



Flipped classroom improves academic achievement, learning retention and attitude towards course: a meta-analysis

Özgür Tural¹ · Taha Yazar²

Received: 25 December 2020 / Revised: 17 May 2021 / Accepted: 24 May 2021 / Published online: 9 June 2021
© Education Research Institute, Seoul National University, Seoul, Korea 2021

Abstract

Numerous studies have been conducted to date on the effectiveness of the flipped classroom. However, the results of studies reporting on the model are generally inconsistent. The aim of this research was to evaluate the results of these primary studies in a comprehensive manner, and to obtain general results to document the effectiveness of flipped classroom, also expose the impact of various study characteristics on this effect. To this end, a meta-analysis of primary studies examining the impact of flipped classroom on academic achievement, learning retention and attitude towards course was conducted. 177 studies for the academic achievement, 9 studies for the learning retention and 17 studies for the attitude towards course variables that meet the inclusion criteria were coded and analyzed. Additionally, moderator analyses were conducted for 8 possible moderator variables. The results of the analysis indicated a moderate main effect size for the effectiveness of flipped classroom on academic achievement ($g = 0.764$) and learning retention ($g = 0.601$) also a modest main effect size on attitude towards course ($g = 0.406$). Through to this meta-analysis we have learned that in terms of academic achievement the flipped classroom phenomenon; (a) has been implemented more effectively in small classes, (b) has been applied most effectively in primary schools, (c) its effectiveness has decreased as the duration of implementation extends, (d) has been implemented effectively in almost all domain subjects, (e) also in terms of attitude towards course and learning retention it has been more effective than traditional lecture-based instruction.

Keywords Flipped classroom · Meta-analysis · Academic achievement · Attitude · Retention

Introduction

Today's learning paradigm argues that learning is actively structured by learners (Birenbaum, 2003). Learning takes place only when learners get out of the passive listener position and become active participants (Hawtrey, 2007). In a related manner, many active learning methods have been developed by educators and researchers who are aware of this situation and also the search for alternative methods aiming to involve students in the learning process is still

continuing (Findlay-Thompson & Mombourquette, 2014; Bhagat et al., 2016).

These days (i.e. during the information age), the developments in information and communication technologies and electronic devices have become an indispensable part of students' life and the potential of technology to improve education have not been ignored. As a result, blended learning models that use technology to achieve learning goals have come to the forefront, among which flipped classroom is a recently developed model exclusively based on active learning (Mohanty & Parida, 2016; Mzoughi, 2015; Strayer, 2012).

Flipped classroom is designed as an inverted instructional model since it reverses the lecture and the practice modules of a course (Lee et al., 2016). In this model, learners perform in-class routines such as listening to the lecture and observing the teacher outside the classroom and listen to and watch pre-prepared video lessons on their own by using technological devices such as computers or smart phones. In doing so, they use their in-class time for

This meta-analysis is an extended and updated version of the paper presented at the Fifth International Congress on Curriculum and Instruction, Muğla, Turkey, October, 2017.

✉ Özgür Tural
ozgurtural@windowsslive.com

¹ Ministry of National Education, Diyarbakır, Turkey

² Department of Educational Sciences, Dicle University, Diyarbakır, Turkey

active learning activities such as discussion, group work, peer instruction and problem solving (Bishop & Verleger, 2013; Yestrebsky, 2016).

The implementation of the flipped classroom model brings about several benefits. For one thing, learners can watch videos as much as they want, at any time and place, which allows them to progress at their own pace and control the speed and time needed to learn the material. As such, with flipped classroom, learners can assume the ownership of their own learning (Bates et al., 2017; Fulton, 2012; Mok, 2014). In the classroom, the model, by decreasing the time used for lecturing, maximizes the time used by the teacher for active-learning activities, thereby enhancing teacher-student interaction, student-student interaction, as well as collaboration. As a result of these, teachers can have more time to respond to students' individual demands and needs (Bergmann & Sams, 2012, 2014; Sota, 2016). Taken together, these notions clearly indicate that flipped classroom provides better opportunities for inclusive, differentiated, and personalized education, thus leaving out traditional classroom practices (Spector, 2016).

Active-learning experiences are highly likely to show the rationale behind the subject being studied (de Caprariis et al., 2001), which thereby increases the level of learning by providing more information processing, better understanding and better retention (Beard & Wilson, 2005; Taylor & MacKenney, 2008). Moreover properly applied student-centered teaching will lead to more positive attitude towards course (Collins & O'Brien, 2003). Attitude is a psychological construct that is seen as an important predictor of individual behavior (Ajzen & Gilbert-Cote, 2008) and has cognitive, affective and behavioral factors (Toraman & Ulubey, 2016). Affective and cognitive factors play an important role in the learning process with their in-depth interaction (Di Martino & Zan, 2011). It is known that academic achievement is directly or indirectly related to many factors. When considered in this context, it can be thought that affective factors such as attitude may affect many factors, especially students' desire and interest towards the course, which may affect students' performance and thus their academic achievement (Kan & Akbaş, 2005). Students' attitudes towards a course may make a difference between their success or failure in that course. For this reason, in many experimental studies, the effect of active learning on students' academic achievement as well as their attitudes towards the course and learning retention was also investigated. These effects have been demonstrated either by comprehensive meta-analyses investigating the effectiveness of active learning (Freeman et al., 2014; Tural, 2019). Considering that active learning enhances academic achievement, provides greater learning retention and leads to more positive attitude towards course, flipped classroom, which is a

recently developed model based on active learning, is likely to have similar effects.

The main purpose of flipped classroom is also to increase student achievement via active-learning activities (Zainuddin & Halili, 2016) and this educational model is theorized to improve learner engagement and retention (Rose et al., 2016). Nevertheless, the results of studies reporting on the effectiveness of the flipped classroom model are generally inconsistent. To illustrate, while many studies indicated that flipped classroom has a positive effect on the variables in question (Ceylan, 2015; Heyborne & Perrett, 2016; Özdemir, 2016; Wiginton, 2013), some studies did not reveal a statistically significant difference (Dixon, 2017; Elakovich, 2018; Makinde & Yusuf, 2017; Perçin, 2019) and some others even reported on a negative effect (Akdeniz, 2019; Carlisle, 2018; Howell, 2013; Johnson & Renner, 2012). Accordingly, these contradictions indicate the need for meta-analyses presenting general conclusions on the effectiveness of the flipped classroom.

