The Effectiveness of the Flipped Classroom on Students' Learning Achievement and Learning Motivation: A Meta-Analysis

Lanqin Zheng[1], Kaushal Kumar Bhagat[2*], Yuanyi Zhen[1] and Xuan Zhang[1]
[1]School of Educational Technology, Faculty of Education, Beijing Normal University, Beijing, China // [2]Centre for Educational Technology, Indian Institute of Technology Kharagpur, India // bnuzhenglq@bnu.edu.cn // kkntnu@hotmail.com // zyyouc@126.com // zhxuan@mail.bnu.edu.cn
[*]Corresponding author

ABSTRACT: The purpose of this study was to examine the overall effectiveness of the flipped classroom on students' learning achievement and motivation. Data were collected from three databases, which include Web of Science, Scopus, and Eric. The present meta-analysis synthesized the findings of 95 studies with 15386 participants published from 2013 to 2019. The results revealed that the flipped classroom approach had a moderate effect size for learning achievement and learning motivation. The effect sizes of 12 moderators, including sample level, sample size, learning domain, the flip classroom model, research design, intervention duration, teaching method in the class, sample region, interactions in a pre-class and face-to-face class, tools in pre-class, and resources in pre-class were also analyzed. The results indicated that sample size, intervention durations, and sample regions significantly moderated the effect sizes. The findings of this study are discussed in-depth, together with the implications for practices on the use of the flipped classroom approach.

Keywords: Flipped classroom, Learning achievement, Learning motivation, Meta-analysis

1. Introduction

The flipped classroom has gained significant attention in recent years. It is also considered as an "inverted classroom" or "reversed instruction" (Bergmann & Sams, 2012). In the flipped classroom, learners watch the content videos at home and solve problems in the class (Tucker, 2012). The flipped classroom switches the in-class time and out-of-class time to enable more interactions between teachers and students in the class (Lai & Hwang, 2016). For example, Bergmann and Sams (2012) mentioned that in a traditional classroom, the main activities consisted of 5 minutes' warm-up activity, 20 minutes' review, 30 minutes' lecture, and 20 minutes' practice or lab activity. On the other hand, the activities of the flipped classroom include 5 minutes' warm-up activity, 10 minutes Q&A time on video, and 75 minutes of practice or lab activity. Class time is mainly used for collaboration among the students, discussion, and personalized learning (Francl, 2014).

Furthermore, several flipped models with different focuses have been proposed and implemented in practice. For example, the conventional flipped classroom emphasized content delivery (Bergmann & Sams, 2012). The FLIPPED model proposed by Chen, Wang, Kinshuk, and Chen (2014) advocated progressive activities, engaging experiences, and diversified platforms. These flipped classroom models are very promising and helpful for both research and practice. Previous studies found the positive effects of the flipped classroom and reported that the use of flipped classroom promoted students' learning performance (Lin, Hwang, Fu, & Chen, 2018) and learning satisfaction (Sergis, Sampson, & Pelliccione, 2018) compared to the traditional classroom. However, some studies indicated that there was no difference between the flipped classroom and the traditional classroom (Sparks, 2013; Strayer, 2012). So far, the effects of the flipped classroom are still debatable among the researchers. Therefore, it is vital to investigate the effects of the flipped classroom through systemic and comprehensive meta-analysis. This study aims to examine the effects of the flipped classroom and provide a clear picture about the mediating effects of moderator variables.

1.1. Previous reviews and meta-analysis of flipped classroom

The specifications for adopting the flipped classroom approach have been documented in previous literature reviews. For example, Seery (2015) analyzed the emerging trends on integrating the flipped learning model in chemistry in higher education. The findings indicated that the flipped learning approach developed an active learning environment

that resulted into a better conceptual understanding of learning engagement. Nederveld and Berge (2015) presented several tools for creating the flipped classroom in the workplace and discussed the benefits as well as challenges of the flipped classroom approach. O'Flaherty and Phillips (2015) conducted a systematic review of the flipped classroom in higher education. They found that the flipped classroom approach can improve academic performance and satisfaction. In another study, Kashada, Li, and Su (2017) analyzed ten studies related to the flipped classroom and examined the effects of the flipped classroom on students' performance in K-12 education. They found a positive impact of the flipped classroom on students' learning achievement. Njie-Carr, Ludeman, Lee, Dordunoo, Trocky, and Jenkins (2017) conducted a comprehensive review of relevant research concerning the flipped classroom model in nursing education. They provided the design and process information as well as the current status of the flipped classroom models through an analysis of 13 studies published in 2016. Lo and Hew (2017) conducted a literature review of the flipped classroom in K-12 education by analysis of 15 articles. They found that the flipped classroom model had a positive or neutral impact on learning achievement in K-12 education. However, some previous studies also reported the limitations of the flipped classroom. For example, Mellefont and Fei (2016) found that students' lack of preparation may hinder the effectiveness of the flipped classroom. Students were easily distracted when they watched the video (Toto & Nguyen, 2009). Besides, the effectiveness of the flipped classroom heavily relied on students' self-motivation (Wang, 2017). It is also difficult for teachers to monitor student comprehension and provide real-time feedback for each student (Milman, 2012).

Furthermore, some researchers conducted meta-analysis studies to examine the effectiveness of the flipped classroom. For example, Rahman et al. (2014) reviewed 15 studies on the flipped classroom. The results showed that the flipped classroom had a positive impact on students' achievement. The researchers conducted only qualitative analysis without calculating the effect size. Hew and Lo (2018) conducted a meta-analysis on 28 studies in the domain of health professionals and found a significant effect size in favor of the flipped classroom as compared to the traditional classroom. A meta-analysis study by Gillette et al. (2018) examined the effect of the flipped classroom in the pharmacy education and found a small positive effect for using the flipped model instead of the traditional lecture-based classroom. In the recent study, Cheng, Ritzhaupt, and Antonenko (2019) studied the overall effect of the flipped classroom approach. They also included different subject domains, student levels, and study durations as the moderator variables. The results indicated a moderate but significant positive effect of the flipped classroom on students' learning achievement.

1.2. The need for this study

Although previous reviews and meta-analysis analyzed the current status of the flipped classroom, there were three significant shortcomings of previous studies. First, very few meta-analysis studies have examined the effect of flipped classrooms compared to traditional classrooms on both learning achievement and learning motivation. Second, a systematic meta-analysis of the flipped classroom based on activity theory has not been published yet. Third, previous meta-analysis studies only analyzed the effects of three moderators or only focused on the specific subject domains, such as health professional or pharmacy education. There is a lack of a comprehensive meta-analysis to examine more moderators and cover all studies from 2013 to 2019. The present study is an attempt to fill the above research gaps.

