	China
Gillette et al., 2018	6	315	6	0.35	0.82*	Random	All	All	United States
Hew and Lo, 2018	28	2,295	28	0.33	0.06	Random	College	Health	Hong Kong
Hu et al., 2018	11	1,180	11	1.19	0.48	Random	College	Pharmacy	China
Kim and Lim, 2021	21		21	0.66	0.09	Random	College	All	Korea
Jang, 2019	29		292	0.56	0.06	Random	Elem	All	Korea
Jang and Kim, 2020	43		153	0.24	0.01	Random	Elem	All	Korea
Kang and Kang, 2021	23		23	1.21	0.20	Random	All	Nursing	Korea
Kang and Shin, 2005	36		288	0.54	0.12*	Fixed	All	All	Korea
Karagöl and Esen, 2019	55		80	0.57	0.07	Random	All	All	Turkey
Lag and Saele, 2019	271		272	0.35	0.05	Random	All	All	Norway
Li et al., 2020	32		32	1.46	0.17	Random	College	Nursing	China
Liu et al., 2018	12	1,440	12	1.18	0.24	Random	College	Nursing	China
Lo et al., 2017	21		21	0.30	0.07	Random	All	Math	Hong Kong
Sola Martínez et al., 2019	12	3,326	12	2.29	0.24	Random	College	All	Spain
Ming, 2017	33		33	0.37	0.35*	Fixed	College	All	China
Orhan, 2019	13		13	0.74	0.08	Fixed	All	All	Turkey
Ralević and Tomašević, 2021	22		22	0.97	0.21	Fixed	All	All	Serbia
Shahnama et al., 2021	69		69	1.24	0.10	Random	All	Second language	Iran
Shi et al., 2020	33	6,957	33	0.53	0.17*	Random	College	All	China
Sparkes, 2019	114		125	0.30	0.18*	Random	College	All	Canada
Strelan et al., 2020	174	33,678	198	0.50	0.04	Random	All	All	Australia
Tan et al., 2017	29	3,694	29	1.13	0.40*	Random	College	Nursing	China
Turan, 2021	18	2845	18	0.63	0.14	Random	All	Science	Turkey
Tutal, 2021	177	17,807	177	0.76	0.16*	Random	All	All	Turkey
van Alten et al., 2019	114		115	0.36	0.07	Random	College	All	Netherlands
Vitta and Al-Hoorie, 2020	56	4,220	61	0.58	0.10	Random	All	Second language	Japan
Wagner et al., 2021	25	2,323	44	0.42	0.11	Random	High	All	Germany
Xu et al., 2019	22	4,295	22	1.79	0.12	Random	College	Nursing	China
Yakar, 2021	45		46	0.51	0.09	Random	All	Math	Turkey
Yoon, 2018	26		26	0.83	0.09	Fixed	High	All	Korea
Zhang, 2018	28		28	0.42	0.07	Fixed	All	Science	Hong Kong
Zhang et al., 2021	20		28	0.66	0.08	Random	College	All	China
Zheng et al., 2020	95	15,386	95	0.44	0.07	Fixed	All	All	China
Zhu, 2021	27		53	0.54	0.10	Random	K-12	All	China
Zhu et al., 2019	25		25	0.56	0.08	Random	K-12	All	China

*SE indicates estimated standard error.

lowest from Western countries ($g = 0.53$; each meta-analysis includes studies across cultures).

Sample size

The sample size of many studies was relatively small. Thirty of the 46 meta-analyses have fewer than 50 studies, and 22 fewer than 30. This points to low power in many meta-analyses, and selective bias in the choice of articles, resulting in major variability in the findings (Hedges and Pigott, 2001).

Intervention length

The effect of intervention length was not consistent. On the one hand, Karagöl and Esen (2019) reported higher effects for shorter than longer interventions (1–4 weeks: $g = 0.69$, 5–8 weeks: $g = 0.58$, 9+ weeks: $g = 0.41$). On the other hand, Cheng et al. (2019) found higher effects for at least a semester ($g = 1.15$) compared to shorter than one semester ($g = 0.35$).

Quality of meta-analysis

Although there is a rich literature on assessing the quality of individual studies (Sipe and Curlette, 1996; Liberati et al., 2009; Higgins and Green, 2011; Higgins, 2018), there is far less on the quality of meta-analyses. One proxy of quality is the impact factor (IF) of the journal in which the meta-analysis is published. The IF of a journal is calculated over 2 years by dividing the number of times articles from the journal were cited by the number of citable articles. For example, if the articles in a journal over 2 years were cited 100 times, and there were 50 articles in the journal in that period, then the $IF = 100/50 = 2$. The IF for 87 of the 115 unique journals was available, resulting in IF for 113 of the studies. The 2020 IF was used and thus relates to the citations from 2018 to 2019. The IF can be seen as a rough proxy for the spread of ideas and the quality of the journal (although there is much debate on this issue; Harzing, 2010).

Nature of pre-class and in-class activities

One of the major arguments in flipped learning is that there should be active learning in-class to complement the primarily passive pre-class instruction. But how this was implemented was far from uniform. As we have noted earlier, even within active learning strategies, there were many variations, including collaborative learning, Internet blogs, case-based and application problems, simulations, interactive demonstrations, and student presentations. For the in-class sessions, some introduced the use of clickers or similar student response methods, laboratories, problem sets, think-pair-share, group work, case studies, and going over pre-assessment activities. Clearly, assuming that all flipped learning implementations were similar is problematic. We needed a more detailed understanding of the nature of the activities in the flipped and control conditions. Therefore, we developed a coding scheme (see the “Methods” section) to examine the nature (active vs. passive) of the various pre-class and in-class activities in the flipped and control conditions. The coding scheme aimed to

characterize the extent to which the pre- and in-class were passive and/or active in a more nuanced way.

Confounds

Extra instructional time

Because of the addition of pre-class on top of in-class time, the total instructional time in flipped learning is often greater than in the traditional method alone, which means that the effects on learning may simply be a function of students spending more time on the learning material [McLaughlin et al. (2014) estimated an increase of 127% more time to develop and deliver a flipped compared to a lecture course]. Indeed, as we report later, more than half of the studies in our meta-analysis gave more time to flipped learning conditions. Proponents of flipped learning may well see it as a good thing that it gets students to spend more time learning the content. However, if instructional time can potentially explain the effects, then might not we achieve these effects with a modest increase in instruction time for the traditional method? Hew et al. (2021) also noted the additional time (127%) needed to develop and manage a flipped course and 57% more time to maintain than a lecture course. Also, students noted that they needed more time, and only about 30–40% of students completed the pre-class work.

Formative assessment and feedback

This is potentially a confound and a moderating factor. The claims for flipped learning often recommend that, after the pre-class activity, students should engage in formative assessment and feedback activities. Students in the traditional method alone typically do not get such formative assessment and feedback after their in-class lectures. Hew and Lo (2018) found the availability of a quiz at the start of the in-class led to an effect of $g = 0.56$ compared to not having a quiz of $g = 0.42$. They concluded that it helped when instructors identified students’ possible misconceptions of the pre-class materials but did not indicate whether they then dealt with these misconceptions. Lag and Saele (2019) also found that a test of preparation led to higher effects ($g = 0.31$ – 0.40) from flipped learning. Given the well-established effects of formative evaluation and feedback on learning (Hattie and Timperley, 2007; Shute, 2008), one must wonder if the effects on learning could well be achieved with the inclusion of such activities in the traditional method itself.

Instructor consistency

A further confound was having different instructors for the flipped and control groups. As teacher quality is among the most critical attributes to successful student learning, this non-comparability of teachers could be a major confound.

Research questions

This review aims to resolve some of these anomalies to answer the following research questions to critically

investigate conditions under which the effects of flipped learning may be realized.

a) RQ1: What is the overall effect of flipped learning over traditional instruction?

b) RQ2: How does publication bias impact the overall effect?

c) RQ3a: How do moderators between studies impact the overall effect? Among the various moderators we identify, we are particularly interested in the nature of the intervention itself. We first characterize the nature (active vs. passive) of flipped learning activities (pre- and in-class) and then examine how the nature of the activities affects student learning in flipped classrooms.

d) RQ3b: How does the presence of confounds within studies impact overall effects?

Methods

A search was conducted for peer-reviewed studies on flipped learning in 26 research databases using a comprehensive set of search terms and reviewing the studies included in [Table 1](#). Relevant quasi-experimental studies examining the learning outcomes of flipped learning were located through the literature published through to and including 2019. We restricted our search until 2019 since articles from the post-COVID era likely require their own focused meta-analysis research as classrooms have undergone radical transformations globally. The major source of the search was *via* educational research databases, and to ensure that a good coverage of studies, 28 electronic research databases were relied upon, and these included the *Academic Search Premier*, *British Education Index*, *Business Source Premier*, *Communication and Mass Media Complete*, *Computer Source*, *eBook Collection (EBSCOhost)*, *EconLit with Full Text*, *Education Source*, *ERIC*, *Google Scholar*, *GreenFILE*, *Hospitality and Tourism Complete*, *Index to Legal Periodicals and Books Full Text (H. W. Wilson)*, *Information Science and Technology Abstracts*, *International Bibliography of Theater and Dance with Full Text*, *Library Literature and Information Science Full Text (H. W. Wilson)*, *Library, Information Science and Technology Abstracts*, *MAS Ultra—School Edition*, *MathSciNet via EBSCOhost*, *MEDLINE*, *MLA Directory of Periodicals*, *MLA International Bibliography*, *Philosopher's Index*, *ProQuest*, *PsycARTICLES*, *PsycCRITIQUES*, *PsycINFO*, *Regional Business News*, *RILM Abstracts of Music Literature (1967 to Present only)*, *SPORTDiscus with Full Text*, *Teacher Reference Center*, and *Web of Science*. The search terms used included “Flipped classroom,” “Flipped instruction.”

“Inverted classroom,” “Reversed instruction,” “Blended learning AND video lecture,” “Blended learning AND web lecture,” combinations of the terms “Video Lecture,” “Web lecture,” “Online lecture” and “Active learning.” To restrict the results produced in Google Scholar to controlled studies with quantifiable outcome measures, the search terms used were “Flipped classroom” AND traditional AND “standard deviation” AND mean AND “course grade” OR examination OR exam AND “same instructor.”