Review of literature

Despite being a relatively new model, flipped classroom has been a major concern among researchers. Tan et al. (2017) analyzed 29 studies in Chinese nursing education to examine learners' academic performance both in knowledge and skills and found strong effect sizes for the effectiveness of flipped classroom on knowledge ($SMD = 1.13$) and skills ($SMD = 1.68$). Hew and Lo (2018) analyzed 28 studies conducted in the realm of health professions education and found that the flipped classroom model had a modest effect on students' academic achievement ($SMD = 0.33$). In addition, the authors also evaluated the effect of seven possible moderator variables on the results and determined that the quiz implementations performed prior to classroom activities significantly increased the effect size. In contrast, Gillette et al. (2018) conducted a meta-analysis to identify the influence of flipped classroom on pharmacy students' educational outcomes by combining the results of five studies. The authors concluded that the model did not lead to a significant difference in students' academic achievement. K. S. Chen et al. (2018) reviewed the results of 46 studies conducted in health sciences and some other disciplines at tertiary level and found a moderate effect ($SMD = 0.47$). Shi et al. (2019) conducted a meta-analysis to determine the effect of the model on college students' cognitive learning outcomes and after combining the results of 33 multidisciplinary studies, the authors indicated a moderate main effect size ($SMD = 0.53$). In the same study, the moderator analyses indicated that only pedagogical approach was statistically significant among the seven potential moderator variables. Orhan (2019) analyzed 13 studies published in Turkey that

conducted in secondary and tertiary levels. The author indicated a moderate main effect size ($SMD = 0.74$). Moderator analyses showed that the study result was not moderated by the education level, type of study and publication year. Martínez et al. (2019) meta-analyzed 12 studies that assessed the effect of flipped classroom on university students' academic performance. In the study which moderator analyses was not performed, it was determined that the mean effect size value was strong ($SMD = 2.29$).

In addition to the studies abovementioned, several recently published meta-analyses evaluated studies conducted in a wide range of disciplines and with students from different educational levels. In one of these, van Alten et al. (2019) analyzed 114 studies conducted at secondary and postsecondary levels to determine the effect of flipped classroom on the assessed and perceived learning outcomes. The authors found modest effects on the assessed ($SMD = 0.36$) and perceived ($SMD = 0.36$) learning outcomes. Additionally, their moderator analyses showed that higher effect size values could be achieved when the time allocated for face-to-face training increased and when the quizzes were performed during treatments. Karagöl and Esen (2019) evaluated the results of 80 studies conducted at primary, high school and tertiary levels between 2012 and 2017. The result of the study indicated that flipped classroom had a moderate effect ($SMD = 0.566$) on students' academic achievement. Moreover, it was also determined that the main effect size value was moderated by the sample size and by the geographical location where the studies were conducted, among the five different possible moderator variables. In their meta-analysis, Cheng et al. (2019) combined the results of 55 studies, the participants of which were graduates, K-12 and university students, published between 2000 and 2016. The authors found that the flipped classroom model had a weak effect ($SMD = 0.193$) on students' learning outcomes. Their moderator analyses also showed that among the four possible moderator variables the effect sizes were moderated only by subject area and the largest effect size detected was in the Arts and Humanities. Låg and Sæle (2019) reviewed the results of 272 studies conducted at primary, secondary and tertiary levels, between 2010 and 2017. Their analysis indicated that flipped classroom had a modest effect ($SMD = 0.35$) on student learning, a small effect on pass rates (odds ratio = 1.55), and a weak effect ($SMD = 0.16$) on student satisfaction. The overall effect on student learning moderated by disciplines significantly in favour of the Humanities. Also, in the analysis, it was concluded that the effects of education level, the test of preparation and social activity on learning outcomes were not statistically significant.

Some of the studies mentioned above are meta-analyses that were conducted in a certain discipline or at a certain level of education with relatively small samples. Such

limitations impede the generalizability of the results of those studies. And extreme values that can be found among the effect sizes of individual studies may also distort the main effect size value and cause it to be estimated lower or higher than it is when meta-analyses are performed with a low number of studies. In addition, it is unlikely that reliable moderator analyzes can be performed, as there will not be enough studies in the subgroups of such meta-analyses. On the other hand, due to the inconsistent findings obtained from the meta-analyses published thus far, there appears to be a literature gap regarding the effectiveness of flipped class on academic achievement. Accordingly, there is need for meta-analyses evaluating the efficacy of flipped classroom on academic achievement from a wider perspective (K. S. Chen et al., 2018; Hughes & Lyons, 2017) also on learning retention and attitude towards course.

Purpose of the present study

In the present meta-analysis, we aimed to eliminate some of the limitations mentioned above by using a set of inclusion criteria and to evaluate the subject matter from a wider perspective. In order to broaden the scope of the meta-analysis, we included studies conducted at primary, secondary, and tertiary levels. Instead of focusing on a single domain, we included all the domains evaluated by the studies reached. The literature on flipped classroom is dominated by the studies published in English language. Additionally, we included the studies published in Turkish language, though few in number. However, we also included the studies published in 2018 and 2019 to perform an up-to-date meta-analysis. Besides, we included experimental and quasi-experimental studies that involved a concurred control group and implemented group equivalence tests. By applying the limitations above, we aimed to increase the statistical power of the meta-analysis, to enable more robust moderator analyses, and to evaluate the subject matter from a wider perspective, thus contributing to the scientific and practical results presented by previous meta-analyses. In order to better understand this relatively new model, although there are few studies in the literature, we also examined its effect on learning retention and attitude towards course, which can be considered closely related to academic achievement.

Accordingly, the primary aim of the present study was to review the results of experimental and quasi-experimental studies that examined the effect of flipped classroom on students' academic achievement, learning retention and attitude towards course by meta-analysis and to reveal the effect of various study characteristics on the results presented by previous studies. To this end, the following questions were addressed in the present study:

- (1) What is the effectiveness of flipped classroom versus traditional lecture-based instruction on students' academic achievement?
- (2) What factors moderate the effectiveness of flipped classroom (if any) on students' academic achievement?
- (3) What is the effectiveness of flipped classroom versus traditional lecture-based instruction on learning retention?
- (4) What is the effectiveness of flipped classroom versus traditional lecture-based instruction on students' attitude towards course?

Method

Literature search

The first step in identifying the target experimental and quasi-experimental studies was performing a literature search in electronic databases (Table 1). The search was conducted three times in July 2017, June 2019 and February 2020.

Given that checking the reference lists leads to both reaching as many individual studies as possible and increasing the validity of the meta-analysis (Brunton et al., 2012), in the second step we manually checked the reference lists of other meta-analyses and systematic reviews ($k=29$) conducted on flipped classroom, which yielded additional 1091 studies. Eventually, the literature search yielded a total of 4463 relevant studies. However, we excluded proceedings since most of them were not indexed in the databases and were not available in full text. The stages of the literature search and the number of studies examined at each phase are presented in the flow diagram below (Fig. 1).