1.3. Research questions

The purpose of the present study is twofold: the first aims to evaluate the effectiveness of the flipped classroom approach. Another is to examine whether moderator variables influence the effects of the flipped classroom on learning achievement. This study examined the effects of 12 moderator variables, including sample levels, sample size, learning domains, flip classroom models, research design, intervention durations, teaching methods in the class, sample regions, interactions in a pre-class and face-to-face class, tools in pre-class, and resources in pre-class. Therefore, the following research questions were proposed:

- What is the overall effectiveness of using the flipped classroom on students' learning achievement and learning motivation compared to the traditional classroom?
- How do various moderator variables influence on the effects of the flipped classroom?

2. Method

2.1. Data source

The data of this study were taken from three databases, including Web of Science, Scopus, and Eric. All of the studies relevant to the flipped classroom published from 2013 to 2019 were downloaded and further analyzed. Two sets of keywords were adopted to search research papers: (1) flipped classroom-related keywords, including flipped classroom, inverted classroom, flipped learning, flipped approach, and flipped-classroom; (2) learning achievement-related and learning motivation-related keywords, including learning outcome, learning achievement, achievement, outcome, learning performance, academic achievement, academic performance, learning motivation, motivation, and self-efficacy. The Boolean operator "AND" was adopted to integrate the two sets of keywords and the "OR" operator was used to connect within the set (Cooper, 2010).

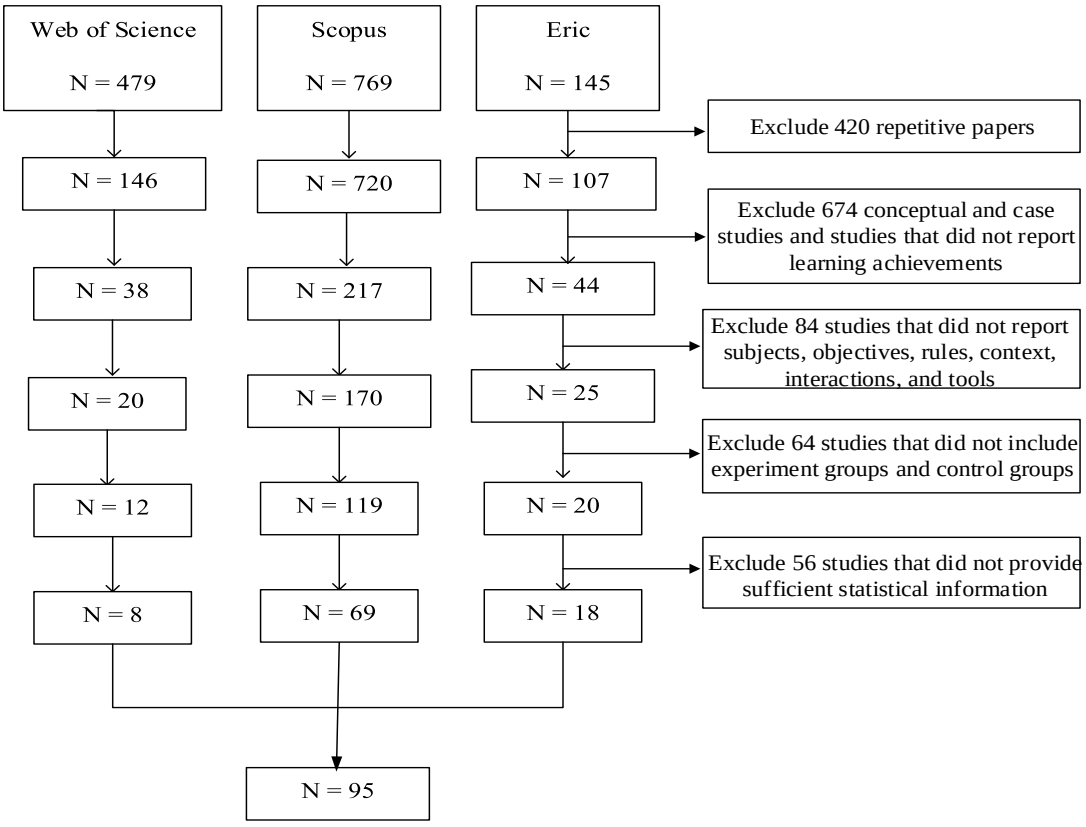

Figure 1. The search results

2.2. Search results

The research paper selection included two stages. The initial search yielded 1393 research papers, including 479 research papers from Web of Science, 769 research papers from Scopus, and 145 research papers from Eric. All of the research papers were examined, according to the following criteria:

- Studies published from 2013 to 2019.
- Research articles reported in English were only included in the present study. Studies not published in peer-reviewed journals (e.g., conference papers, book reviews, news, abstracts, and editorials) were excluded.
- The quasi-experimental or true-experimental studies were included. The conceptual studies were excluded. In addition, the selected studies should adopt the flipped classroom approach and report learning achievement and learning motivation.
- The selected studies should report how to implement the flipped classroom, including subjects, objectives, rules, context, interactions, and tools.

- The selected studies should include the experimental group and the control group. The studies should adopt the pretest to examine the equivalence of prior knowledge between the experimental and control groups. In addition, the instructors and learning content should be the same for the experimental and control groups.
- The selected studies should provide sufficient statistical information about learning achievement and motivation to calculate the effect size, such as means, standard deviations, *t* or *F* values, and the number of participants in each group.

Finally, 95 research papers were included in the present study for further analysis based on the above criteria. Figure 1 shows the search process and results.

2.3. Coding scheme

This study adopted activity theory as a model to analyze the features of the flipped classroom studies and the effects of moderator variables. Activity theory includes six components, including subject, object, mediating artifact, rules, community, and division of labor (Engeström, 1987). Engeström (2001) believed that activity theory represented the elements of learning activities and how learning activities occur. Moreover, previous studies also adopted activity theory to analyze different learning activities (Chung, Hwang, & Lai, 2019; Zheng et al., 2019). Figure 2 shows the adapted framework based on previous studies (Engeström, 1987; Sung, Yang, & Lee, 2017), and it includes six elements: subjects, objectives, rules, context, communication, and tools.