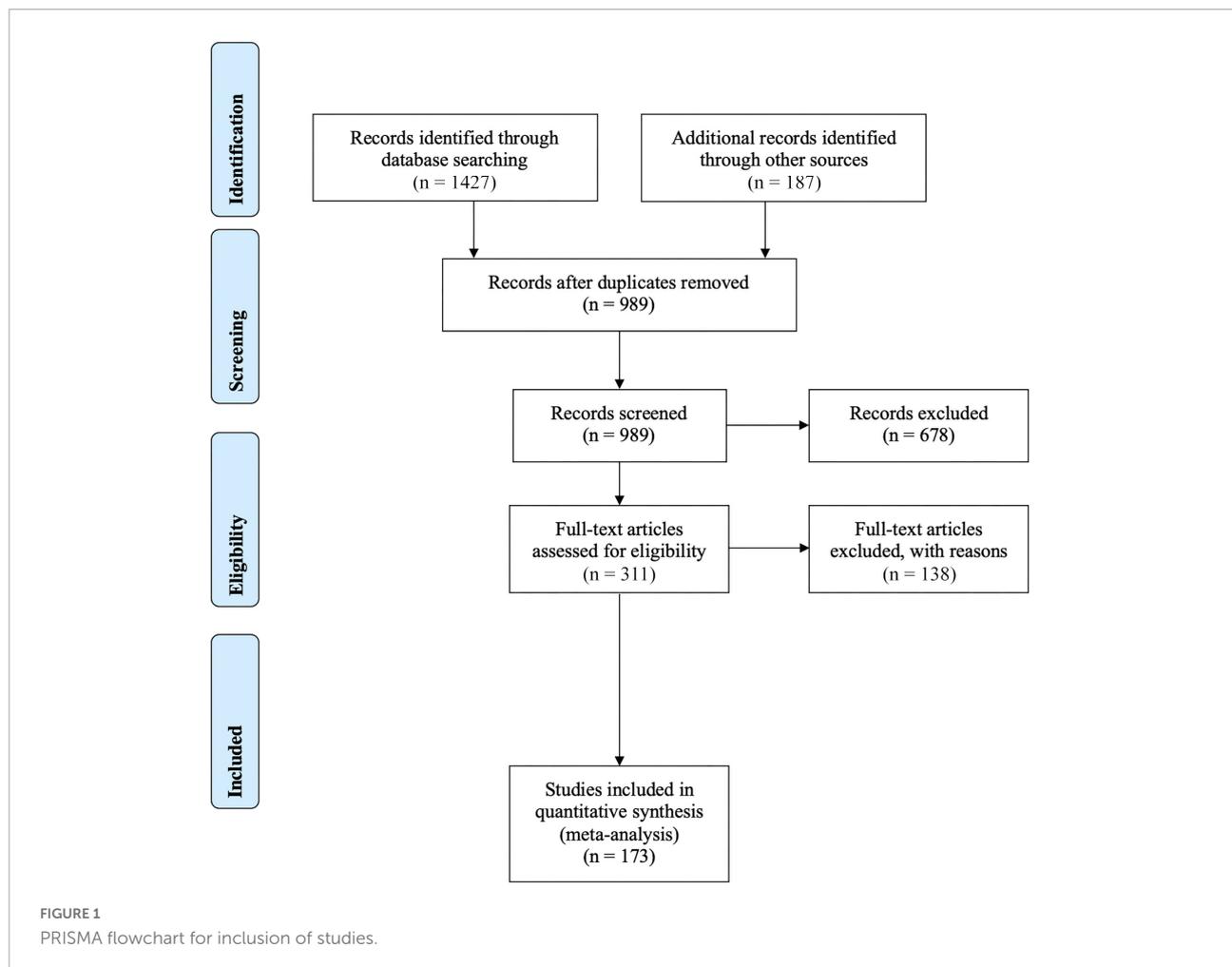
From an initial list of 1,477 English articles that came from peer-reviewed journals, we shortlisted 311 unique studies that were to be further screened for relevance. The criteria for inclusion in this review include (a) being published in a peer-reviewed journal, (b) having a quasi- or controlled-experimental design comparing flipped learning with a traditionally taught counterpart, and (c) examining student learning outcomes, with adequate information about statistical data, procedures, and inference. Studies that focused only on subjective student perceptions were excluded because they do not necessarily correlate with learning ([Reich, 2015](#)). Given these criteria, 173 studies were shortlisted as meeting all criteria. [Figure 1](#) presents the corresponding PRISMA flowchart ([Moher et al., 2009](#)). Details of the shortlisted studies that met the criteria can be found in [Supplementary Table 1](#).

The nature of the pre-class and in-class sessions was coded. The nature of the pre-class included readings, video + quiz, video + PowerPoint, video +, recordings, and online lectures. The in-class sessions included lectures, labs, problem sets, problem-based methods, debates, Socratic questioning, use of clickers, class discussions, role plays, case studies, group work, reviewing assignments, and student presentations. The various forms of assessment included: assessments that led to pre-class modifications, post-lecture quizzes, in-class post-lecture quizzes, and other quizzes throughout the classes. [Table 2](#) provides more explanation.

The three authors with expertise in flipped learning, meta-analysis, or both coded all studies. Discrepancies were resolved collectively. To evaluate the reliability of the coding, an external independent coder independently coded all variables. Cohen's kappa was calculated for each variable from this independent and our coding. [Cohen \(1960\)](#) recommended that values > 0.20 were fair, > 0.4 moderate, > 0.6 substantial, and > 0.8 almost perfect agreement. The average kappa for the background variables = 0.54, for the control implementation variables = 0.40, for the flipped implementation variables = 0.58, and overall variables = 0.51. These are sufficiently high to have confidence in the coding.

Meta-analytic procedures and statistical analyses

Given the wide range of research participants, subject areas, flipped learning interventions, and study measures, each



outcome is unlikely to represent an approximation of a single true effect size. Thus, we utilized a random effects model (Borenstein et al., 2011). Analyses were conducted using the *Letan* package of R. Unlike the fixed model, where it is assumed that all the included studies share one true effect size, the random model allows that the true effect could vary from study to study. The studies included in the meta-analysis are assumed to be a random sample of the relevant distribution of effects, and the combined effect estimates the mean effect in this distribution. Because the weights assigned to each study are more balanced, large studies are less likely to dominate the analysis, and small studies are less likely to be trivialized. In all cases, Hedges *g* was calculated for each comparison of flipped to control groups.

Results

The inclusion criteria were met by 173 articles, which reported results from 192 independent samples with 532 effect sizes obtained from 43,278 participants (45% in flipped groups,

55% in controls). All included articles were published between 2006 and 2019 with a median publication year of 2016. The coded effect sizes ranged from -3.85 to 3.45 . The sample sizes ranged from 13 to 4,283 with a mean of 39, which resulted in low statistical power, and the average effects may likely be somewhat overestimated.

The majority of students were university-based (90%), and the other 10% were from mainly secondary schools with one primary school. The academic domains include Science (27.4%), Engineering (15.9%), Medicine (13.3%), Humanities (12.6%), Mathematics (12.2%), Business (7.5%), Computing (4.5%), Nursing (4.1%), and Education (2.4%). Thus, flipped learning has been most studied in the Sciences domain (73%). The typical research design was a multiple group design (flipped, control = 74%, and the remainder pre- and post-group design). The typical length was less than a month (18.8%), semester (77%), or year-long (4.7%). The instructor(s) were the same in 84.1% of the studies, and about half of the studies (50.3%) gave the students extra time in the flipped condition, and 35.1% provided feedback from the activities that were part of the flipped model.

Sample mean ages ranged from 11 to 42 ($M = 18.5$, $SD = 3.86$). Of the included effect sizes, 3% were from primary school, 42% from secondary school, and 55% from higher education. The mean percentage of females in the included samples was 55.4%. In most studies, the predominant ethnicity

was White/Caucasian (30%). However, 66% of the included studies did not report the ethnicity of their sample. 72% of studies were based on samples educated in North America, 11.1% from Asia, 10.6% from Europe, 2.8% from North and Central America (Canada, Trinidad), 1.6% from Africa, and 1.1% from Australia.

Nearly all (93%) of the measures of academic achievement were mid-term or final quizzes or examinations. The remainder included clinical evaluations, group projects, homework exercises, lab grades, clinical rotation measures, or quality of treatment plans.

TABLE 2 Nature of pre- and in-class activities, description, and two example studies.

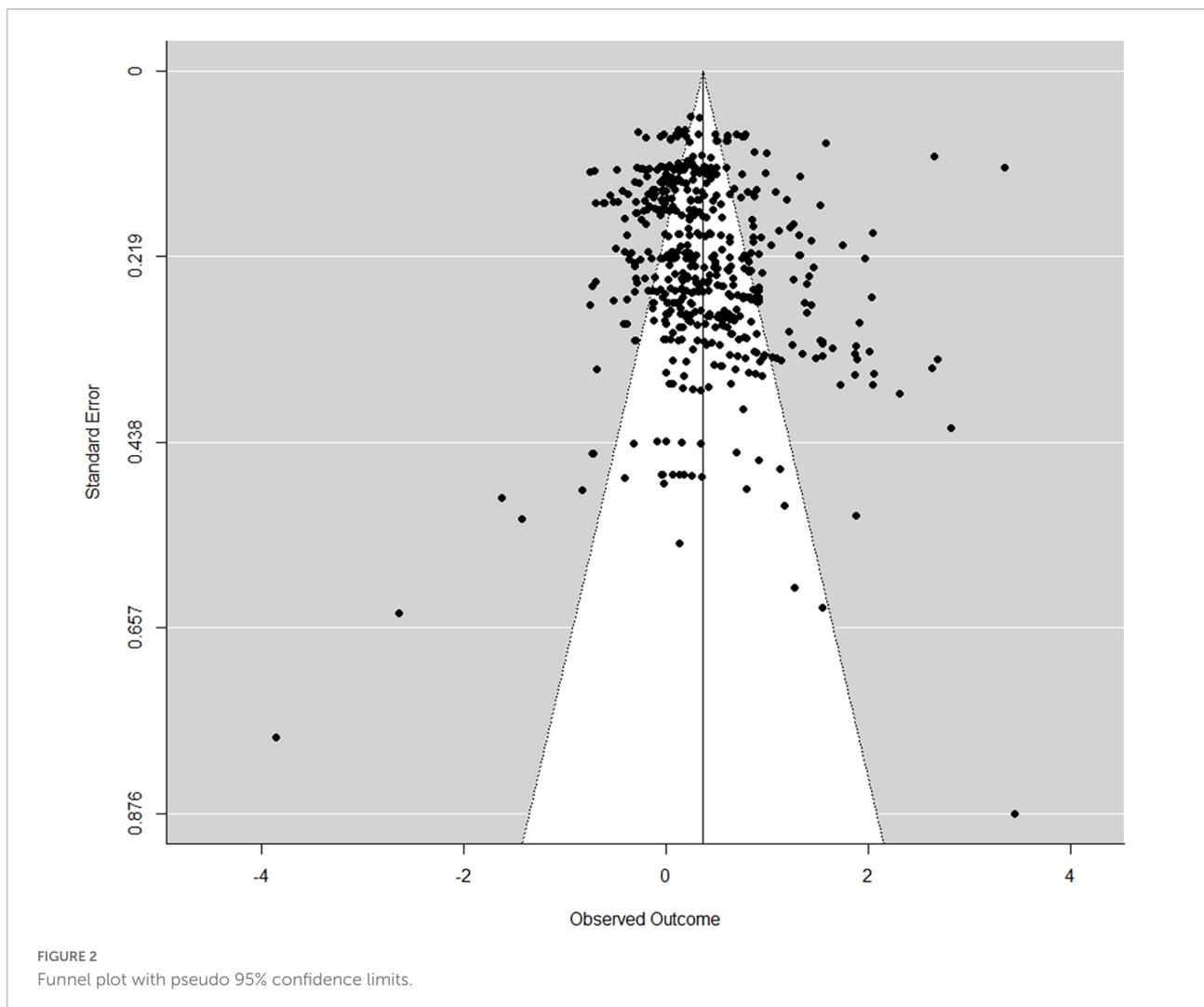
	Description	Example study
Pre-class activities		
Readings	Provision of articles, text books, and written materials	Ferreri and O'Connor, 2013; Nielsen et al., 2018
Video	Provision of a video	Harrington et al., 2015; Hung, 2015
Video + quiz	Provision of a video that includes a quiz	Murphy et al., 2016; Eichler and Peeples, 2016
Video + PowerPoint	Provision of a video and a PowerPoint	Lewis and Harrison, 2012; Moffett and Mill, 2014
Video + readings	Provision of video and readings	Flynn, 2015; Tang et al., 2017
Video +	Video, LMS, extra material	Aidinopoulou and Sampson, 2017; Blázquez et al., 2019
Online lectures	Provision of an online lecture	Guerrero et al., 2015; Kiviniemi, 2014
Online modules	Provision of modules relating to the in-class to view or complete	Bonnes et al., 2017; Sezer, 2017
Assignments/exams	Provision of assignment or exam to complete before class	Hudson et al., 2015; Jensen et al., 2015
PowerPoints	Provision of PowerPoint	Lewis and Harrison, 2012; Oki, 2016
In-class activities		
Lectures	In-class lecture	He et al., 2016; Moffett and Mill, 2014
Laboratories	Laboratories or practicals	Kakosimos, 2015; Hung, 2015
Demonstrations	Demonstration of experiments	Beale et al., 2014; Krueger and Storlie, 2015
Problem-based activities	Involving students in problem-based learning	Lewis and Harrison, 2012; Boyraz and Ocak, 2017
Use of clickers	Using clickers to gather students' reactions and understanding of the lecture	Kostaris et al., 2017; Kennedy et al., 2015
Problem-solving	Engaging in solving problems in class	O'Connor et al., 2016; Lee and Liu, 2016
Group work	Conducting group work in class	Boyraz and Ocak, 2017; Lewis and Harrison, 2012
Review assignments	Reviewing pre-set assignments	Gundlach et al., 2015; Whitman Cobb, 2016
In-class quizzes	Conducting in-class quizzes	González-Gómez et al., 2016; Jensen, 2011
Case studies	Engaging students in case studies	Gundlach et al., 2015; Boysen-Osborn et al., 2016
Student presentations	Having students present sessions to the class	O'Connor et al., 2016; Wasserman et al., 2017