Inclusion criteria and coding process

The following criteria were implemented for the primary studies to be included in the meta-analysis:

- (1) Examining the effect of flipped classroom on students' academic achievement, learning retention or attitude towards course
- (2) Having an experimental or quasi-experimental design and a pretest-posttest control group design
- (3) Theses, dissertations or articles published in peer-reviewed journals
- (4) Published in English or Turkish languages
- (5) Published before 2020
- (6) Containing sufficient statistical information (such as X , N , SD , t or p values) for the calculation of effect sizes
- (7) Including participants at primary to tertiary levels
- (8) Published in full text

Prior to the coding process, the first author excluded 1606 duplicates. Subsequently, a total of 2129 other studies were eliminated since they were not experimental or quasi-experimental, not published in English or Turkish, not a thesis, dissertation or article, not available in full text, their participants were not at elementary to tertiary levels, or did not address the purpose of the present study. As a result, the remaining 728 full-text studies were evaluated and coded by the first author and a second researcher who has a master's degree in biology education (Table 2).

After the coding training in which the coders were provided detailed definitions, descriptions of codes, and practice coding, the coders coded sixteen different variables for each individual study. As per the inclusion criteria, they selected studies which contained sufficient statistical information for

Table 1 Overview of literature search

Database	Keywords searched	Results
ERIC	"flipped classroom", "flipped learning", "inverted classroom", "inverted learning", "flip classroom", "flip learning", "invert classroom", "invert learning"	1096
Web of Science		1400
ProQuest Dissertation and Theses*		248
Yök Ulusal Tez Merkezi	"ters yüz", "ters düz", "tersine eğitim", "çevrilmiş öğrenme", "flipped classroom", "flipped learning", "inverted classroom", "inverted learning"	253
Ulakbim		71
Asos İndeks		7
Türk Eğitim İndeksi		22
Akademik Dizin		15
Araştırmaz		25
Yök Akademik		123
Sobiad		112
Total		3372

*To reduce the number of irrelevant studies, we used several descriptors including "academic achievement", "student achievement", "achievement", "student success", "success" and "academic performance" in the ProQuest Dissertation and Theses database

Fig. 1 PRISMA flow diagram for the literature search process. Some of the studies have reported multiple variables

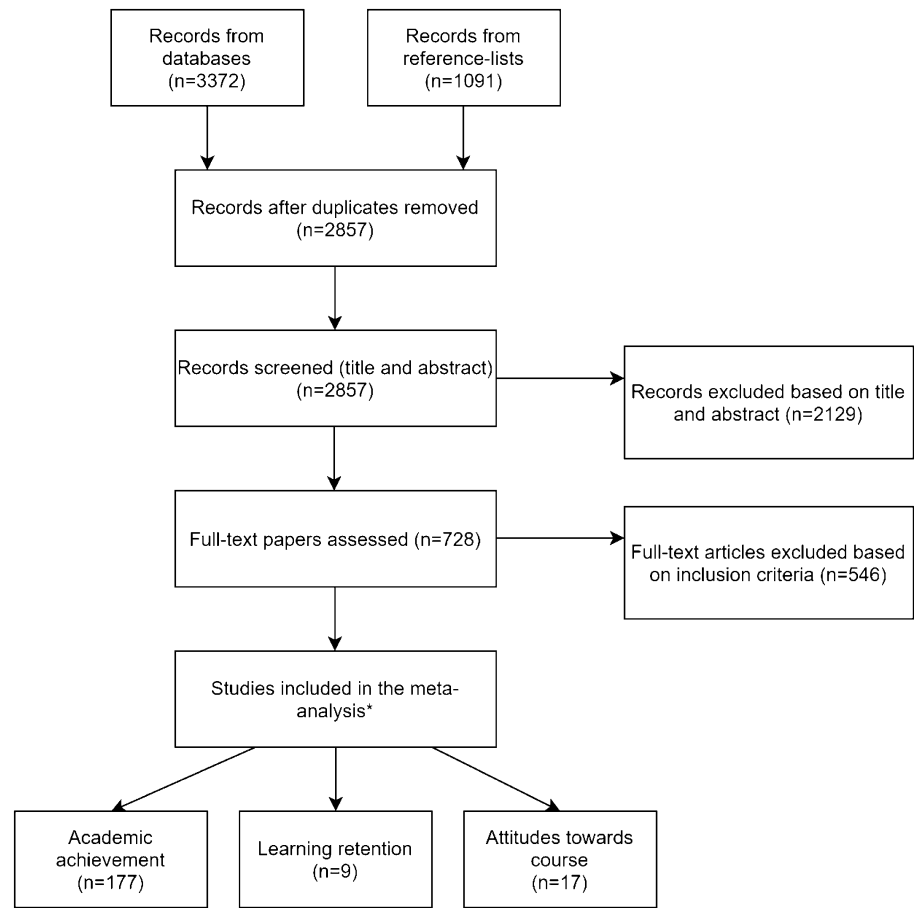


Table 2 Rationales for including and excluding studies

Criterion	Rationale for inclusion	Rationale for exclusion	Number of studies excluded
Research design	Experimental or quasi-experimental with pretest-posttest and control group	Non-experimental or without pretest or control group	1535
Statistical information	Sufficient statistical information to calculate effect size	Insufficient statistical information	67
Language	Published in English or Turkish	Published in languages other than English or Turkish	75
Publication type	Thesis, dissertation or journal article	Book chapters, reports, proceedings papers	627
Availability	Full-text studies available on the internet or university library systems	Title or abstract only	63
Variables	Dependent variable: academic achievement, learning retention or attitudes towards course; independent variable: flipped classroom	Studies with different dependent or independent variables	274
Participants	Elementary, middle school, high school or college students	Participants pre-school students, postgraduates or teachers	34

the calculation of effect sizes and a pretest-posttest control design (Table 2).

It is possible for a study to report more than one effect size. The procedure followed by coders for studies with multiple effect sizes was as follows. In cases where a study reported multiple measurements (quizzes, midterm exams, post-test), the coders chose only the post-test if reported. However, if studies ($k = 19$) reported multiple post-test results (post-tests of interventions in different units or post-tests whose results were not reported as total scores, reported based on different dimensions such as reading, writing, listening, and speech), each effect size was included separately. In studies ($k = 3$) using multiple independent samples (e.g., students from different disciplines, genders or achievement levels), each sample group's effect size was included separately. Similarly, each comparison's estimated effect size was included separately for comparisons where there were multiple experimental groups versus a single control group ($k = 6$).

There are different ways of coding studies with multiple effect sizes in the literature. It can be done by selecting a random one of the different effect sizes in a study or averaging these effect sizes. However, this way is not recommended, as it leads to loss of data and statistical power (Cheung, 2014; Marín-Martínez & Sánchez-Meca, 1999; Scammacca et al., 2014). Also, because it combines conceptually different outcome measures that may be statistically irrelevant (Littell et al., 2008). Another way is incorporating all effect sizes into the meta-analysis separately. But this situation also has some drawbacks. When this method is preferred, statistical dependence occurs between the estimated effect sizes and ignoring such dependencies may lead to biases or lack of efficiency in statistical inferences (Gleser & Olkin, 2009). Nevertheless, if few studies have more than one effect size, then it may not be a problem to assume independence (Hedges, 2007). The researcher who chooses this way should additionally perform sensitivity analysis (Greenhouse & Iyengar, 2009). In the current review, the dependent effect sizes were included separately in the analysis in some cases. The reason for this is both to minimize data loss and the low number of studies ($k = 25$) in this situation, and the fact that it can avoid dependency with publication bias, sensitivity and outlier analyses performed.