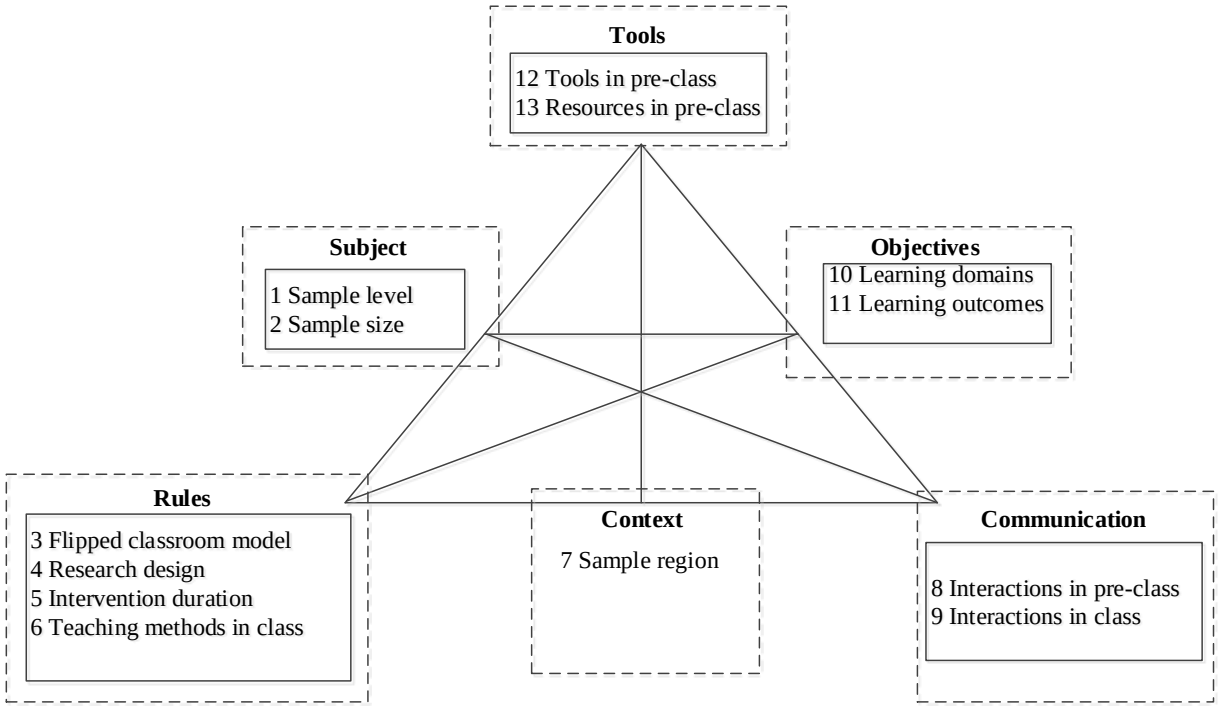

Figure 2. The analysis framework for flipped classroom based on activity theory

Table 1 shows the coding scheme in detail. Regarding learning outcomes, it includes learning achievement and learning motivation. Learning achievement is usually measured by standardized, teacher-made, or research-made tests to evaluate learners' knowledge acquisition or utilization (Sung, Yang, & Lee, 2017). Learning motivation was conceptualized as an established pattern of pursuing goals, beliefs, and emotions (Ford, 1992). In addition, the flipped classroom model included the traditional flipped classroom model and the innovative flipped classroom model. The traditional flipped classroom model refers to a teaching strategy that reverses what is done inside the classroom and outside the classroom (Abeysekera & Dawson, 2015). Innovative flipped classroom model refers to a new teaching strategy that integrating traditional flipped classroom model into other learning approaches such as social inquiry learning approach, problem-based learning, and so on. The coding scheme was developed based on the studies conducted by Zheng (2016), Zheng et al. (2019), and Sung, Chang, & Liu (2016). The coding process

included three steps proposed by Cooper (2010). First, three coders achieved a consensus about the definition of all entries by analyzing two papers. Second, three coders selected ten papers, independently coded, and negotiated until they achieve the consensus. Third, all of the rest papers were analyzed by three coders. The inter-coder Kappa reliability was 0.91.

Table 1. The coding scheme

Super-dimensions	Sub-dimensions	Coding scheme
Subjects	Sample level	(1) Primary school; (2) Junior and Senior High School; (3) Higher education.
	Sample size	(1) 1–50; (2) 51–100; (3) 101–300; (4) More than 300.
Objectives	Learning domain	(1) Natural Science (including science, mathematics, physics, biology, geography); (2) Social Science (including politics, education, psychology, linguistics); (3) Engineering & Technological Science (including engineering, computer science, educational technology); (4) Medical Science (including health and medicine).
	Learning outcome	Learning achievement; Learning motivation.
Rules	Flipped classroom model	(1) Traditional flipped classroom model; (2) Innovative flipped classroom model (e.g., technology enhanced flipped classroom, "Flipped" social inquiry learning model, clicker-aided flipped classroom, modern flipped classroom model, partial flipped classroom, flipped-blended classroom, in-flipped classroom, problem-based learning with flipped classroom).
	Research design	(1) True experimental design; (2) Quasi-experimental design.
	Intervention duration	(1) 2-4 weeks; (2) 5-8 weeks; (3) 9-24 weeks; (4) More than 24 weeks.
	Teaching method in F2F class	(1) One teaching method (e.g., problem-based learning or collaborative learning or self-directed study); (2) Two teaching methods (e.g., project-based learning and collaborative learning); (3) Three or more than three types of teaching methods (e.g., problem-based learning, collaborative learning, and inquiry-based learning).
Context	Sample region	(1) Africa; (2) Asia; (3) Europe; (4) North America; (5) Mixed region (e.g., China and US)
Communication	Interaction in pre-class	(1) Reading learning materials (one kind of interaction); (2) Watching the teaching videos (one kind of interaction); (3) Two types of interactions (e.g., watching the teaching videos, reading materials); (4) Three or more than three types of interactions (e.g., watch videos, reading learning materials, self-test)
	Interaction in F2F class	(1) Two types of interactions (e.g., group discussion and problem solving); (2) Three or more than three types of interactions (e.g., group discussion, presentation, and quiz).
Tools	Tools after class	(1) Online learning platform; (2) Others (Online discussion forum or game).
	Resources after class	(1) Video recordings; (2) Two types of resources (e.g., video recordings, readings); (3) Three or more than three types of resources (e.g., lectures, readings, video recordings).