Overall effect (RQ1)

The overall mean effect size was $g = 0.37$ ($SE = 0.025$) with a 95% confidence interval ranging from 0.32 to 0.41 (and the uncorrected effect was 0.38). When the multiple effects within each study were averaged for a per study effect, the mean was $g = 0.41$ ($SE = 0.039$) with a 95% confidence interval from 0.33 to 0.48. A three-level meta-analysis, accounting for nesting multiple effects within studies, produced a similar result ($g = 0.41$, $SE = 0.037$). The Q-statistics showed that there is significant heterogeneity among the effect sizes, and thus the flipped interventions did not share the same true effect size ($Q = 5222.22$, $df = 531$, $p < 0.001$). The percentage of variance across studies due to heterogeneity rather than chance is very large ($I^2 = 91.27\%$, $s^2 = 11.45$), indicating substantial heterogeneity, implying that the relationship between academic achievement and flipped learning is moderated by important variables.

Publication bias (RQ2)

An important concern is to detect and mitigate the risks of publication bias and be confident that there do not exist unpublished papers that, if found and included, would alter the overall effect size. We inspected a funnel plot, the trim and fill approach (Borenstein et al., 2017), and the Egger's regression test (Egger et al., 1997). They all examine the distribution of effect size estimates relative to standard error, assessing whether there is symmetry. Figure 2 presents the funnel plot showing the relationship between the standard deviation and the effect size. In funnel plots, the larger studies appear toward the top of the graph and cluster near the mean effect size. The smaller studies appear toward the bottom of the graph and (since there is more sampling variation in effect size estimates in the smaller studies) can be seen to be dispersed across a range of values. If there is no publication bias, the studies will be distributed symmetrically about the combined effect size. As can be seen, as desired, the effects of the studies with larger standard errors (typically the smaller studies) scatter more widely at the bottom,



with the spread narrowing among larger studies. The majority of effects are within the funnel, with a slight bias toward larger than anticipated effects, suggesting possible an overestimation of the overall effects.

To further check this interpretation, we carried out the Egger's test for funnel plot asymmetry (Egger et al., 1997), testing the null hypothesis that symmetry in the funnel plot exists. The Egger test failed to find a non-significant intercept near 0, thus indicating the presence of publication bias on effect size level ($z = 3.41$, $p < 0.001$). The correlation between the sample size within each study with the overall effect size was $r = -0.11$; indicating that smaller size studies were more related to higher effects, and it is noted that there is a preponderance of studies with small sample sizes (almost 50% of the studies had $N < 100$). There need to be at least 6,424 unpublished studies with a mean of 0; however, to overturn the claim that flipped learning has a positive impact on student learning. This is in line with the findings in the meta-analysis by Lag and Saele (2019), who also found that smaller studies often produced higher

effect sizes which may cause the benefit of flipped instruction to be overestimated. Once smaller, underpowered studies were excluded, flipped instruction was still found to be beneficial; however, the benefit was slightly more modest (Lag and Saele, 2019).

Moderators (RQ3a)

There were similar effects across instructional domains: in the non-sciences ($g = 0.44$) and lowest in the science domains ($g = 0.37$; Table 3). The few effects from school-age children ($g = 0.68$, $N = 53$) were almost twice the effect from university students ($g = 0.35$, $N = 480$). The effects were more than doubled in the developing countries ($g = 0.81$) when compared with the developed countries ($g = 0.40$, Table 4). Shorter interventions had higher effects than longer ones: short (less than 1 month, $g = 0.54$, $SE = 0.071$, $N = 99$), a semester ($g = 0.33$, $SE = 0.032$, $N = 408$) or a year-long ($g = 0.39$, $SE = 0.084$, $N = 25$). There

TABLE 3 Hedges g effect-size moderated by domain.

Domain	Hedges g	SE	Lower CI	Higher CI	No. effects
Sciences					
Computing	0.81	0.16	0.49	1.13	24
Engineering	0.25	0.10	0.06	0.45	85
Math	0.26	0.067	0.13	0.39	65
Medicine	0.27	0.063	0.14	0.40	71
Nursing	0.25	0.12	0.00	0.50	22
Science	0.39	0.05	0.30	0.48	146
Non-Sciences					
Business	0.37	0.070	0.23	0.51	40
Education	0.32	0.091	0.13	0.50	13
Humanities	0.63	0.080	0.47	0.79	66

$Q_{Between} = 251.25, df = 9, p < 0.001$.

were very low correlations between the effect-size and the year of publication ($r = -0.05, df = 530, p = 0.190$), the number in the flipped ($r = -0.05, df = 530, p = 0.277$), control ($r = -0.01, df = 530, p = 0.774$), and the total sample size ($r = 0.08, df = 530, p = 0.541$); although as noted above, the sample sizes in general were relatively small. The effects from pre- and post-tests ($g = 0.41, SE = 0.071, N = 137$) were similar to those from multiple group designs ($g = 0.37, SE = 0.029, N = 396$).

For 113 of the studies, it was possible to locate the impact factor of the journal in which the study was published. There was no correlation between the study effect-size and the impact factor (average IF = 2.13, $r = -0.01, df = 87, p = 0.947$). However, there needs to be some care in interpreting, as journals that are more cited do not necessarily and always publish the highest quality studies (Munafò and Flint, 2010; Fraley and Vazire, 2014; Szucs and Ioannidis, 2017).

Next, we turn to the main moderator of interest: the nature of flipped and traditional instruction activities. To better understand the overall effect and the variability, we used our coding scheme to characterize the nature (active vs. passive) of flipped learning activities (pre- and in-class). This allowed us to examine the claim whether flipped learning is indeed implemented as theoretically claimed to be, that is, a passive learning pre-class followed by active learning in-class. Following that, we examined how the nature of such activities affects student learning in flipped classrooms.

Not surprisingly, most tasks in the pre-class part were passive, with little involvement of students in actively interrogating the ideas via quizzes, problem questions, or opportunities to express what they do and do not understand, alone or with their peers. According to our coding scheme in Table 2, passive tasks included those involving students reading and watching videos, or presentations. In contrast, active tasks involved students engaging in dialog with others including the lecturer, practical activities, use of response clickers, group

TABLE 4 Hedges g effect-size moderated by country.

Country	Hedges g	SE	Lower CI	Higher CI	No. effects
Developing Countries					
Iran	1.53	0.20	1.13	1.92	7
India	1.44				1
Nigeria	1.38	0.40	0.58	2.19	3
Malaysia	1.06	0.62	-0.17	2.29	4
Turkey	0.77	0.16	0.45	1.10	11
Qatar	0.36	0.14	0.08	0.64	4
Serbia	0.36	0.080	0.20	0.52	2
Kuwait	0.34	0.048	0.24	0.44	4
Brazil	0.32	0.15	0.01	0.62	6
Saudi Arabia	0.07	0.050	-0.03	0.17	2
Trinidad and Tobago	0.02				1
All Developing countries	0.70	0.21	0.28	1.10	45
Developed Countries					
Cyprus	1.87				1
Spain	0.97	0.20	0.58	1.37	4
Norway	0.93	0.08	0.77	1.09	3
Finland	0.75	0.19	0.38	1.12	8
China	0.74	0.32	0.10	1.37	6
Taiwan	0.70	0.15	0.39	1.00	15
Greece	0.63	0.12	0.39	0.87	20
Canada	0.34	0.07	0.19	0.48	7
United States	0.30	0.031	0.24	0.37	407
Korea	0.22	0.072	0.08	0.36	4
Australia	-0.02	0.13	-0.27	0.23	5
United Kingdom	-0.45	0.09	-0.63	-0.27	7
All Developed countries	0.33	0.13	0.20	0.73	487

$Q_{Between} = 362.857, df = 23, p < 0.001$. Country classifications are from the International Monetary Fund (2018).

work, case studies, problem-based group tasks, field trips, socratic questioning, and student presentations.

Table 5 focuses on the nature of the tasks in the pre-class and in the in-class part of the flipped classroom. The table lists the number of studies that included various activities in the pre-class and in-class parts of the lesson, and the corresponding mean effect size from those studies that included this activity. For example, the row for “Readings” indicates that eight studies involved a pre-class structure involving readings. The mean of 0.75 and this task had the largest effect of all pre-tasks; followed by combining watching a video and a video quiz or watching a video, PowerPoint, or readings. The lowest pre-class activity was watching a PowerPoint, completing an assignment, or watching a video. The largest effect in-class included demonstrations, problem-based methods, and the use of response clickers, and the lowest were student presentations, case studies, and in-class quizzes. The combination of activities pre-class seems to matter, and this could merely be amplifying the time-on-task that students engage with the material. Indeed, this can be

viewed as a positive if it also allows students to appreciate better the gaps in their knowledge and misunderstandings (which is a claim in favor of flipping).

There is similarly a high level of passive learning during the in-class. In all cases, the in-class included a lecture. There were higher effects for the more active components (demonstration, problem-based, use of clicker, problem-solving, group work, practical laboratory) than the more passive (reviewing assignments, in-class quizzes, and class studies). However, there were no differences related to the presence ($g = 0.42$, $SE = 0.036$, $N = 271$) or absence ($g = 0.33$, $SE = 0.044$, $N = 421$) of active learning in the in-class sessions ($t = 4.11$, $df = 690$, $p < 0.001$).

Next, we evaluated the impact of the type of traditional classroom and the type of flipped classroom on the overall effect size. We created a classification system for the flipped compared to the control groups' class activities. There was more than one activity in many cases, so **Table 6** compares whenever an activity was present in the flipped classes (including the pre and in-class sessions) compared to when this activity was present in control traditional classes only. In the control classes, the majority involved a lecture. Flipped classes often also involved lectures in-class; however, these were shorter than their traditional counterparts (10–20 min) and often involved

reviews of the pre-class activities, gave students opportunities to ask questions about the pre-class, and sometimes targeted areas of need identified through a review of student performance on the pre-class quiz.