In the last instance, the coding process resulted in the selection of 177 studies (234 effect sizes) for academic achievement, 9 studies (10 effect sizes) for learning retention and 17 studies (17 effect sizes) for attitude towards course (Appendix A). Inter-coder reliability was found to be 97% based on the formula proposed by Miles and Huberman (1994): $\text{Coder reliability} = \frac{\text{number of agreements}}{\text{total number of agreements} + \text{disagreements}} \times 100$. Disagreements noted by the coders in the coding period were discussed and resolved by consensus.

Model choice, heterogeneity test and moderator analysis

The issue of which model should be used while performing meta-analysis is considered very important and discussed. According to some researchers (Card, 2012; Cumming, 2012; Dinçer, 2014; Lipsey & Wilson, 2001), the model selection is decided based on the results of the heterogeneity test. However, this approach is criticized and not accepted much today. Deciding on the model selection according to the results of the statistical tests is seen as an erroneous approach (Borenstein et al., 2009; Schroll et al., 2011). Model selection in meta-analysis should be based on researchers' beliefs about the nature of basic data (Rothstein et al., 2013). The most important issue in choosing the statistical model should be the nature of the result to be achieved, and the model should be determined according to the type of inference the researcher wants to make (Hedges & Vevea, 1998; Konstantopoulos & Hedges, 2009). Prior to the analysis, the researcher must make the decision on model selection according to the scope of the studies, the nature of the variables considered and the design used in the studies (Başol, 2016). The primary studies examined within the scope of the current meta-analysis differed in terms of settings, course types, measurement tools, experimental designs, participants' grades and age groups. Moreover, the studies did not have the same universe parameters and thus were not homogeneous. Therefore, it was decided to estimate the main effect size using the random effects model prior to the analysis.

Heterogeneity is an assumption of the random effects model and the assessment of heterogeneity is a primary goal of meta-analysis since the presence of heterogeneity among the effect sizes of primary studies indicates the presence of moderator variables (Huedo-Medina et al., 2006). Accordingly, we conducted the heterogeneity analysis to assess heterogeneity between the effect sizes of the primary studies examined in the meta-analysis and to determine the possible moderator variables that could affect the results of the research in case of heterogeneity (Table 3).

The analysis of the values of Q , I^2 and p statistics indicated heterogeneity between the effect sizes of the studies and also indicated the presence of possible moderator variables. In this meta-analysis, several study characteristics were predicted to be possible moderator variables for academic

Table 3 Findings of heterogeneity analyses

	df	Q	X^2	p	I^2
Academic achievement	233	1923.09	277.138	0.000	87.88
Learning retention	9	16.739	16.919	0.053	46.235
Attitude towards course	16	21.163	26.296	0.172	24.39

achievement including (a) publication status, (b) domain of subject, (c) educational level, (d) duration of experiment, (e) implementer, (f) research design, (g) published domestically vs. abroad and (h) sample size (total number of students in the experiment and control group). On the other hand, Borenstein et al. (2009) and Hedges and Olkin (1985) suggest that each sub-group should include at least ten studies to get robust results from moderator analyses. Hence moderator analyses could not be performed for learning retention and attitude towards course variables because the test results showed low heterogeneity and the small number of primary studies testing the impact of the model on them.

Outliers, publication bias and sensitivity analysis

There is a possibility of outliers in the data sets of meta-analyses as well as in primary studies (Arthur et al., 2001). To detect outliers, we weighted the effect sizes of individual studies with the inverse of the variance and then ordered them linearly. Subsequently, we defined the effect sizes with a difference between two consecutive weighted effect sizes equal to or greater than the standard deviation (SD) of the distribution as outliers (see McLeod & Weisz, 2004; Hittner & Swickert, 2006; Swanson et al., 2009).

In some cases, researchers may not choose to publish studies without statistically significant results (Becker, 2005). Even if they wish to publish, journals usually are unlikely to accept studies with negative or unexpected results (Bronson & Davis, 2012). Clearly, this means that the results of published and unpublished research can be systematically different (Clarke, 2007). In such cases, systematic reviews will tend to take a sample of positive studies that will include publication bias, i.e. an inflated estimate of application or method impact (Torgerson, 2003). Publication bias can affect any kind of research synthesis. However, there are some methods that enable the determination of publication bias in the context of meta-analysis (Brunton et al., 2012). In the present meta-analysis, we used the funnel plot to investigate publication bias visually. Then, we tested publication bias by using the classic fail-safe N (Rosenthal, 1979), Orwin's fail-safe N (Orwin, 1983), the trim-and-fill method (Duval & Tweedie, 2000a, b), and the weight-function model (Vevea & Hedges, 1995), which is recommended because it gives better results in studies with large samples (Vevea & Woods, 2005).

A number of decisions are taken that are likely to affect the conclusions in research syntheses, such as meta-analysis (Greenhouse & Iyengar, 2009). In such cases, sensitivity analyses are used to evaluate the robustness of combined estimates to different assumptions (Hanji, 2017). In the present review, we decided to include dependent effect sizes separately in the analysis in order to reduce data and statistical power loss. To see how this decision

affects the result, we conducted a sensitivity analysis by comparing the overall effect size obtained by analyzing multiple effect sizes separately and the overall effect size obtained by combining multiple effect sizes.

Effect size estimation

The effect size estimation included four steps suggested by Borenstein et al. (2009). (1) effect size of each study was computed, (2) the effect sizes of all studies was integrated to estimate the mean effect sizes by Hedges's g , (3) the confidence interval (CI) for the overall mean effect size was calculated by the random effects model, (4) whether the effect size influenced by moderator variables was examined through the Q_b value. The effect sizes were estimated via Comprehensive Meta-analysis software. To estimate the effect sizes, we used the standardized mean difference index developed by Hedges (1982):

$$\text{Hedges} = M_1 - M_2 / SD_{\text{pooled}}, \quad (1)$$

where M_1 represents the mean score of the experiment group, M_2 represents the mean score of the control group, and SD_{pooled} represents the weighted average of the SD value of the groups. Using this formula, we subtracted the post-test mean of the control group from the post-test mean of the flipped classroom intervention group and divided the difference by their pooled SD. When no SD emerged, we used the p values of independent groups ($k = 11$). Accordingly, positive effect sizes indicated that the students in the experimental group had more positive outcomes than the students in the control group.

Results

Results for academic achievement

A total of 177 individual studies were included in the meta-analysis to determine the effect of the flipped classroom on the academic achievement of the students. In these individual studies, there were 234 separate effect sizes and 17,807 participants. The studies were published between 2012 and 2019 and, as seen in Table 4, were mostly published as journal articles (61.1%), were conducted in the domain of languages (27%), focused on tertiary education (63.6%), involved an experimental duration of 5–8 weeks (27.5%), were implemented by course instructors (56.1%), had a quasi-experimental design (87.9%), were conducted abroad (65.3%), and had 50–100 participants (44.4%).