2.4. Effect size calculation

The effect size calculation included four steps proposed by Borenstein, Hedges, Higgins, and Rothstein (2009). First, calculate the effect size of each study. Second, integrate the effect sizes of all studies to compute the overall weighted mean effect size by Hedges's *g*. Third, calculate the confidence interval for the overall mean effect size by the random effect model. Fourth, examine whether the moderator variables influenced the effect size through the Q_B value. A random-effect model was adopted to examine the impacts of moderator variables. The effect size was calculated using the Comprehensive Meta-analysis software. The publication bias was examined by the classic fail-safe N and Orwin's fail-safe N (Rosenthal, 1979). If the fail-safe N is above 5n+10 (n represents the number of studies), then it is unlikely to influence the effect size by the unpublished studies.

3. Results

3.1. Descriptive information

The present study analyzed the demographics of 95 studies and the features of the flipped classroom. The following sections will describe the results in detail. Table 2 presents the descriptive information of moderator variables and their percentages. There were 95 articles with 15,386 participants. With respect to subjects, the largest proportion of studies selected higher education and 51-300 participants. With regard to objectives, the most frequently selected learning domains were social science, followed by natural science and engineering as well as technological science. In terms of rules, most of the studies adopted quasi-experimental design to conduct studies for 9-24 weeks using the traditional flipped classroom model and two types of teaching methods. As for context, most studies implemented the flipped classroom in North America, followed by Asia. Concerning communication, most studies engaged participants in two types of interactions in pre-class and three or more than three types of interactions in class. Regarding tools, most studies adopted the online learning platform and two types of resources in the pre-class.

Table 2. The moderator variables categories and proportion of 95 studies

Variable	Category	No. of studies (k)	Proportion of studies
Sample levels	(1) Primary school	3	3.16%
	(2) Junior and Senior High School	14	14.74%
	(3) Higher education	78	82.1%
Sample size	(1) 1-50	13	13.69%
	(2) 51-100	36	37.89%
	(3) 101-300	36	37.89%
	(4) More than 300	10	10.53%
Subject domains	(1) Natural Science	30	31.58%
	(2) Social Science	34	35.79%
	(3) Engineering and Technological Science	16	16.84%
	(4) Medical Science	15	15.79%
Learning outcomes	(1) Learning achievements	95	100%
	(2) Learning motivation	9	9.47%
Research design	(1) Quasi-experimental design	90	94.74%
	(2) True experimental design	5	5.26%
Intervention durations	(1) 2-4 weeks	7	7.37%
	(2) 5-8 weeks	13	13.68%
	(3) 9-24 weeks	57	60.00%
	(4) More than 24 weeks	18	18.95%
Flipped classroom models	(1) Traditional flipped classroom model	81	85.26%
	(2) Innovative flipped classroom model	14	14.74%
Teaching methods in F2F class	(1) One teaching method	20	21.05%
	(2) Two teaching methods	60	63.16%
	(3) Three or more than three types of teaching methods	15	15.79%
Sample regions	(1) Africa	3	3.16%
	(2) Asia	42	44.21%

	(3) North America	43	45.26%
	(4) Europe	6	6.32%
	(5) Mixed region	1	1.05%
Interactions in pre-class	(1) Reading learning materials	5	5.26%
	(2) Watching the teaching videos	23	24.21%
	(3) Two types of interactions	43	45.27%
	(4) Three or more than three types of interactions	24	25.26%
Interactions in F2F class	(1) Two types of interactions	29	30.53%
	(2) Three or more than three types of interactions	66	69.47%
Tools in pre-class	(1) Online learning platform	89	93.68%
	(2) Others (Online discussion forum or game).	6	6.32%
Resources in pre-class	(1) Videos	25	26.32%
	(2) Two types of resources	49	51.57%
	(3) Three or more than three types of resources	21	22.11%

3.2. Overall effect size

Table 3 and Table 4 shows the overall effect sizes for learning achievement and learning motivation respectively. Based on the procedure of Borenstein et al. (2009), a random effect model was adopted to calculate the effect sizes of 95 selected studies. The results indicated that the overall effect size for learning achievement was 0.663, with a 95% confidence interval of 0.544-0.783. The effect sizes of 0.80, 0.50, and 0.20 were regarded as a larger, medium, and small effect size respectively based on Cohen's (1992) finding. Therefore, the flipped classroom approach had a medium effect size on students' learning achievement. The test of heterogeneity revealed that the effect sizes were heterogeneous in the present study (Q_{total} = 1192.145, z = 10.877, p < 0.001). In addition, the flipped classroom approach had a medium effect size on students' learning motivation (ES = 0.661). The results of heterogeneity analysis indicated the effect sizes were heterogeneous in this study (Q_{total} = 70.95, z = 2.999, p < 0.005). These findings also revealed that the significant differences among the effect sizes were due to sources other than subject-level sample error (Sung, Yang, & Lee, 2017).

Table 3. Overall effect sizes of learning achievement

	k	ES	*SE*	σ^2	95% CI		Test of mean		Test of heterogeneity		
					Lower	Upper	*Z*	*p*	*Q*	*df(Q)*	*p*
Fixed	95	0.501	0.016	0.000	0.468	0.533	30.473	.000	1192.145	94	.000
Random	95	0.663	0.061	0.004	0.544	0.783	10.877	.000			

Table 4. Overall effect sizes of learning motivation

	k	ES	*SE*	σ^2	95% CI		Test of mean		Test of heterogeneity		
					Lower	Upper	*Z*	*p*	*Q*	*df(Q)*	*p*
Fixed	9	0.437	0.071	0.005	0.299	0.576	6.196	.000	70.950	8	.000
Random	9	0.661	0.220	0.049	0.229	1.093	2.999	.003			

3.3. Effect sizes of learning achievements for moderator variables

The random-effect model was adopted to analyze the effect size of each moderator variable. Table 5 shows the results of twelve moderator variables.

3.3.1. Subjects

It was found that the flipped classroom studies implemented in junior and senior high school produced the largest effect size, followed by higher education and primary school. However, Q_B did not achieve statistical significance. Regarding the sample size, it was found that the sample size of 1-50 produced the largest effect size, followed by 51-100, 101-300, and more than 300. In addition, Q_B reached statistical significance (Q_B = 11.290, df = 3, p = .010), showing that the effect sizes of different sample sizes differed significantly.

3.3.2. Objectives

Table 5 demonstrated that the effect size for natural science domain achieved the highest effect size, followed by engineering & technological science, medical science, and social science. However, Q_B did not achieve statistical significance, which means that there was no significant difference among different subject domains.