As shown in **Table 6**, we classified under 'Students active involvement and responding' group work, student presentations, class discussions, role plays, and use of clickers. 'Problem-solving' included completing problem sets, and problem-based learning. 'Student involvement but more structured and content related' included case studies, Socratic questioning, and debates.' Assessments included modified lectures based on assessment tasks, pre-class quizzes, in-class quizzes, and class quizzes at the end of the session. It was rare to have estimates of reliability to correct the effect-size estimates, and it could well be that the quizzes were lower in quality, but then so too may be the final assignment scoring.

Table 6 shows that, in all but four instances, the two means were not statistically significantly different ($P > 0.01$). It did not matter the teaching activities in the flipped or in-class as the effects did not significantly differ. But when lectures were included in the flipped and/or in-class, the effects (0.54) were higher than lectures in the control classes (0.33). Furthermore, when using Labs, Practicals, case studies, or regular quizzes then the effects were greater in the control group (that is, where there were no pre-class assigned activities).

Four findings were salient. First, the greatest impact of flipped learning was when the flipped in-class included a lecture but with pre-assigned preparation for this lecture. The lecture in class was not the first time hearing it, or pre-material to the lecture had been provided. Second, pre-rehearsals or learning for labs, practicals, or case studies, if anything as part of the pre-class flipping had lower effects, possibly because there was little learning in repeating these activities. Third, with more active learning in traditional instructions, the effect sizes were not that different between the flipped and traditional instructions. In fact, for active learning strategies such as case studies, Socratic questioning, and debates, the effects were greater when these activities were carried out in traditional instructions. Fourth, the effect of having regular quizzes was greater in traditional instruction than in flipped learning, possibly because participation in quizzes may be obliged in-class but not mandatory in flipped learning.

Confounds withing studies (RQ3b)

There was no correlation between the effect size and the class time (in min, $r = 0.08$), but when students are given extra time (pre-class + in-class), the effect is higher (0.38) than when they are compared (pre-class + in-class) with equal in-class time (0.22).

Some of the flipped learning interventions included feedback or formative evaluations of the pre-class, and some

TABLE 5 Summary statistics for activities in the pre-class and in-class parts of the lesson.

	N	M	SE	Low CI	High CI
Pre- class activities					
Readings	8	0.75	0.23	0.30	1.20
Video + quiz	15	0.66	0.23	0.21	1.11
Video + PowerPoint	26	0.31	0.11	0.09	0.53
Video + readings	24	0.45	0.06	0.33	0.57
Online lectures	336	0.40	0.04	0.32	0.48
Online modules	45	0.38	0.07	0.24	0.52
Video	29	0.22	0.09	0.04	0.40
Video +	18	0.18	0.08	0.02	0.34
Assignments/Exams	10	0.18	0.11	-0.04	0.40
PowerPoint	13	0.15	0.07	0.01	0.29
In- class activities					
Demonstration	30	0.57	0.15	0.28	0.86
Problem based	19	0.44	0.15	0.15	0.73
Use of clicker	49	0.43	0.10	0.23	0.63
Lectures	532	0.38	0.04	0.30	0.46
Problem-solving	203	0.38	0.06	0.26	0.50
Group work	175	0.34	0.06	0.22	0.46
Practical laboratory	139	0.25	0.07	0.11	0.39
Reviews assignments	29	0.25	0.06	0.13	0.37
In class quizzes	132	0.20	0.06	0.08	0.32
Case studies	73	0.19	0.10	-0.01	0.39
Student presentation	8	0.00	0.07	-0.14	0.14

instructors then used this feedback to modify the in-class experience. There were no effects relating to the presence or absence of formative assessment and feedback, or whether it was used or not to influence in-class experiences (Table 7).

In most studies, the traditional and flipped conditions were taught by the same instructors. But, it seems to not matter whether the instructors are the same ($g = 0.37$, $SE = 0.031$, $N = 455$) or not ($g = 0.40$, $SE = 0.068$, $N = 77$) in the two conditions.

Discussion

Flipped learning remains a well-promoted, well-resourced, and well-researched topic. It has become popular with many researchers, teachers, professional learning companies, bloggers, and commercial agencies. Indeed, we located 46 meta-analyses of flipped learning showing the spread and interest in the method but noted the remarkable variability in the findings of these syntheses of the research.

We found the overall effects from 0.19 to 2.29 from the 46 meta-analyses, which suggests that there is much

TABLE 7 Hedges g effect-size moderated by the role of formative assessment.

Domain	Hedges g	SE	Lower CI	Higher CI	No.
Included formative assessment	0.36	0.052	0.26	0.46	187
No formative assessment	0.38	0.033	0.32	0.44	345
Formative assessment included from pre-class	0.38	0.051	0.28	0.48	211
No formative assessment included from pre-class	0.37	0.033	0.31	0.43	321

more to understanding this phenomenon. There is likely no one method of flipping, and perhaps not even an agreed understanding of its comparison (the “traditional” classroom). Glass (2019), in an interview about meta-analyses, noted that an overall average effect can help make a claim about whether the method is promising or not, and then more critically noting that it can still be done well or poorly. There is so much information in the variance around the mean,

TABLE 6 Means, standard deviations of effect sizes when an activity was present in the flipped and the control classrooms, and the effect-size difference between the flipped and control.

	Activity present in flipped + in-class			Activity present in control classes			Effect size			
	<i>M</i>	<i>SE</i>	No.	<i>M</i>	<i>SE</i>	No.	ES	<i>t</i> -test	df	<i>p</i>
Content delivery										
Lectures	0.54	0.06	118	0.33	0.03	414	0.21	3.25	530	0.001
Labs/Practical	0.25	0.08	122	0.41	0.03	410	-0.16	-2.28	530	0.023
Review assignment	0.42	0.06	199	0.35	0.03	333	0.07	1.15	530	0.250
Student active involvement and responding										
Group Work	0.40	0.04	395	0.29	0.04	137	0.11	1.54	530	0.128
Student presentations	0.46	0.09	116	0.35	0.03	416	0.11	1.49	530	0.137
Class discussions	0.37	0.04	314	0.38	0.04	218	-0.01	-0.17	530	0.864
Role plays	0.40	0.11	42	0.37	0.03	490	0.03	0.28	530	0.779
Clickers	0.41	0.09	58	0.37	0.03	474	0.04	0.44	530	0.661
Problem-Solving										
Problem sets	0.40	0.04	410	0.30	0.05	122	0.10	1.28	530	0.201
Problem based	0.30	0.08	48	0.38	0.03	484	-0.08	-0.81	530	0.417
Student involvement—structured and content related										
Case studies	0.27	0.06	156	0.42	0.03	376	-0.15	-2.48	530	0.013
Socratic questioning	0.14	0.13	9	0.38	0.03	523	-0.24	-1.04	530	0.300
Debates	0.34	0.09	29	0.38	0.03	503	-0.04	-0.32	530	0.752
Assessment										
Pre-class Modification	0.26	0.09	83	0.40	0.03	449	-0.14	-1.75	530	0.081
Post Lecture Quiz	0.36	0.05	231	0.39	0.03	301	-0.03	-0.72	530	0.470
In class Post-Lecture Quiz	0.41	0.04	76	0.37	0.03	456	0.04	0.53	530	0.595
Other Regular Quiz	0.22	0.07	114	0.42	0.03	418	-0.21	-2.95	530	0.003

requiring the need to identify and interrogate moderators or confounds to understand this variance better. Glass also noted that in education, unlike medicine, we can rarely control the dosage, fidelity, and quality of intervention. In medicine, the implementation could be an “intravenous injection of 10 mg of Nortriptyline” and this is uniform and well defined, “whereas even interventions that carry the same label in education are subject to substantial variation from place to place, or time to time” (p. 4). Exposing students to flipped learning, to paraphrase Glass, can take many different forms, some of which are effective and some of which are not.

Thus, a novel contribution of our work is the identification of the major and critical moderators and confounds, of which the nature of the implementation seemed to be the most critical. More significantly, we show that accounting for these moderators and confounds changes the interpretation of evidence, much of which appeared to be problematic from previous meta-analyses. We summarize four key findings:

- a) The quality of implementation of many of the flipped studies was not consistent with its core claim of active learning being critical to its success. Not only did we find a low prevalence of active learning in flipped learning implementations, but also that active learning when present did not add to the effect. Therefore, the effects of flipped learning cannot be attributed to the presence of active learning because active learning was largely absent in the flipped learning implementations. This questions the quality of implementation of many flipped studies, as one of the core claims is that active learning is critical to its success. It seems not.
- b) The greatest impact of flipped learning was when the in-class included a lecture. This suggests that the major advantage of flipped learning may be the double or extended exposure to the lecturers’ interpretation of the knowledge and understanding, and not necessarily the active involvement of the students. A counter-argument (noted by a reviewer) is that lectures can focus on content. If the aim of the flipped and class is content and less deep relational or transfer thinking, then perhaps this finding is less surprising. An analysis of the confounds further supported this finding that it was likely the increase in exposure, additional time, the practice or repetition effect that allowed flipped learning to make an increase in student learning.
- c) As more active learning was incorporated in traditional instruction, the effect of flipped learning over traditional instruction tended to reduce, and in several cases, even reverse. This suggests that effects on learning are not due to flipped learning or traditional instruction, but due to active learning. Active learning, when designed well, be it

in flipped or traditional instruction, is effective, and we should focus on that more squarely; and

- d) Problem-solving as an active learning strategy when carried out prior to instruction had a positive impact on learning in both traditional and flipped learning, although the effect was greater in flipped learning. Making modifications based on such problem-solving was also effective in both traditional and flipped learning, although this time effect was greater in traditional instruction. This suggests that using active learning activities such as problem-solving prior to a lecture (online or in-class) and modification based on such problem-solving can be effective.

A major message is that the effects of flipped must consider the nature of the implementation and not generalize to “flipping” as if there was a singular interpretation of the activities in flipped and control classes. Many of the activities deemed critical to flipping also occur regularly in the traditional classroom, and the major effect of flipping seems to be increasing exposure to passive learning and the time on task. The largest impact of active learning is when it precedes in-class instruction or is undertaken as part of traditional instruction, contrary to the core claim of flipped instruction.

Taking together, these findings seem to suggest, somewhat paradoxically, that the effectiveness of flipped over traditional instruction results largely from perpetuating passive than active learning, whereas the efficacy of traditional instruction over flipped learning results largely from incorporating active learning prior to in-class instruction.

One way to interpret these findings is that there are major missed opportunities in the typical implementation of many flipped classes. The chance to have students prepare, be exposed to a new language, understand what they already know and do not know, and then capitalize on these understandings and misunderstandings was the promise of flipped classes—but hardly seems to be realized in reality. In the few cases where we located teachers making students complete quizzes after the pre-class and then modifying their instruction, there was no evidence that this improved their subsequent performance. The nature of what was undertaken in the pre-class needs to be questioned. It is likely that these activities did not provide data for instructors to detect what is needed to support their exposure to the ideas. [Rodríguez and Campión \(2016\)](#) have provided methods and rubrics to assess the quality of pre- and in-class sessions, for example, by evaluating whether interactions with the teacher/classmates are more frequent and positive, whether students can work at their own pace, have opportunities to show the teacher/classmate what has been learned, whether there is greater participation in classroom decision-making by collaborating with other classmates, etc. More attention to the effectiveness of the two parts is critical.

A second way to interpret these findings is that flipped instruction should not replace the role of a teacher as a provider of didactic instruction (Lai and Hwang, 2016). Instead, learning outcomes are best when the pre-class is reinforced by a targeted mini-lecture in-class, particularly when it is based on the instructor being aware of what the students are grappling with from their exposure to the vocabulary and ideas in the pre-class session.