Table 4 Categories and effect sizes based on moderator variables

Moderator	<i>k</i>	Effect size (<i>g</i>)	95% CI		<i>Q_b</i>	<i>p</i>
			Lower L	Upper L		
Publication type	234				0.566	0.754
Master thesis	52	0.832	0.634	1.030		
Doctoral dissertation	39	0.738	0.514	0.961		
Article	143	0.749	0.634	0.864		
Domain subject	233 ^a				25.382	0.001*
Science	40	0.680	0.471	0.889		
Mathematics	38	0.524	0.310	0.738		
Social sciences	21	0.712	0.420	1.005		
Educational sciences	18	0.840	0.518	1.162		
Computer and information technology	26	0.724	0.466	0.983		
Language	63	0.952	0.778	1.127		
Health-care programs	12	0.970	0.589	1.350		
Engineering	9	0.196	-0.229	0.622		
Physical education	6	1.471	0.951	1.992		
Educational level	234				8.914	0.030*
Elementary school	11	1.162	0.739	1.585		
Middle school	41	0.671	0.454	0.887		
High school	33	1.020	0.768	1.272		
Tertiary	149	0.711	0.598	0.824		
Experimental duration	207 ^b				14.793	0.005*
1–4 weeks	44	0.908	0.696	1.120		
5–8 weeks	57	0.995	0.807	1.184		
9–12 weeks	34	0.783	0.549	1.018		
13–16 weeks	55	0.582	0.395	0.768		
≥ 16 weeks	17	0.434	0.095	0.772		
Implementer	187 ^c				1.491	0.474
Researcher	72	0.778	0.623	0.932		
Course instructor	105	0.663	0.537	0.788		
Both	10	0.605	0.175	1.035		
Research design	215 ^d				0.004	0.950
Experimental	26	0.776	0.506	1.046		
Quasi-experimental	189	0.767	0.665	0.869		
Published domestically/abroad	234				16.096	0.000*
Turkey	81	1.021	0.866	1.175		
Abroad	153	0.635	0.527	0.743		
Sample size	234				17.002	0.000*
< 50	85	1.008	0.853	1.162		
≥ 50. < 100	104	0.701	0.570	0.832		
≥ 100	45	0.515	0.325	0.706		

Random effects model, * $p < 0.05$

^aOne study was not specified and thus was extracted from the analysis

^bTwenty seven studies were not specified and thus were extracted from the analysis

^cForty seven studies were not specified and thus were extracted from the analysis

^dNineteen studies were not specified and thus were extracted from the analysis

Overall effect size

In the research, the sample sizes, means, mean differences, SDs, and p values of the experimental and control groups

were used to estimate the effect sizes. Statistical significance level of the studies was accepted as $p = 0.05$. Information about individual studies included in the meta-analysis related to academic achievement is presented in Table S1,

and the forest plot is presented in Figure S1. The smallest effect size value was -2.314 and the largest effect size value was 10.483 , whereby 24 of the 234 effect size values included in the meta-analysis were negative and 210 were positive. According to the classification proposed by Cohen et al. (2007), 25 of the positive studies had a weak, 43 of them had a modest, 74 of them had a moderate, and 68 of them had a strong effect.

It is commonly known that excessively small or large effect sizes of primary studies may distort the results of the research by creating an extreme effect on the main effect size of the meta-analysis. Therefore, prior to the estimation of the main effect size, the effect size values in the present study were examined to investigate the presence of outliers among the effect sizes. The 234 effect size values analyzed in the meta-analysis were initially weighted with the inverse of their variance and then the weighted effect size values were ordered linearly (Table S2). Subsequently, the SD of the distribution was calculated as $SD = 1.522$. The difference between the weighted effect sizes in the first (Howell, 2013) and second (Kennedy et al., 2015) order of the distribution was found to be greater than the SD of the distribution. Therefore, the effect size of Howell (2013) was defined as outlier and was winsorized to the preceding effect size (Kennedy et al., 2015) in the distribution. After winsorizing the outlier, the effect size value of Howell (2013) increased from -2.314 to -1.116 . The differences between the weighted effect sizes fell below the new SD (1.518) of the distribution and further analysis was performed by using winsorized effect size.

At the end of the analysis, the main effect size value was estimated as 0.764 with a standard error of 0.046 under the random effects model ($p = 0.000$). At a 95% CI, the lower limit of the main effect size was 0.674 and the upper limit was 0.855 . Accordingly, this result could be accepted as statistically significant. The positive mean effect size value ($g = 0.764$) indicated that the effect of the process was in favor of the experimental group, implicating that flipped classroom is more effective on students' academic achievement than the traditional lecture-based instruction. However, this effect was found to be moderate according to the classification proposed by Cohen et al. (2007).

In the next step, sensitivity analysis was performed to determine whether the estimated overall effect size would differ substantially if combined effect sizes were included from the same samples. For this, a single combined effect size value was estimated for each of the 25 studies with multiple effect size values. 182 independent effect sizes were combined using the random-effects model, and a different overall effect size value was estimated. It was observed that the difference between the overall effect size ($g = 0.764$) estimated by combining 234 effect sizes in the first case and the overall effect size ($g = 0.749$) estimated by combining

182 effect sizes in the second case was found to be trivial (Table S3).

Evaluation of publication bias

To determine whether the estimated overall effect size value was affected by publication bias, we initially conducted a comparison between published and unpublished studies based on the recommendations of Banks et al. (2012). The result of the comparison indicated that there was no significant difference between the effect sizes of published ($g = 0.747$) and unpublished ($g = 0.795$) studies ($p = 0.617$). Subsequently, we conducted a visual inspection based on funnel plot. When the funnel plot was examined, it was seen that most of the effect sizes have low standard error and were close to the main effect size, and as a result, they accumulated at the top of the funnel. It was not possible to determine whether the effect sizes were symmetrically distributed or not, that is, whether the main effect size value was affected by publication bias according to visual inspection of the funnel plot's current version (Figure S2.). Therefore, the other statistical methods used in determining publication bias were applied. However, the result of classic fail-safe N (Rosenthal, 1979) showed that to make the p value (0.000) non-significant, a minimum number of 102,532 additional studies showing a null effect of flipped classroom on students' academic achievement should be added to the analysis (Table S4). The number of additional studies resulting from this calculation exceeded the "5k + 10" limit (Rosenthal, 1979). As shown in Table S5, the result of Orwin's fail-safe N (Orwin, 1983) suggested that approximately 134,896 studies with null results were needed to bring the overall effect size to a trivial level ($g = 0.001$). The fact that 134,896 studies with an impact size of 0 were unlikely to be included in the present study, which included 234 studies only, is an indication that the analysis results were not affected by publication bias. The result of the trim-and-fill method (Duval & Tweedie, 2000a, 2000b) revealed that the observed and adjusted effect size values were the same and thus no studies needed to be trimmed (Table S6). Moreover, since the number of effect sizes analyzed for academic achievement in the review is quite high, the weight function model (Vevea & Hedges, 1995), which is a more sophisticated technique used to determine publication bias in meta-analyses with large samples, was used. The results obtained from the analysis (Table S7) showed that the adjusted and unadjusted likelihood estimates (106.4 and 106.5 , respectively) were quite similar and the difference between them was not statistically significant ($p = 0.684$). When the results of comparison between published and unpublished studies, the classic fail-safe N , Orwin's fail-safe N , trim-and-fill, and weight function model are evaluated together, it is possible to say that the current meta-analysis is not affected by publication bias.