3.3.3. Rules

Table 5 also indicated that the traditional flipped classroom model produced a larger effect size than the innovative flipped classroom model. However, the test of heterogeneity indicated that there was no significant difference between the two types of flipped classroom models. With respect to research design, the findings revealed that the true experimental design had a higher effect size and the quasi-experimental design had the lowest effect size. Both the two types of research design showed significant effect sizes. However, the Q_B did not achieve the significance, showing that the average effect sizes did not significantly differ between the true-experimental and quasi-experimental design.

With regard to the intervention duration, the findings indicated that interventions of 5-8 weeks had the largest effect size, followed by interventions of 2-4 weeks, interventions of 9-24 weeks, interventions of more than 24 weeks. Additionally, the Q_B was significant (Q_B = 9.458, df = 3, p = .024), which suggested that the average effect size differed significantly within the four types of intervention durations.

In terms of teaching methods in a face-to-face classroom, the results indicated that one teaching method had the largest effect size, followed by three or more than three types of teaching methods, and two teaching methods. However, the Q_B did not achieve the significance, showing that the average effect sizes did not significantly differ among different types of teaching methods.

3.3.4. Context

Table 5 indicated that the flipped classroom approach produced the largest effect size in Africa, followed by mixed region, Asia, Europe, and North America. The test of heterogeneity indicated that there was a significant difference among five types of sample regions (Q_B = 21.066, df = 4, p = .000).

3.3.5. Communications

This study analyzed two types of communications for flipped classroom studies. One was interaction in pre-class and another was interaction within class. It was found that watching the teaching videos yielded the largest effect size, followed by reading learning materials, two types of interactions, and three or more than three types of activities. However, the Q_B did not achieve the significance, showing that the average effect sizes did not significantly differ among different types of interactions. Concerning interactions in a face-to-face class, the results indicated that two types of interactions had the highest effect size and three or more than three types of interactions had the lowest effect size. However, the Q_B did not achieve the significance, showing that the average effect sizes did not significantly differ.

3.3.6. Tools

It was found that online discussion forum or game produced a larger effect size than the online learning platform. The test of heterogeneity indicated that there was no significant difference between the two types of tools in pre-class. In addition, the data given in Table 5 demonstrated that the effect size for video recordings achieved the highest effect size, followed by three or more than three types of resources. However, the Q_B did not achieve the significance, showing that the average effect sizes did not differ significantly.

Table 5. The analysis results for moderator variables

Category	*k*	*g*	*z*	95% CI	Q_B	*df*
Sample levels					0.828	2
1. Primary school	3	0.541	1.515	[-0.159,1.241]		
2. Junior and Senior High School	14	0.793	4.905***	[0.476,1.110]		
3. Higher education	78	0.646	9.576***	[0.513,0.778]		
Sample size					11.290**	3
1. 1-50	13	0.953	5.340***	[0.603,1.303]		
2. 51-100	36	0.830	8.250***	[0.633,1.028]		
3. 101-300	36	0.534	5.623***	[0.348,0.720]		
4. More than 300	10	0.312	1.781	[-0.031,0.655]		
Learning domains					1.266	3
1. Engineering& Technological Science	16	0.693	4.635***	[0.400,0.985]		
2. Medical Science	15	0.662	4.525***	[0.375,0.948]		
3. Natural Science	30	0.740	6.948***	[0.531,0.948]		
4. Social Science	34	0.576	5.602***	[0.375,0.778]		
Flipped classroom models					0.405	1
1. Traditional flipped classroom model	81	0.680	10.230***	[0.550,0.810]		
2. Innovative flipped classroom model	14	0.570	3.567***	[0.257,0.883]		
Research design					0.240	1
1. True experimental design	5	0.793	2.918**	[0.260,1.326]		
2. Quasi-experimental design	90	0.657	10.459***	[0.534,0.780]		
Intervention Durations					9.458*	3
1. 2-4 weeks	7	0.774	3.322***	[0.317,1.230]		
2. 5-8 weeks	13	1.112	6.439***	[0.774,1.451]		
3. 9-24 weeks	57	0.626	7.893***	[0.471,0.781]		
4. More than 24 weeks	18	0.458	3.346***	[0.190,0.726]		
Teaching methods in F2F class					4.753	2
1. One teaching method	20	0.891	6.721***	[0.631,1.150]		
2. Two teaching methods	60	0.570	7.561***	[0.422,0.717]		
3. Three or more than three types of teaching methods	15	0.743	4.785***	[0.439,1.047]		
Sample regions					21.066***	4
1. Africa	3	1.352	3.725***	[0.641,2.063]		
2. Asia	42	0.913	10.013	[0.734,1.091]		
3. Europe	6	0.627	2.668**	[0.166,1.088]		
4. North America	43	0.397	4.638***	[0.229,0.565]		
5. Mixed region	1	0.993	1.667	[0.734,1.091]		
Interactions in pre-class					2.712	3
1. Reading learning materials	5	0.698	2.707**	[0.193,1.202]		
2. Watching the teaching videos	23	0.774	6.349***	[0.535,1.014]		
3. Two types of interactions	43	0.685	7.733***	[0.512,0.859]		
4. Three or more than three types of interactions	24	0.504	4.238***	[0.271,0.737]		
Interactions in F2F class					0.109	1

1. Two types of interactions	29	0.695	6.190***	[0.475,0.915]		
2. Three or more than three types of interactions	66	0.651	8.839***	[0.506,0.795]		
Tools in pre-class					0.123	1
1. Online learning platform	89	0.658	10.448***	[0.534,0.781]		
2. Others	6	0.746	3.058**	[0.268,1.224]		
Resources in pre-class					3.521	2
1. Video recordings	25	0.838	7.308***	[0.614,1.063]		
2. Two types of resources	49	0.576	7.150***	[0.418,0.734]		
3. Three or more than three types of resources	21	0.647	5.201***	[0.403,0.891]		

Note. $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$.

3.4. Publication bias

The publication bias was evaluated by the funnel plot, the classic fail-safe N, and Orwin's fail-safe N. As shown in Figure 3, it was found that the funnel plot had symmetrical distribution. Therefore, there was no publication bias in the present meta-analysis. As shown in Table 6, the results of the classic fail-safe N indicated that 4885 missing studies would be needed to nullify the effect size, which was far larger than 485 (5n+10). Furthermore, the result of Orwin's fail-safe N revealed that 4662 missing studies would be needed to reduce Hedges's g to a trivial level (see Table 7). Therefore, the findings indicated that this meta-analysis was not affected by publication bias.