Finally, and building on the second way, a third way to interpret is to argue that it is not the presence of learning from what the students do or do not know prior to class and adjusting the teaching. It is not engaging them in group work, class discussions, role plays, clickers, or response methods, and it is certainly not providing case studies, Socratic questioning, or debates in the flipped classrooms. Indeed, many of these activities are also included in the traditional class. It is flipping such that students do pre-work and then followed the traditional instruction—provided this traditional instruction is not seen as solely lecturing—that makes the difference. As our findings show, if such pre-class involves problem-solving assessments, then students gain more from the subsequent in-class instruction. This was an unexpected finding but consistent with the large body of research on the effectiveness of problem-solving followed by the instruction (Loibl et al., 2017; Sinha and Kapur, 2021a).

Considering the above-unexpected finding, it is worth returning to some of the fundamental claims underpinning flipped learning and making a case for a variant of the flipped learning model. A major claim for pre-class is to familiarize students with the language, show them what they do and do not know, and prepare them to learn from subsequent instruction (Schwartz and Bransford, 1998). Our fourth main finding was consistent with and points to a connection with a robust body of evidence that preparatory activities such as generating solutions to novel problems prior to instruction can help students learn better from the instruction (Kapur, 2016; Loibl et al., 2017). Research shows that students often fail to solve the problem correctly because they have not learned the concepts yet. However, to the extent they can generate and explore multiple solutions to the problem, even if they are suboptimal or incorrect, this failure prepares them to learn from subsequent instruction (Sinha and Kapur, 2021a). This is called the Productive Failure effect (Kapur, 2008). Taken together, and even though we were not expecting this from the outset, we view the results from the meta-analyses as consistent with productive failure. More importantly, we connect findings from our meta-analysis on flipped learning with the findings from research on productive failure to derive an alternative model for flipping. We briefly describe research on productive failure first before connecting and deriving the alternative model next.

Productive failure

Over the past two decades, there has been considerable debate about the design of initial learning: When learning a new concept, should students engage in problem-solving followed by instruction, or instruction followed by problem-solving? Evidence for Productive Failure comes not only from quasi-experimental studies conducted in the real ecologies of classrooms (e.g., Schwartz and Bransford, 1998; Schwartz and Martin, 2004; Kapur, 2010, 2012; Westermann and Rummel, 2012; Song and Kapur, 2017; Sinha et al., 2021) but also from controlled experimental studies (e.g., Roll et al., 2011; Schwartz et al., 2011; DeCaro and Rittle-Johnson, 2012; Schneider et al., 2013; Kapur, 2014; Loibl and Rummel, 2014a,b; Sinha and Kapur, 2021b).

Sinha and Kapur's (2021a) meta-analysis of 53 studies with 166 comparisons that compared the PS-I design with the I-PS design showed a significant effect in favor of starting with problem-solving followed by instruction [Hedge's g 0.36 (95% CI 0.20; 0.51)]. The effects were even more substantial (Hedge's g ranging between 0.37 and 0.58) when problem-solving followed by instruction was implemented with high fidelity to the principles of productive failure (PF; Kapur, 2016), a variant that designs the initial problem-solving to lead to failure deliberately. Estimation of true effect sizes after accounting for publication bias suggested a strong effect size in favor of PS-I (Hedge's g 0.87).

Not only does learning from productive failure work better than the traditional instruction-first method (Sinha and Kapur, 2021a), we understand the confluence of learning mechanisms that explains why that is the case. First, preparatory problem-solving helps activate relevant prior knowledge even if students produce sub-optimal or incorrect solutions (Siegler, 1994; Schwartz et al., 2011; DeCaro and Rittle-Johnson, 2012). Second, prior knowledge activation makes students notice their inconsistencies and misconceptions (Ohlsson, 1996; DeCaro and Rittle-Johnson, 2012), which in turn makes them aware of the gaps and limits of their knowledge (Loibl and Rummel, 2014b). Third, prior knowledge activation affords students opportunities to compare and contrast their solutions with the correct solutions during subsequent instruction, thereby increasing the likelihood of students' noticing and encoding critical features of the new concept (Schwartz et al., 2011; Kapur, 2014). Finally, besides the cognitive benefits, problem-solving prior to instruction also has affective benefits of providing greater learner agency, engagement, and motivation to learn the targeted concept (Belenky and Nokes-Malach, 2012), as well as naturally triggering moderate levels of negative emotions (e.g., shame, anger) that can act as catalysts for problem space exploration (Sinha, 2022). Suppose these considerations are part of designing the flipped experience, and the subsequent teaching considerate of productive failure during and from the flipping.

In that case, it is likely that student learning will be more enhanced.

In the light of the findings and mechanisms of productive failure, it seems worthwhile for instructors to receive feedback from students about their initial pre-class problem-solving attempts to focus on these notions, to build on what is understood, and thus tailor the class to deal with these lesser or unknown notions. It is also a chance to make relations between previous classes, what the students know already, and link this to new knowledge and understanding (Hattie and Donoghue, 2016).

The nature of the in-class activities in flipped learning also matters. For example, in none of the studies was the quality of the teaching evaluated, other than student satisfaction. It would be valuable to create measures of quality apropos whether the classes covered the material less or not known from the pre-class activities, ensuring students have both the surface (or knowing that) and deeper (or knowing how) knowledge and understanding. It would be further valuable to determine retention and transfer to near and far situations.

The major claim here is constructive alignment. Biggs (1996) considered effective teaching to be an internally aligned system working to a state of stable equilibrium. For example, suppose the assessment tasks address lower-level surface activities than the espoused curriculum objectives. In that case, equilibrium will be achieved at the lowest level as the system will be driven by the backwash from testing (and less from the curriculum). Or students with deep learning motives and strategies will perform poorly under mastery learning if the learning is based on narrow, low cognitive level goals (Lai and Biggs, 1994). Thus, good teaching needs to address all the parts of the teaching experience—the goals, curriculum goals, teaching methods, class experiences, and particularly assessment tasks and grading. His constructive alignment notion asks teachers to be clear about what they want their students to learn, and how they would manifest that learning in terms of “performances of understanding.” Students need to be exposed to knowledge and understanding relative to these goals, placed in situations that are likely to elicit the required learnings, and the assessment and in-class activities aligned with the criteria of success. Otherwise, particularly at the college level, there is safety in resorting to knowing lots, repeating what has been said, and privileging surface-level knowledge.

This notion of constructive alignment was not so present in many of the studies on flipped learning. The focus was more on engaging students in repetitive, passive activities—the same in the pre-class repeated in the in-class, usually *via* asking students to pre-review videos of classes, pre-review the PowerPoints then used in class, or listening to a teacher repeat material already exposed to the students. There is no reason to claim these are not worthwhile activities, but it does not seem to be consistent with the claims of flipped learning for deepening understanding. Assuring that the pre-class, the in-class, the assignments, the

assessments, and the grading are aligned with such claims seems necessary if flipped learning is appropriately evaluated.

An alternative model to flipped learning

Constructive alignment allowed us to connect the findings of our meta-analysis with research on productive failure, and to build on the two-phase flipped learning model to an alternative four-phase model. We outline the alternative model before describing its derivation.

- a) Fail—providing opportunities for the instructor and the student to diagnose, check, and understand what was and was not understood. This proposal is a direct consequence of our key finding, finding d, that problem-solving when included prior to instruction be it in flipped or traditional instruction had a positive impact on learning. Situating this finding in the broader research on productive failure only strengthens our proposal for starting with the Fail phase.
- b) Flip—pre-exposure to the ideas in the upcoming class (as simple as providing a video of the class). This proposal is consistent with the logic of the class is flipped learning model, but even more so, when it is preceded by a Fail phase.
- c) Fix—a class where these misconceptions are explored and opportunity to re-engage in learning the ideas and a traditional lecture is efficient to accomplish this. This proposal follows the key finding, finding b, that the greatest impact of flipped learning was when the in-class included a lecture, thereby allowing the instructor to re-engage misconceptions and assemble them into robust learning.
- d) Feed—feedback to the students and instructor about levels of understanding and “Where to next” directions. Feedback, especially formative assessment, is an essential component of active learning, as indeed our key findings, findings a and c, suggest. Finding a suggests a lack of such opportunities in flipped learning, and finding c suggests the inclusion of such activities improves learning outcomes.

As noted throughout, flipped learning comprises two phases: a flipped or pre-class (online) lecture followed by an in-class discussion and elaboration. Our findings have revealed that such a two-phase model is not any more effective than a traditional model once the nature of implementations is considered. What matters more is the inclusion of active learning. And one particular active learning strategy that makes a difference is engaging students in problem-solving prior to instruction. Given that the positive effects of problem-solving prior to instruction was in fact our key finding, finding d, and one that already has strong theoretical and empirical support

in the literature as we have outlined earlier, we propose starting with precisely such problem-solving activities. We call this phase the “Fail” phase.

The aim of the Fail phase is to help students understand what they do not know using problem-solving activities based on principles of productive failure. That is, when learning a new concept, instead of first viewing an online lecture (and similar), students start with a preparatory problem-solving activity designed to activate their knowledge about what they are going to learn in the lecture. It is by assisting the students to orient to what they do not and need to know that subsequent learning is maximized.

The Fail phase can then be followed by the second phase, the Flip, where students proceed to view the online lectures to learn the targeted concepts, as would be typical in the first phase of the two-phase flipped learning model currently.

Students can then move to the third phase, Fix, where they convene for in-class activities to consolidate what they have generated, compare and contrast student-generated and canonical solutions, attend to the critical concept features, and observe how these features are organized and assembled. As noted, our key finding b supported this phase.

Finally, the Feed stage, where the students and the teachers learn what has been learned, who has accomplished this learning, and the magnitude or strength of this learning. As noted earlier, formative feedback and assessment are essential components of active learning strategies. In finding a, we noted the evidence of such strategies, leading us to suggest their inclusion. In finding c, our review showed how the inclusion of such strategies improved learning outcomes. Finally, we did find evidence for the use of regular assessments, mostly formative, to help both students and teachers understand the progress and adapt accordingly.

The alternative model is a principle-based model setting out the goals of the design of each phase. It is meant to be a prescriptive model that describes the design of the constituent activities, instruction, feedback, and assessments in various phases. Future research is needed to investigate the validity and reliability of its implementation.

Limitations and future work

Other aspects of studies included in our meta-analyses need closer attention. The sample sizes in too many studies were relatively small; the median sample in the flipped and in the traditional classes was 40, and a quarter less than 25. In most studies, the unit of analysis was the student, but it should be the class or instructor so that hierarchical modeling can be used to account for students nested within classes. The cultural context of the study is most critical, and this behooves more information about the traditional

classes prior to implementing flipped learning. Indeed, if the flipped learning intervention was novel, the effect may well be due to the Hawthorne effect. This is not to say that there could not be improvements in learning, but caution is needed when making claims about causal attributions of these improvements.

Our meta-analysis did not include dissertations, conference papers, or gray literature, which may have led to a bias. However, for flipped learning studies, the direction of the bias remains unclear. For example, on the one hand, [Bredow et al. \(2021\)](#) found much lower effects for dissertations ($g = 0.14$) than for peer-reviewed journals ($g = 0.42$) and conferences papers ($g = 0.34$). [Cheng et al. \(2019\)](#) and [Tutal \(2021\)](#) reported similar effects of dissertations, conference, and peer-reviewed articles. On the other hand, [Jang \(2000\)](#) found higher effects for dissertations (0.61) than journal articles (0.29).