Moderator analyses

As it seen in Table 3, the heterogeneity value of the studies related with academic achievement included in the meta-analysis was $Q = 1923.09$. The critical value at 95% significance level of 233 degrees of freedom was $X^2 = 277.138$. The Q value exceeded the critical value of the chi-square distribution and thus revealed heterogeneity among the studies ($p < 0.001$). Based on these assumptions, we conducted several moderator analyses under the random effects model to offer possible explications to the significant heterogeneity detected in the analysis (Table 4).

The findings of moderator analysis indicated that the effect sizes of master theses ($g = 0.832$), doctoral dissertations ($g = 0.738$), and articles ($g = 0.749$) were close to each other and all three of them had a moderate effect. Moreover, the between-groups heterogeneity statistics ($Q_b = 0.566$) showed no significant difference between groups with regard to publication type ($p > 0.05$).

When the effect sizes of the studies were compared with regard to the domain subjects, flipped classroom had a strong effect in the studies conducted on physical education ($g = 1.471$), a moderate effect in studies conducted on health-care programs ($g = 0.970$), languages ($g = 0.952$), educational sciences ($g = 0.840$), computer and information technology ($g = 0.724$), social sciences ($g = 0.712$), and mathematics ($g = 0.524$), and a weak effect in studies conducted on engineering ($g = 0.196$). Otherwise the difference between the effect sizes of the domain subject groups was statistically significant ($Q_b = 25.382$, $p < 0.05$).

The between-group heterogeneity statistics also showed statistical significance for educational level ($Q_b = 8.914$, $p < 0.05$). While middle school ($g = 0.671$) and tertiary education ($g = 0.711$) had moderate effect sizes, elementary school education ($g = 1.162$) and high school ($g = 1.020$) had strong effect sizes.

Another point of comparison examined by moderator analysis was experimental duration. The results indicated that studies with an experimental duration of 1–4 weeks ($g = 0.908$), 5–8 weeks ($g = 0.995$), 9–12 weeks ($g = 0.783$) and 13–16 weeks ($g = 0.582$) had a moderate effect size, whereas the studies with an experimental duration of over 16 weeks ($g = 0.434$) had a modest effect size. This finding implicates that the effect of flipped classroom on academic achievement gradually decreases in programs longer than 8 weeks and decreases to the lowest level when the experimental duration exceeds 16 weeks. Heterogeneity analysis also indicated a significant difference among studies with regard to experimental duration ($Q_b = 14.793$, $p < 0.05$).

The findings given in Table 4 indicated that the studies in which the intervention was implemented by researchers ($g = 0.778$), course instructors ($g = 0.663$) and researchers and course instructors together ($g = 0.605$) had a moderate

effect size. However, the heterogeneity value among the groups was statistically insignificant ($Q_b = 1.491$, $p > 0.05$).

When the effect sizes of the studies were compared with regard to research design, flipped classroom had a moderate effect both in experimental ($g = 0.776$) and quasi-experimental ($g = 0.767$) studies and a significant difference was not found between the two types ($Q_b = 0.004$, $p > 0.05$).

The moderator analysis of regional differences revealed a strong effect for the studies conducted in Turkey ($g = 1.021$) and a moderate effect for the studies conducted abroad ($g = 0.635$). The results of heterogeneity analysis indicated that the difference between the effect sizes of both groups was statistically significant ($Q_b = 16.096$, $p < 0.05$).

Another moderator that emerged as a possible explanation to heterogeneity was sample size (total number of students in the experiment and control group). While the studies with fewer than 50 participants ($g = 1.008$) had a strong effect, the studies with 50–100 participants ($g = 0.701$) and with more than 100 participants ($g = 0.515$) had a moderate effect. Moreover, the heterogeneity value among the groups was statistically significant ($Q_b = 17.002$, $p < 0.05$).

Results for learning retention

A total of nine individual studies were included in the meta-analysis to determine the effectiveness of the model on the learning retention. In these individual studies, there were 10 separate effect sizes and 716 participants.

Overall effect size

Information on individual studies included in the meta-analysis regarding learning retention is presented in Table S8 and the forest plot is illustrated in Figure S3. The smallest effect size value was -0.252 and the largest effect size value was 1.126 , whereby 1 of the 10 effect size values included in the meta-analysis was negative and nine were positive. According to the classification of Cohen et al. (2007), seven of the positive studies had a moderate and two of them had a strong effect.

Prior to the estimation of the main effect size, the individual effect size values were examined for outliers. The ten effect size values analyzed in the meta-analysis were initially weighted with the inverse of their variance and then the weighted effect size values were ordered linearly (Table S9). After that, the SD of the distribution was calculated as $SD = 3.576$. The difference between the weighted effect sizes in the first (Akdeniz, 2019; Koç-Deniz, 2019) and last two (Ceylan, 2015; Makinde & Yusuf, 2017) order of the distribution were found to be greater than the SD of the distribution. Therefore, the effect size of Akdeniz (2019) and Makinde and Yusuf (2017) were defined as outliers and were winsorized to the preceding effect sizes (Ceylan, 2015;

Koç-Deniz, 2019, respectively) in the distribution. After winsorizing the outliers, the effect size value of Akdeniz (2019) increased from -0.252 to 0.422 and Makinde and Yusuf (2017) decreased from 0.908 to 0.355 . The differences between the weighted effect sizes fell below the new SD (3.23) of the distribution and further analysis was performed by using winsorized effect sizes.

The main effect size value was estimated as 0.601 with a standard error of 0.085 under the random effects model ($p=0.000$). At a 95% CI, the lower limit of the main effect size was 0.436 and the upper limit was 0.767 . Consequently, this result could be accepted as statistically significant. The positive mean effect size value ($g=0.601$) indicated that the effect of the implication was in favor of the experimental group, implicating that flipped classroom is more effective on learning retention than the traditional lecture-based instruction. However, this effect was found to be moderate according to the classification of Cohen et al. (2007).