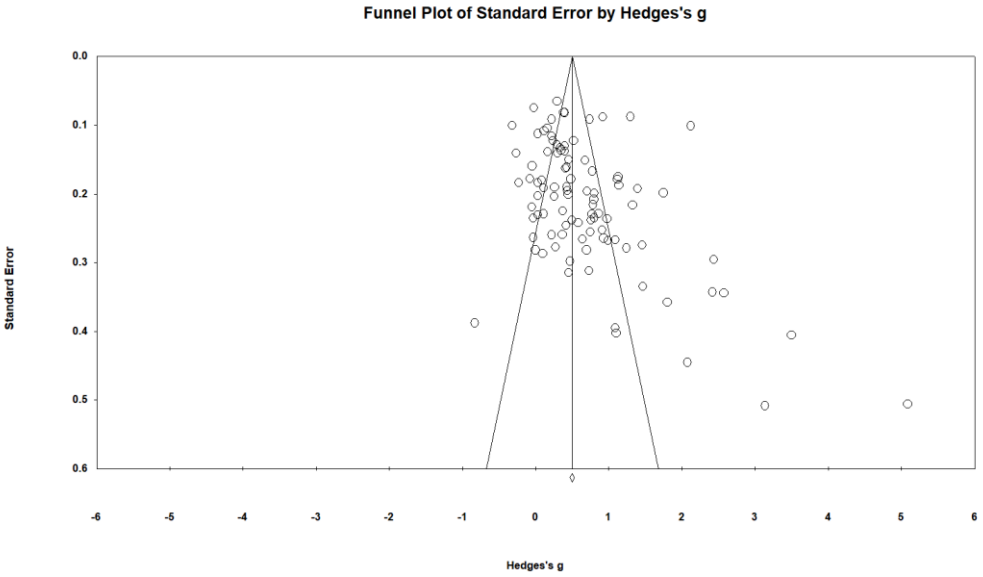

Figure 3. Funnel plot of standard error by effect size

Table 6. Classic fail-safe N

Items	Value
Z value for observed studies	31.845
p value for observed studies	0.000
Alpha	0.050
Tails	2.000
Z for alpha	1.960
Number of observed studies	95
Number of missing studies that would bring p value to > alpha	4885

Table 7. Orwin's fail-safe N

Items	Value
Hedges's g in observed studies	0.501
Criterion for a 'trivial' Hedges's g	0.010
Mean Hedges's g in missing studies	0.000
No. of missing studies needed to reduce Hedges's g to <0.01	4662

4. Discussion

This study examined the effects of the flipped classroom approach on students' learning achievement and learning motivation compared to traditional lecture-based instruction. Based on a total of 95 eligible studies with a total of 15386 students, it was found that the flipped classroom approach had an overall positive effect on students' learning achievement and learning motivation. The finding expanded the previous studies and revealed that the use of the flipped classroom had a significant impact on learning motivation through a comprehensive meta-analysis. The present study also provided substantial evidence on how the use of the flipped classroom was moderated by 12 variables, including sample levels, sample size, learning domains, flip classroom models, research design, intervention durations, teaching methods in the class, sample regions, interactions in a pre-class and face-to-face class, tools in pre-class, and resources in pre-class.

4.1. Sample level and sample size

For the sample level, it was found that there was no significant difference among the three sample levels. This result was similar to the findings of Cheng et al. (2019) in which they did not find any significant effect of sample level in the flipped classroom. The studies conducted at the junior and senior high school showed larger effects as compared to higher education. There was no significant effect size for the primary level. This may be because very few studies used the primary level as the sample for their study. Furthermore, this study also found that the small sample size had the largest effect size. The main reason was that the small sample size produced the less variation source, which led to the larger effect size (Slavin & Smith, 2009).

4.2. Learning domains

For learning domains, there was no significant difference among different subject domains. This finding indicated that learning domains did not have a significant impact on the effectiveness of the flipped classroom. This result might be explained by the fact that the appropriate use of flipped classroom would be effective in any learning domains that include real-world problems, design effective in-class learning activities, facilitate efficient interactions through information technologies, and integrate other pedagogical models according to the characteristics of different learning domains. Furthermore, natural science, engineering and technological science, medical science, and social science showed positive and medium effects size. However, Cheng et al. (2019) found that there was a significant difference in learning domains. The possible reason could be that the data sources and statistical information were different between the two studies.

4.3. Interventions

The findings revealed that there was no significant difference in the flipped classroom models. Therefore, the practitioners can select either the traditional flipped classroom or innovative classroom model. In addition, it was found that 94.7% of the studies in the present meta-analysis employed quasi-experimental design, and only 5.3% selected true experiments. The effect size of the true experimental design was larger than the quasi-experimental design. Therefore, more true-experimental studies need to be conducted in the flipped classroom research. Moreover, the present study revealed that the medium intervention duration produced the largest effect size. The main reason might be that too long durations will produce potential variation, and too short durations cannot validate the effectiveness of the flipped classroom. In terms of teaching methods in a face-to-face classroom, it was found that

there was no significant difference among different types of teaching methods. Thus, teachers and practitioners can select appropriate teaching methods based on instructional objectives and content.

4.4. Sample regions

The results indicated that there was a significant difference among the five types of sample regions. The studies conducted in the Africa region showed the most significant effects of the flipped classroom. The reason may be that the flipped classroom model is helping the developing countries to enhance the learning achievements and motivation of the students, which is a very significant output.

4.5. Interaction types

The findings revealed that different types of interactions in pre-class and the face-to-face class did not differ significantly. However, it was found that watching the teaching videos yielded the largest effect size. This could be explained that watching teaching videos is very important for a better understanding of learning content in a flipped classroom. Therefore, it is strongly recommended to develop high-quality videos that include recordings with elaborated instructional design, clear pictures, content-rich learning materials, and high-degree interactions, which can engage the learners prior to class.

4.6. Tools

The results indicated that the online discussion forum or game produced a larger effect size than the traditional online learning platform. The reason may be that the effective application of advanced technologies in the flipped classroom can promote learning achievement and learning motivation (Lin, 2019). In terms of resources in pre-class, it was found that video recordings achieved the highest effect size. Therefore, it is recommended to develop high-quality video recordings to facilitate the flipped classroom.

4.7. Implications

The present study has several implications for implementation of the flipped classroom, which are described and analyzed below.

4.7.1. Enhancing the research design quality for the flipped classroom interventions

The present meta-analysis found that different sample regions, sample sizes, and intervention durations had significant impacts on effect size. In order to enhance the research design quality, the following aspects may be considered by researchers and practitioners before the flipped classroom implementation. First, the characteristics of participants should be taken into account before the implementation of the flipped classroom approach. Students' experiences, prior knowledge, information and communication technology skills, and attitude toward the flipped classroom had great impacts on the effectiveness of the flipped classroom. Furthermore, if the participants come from mixed regions, their cultural background is another important factor for the flipped classroom interventions.