We also note that many studies have been published beyond our-2019 limit for the meta-analyses and that other meta-analyses also have been published since that time (and they are included in [Table 1](#)) which show similar confound issues, lack of attention to coding for implementation processes, and wide variation in results. It is recommended that future flipped studies and meta-analyses pay more attention to the implementation processes (e.g., dosage, fidelity, quality, adaptations), and be clear about the pre and in-class components.

It is also not clear why shorter (≤ 1 month) interventions are more effective, but it could relate to a reduction in the power of the Hawthorne effect. It could be that not every focus of teaching is amenable to being flipped, or it could be the quality of the pre-class and the in-class interactions. Constructive alignment with the major assessment tasks and level of cognitive complexity might demand a variety of teaching and learning strategies. Overexposure to one method may not be the most conducive method.

As discussed in [Lag and Saele \(2019\)](#), another limitation of the investigation of flipped learning interventions is that random assignment is rare. Furthermore, flipped learning interventions are often conducted after the reference group, following a class redesign. Random assignment is difficult in real-world conditions, and teaching two formats simultaneously requires more resources, that are often unavailable. Further studies should be conducted to investigate how the order of the interventions impacts academic outcomes.

All these claims require further research, and the plea is to be more systematic in controlling and studying the moderators, especially the nature of implementation of the flipped classroom. It was the nature of the implementation that matters significantly. Future studies need to be quite specific what is involved in both the traditional classes (as many also included active learning as part of the more formal lecture) and in the flipped classes.

Conclusion

In the final analysis, while there may be other reasons for advocating flipped learning as it is currently implemented, it is clear that robust scientific evidence in the quality of implementation or its effectiveness over traditional instruction is not among them. Indeed, it seems that implementations of flipped learning perpetuate the things they claim to reduce, that is, passive learning. It is passive learning as opposed to active learning that seems to have the greatest impact on the overall effects. However, the effectiveness of incorporating active learning seems more consistent with research on productive failure, a connection we made to derive an alternative model. Together, this must at the very least force us to rethink the overenthusiasm for flipped learning and be cautious about the conditions that are needed to make it work (and that are often not there, as the present study points out). More critically, in the light of recent advances in the learning sciences, the underlying commitment to the instruction-first paradigm seems fundamentally problematic. Instead, we invite research relating to Fail, Flip, Fix, and Feed.

Data availability statement

The original contributions presented in this study are included in the article/[Supplementary material](#), further inquiries can be directed to the corresponding author.

Author contributions

MK: conceptualization, data curation, funding acquisition, writing—original draft, and writing—review and editing. JH: conceptualization, funding acquisition, methodology, writing—original draft, and writing—review and editing. IG: methodology, coding, writing—original draft, and writing—review and editing. TS: methodology, project administration, and writing—review and editing. All authors contributed to the article and approved the submitted version.

References

- Aidinopoulou, V., and Sampson, D. G. (2017). An action research study from implementing the flipped classroom model in primary school history teaching and learning. *J. Educ. Technol. Soc.* 20, 237–247. doi: 10.2307/jeductechsoci.20.1.237
- Algarni, B. (2018). “A meta-analysis on the effectiveness of flipped classroom in mathematics education,” in *EDULEARN18 Proceedings*, (Palma, Spain: 10th International Conference on Education and New Learning Technologies), 7970–7976. doi: 10.21125/edulearn.2018.1852
- Aydin, M., Okmen, B., Sahin, S., and Kilic, A. (2020). The meta-analysis of the studies about the effects of flipped learning on students' achievement. *Turkish Online J. Distance Educ.* 22, 33–51.

Funding

The research was funded by ETH Zürich, as well as a Ministry of Education (Singapore) grant DEV03/14MK to MK; and funded as part of the Science of Learning Research Centre, a Special Research Initiative of the Australian Research Council Project Number SR120300015 to JH, and open access funding provided by ETH Zürich.

Acknowledgments

MK thank his research assistants, June Lee and Wong Zi Yang, for their help with the study. We thank Raul Santiago and Detlef Urhahne for constructive improvements.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/educ.2022.956416/full#supplementary-material>

- Beale, E. G., Tarwater, P. M., and Lee, V. H. (2014). A retrospective look at replacing face-to-face embryology instruction with online lectures in a human anatomy course. *Anat. Sci. Educ.* 7, 234–241. doi: 10.1002/ase.1396

- Belenky, D. M., and Nokes-Malach, T. J. (2012). Motivation and transfer: The role of mastery-approach goals in preparation for future learning. *J. Learn. Sci.* 21, 399–432. doi: 10.1080/10508406.2011.651232

- Bergmann, J., and Sams, A. (2012). *Flip your classroom: Reach every student in every class every day*. USA: International Society for Technology in Education. ITSE, ACDE.

- Biggs, J. (1996). Enhancing teaching through constructive alignment. *High. Educ.* 32, 347–364. doi: 10.1007/BF00138871
- Blázquez, B. O., Masluk, B., Gascon, S., Díaz, R., Aguilar-Latorre, A., Magallón, I., et al. (2019). The use of flipped classroom as an active learning approach improves academic performance in social work: A randomized trial in a university. *PLoS One* 14:e0214623. doi: 10.1371/journal.pone.0214623
- Bonnes, S. L., Ratelle, J. T., Halvorsen, A. J., Carter, K. J., Hafidahl, L. T., Wang, A. T., et al. (2017). Flipping the quality improvement classroom in residency education. *Acad. Med.* 92, 101–107. doi: 10.1097/ACM
- Bong-Seok, J. (2018). A meta-analysis of the effect of flip learning on the development and academic achievement of elementary school students. *Korea Curric. Eval. Institute* 21, 79–101.
- Borenstein, M., Hedges, L. V., Higgins, J. P., and Rothstein, H. R. (2011). *Introduction to meta-analysis*. New Jersey: John Wiley & Sons.
- Borenstein, M., Higgins, J. P. T., Hedges, L. V., and Rothstein, H. R. (2017). Basics of meta-analysis: I2 is not an absolute measure of heterogeneity. *Res. Synth. Methods* 8, 5–18. doi: 10.1002/jrsm.1230
- Boyras, S., and Ocak, G. (2017). The implementation of flipped education into Turkish EFL teaching context. *J. Lang. Linguist. Stud.* 13, 426–439.
- Boysen-Osborn, M., Anderson, C. L., Navarro, R., Yanuck, J., Strom, S., McCoy, C. E., et al. (2016). Flipping the advanced cardiac life support classroom with team-based learning: Comparison of cognitive testing performance for medical students at the University of California, Irvine, United State. *J. Educ. Eval. Health Prof.* 13:11. doi: 10.3352/jehp.2016.13.11
- Bredow, C. A., Roehling, P. V., Knorp, A. J., and Sweet, A. M. (2021). To flip or not to flip? A meta-analysis of the efficacy of flipped learning in higher education. *Rev. Educ. Res.* 91, 878–918. doi: 10.3102/00346543211019122
- Butzler, K. B. (2016). The synergistic effects of self-regulation tools and the flipped classroom. *Comp. Schools* 33, 11–23. doi: 10.1080/07380569.2016.1137179
- Chen, K. S., Monrouxe, L., Lu, Y. H., Jenq, C. C., Chang, Y. J., Chang, Y. C., et al. (2018). Academic outcomes of flipped classroom learning: A meta-analysis. *Med. Educ.* 52, 910–924. doi: 10.1111/medu.13616
- Chen, L. L. (2016). Impacts of flipped classroom in high school health education. *J. Educ. Technol. Syst.* 44, 411–420. doi: 10.1177/0047239515626371
- Cheng, L., Ritzhaupt, A. D., and Antonenko, P. (2019). Effects of the flipped classroom instructional strategy on students' learning outcomes: A meta-analysis. *Educ. Technol. Res. Dev.* 67, 793–824. doi: 10.1007/s11423-018-9633-7
- Cho, B., and Lee, J. (2018). A meta analysis on effects of flipped learning in Korea. *J. Digit. Conver.* 16, 59–73. doi: 10.14400/JDC.2018.16.3.059
- Cohen, J. A. (1960). A coefficient of agreement for nominal scales. *Educ. Psychol. Measure.* 20, 37–46. doi: 10.1177/001316446002000104
- DeCaro, M. S., and Rittle-Johnson, B. (2012). Exploring mathematics problems prepares children to learn from instruction. *J. Exp. Child Psychol.* 113, 552–568. doi: 10.1016/j.jecp.2012.06.009
- Deshen, L., and Yu, T. (2021). The influence of flipping classroom on the academic achievements of students at higher vocational colleges: Meta-analysis evidence based on the random effect model. *J. ABA Teach Univ.* 38, 100–107.
- Doğan, Y., Batdı, V., and Yaşar, M. D. (2021). Effectiveness of flipped classroom practices in teaching of science: A mixed research synthesis. *Res. Sci. Technol. Educ.* 1–29. doi: 10.1080/02635143.2021.1909553
- Egger, M., Smith, G. D., Schneider, M., and Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ* 315, 629–634. doi: 10.1136/bmj.315.7109.629
- Eichler, J. F., and Peeples, J. (2016). Flipped classroom modules for large enrollment general chemistry courses: A low barrier approach to increase active learning and improve student grades. *Chem. Educ. Res. Pract.* 17, 197–208. doi: 10.1039/C5RP00159E
- Farmus, L., Cribbie, R. A., and Rotondi, M. A. (2020). The flipped classroom in introductory statistics: Early evidence from a systematic review and meta-analysis. *J. Stat. Educ.* 28, 316–325. doi: 10.1080/10691898.2020.1834475
- Ferreri, S. P., and O'Connor, S. K. (2013). Redesign of a large lecture course into a small-group learning course. *Am. J. Pharm. Educ.* 77, 1–13. doi: 10.5688/ajpe77113
- Flynn, A. B. (2015). Structure and evaluation of flipped chemistry courses: Organic & spectroscopy, large and small, first to third year, English and French. *Chem. Educ. Res. Pract.* 16, 198–211. doi: 10.1039/C4RP00224E
- Fraleigh, R. C., and Vazire, S. (2014). The N-pact factor: Evaluating the quality of empirical journals with respect to sample size and statistical power. *PLoS One* 9:e109019. doi: 10.1371/journal.pone.0109019
- Ge, L., Chen, Y., Yan, C., Chen, Z., and Liu, J. (2020). Effectiveness of flipped classroom vs traditional lectures in radiology education: A meta-analysis. *Medicine* 99:e22430. doi: 10.1097/MD.00000000000022430
- Gillette, C., Rudolph, M., Kimble, C., Rockich-Winston, N., Smith, L., and Broedel-Zaugg, K. (2018). A meta-analysis of outcomes comparing flipped classroom and lecture. *Am. J. Pharm. Educ.* 82:6898. doi: 10.5688/ajpe6898
- Glass, G. (2019). *The promise of meta-analysis for our schools: A Q&A with Gene V. Glass Nomanis*. Available online at: <https://www.nomanis.com.au/post/the-promise-of-meta-analysis-for-our-schools-a-q-a-with-gene-v-glass> (accessed on May 11, 2020).
- González-Gómez, D., Jeong, J. S., Airado Rodríguez, D., and Cañada-Cañada, F. (2016). Performance and perception in the flipped learning model: An initial approach to evaluate the effectiveness of a new teaching methodology in a general science classroom. *J. Sci. Educ. Technol.* 25, 450–459. doi: 10.1007/s10956-016-9605-9
- Guerrero, S., Beal, M., Lamb, C., Sonderegger, D., and Baumgartel, D. (2015). Flipping undergraduate finite mathematics: Findings and implications. *Primus* 25, 814–832. doi: 10.1080/10511970.2015.1046003
- Gundlach, E., Richards, K. A. R., Nelson, D., and Levesque-Bristol, C. (2015). A comparison of student attitudes, statistical reasoning, performance, and perceptions for web-augmented traditional, fully online, and flipped sections of a statistical literacy class. *J. Stat. Educ.* 23, 1–33.
- Harrington, S. A., Bosch, M. V., Schoofs, N., Beel-Bates, C., and Anderson, K. (2015). Quantitative outcomes for nursing students in a flipped classroom. *Nurs. Educ. Perspect.* 36, 179–181. doi: 10.5480/13-1255
- Harzing, A. W. (2010). *The publish or perish book*. Melbourne: Tarma Software Research Pty Limited.
- Hattie, J., and Timperley, H. (2007). The power of feedback. *Rev. Educ. Res.* 77, 81–112. doi: 10.3102/003465430298487
- Hattie, J. A., and Donoghue, G. M. (2016). Learning strategies: A synthesis and conceptual model. *npj Sci. Learn.* 1:16013. doi: 10.1038/npjscilearn.2016.13
- He, W., Holton, A., Farkas, G., and Warschauer, M. (2016). The effects of flipped instruction on out-of-class study time, exam performance, and student perceptions. *Learn. Instr.* 45, 61–71. doi: 10.1016/j.learninstruc.2016.07.001
- Hedges, L. V., and Pigott, T. D. (2001). The power of statistical tests in meta-analysis. *Psychol. Methods* 6, 203–217. doi: 10.1037/1082-989X.6.3.203
- Hew, K. F., Bai, S., Dawson, P., and Lo, C. K. (2021). Meta-analyses of flipped classroom studies: A review of methodology. *Educ. Res. Rev.* 33:100393. doi: 10.1016/j.edurev.2021.100393
- Hew, K. F., and Lo, C. K. (2018). Flipped classroom improves student learning in health professions education: A meta-analysis. *BMC Med. Educ.* 18:38. doi: 10.1186/s12909-018-1144-z
- Higgins, J. P., and Green, S. (2011). The cochrane collaboration: The cochrane collaboration; 2011. *Cochrane handbook for systematic reviews of interventions*. Available online at: <http://handbook-5-1.cochrane.org/> (accessed May 11, 2020).
- Higgins, S. (2018). *Improving learning: Meta-analysis of intervention research in education*. Cambridge, UK: Cambridge University Press.
- Hu, R., Gao, H., Ye, Y., Ni, Z., Jiang, N., and Jiang, X. (2018). Effectiveness of flipped classrooms in Chinese baccalaureate nurse education: A meta-analysis of randomized controlled trials. *Int. J. Nurs. Stud.* 79, 94–103. doi: 10.1016/j.ijnurstu.2017.11.012
- Hu, X., Zhang, H., Song, Y., Wu, C., Yang, Q., Shi, Z., et al. (2019). Implementation of flipped classroom combined with problem-based learning: An approach to promote learning about hyperthyroidism in the endocrinology internship. *BMC Med. Educ.* 19:290. doi: 10.1186/s12909-019-1714-8
- Hudson, D. L., Whisenhunt, B. L., Shoptaugh, C. F., Visio, M. E., Cathey, C., and Rost, A. D. (2015). Change takes time: Understanding and responding to culture change in course redesign. *Scholarsh. Teach. Learn. Psychol.* 1, 255–268. doi: 10.1037/stl0000043
- Hung, H.-T. (2015). Flipping the classroom for English language learners to foster active learning. *Comput. Assist. Lang. Learn.* 28, 81–96. doi: 10.1080/09588221.2014.967701
- International Monetary Fund (2018). *World economic and financial surveys. World economic outlook. Database—WEO groups and aggregates information*. Available online at: <https://www.imf.org/external/pubs/ft/weo/2018/02/weodata/groups.htm> (accessed May 11, 2020).
- Jang, B. S. (2019). Meta-analysis of the effects of flip learning on the development and academic performance of elementary school students. *J. Curric. Eval.* 21, 79–101.