One of the nine individual studies that were brought together for learning retention had two effect sizes and we included these effect sizes separately in the analysis. However, we conducted a sensitivity analysis to determine whether combining these effect sizes would cause a substantial difference in the overall effect size. According to the sensitivity analysis findings, there was a trivial difference between the overall effect size ($g=0.601$) obtained from the combination of ten effect sizes and the overall effect size ($g=0.616$) obtained from the combination of nine effect sizes (Table S10).

Evaluation of publication bias

To determine whether the estimated overall effect size value was affected by publication bias, we initially compared the summary effect sizes of published and unpublished studies. The result showed that there was no significant difference between the effect sizes of published ($g=0.438$) and unpublished ($g=0.736$) studies ($p=0.05$). Then we conducted a visual inspection based on funnel plot (Figure S4). When the funnel plot was examined, it was observed that the effect sizes of the studies did not show an excessive asymmetry. However, other statistical methods were used to determine publication bias to make a more objective evaluation than visual inspection. The result of classic fail-safe N (Rosenthal, 1979) showed that in order to make the p value (0.000) non-significant, a minimum number of 154 additional studies showing a null effect of flipped classroom on learning retention should be added to the analysis (Table S11). The number of additional studies resulting from this calculation exceeded the “ $5k + 10$ ” limit (Rosenthal, 1979). As shown in Table S12, the result of Orwin’s fail-safe N (Orwin, 1983) indicated that almost 5722 studies with null results were needed to bring the overall effect size to a

insignificant level ($g=0.001$). Essentially 5722 studies with an impact size of zero were improbable to be included in the present study, which included ten studies only, is an indication that the analysis results were not affected by publication bias. Moreover, the result of the trim-and-fill method (Duval & Tweedie, 2000a, 2000b) revealed that the observed and adjusted effect size values were the same and therefore no studies needed to be trimmed (Table S13). When results of all the analyses are evaluated together, it is possible to say that the current meta-analysis is not affected by publication bias.

Results for attitude towards course

A total of 17 primary studies, which have 17 discrete effect sizes and 1345 participants, were included in the meta-analysis to determine the effectiveness of the model on the attitude towards course. In these primary studies, attitude scales were used to measure students’ attitudes towards the course. In the aforementioned Likert-type scales, there are items that contain attitude expressions such as; “I like English; Physics lessons are boring for me; It is not interesting for me to try solving mathematics problems; I enjoy learning how to use chemistry in daily life”. The scales were applied as a pre-test before the intervention, including the flipped classroom model and as a post-test after the intervention to the students who were the study participants. By comparing the post-test average scores of the experimental and control group students, it has been tried to determine to what extent the flipped classroom model changes the students’ attitudes towards the course.

Overall effect size

Information about individual studies included in the meta-analysis related to attitude towards course is presented in Table S14, and the forest plot is illustrated in Figure S5. The effect sizes ranged between 0.016 and 1.035 , whereby all of the effect size values included in the meta-analysis were positive. According to the classification of Cohen et al. (2007), five of the positive studies had a weak, six of them had a modest, five of them had a moderate and one of them had a strong effect.

Prior to the estimation of the main effect size, the individual effect size values were examined for outliers. As a result of the outlier analysis, all of the differences between the weighted effect sizes were found to be lower than the SD (3.35) of the distribution (Table S15). After realizing that no effect size was outlier further analysis was performed using effect sizes without winsorizing.

As a result of the analysis, the main effect size value was determined as 0.406 with a standard error of 0.068 by random effects model ($p=0.000$). At a 95% CI, the lower limit

of the main effect size was 0.274 and the upper limit was 0.538. The result was statistically significant and indicated that the effect of the treatment was in favor of the experimental group, implicating that flipped classroom is more effective on attitude towards course than the traditional lecture-based instruction. Nonetheless, this effect was found to be modest according to the classification of Cohen et al. (2007).

Evaluation of publication bias

To determine whether the estimated overall effect size value was affected by publication bias, firstly we compared the summary effect sizes of published and unpublished studies. The result of the comparison showed that there was no significant difference between the effect sizes of published ($g = 0.425$) and unpublished ($g = 0.399$) studies ($p = 0.858$). Next, we conducted a visual inspection based on funnel plot (Figure S6). When the funnel plot was examined, it was observed that the effect sizes of the studies did not show an excessive asymmetry. However, for make a more objective evaluation the other statistical methods used. The result of classic fail-safe N (Rosenthal, 1979) indicated that in order to make the p value (0.000) insignificant, a minimum number of 205 additional studies showing a null effect of flipped classroom on attitude towards course should be added to the analysis (Table S16). The number of additional studies resulting from this estimation were more than the “ $5k + 10$ ” limit (Rosenthal, 1979). As shown in Table S17, the result of Orwin’s fail-safe N (Orwin, 1983) indicated that almost 6550 studies with null results were needed to bring the main effect size to a trivial level ($g = 0.001$). In fact 5722 studies with an impact size of zero were unlikely to be included in the present study, which included 17 studies only, is an indication that the analysis results were not affected by publication bias. Besides, the result of the trim-and-fill method (Duval & Tweedie, 2000a, 2000b) revealed that the observed and adjusted effect size values were the same and thus no studies needed to be trimmed (Table S18).

When results of all the analyses are evaluated together, it is conceivable to say that the current meta-analysis is not affected by publication bias.

Discussion

The present meta-analysis reviewed a total of 177 studies that investigated the effectiveness of flipped classroom on academic achievement. The meta-analysis included individual studies conducted in multiple disciplines, educational levels, and in different countries. In this respect, to our knowledge, the present study is one of the most comprehensive studies conducted on this subject to date. The mean effect size calculated in the study ($g = 0.764$) revealed

that flipped classroom is more effective than traditional lecture-based instruction in increasing students’ academic achievement. From this point the result was consistent with the findings of previous meta-analyses. Except one (Gillette et al., 2018) in all of the meta-analyses conducted on this subject (Tan et al., 2017; Hew & Lo, 2018; K. S. Chen et al., 2018; Cheng et al., 2019; Karagöl & Esen, 2019; Låg & Sæle, 2019; Martínez et al., 2019; Orhan, 2019; Shi et al., 2019; van Alten et al., 2019) it was concluded that flipped classroom was more effective than traditional lecture-based instruction. However, the study results were inconsistent in terms of the level of effectiveness of flipped classroom (Ranged from 0.19 to 2.29). The inconsistencies in the mean effect sizes of the meta-analyses in the literature may have resulted from the fact that some of them included a relatively low number of studies at a particular discipline or level of education. In meta-analyses conducted with a small number of studies, it is possible that extreme values in individual effect sizes can distort the mean effect size so that it may cause the mean effect size to be estimated higher or lower than its actual value. We believe that in our meta-analysis not only the presence of a large number of individual effect sizes but also the winsorizing of extreme values through our outlier analysis allows more literal results to be achieved by preventing distortions of the main effect size value.