Second, the present meta-analysis indicated that less than 300 participants could produce a large effect size. The largest effect size was produced by less than 50 participants in this study. Previous studies reported that the appropriate sample size could ensure unbiased findings and estimates (McNeish & Stapleton, 2016). Therefore, it is suggested that the sample size should be less than 300 participants to decrease the potential variation source.

Third, the midterm intervention duration is more appropriate than shorter or longer intervention duration. Previous studies revealed that the intervention duration affected the reliability and validity of the research (Sung, Chang, & Liu, 2016). The present meta-analysis found that the midterm intervention duration (5-8 weeks) produced the largest effect size. It is very difficult to yield any effects for a too short duration. In addition, it will take lots of time to

introduce long-term flipped classroom implementation. Therefore, the teachers and practitioners may adopt the midterm intervention duration to implement the flipped classroom.

4.7.2. Integrating other pedagogical models with the flipped classroom approach

The appropriate pedagogy can improve the effectiveness of the flipped classroom. The traditional flipped classroom ignored the activity delivery and students' experiences (Chen, Wang, Kinshuk, & Chen, 2014). By integrating other pedagogical models such as collaborative learning, inquiry-based learning, and problem-based learning into the flipped classroom, the effectiveness of the flipped classroom can be maximized. These pedagogical models included modified flipped classroom (Scott, Green, & Etheridge, 2016), flipped social inquiry learning approach (Jong, 2017), clicker-aided flipped classroom (Yu & Wang, 2016), in-flipped classroom (Chiang, & Wang, 2015), problem-based learning with the flipped classroom (Tsai, Shen, & Lu, 2015), and so on. Therefore, it is suggested that educators and practitioners can harness an innovative pedagogy to implement the flipped classroom.

5. Conclusions

This meta-analysis provided substantial evidence for the positive effect of adopting the flipped classroom and how those effects were influenced by different moderator variables. The main findings were summarized as below:

- The flipped classroom approach had a moderate effect size of 0.663 for learning achievement and a moderate effect size of 0.661 for learning motivation.
- The results indicated that sample size, intervention durations, and sample regions significantly moderated the effect sizes.
- The small sample size (1-50) had a larger effect size than the large sample size (more than 50).
- The true experimental design had better effects than the quasi-experimental design.
- The midterm interventions (5-8 weeks) produced better effects than short duration (shorter than five weeks) and long-term intervention duration (longer than eight weeks).
- Watching the teaching videos yielded the largest effect size in pre-class, and videos produced a better effect size than other resources in pre-class.

These findings are very promising and provide insight into the implementation of the flipped classroom in the future. However, this study has several limitations. First, due to the limited empirical studies on the flipped classroom approach, only 95 empirical studies reported sufficient statistical information and descriptive information about the flipped classroom. In the future study, the data source, including grey literature and unpublished studies, needs to be expanded further to get a more comprehensive understanding of the flipped classroom. Second, the present study analyzed the effects of 12 moderators. Further studies are needed to explore the effect sizes of other moderators. For example, achievement indicators (standardized achievement test or self-reported grades) are also a source of variance that can be analyzed as a moderator. Finally, this study only analyzed the effects of the flipped classroom approach on learning achievement and learning motivation. Future studies may examine the effects of the flipped classroom approach on other dependent variables, such as learning behavior or learning attitude.

Acknowledgement

This study is funded by the youth project of Humanities and Social Science Research in the Ministry Education (19YJC880141) and National Natural Science Foundation of China (61907003).

References

Abeysekera, L., & Dawson, P. (2015). Motivation and cognitive load in the flipped classroom: Definition, rationale and a call for research. *Higher Education Research and Development*, *34*, 1–14. doi:10.1080/07294360.2014.934336

Bergmann, J., & Sams, A. (2012). *Flip your classroom: Reach every student in every class every day*. Washington, DC: Internal Society for Technology in Education.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, UK: Wiley.

Chen, Y., Wang, Y., Kinshuk, & Chen, N. S. (2014). Is FLIP enough? Or should we use the FLIPPED model instead? *Computers & Education*, *79*, 16–27. doi:10.1016/j.compedu.2014.07.004

Cheng, L., Ritzhaupt, A. D., & Antonenko, P. (2019). Effects of the flipped classroom instructional strategy on students' learning outcomes: A Meta-analysis. *Educational Technology Research and Development, 67*(4), 793–824. doi:10.1007/s11423-018-9633-7

Chiang, Y. H., & Wang, H. C. (2015). Effects of the in-flipped classroom on the learning environment of database engineering. *International Journal of Engineering Education*, *31*(2), 454–460.

Chung, C. J., Hwang, G. J., & Lai, C. L. (2019). A review of experimental mobile learning research in 2010–2016 based on the activity theory framework. *Computers & Education*, *129*, 1–13. doi:10.1016/j.compedu.2018.10.010

Cohen, J. (1992). A Power primer. *Psychological Bulletin*, *112*, 155–159. doi:10.1037/0033-2909.112.1.155

Cooper, H. (2010). *Research synthesis and meta-analysis: A step-by-step approach* (4th ed.). London, UK: Sage Publications.

Engeström, Y. (1987). *Learning by expanding: An activity-theoretical approach to developmental research*. Helsinki, Finland: Prienta-Konsultit Oy.

Engeström, Y. (2001). Expansive learning at work: Toward an activity-theoretical reconceptualization. *Journal of Education and Work*, *14*(1), 133–156. doi:10.1080/13639080020028747

Ford, M. E. (1992). *Human motivation: Goals, emotions, and personal agency beliefs*. Newbury Park, CA: Sage.

Francl, T. J. (2014). Is flipped learning appropriate. *Journal of Research in Innovative Teaching, 71*, 119–128.

Gillette, C., Rudolph, M., Kimble, C., Rockich-Winston, N., Smith, L., & Broedel-Zaugg, K. (2018). A Meta-analysis of outcomes comparing flipped classroom and lecture. *American journal of pharmaceutical education, 82*(5), 6898–6898. doi:10.5688/ajpe6898

Hew, K. F., & Lo, C. K. (2018). Flipped classroom improves student learning in health professions education: A Meta-analysis. *BMC Medical Education, 18*: 38. doi:10.1186/s12909-018-1144-z

Jong, M. S. Y. (2017). Empowering students in the process of social inquiry learning through flipping the classroom. *Educational Technology & Society, 20*(1), 306–322.