- Jang, H. Y., and Kim, H. J. (2020). A meta-analysis of the cognitive, affective, and interpersonal outcomes of flipped classrooms in higher education. *Educ. Sci.* 10:1115. doi: 10.3390/educsci10040115
- Jensen, J. L., Kummer, T. A., and Godoy, P. D. D. M. (2015). Improvements from a flipped classroom may simply be the fruits of active learning. *CBE Life Sci. Educ.* 14:ar5. doi: 10.1187/cbe.14-08-0129
- Jensen, S. A. (2011). In-class versus online video lectures: Similar learning outcomes, but a preference for in-class. *Teach. Psychol.* 38, 298–302. doi: 10.1177/0098628311421336
- Kakosimos, K. E. (2015). Example of a micro-adaptive instruction methodology for the improvement of flipped-classrooms and adaptive-learning based on advanced blended-learning tools. *Educ. Chem. Eng.* 12, 1–11. doi: 10.1016/j.ece.2015.06.001
- Kang, M. J., and Kang, K. J. (2021). The effectiveness of a flipped learning on Korean nursing students. A meta-analysis. *J. Digit. Converg.* 19, 249–260. doi: 10.14400/JDC.2021.19.1.249
- Kang, S., and Shin, I. (2005). The effect of flipped learning in Korea: Meta-analysis. *Soc. Digit. Policy Manage.* 16, 59–73. doi: 10.14400/JDC2018.16.3.059
- Kapur, M. (2010). Productive failure in mathematical problem solving. *Instruct. Sci.* 38, 523–550. doi: 10.1007/s11251-009-9093-x
- Kapur, M. (2012). Productive failure in learning the concept of variance. *Instruct. Sci.* 40, 651–672. doi: 10.1007/s11251-012-9209-6
- Kapur, M. (2014). Productive failure in learning math. *Cognitive Sci.* 38, 1008–1022. doi: 10.1111/cogs.12107
- Kapur, M. (2016). Examining productive failure, productive success, unproductive failure, and unproductive success in learning. *Educ. Psychol.* 51, 289–299. doi: 10.1080/00461520.2016.1155457
- Kapur, M. (2008). Productive failure. *Cogn. Instr.* 26, 379–424. doi: 10.1080/07370000802212669
- Karagöl, I., and Esen, E. (2019). The effect of flipped learning approach on academic achievement: A meta-analysis study. *H.U. J. Educ.* 34, 708–727. doi: 10.16986/HUJE.2018046755
- Kennedy, E., Beaudrie, B., Ernst, D. C., and St. Laurent, R. (2015). Inverted pedagogy in second semester calculus. *Primus* 25, 892–906. doi: 10.1080/10511970.2015.1031301
- Kim, S. H., and Lim, J. M. (2021). A systematic review and meta-analysis of flipped learning among university students in Korea: Self-directed learning, learning motivation, efficacy, and learning achievement. *J. Korean Acad. Soc. Nurs. Educ.* 27, 5–15. doi: 10.5977/jkasne.2021.27.1.5
- Kirschner, P. A., Sweller, J., and Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educ. Psychol.* 41, 75–86.
- Kiviniemi, M. T. (2014). Effects of a blended learning approach on student outcomes in a graduate-level public health course. *BMC Med. Educ.* 14:47. doi: 10.1186/1472-6920-14-47
- Kostaris, C., Stylianou, S., Sampson, D. G., Giannakos, M., and Pelliccione, L. (2017). Investigating the potential of the flipped classroom model in K-12 ICT teaching and learning: An action research study. *Educ. Technol. Soc.* 20, 261–273. doi: 10.2307/jeductechsoci.20.1.261
- Krueger, G. B., and Storlie, C. H. (2015). *Evaluation of a flipped classroom format for an introductory-level marketing class. Journal of higher education theory & practice.* 15. Available online at: <http://www.na-businesspress.com/jhetpopen.html> (accessed May 11, 2020).
- Lag, T., and Saele, R. G. (2019). Does the flipped classroom improve student learning and satisfaction? A systematic review and meta-analysis. *AERA Open* 5, 1–17. doi: 10.1177/2332858419870489
- Lai, C. L., and Hwang, G. J. (2016). A self-regulated flipped classroom approach to improving students' learning performance in a mathematics course. *Comp. Educ.* 100, 126–140. doi: 10.1016/j.compedu.2016.05.006
- Lai, P., and Biggs, J. (1994). Who benefits from mastery learning? *Contemp. Educ. Psychol.* 19, 13–23. doi: 10.1006/ceps.1994.1002
- Lancaster, J. W., and McQueeney, M. L. (2011). From the podium to the PC: A study on various modalities of lecture delivery within an undergraduate basic pharmacology course. *Res. Sci. Technol. Educ.* 29, 227–237. doi: 10.1080/02635143.2011.585133
- Lee, A. M., and Liu, L. (2016). Examining flipped learning in sociology courses: A quasi-experimental design. *Int. J. Technol. Teach. Learn.* 12, 47–64. doi: 10.1007/s11423-018-9587-9
- Lewis, J. S., and Harrison, M. A. (2012). Online delivery as a course adjunct promotes active learning and student success. *Teach. Psychol.* 39, 72–76. doi: 10.1177/0098628311430641
- Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gotzsche, P. C., Ioannidis, J. P., et al. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *J. Clin. Epidemiol.* 62, e1–e34. doi: 10.1016/j.jclinepi.2009.06.006
- Liu, Y. Q., Li, Y. F., Lei, M. J., Liu, P. X., Theobald, J., Meng, L. N., et al. (2018). Effectiveness of the flipped classroom on the development of self-directed learning in nursing education: A meta-analysis. *Front. Nurs.* 5, 317–329. doi: 10.1515/fon-2018-0032
- Li, B. Z., Cao, N. W., Ren, C. X., Chu, X. J., Zhou, H. Y., and Guo, B. (2020). Flipped classroom improves nursing students' theoretical learning in China: A meta-analysis. *PLoS One* 15:e0237926. doi: 10.1371/journal.pone.0237926
- Lo, C. K., Hew, K. F., and Chen, G. (2017). Toward a set of design principles for mathematics flipped classrooms: A synthesis of research in mathematics education. *Educ. Res. Rev.* 22, 50–73. doi: 10.1016/j.edurev.2017.08.002
- Loibl, K., Roll, I., and Rummel, N. (2017). Towards a theory of when and how problem solving followed by instruction supports learning. *Educ. Psychol. Rev.* 29, 693–715. doi: 10.1007/s10648-016-9379-x
- Loibl, K., and Rummel, N. (2014a). Knowing what you don't know makes failure productive. *Learn. Instruct.* 34, 74–85. doi: 10.1016/j.learninstruc.2014.08.004
- Loibl, K., and Rummel, N. (2014b). The impact of guidance during problem-solving prior to instruction on students' inventions and learning outcomes. *Instruct. Sci.* 42, 305–326. doi: 10.1007/s11251-013-9282-5
- McLaughlin, J. E., Roth, M. T., Glatt, D. M., Gharkholonarehe, N., Davidson, C. A., Griffin, L. M., et al. (2014). The flipped classroom: A course redesign to foster learning and engagement in a health professions school. *Acad. Med.* 89, 236–243. doi: 10.1097/ACM.0000000000000086
- Ming, Y. L. (2017). *A Meta-analysis on Learning Achievement of Flipped Classroom Flip classroom learning effectiveness analysis.* (Ph.D.thesis). Taiwan: Chung Yuan University.
- Moffett, J., and Mill, A. C. (2014). Evaluation of the flipped classroom approach in a veterinary professional skills course. *Adv. Med. Educ. Pract.* 5, 415–425. doi: 10.3928/01484834-20130919-03
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., and The Prisma Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The prisma statement. *PLoS Med.* 6:e1000097. doi: 10.1371/journal.pmed1000097
- Munafo, M. R., and Flint, J. (2010). How reliable are scientific studies? *Br. J. Psychiatr.* 197, 257–258. doi: 10.1192/bjp.bp.109.069849
- Murphy, J., Chang, J.-M., and Suaray, K. (2016). Student performance and attitudes in a collaborative and flipped linear algebra course. *Int. J. Math. Educ. Sci. Technol.* 47, 653–673. doi: 10.1080/0020739X.2015.1102979
- Nielsen, P. L., Bean, N. W., and Larsen, R. A. A. (2018). The impact of a flipped classroom model of learning on a large undergraduate statistics class. *Stat. Educ. Research J.* 17, 121–140.
- O'Connor, E. E., Fried, J., McNulty, N., Shah, P., Hogg, J. P., Lewis, P., et al. (2016). Flipping radiology education right side up. *Acad. Radiol.* 23, 810–822. doi: 10.1016/j.acra.2016.02.011
- Ohlsson, S. (1996). Learning from performance errors. *Psychol. Rev.* 103, 241–262. doi: 10.1037/0033-295X.103.2.241
- Oki, Y. (2016). *Flipping a content-based ESL course: An action research report. Hawaii pacific university TESOL working paper series.* Vol. 14. 62–75. Available online at: <http://www.hpu.edu/> (accessed May 11, 2020).
- Orhan, A. (2019). The effect of flipped learning on students' academic achievement: A meta-analysis study. *Cukurova Universitesi Egitim Fakultesi Dergisi* 48, 368–396.
- Ralević, L., and Tomašević, B. (2021). Comparing the effectiveness of the flipped classroom model and the traditional instruction model: A meta-analysis. *Nastava Vaspitanje* 301–318.
- Reich, J. (2015). Rebooting MOOC research. *Science* 347, 34–35. doi: 10.1126/science.1261627
- Rodríguez, D. M., and Campión, R. S. (2016). "Flipped Learning" en la formación del profesorado de secundaria y bachillerato. Formación para el cambio. *Context. Educat.* 1, 117–134.
- Roll, I., Alevén, V., McLaren, B. M., and Koedinger, K. R. (2011). Improving students' help-seeking skills using metacognitive feedback in an intelligent tutoring system. *Learn. Instruct.* 21, 267–280. doi: 10.1016/j.learninstruc.2010.07.004

- Schneider, B., Wallace, J., Blikstein, P., and Pea, R. (2013). Preparing for future learning with a tangible user interface: The case of neuroscience. *IEEE Trans. Learn. Technol.* 6, 117–129. doi: 10.1109/TLT.2013.15
- Schwartz, D. L., and Bransford, J. D. (1998). A time for telling. *Cogn. Instruct.* 16, 475–5223. doi: 10.1207/s1532690xci1604_4
- Schwartz, D. L., Chase, C. C., Oppezzo, M. A., and Chin, D. B. (2011). Practicing versus inventing with contrasting cases: The effects of telling first on learning and transfer. *J. Educ. Psychol.* 103, 759–775. doi: 10.1037/a0025140
- Schwartz, D. L., and Martin, T. (2004). Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction. *Cogn. Instruct.* 22, 129–184. doi: 10.1207/s1532690xci2202_1
- Sezer, B. (2017). The effectiveness of a technology- enhanced flipped science classroom. *J. Educ. Comput. Res.* 55, 471–494. doi: 10.1177/0735633116671325
- Shahnama, M., Ghonsooly, B., and Shirvan, M. E. (2021). A meta-analysis of relative effectiveness of flipped learning in English as second/foreign language research. *Educ. Technol. Res. Dev.* 69, 1355–1386. doi: 10.1007/s11423-021-09996-1
- Shi, Y., Ma, Y., MacLeod, J., and Yang, H. H. (2020). College students' cognitive learning outcomes in flipped classroom instruction: A meta-analysis of the empirical literature. *J. Comp. Educ.* 7, 79–103. doi: 10.1007/s40692-019-00142-8
- Shute, V. J. (2008). Focus on formative feedback. *Rev. Educ. Res.* 78, 153–189. doi: 10.3102/0034654307313795
- Siegler, R. S. (1994). Cognitive variability: A key to understanding cognitive development. *Curr. Direct. Psychol. Sci.* 3, 1–5.
- Sinha, T. (2022). Enriching problem-solving followed by instruction with explanatory accounts of emotions. *J. Learn. Sci.* 31, 151–198. doi: 10.1080/10508406.2021.1964506
- Sinha, T., and Kapur, M. (2021a). When problem solving followed by instruction works: Evidence for productive failure. *Rev. Educ. Res.* 91, 761–798. doi: 10.3102/00346543211019105
- Sinha, T., and Kapur, M. (2021b). Robust effects of the efficacy of explicit failure-driven scaffolding in problem-solving prior to instruction: A replication and extension. *Learn. Instruct.* 75:101488. doi: 10.1016/j.learninstruc.2021.101488
- Sinha, T., Kapur, M., West, R., Catasta, M., Hauswirth, M., and Trninic, D. (2021). Differential benefits of explicit failure-driven and success-driven scaffolding in problem-solving prior to instruction. *J. Educ. Psychol.* 113, 530–555. doi: 10.1037/edu0000483
- Sipe, T. A., and Curlette, W. L. (1996). A meta-synthesis of factors related to educational achievement: A methodological approach to summarizing and synthesizing meta-analyses. *Int. J. Educ. Res.* 25, 583–698. doi: 10.1016/S0883-0355(96)80001-2
- Sola Martínez, T., Aznar Díaz, I., Romero Rodríguez, J. M., and Rodríguez-García, A. M. (2019). Eficacia del método flipped classroom en la universidad: Meta- análisis de la producción científica de impacto. *REICE Rev. Iberoam. Calid. Eficacia Cambio Educ.* 17, 25–38. doi: 10.15366/reice2019.17.1.002
- Song, Y., and Kapur, M. (2017). How to flip the classroom—"productive failure or traditional flipped classroom" pedagogical design? *J. Educ. Technol. Soc.* 20, 292–305. doi: 10.2307/jeductechsoci.20.1.292
- Sparkes, C. N. (2019). *Flipped classrooms versus traditional classrooms: A systematic review and meta-analysis of student achievement in higher education*. Doctoral dissertation. Montréal, QC: Concordia University.
- Strelan, P., Osborn, A., and Palmer, E. (2020). The flipped classroom: A meta-analysis of effects on student performance across disciplines and education levels. *Educ. Res. Rev.* 30:100314. doi: 10.1016/j.edurev.2020.100314
- Szucs, D., and Ioannidis, J. P. (2017). Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. *PLoS Biol.* 15:e2000797. doi: 10.1371/journal.pbio.2000797
- Tan, C., Yue, W. G., and Fu, Y. (2017). Effectiveness of flipped classrooms in nursing education: Systematic review and meta-analysis. *Chinese Nurs. Res.* 4, 192–200. doi: 10.1016/j.cnre.2017.10.006
- Tang, F., Chen, C., Zhu, Y., Zuo, C., Zhong, Y., Wang, N., et al. (2017). Comparison between flipped classroom and lecture-based classroom in ophthalmology clerkship. *Med. Educ. Online* 22:1395679. doi: 10.1080/10872981.2017.1395679
- Turan, Z. (2021). Evaluating whether flipped classrooms improve student learning in science education: A systematic review and meta-analysis. *Scand. J. Educ. Res.* 19, 1–19. doi: 10.1080/00313831.2021.1983868
- Tutal, Ö. (2021). Flipped classroom improves academic achievement, learning retention and attitude towards course: A meta-analysis. *Asia Pacific Educ. Rev.* 22, 655–673. doi: 10.1007/s12564-021-09706-9
- van Alten, D. C. D., Phielix, C., Jannssen, J., and Kester, L. (2019). Effects of flipped the classroom on learning outcomes and satisfaction: A meta-analysis. *Educ. Res. Rev.* 28:100281. doi: 10.1016/j.edurev.2019.05.003
- Vitta, J. P., and Al-Hoorie, A. H. (2020). The flipped classroom in second language learning: A meta-analysis. *Lang. Teach. Res.* 1–25. doi: 10.1177/1362168820981403
- Wagner, M., Gegenfurtner, A., and Urhahne, D. (2021). Effectiveness of the flipped classroom on student achievement in secondary education: A meta-analysis. *Zeitschrift für Pädagogische Psychologie* 35, 11–31. doi: 10.1024/1010-0652/a000274
- Wasserman, N. H., Quint, C., Norris, S. A., and Carr, T. (2017). Exploring flipped classroom instruction in Calculus III. *Int. J. Sci. Math. Educ.* 15, 545–568. doi: 10.1007/s10763-015-9704-8
- Westermann, K., and Rummel, N. (2012). Delaying instruction: Evidence from a study in a university relearning setting. *Instruct. Sci.* 40, 673–689. doi: 10.1007/s11251-012-9207-8
- Whitman Cobb, W. N. (2016). Turning the classroom upside down: Experimenting with the flipped classroom in American government. *J. Polit. Sci. Educ.* 12, 1–14. doi: 10.1080/15512169.2015.1063437
- Xu, P., Chen, Y., Nie, W., Wang, Y., Song, T., Li, H., et al. (2019). The effectiveness of a flipped classroom on the development of Chinese nursing students' skill competence: A systematic review and meta-analysis. *Nurse Educ. Today* 80, 67–77. doi: 10.1016/j.nedt.2019.06.005
- Yakar, Z. Y. (2021). The effect of flipped learning model on primary and secondary school students' mathematics achievement: A meta-analysis study. *Cukurova Univ. Faculty Educ. J.* 50, 1329–1366.
- Yarbro, J., Arfstrom, K. M., McKnight, K., and McKnight, P. (2014). *Extension of a review of flipped learning*. Available online at: <https://flippedlearning.org/wp-content/uploads/2016/07> (accessed on May 11, 2020).
- Yoon, S. H. (2018). A meta-analysis for effects of flipped learning on secondary school students. *J. Educ. Cult.* 24, 459–476. doi: 10.24159/joec.2018.24.2.459
- Zhang, Q., Cheung, E. S. T., and Cheung, C. S. T. (2021). The impact of flipped classroom on college students' academic performance: A meta-analysis based on 20 Experimental Studies. *Sci. Insights Educ.* 8, 1059–1080. doi: 10.2139/ssrn.3838807
- Zhang, S. (2018). *A systematic review and meta-analysis on flipped learning in science education*. (Ph.D.thesis). Pokfulam, Hong Kong: University of Hong Kong.
- Zheng, L., Bhagat, K. K., Zhen, Y., and Zhang, X. (2020). The effectiveness of the flipped classroom on students' learning achievement and learning motivation: A meta-analysis. *Educ. Technol. Soc.* 23, 1–15.
- Zhu, G., Thompson, C., Suarez, M., and Peng, Z. (2019). *A meta-analysis on the effect of flipped instruction on K-12 students' academic achievement*. Toronto, Canada: Paper presentation at the American Educational Research Association Annual Meeting.
- Zhu, G. (2021). Is flipping effective? A meta-analysis of the effect of flipped instruction on K-12 students' academic achievement. *Educ. Technol. Res. Dev.* 69, 733–761.