Kashada, A., Li, H., & Su, C. (2017). Adoption of flipped classrooms in K-12 education in developing countries: Challenges and obstacles. *International Journal of Emerging Technologies in Learning*, *12*(10), 147–157. doi:10.3991/ijet.v12i10.7308

Lai, C.-L., & Hwang, G.-J. (2016). A Self-regulated flipped classroom approach to improving students' learning performance in a mathematics course. *Computers & Education, 100*, 126–140. doi:10.1016/j.compedu.2016.05.006

Lin, C. J., Hwang, G. J., Fu, Q. K., & Chen, J. F. (2018). A Flipped contextual game-based learning approach to enhancing EFL students' English business writing performance and reflective behaviors. *Journal of Educational Technology & Society*, *21*(3), 117–131.

Lin, Y. T. (2019). Impacts of a flipped classroom with a smart learning diagnosis system on students' learning performance, perception, and problem solving ability in a software engineering course. *Computers in Human Behavior*, *95*, 187–196. doi:10.1016/j.chb.2018.11.036

Lo, C. K., & Hew, K. F. (2017). A Critical review of flipped classroom challenges in K-12 education: Possible solutions and recommendations for future research. *Research and Practice in Technology Enhanced Learning, 12*, 4. doi:10.1186/s41039-016-0044-2

McNeish, D. M., & Stapleton, L. M. (2016). The Effect of small sample size on two-level model estimates: A Review and illustration. *Educational Psychology Review, 28*(2), 295–314. doi:10.1007/s10648-014-9287-x

Mellefont, L. A., & Fei, J. G. (2016). Student perceptions of 'flipped' microbiology laboratory classes. *International Journal of Innovation in Science and Mathematics Education*, *24*(1), 24–35.

Milman, N. (2012). The Flipped classroom strategy: What is it and how can it be used? *Distance Learning, 9*(3), 85–87.

Nederveld, A., & Berge, Z. L. (2015). Flipped learning in the workplace. *Journal of Workplace Learning*, *27*(2), 162–172. doi:10.1108/JWL-06-2014-0044

Njie-Carr, V. P., Ludeman, E., Lee, M. C., Dordunoo, D., Trocky, N. M., & Jenkins, L. S. (2017). An Integrative review of flipped classroom teaching models in nursing education. *Journal of Professional Nursing*, *33*(2), 133–144. doi:10.1016/j.profnurs.2016.07.001

O'Flaherty, J., & Phillips, C. (2015). The Use of flipped classrooms in higher education: A Scoping review. *The Internet and Higher Education*, *25*(4), 85–95. doi:10.1016/j.iheduc.2015.02.002

Rahman, A. A., Aris, B., Mohamed, H., & Zaid, N. M. (2014). The Influences of flipped classroom: A Meta-analysis. In *Proceedings of IEEE 6th Conference on Engineering Education (ICEED)* (pp. 24–28). doi:10.1109/ICEED.2014.7194682

Rosenthal, R. (1979). The File drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638–641. doi:10.1037/0033-2909.86.3.638

Scott, C. E., Green, L. E., & Etheridge, D. L. (2016). A Comparison between flipped and lecture-based instruction in the calculus classroom. *Journal of Applied Research in Higher Education*, *8*(2), 252–264. doi:10.1108/JARHE-04-2015-0024

Seery, M. K. (2015). Flipped learning in higher education chemistry: emerging trends and potential directions. *Chemistry Education Research and Practice*, *16*(4), 758–768. doi:10.1039/C5RP00136F

Sergis, S., Sampson, D. G., & Pelliccione, L. (2018). Investigating the impact of Flipped Classroom on students' learning experiences: A Self-Determination Theory approach. *Computers in Human Behavior*, *78*, 368–378. doi:10.1016/j.chb.2017.08.01

Slavin, R. E., & Smith, D. (2009). Effects of sample size on effect size in systematic reviews in education. *Educational Evaluation and Policy Analysis*, *31*(4), 500–506. doi:10.3102/0162373709352369

Sparks, R. J. (2013). Flipping the classroom: An Empirical study examining student learning. *Journal of Learning in Higher Education*, *9*(2), 65–70.

Strayer, J. F. (2012). How learning in an inverted classroom influences cooperation, innovation and task orientation. *Learning environments research*, *15*(2), 171–193. doi:10.1007/s10984-012-9108-4.

Sung, Y. T., Chang, K. E., & Liu, T. C. (2016). The Effects of integrating mobile devices with teaching and learning on students' learning performance: A Meta-analysis and research synthesis. *Computers & Education*, *94*, 252–275. doi:10.1016/j.compedu.2015.11.008

Sung, Y. T., Yang, J. M., & Lee, H. Y. (2017). The Effects of mobile-computer-supported collaborative learning: Meta-analysis and critical synthesis. *Review of Educational Research*, *87*(4), 768–805. doi:10.3102/0034654317704307.

Toto, R., & Nguyen, H. (2009). Flipping the work design in an industrial engineering course. In *Proceedings of 39th IEEE Frontiers in Education Conference* (pp. 1–4). San Antonio, TX: IEEE. doi:10.1109/FIE.2009.5350529.

Tsai, C. W., Shen, P. D., & Lu, Y. J. (2015). The Effects of problem-based learning with flipped classroom on elementary students' computing skills: A case study of the production of ebooks. *International Journal of Information and Communication Technology Education*, *11*(2), 32–40. doi:10.4018/ijicte.2015040103

Tucker, B. (2012). The flipped classroom. *Education Next*, *12*(1), 82–83.

Wang, T. (2017). Overcoming barriers to 'flip': Building teacher's capacity for the adoption of flipped classroom in Hong Kong secondary schools. *Research and practice in technology enhanced learning*, *12*(1), 6. doi:10.1186/s41039-017-0047-7

Yu, Z., & Wang, G. (2016). Academic achievements and satisfaction of the clicker-aided flipped business English writing class. *Journal of Educational Technology & Society*, *19*(2), 298–312.

Zheng, L., Chen, N. S., Cui, P., & Zhang, X. (2019). A Systematic review of technology-supported peer assessment research. *The International Review of Research in Open and Distributed Learning*, *20*(5), 168–191. doi:10.19173/irrodl.v20i5.4333

Zheng, L. (2016). The Effectiveness of self-regulated learning scaffolds on academic performance in computer-based learning environments: A Meta-analysis. *Asia Pacific Education Review*, *17*(2), 187–202. doi:10.1007/s12564-016-9426-9