Apart from these, the overall effect size estimated for academic achievement in the present meta-analysis was also different from other comprehensive meta-analyses (Låg & Sæle, 2019; van Alten et al., 2019). This is probably due to these meta-analyses’ inclusion criteria and, therefore, the difference in the primary studies they include. To illustrate, Låg and Sæle (2019) review also included studies using non concurred control groups, written in German and Scandinavian languages or based on pass/failure rates, contrary to the existing meta-analysis. In the van Alten et al. (2019) review, contrary to our meta-analysis, individual studies in which previous cohorts were used as a control group, participants were graduates, written only in English or conference papers were included in the analysis.

The benefits of flipped classroom such as allowing learners to watch videos multiple times at anytime and anywhere in accordance with their own pace (Bates et al., 2017; Fulton, 2012; Mok, 2014) can be attributed to the increase in students’ performance. Another explanation to this finding may be that flipped classroom has the potential to transform students into active learners and to support hands-on learning (Nederveld & Berge, 2015). As is commonly known, the flipped classroom model offers more in-class time for active learning activities (Bergmann & Sams, 2012, 2014; Sota, 2016) and active learning, in turn, increases the level of learning by providing learners more information processing and better understanding (Beard & Wilson, 2005; Taylor & MacKenney, 2008).

The homogeneity test indicated heterogeneity among the effect sizes of the studies included in the present meta-analysis. In line with this heterogeneity, the subgroup comparisons of eight different possible moderator variables showed significant results for five moderators. The findings indicated that studies with a small number of participants had a greater effect size than studies with a large number of participants. Furthermore, the average effect size value decreased as the number of students in the groups increased, implicating that flipped classroom can be more effective on learners' academic achievement in smaller classes. This finding was found to be consistent with the findings presented by previous meta-analyses conducted on flipped classroom. Gillette et al. (2018) and Karagöl and Esen (2019) concluded that flipped classroom implementations were more effective in small classes. Likewise, not only the studies investigating flipped classroom but also the meta-analysis evaluating the effectiveness of active learning approach (Tutal, 2019; Freeman et al., 2014) and the strategies that support active learning (Shin & Kim, 2013; Kalaian & Kasim, 2014; Dağyar & Demirel, 2015; M. Chen et al., 2018) reported that the effect size values of the groups with fewer students were greater in terms of academic achievement. Moreover, it has also been reported that the implementation of flipped classroom in small classes provides better classroom management for teachers (Karagöl & Esen, 2019), thus leading to higher success among students. On the other hand, small classes have some other advantages over crowded classes. To illustrate, small classrooms have fewer distractions and the teacher has more time to take care of each student (Mosteller, 1995), which in turn enhances the teacher-student interaction. Moreover, as teachers in non-crowded classrooms have higher morale, they are more likely to provide a supportive learning environment for their students (Biddle & Berliner, 2002).

Another moderator variable evaluated in the analysis was the geographical location where the studies were conducted. Accordingly, it was revealed that the studies conducted in Turkey had a larger effect size compared to those conducted abroad. Similarly, in a previous meta-analysis, Karagöl and Esen (2019) concluded that the studies conducted in Turkey had greater impact size values than those conducted abroad. On the other hand, when we examined the individual studies included in the meta-analysis based on their sample size, which was another moderator variable, we found that the proportion of studies conducted in Turkey with a sample size of less than 50 (51.7%) was greater than the proportion of studies conducted abroad (25.4%). Additionally, the proportion of studies conducted in Turkey with more than 100 participants (3.5%) was lower than those conducted abroad (29.4%) and the average sample size was 54 in the studies conducted in Turkey as opposed to 89 in the studies conducted abroad. These findings suggest that the studies

conducted in Turkey may have greater impact sizes since they had smaller samples.

The comparison of educational levels revealed that the summary effect size value of the studies conducted at elementary and high schools were higher than those in other groups. This finding contradicts the conclusions of other meta-analyses (Cheng et al., 2019; Karagöl & Esen, 2019; Låg & Sæle, 2019; Orhan, 2019) performing a similar subgroup comparison that academic achievement was not moderated by the student level. Considering the fact that flipped classroom, as it is known today, is first being used in high school classes (Bhatnaga & Bhatnagar, 2020; Divyashree, 2018), and is most commonly applied in universities (63.6% in this study), it is important for us to see that the largest effect size was at primary school level in the current meta-analysis. Besides self-regulated learning is a key aspect for the success in flipped classroom (Lai & Hwang, 2016; Sun, 2015) and it is related to learners grade level (Zimmerman & Martinez-Pons, 1990). There are biological limits at different ages for self-regulation. Younger children may have less capacity to regulate their actions (Wigfield et al., 2011), with increasing grade level the self-regulation capacity increases also (Zimmerman & Martinez-Pons, 1990). Considering this situation regarding self-regulated learning, the mentioned finding of meta-analysis is surprising.

One explanation for this result may be that primary school students' abilities are different (Bergmann & Sams, 2016), so they need more teacher guidance than students at other levels. Flipped classroom allows teachers more in-class time (Akçayır & Akçayır, 2018), enabling more effective teacher-student interaction (Bergmann & Sams, 2012) and guidance. Another explanation is that in flipped classroom the motivation for learning increases as being an internal source of excitement. Because in lecture-based instruction, excitement depends on the teacher's expression and behaviour; thus, the source of excitement is external (Elian & Hamaidi, 2018). According to Şık (2019), young students state that they find flipped classroom different, interesting, entertaining and efficient. Therefore, the enthusiastic participation and motivation of the students affect their learning positively. Also, preparing for lessons by watching videos at home for young students with short attention spans may have resulted in less cognitive load and better concentration on the lesson. We hope this result, which implicates the model can be most effectively used in elementary school classes, will be encouraging for teachers who think flipped classroom is for older students and their students are not mature enough to take responsibility.

When the effect sizes of the studies included in the meta-analysis were compared according to the experimental durations applied in the studies, the difference between the summary effect sizes of the groups was found to be statistically significant. Although it was concluded that the summary

effect size values of the groups did not differ significantly according to the experimental duration in other meta-analyses that made a similar comparison (Cheng et al., 2019; Karagöl & Esen, 2019; Shi et al., 2019), we think, this is due to the time intervals determined in the analyses. To illustrate, in two of the meta-analyses (Cheng et al., 2019; Shi et al., 2019), studies shorter and longer than one semester were compared, while in the other (Karagöl & Esen, 2019), studies that lasted between 1 and 4, 5–8 weeks and longer than 9 weeks were compared. On the contrary, in the current meta-analysis, a comparison of the studies that lasted between 1 and 4, 5–8, 9–12, 13–16 weeks and longer than 16 weeks was used. Determining the time intervals in this way allowed for more sensitive comparison. As a result, we observed that flipped classroom could be used most effectively in implementations shorter than 8 weeks, and this effect gradually decreased as time goes on. In this regard, it may be suggested that the educators implementing the model should take this finding into account when determining the duration of implementation.

Another moderator variable that differs significantly according to meta-analysis results was the domain subject. Findings indicated that studies conducted on physical education had a strong, studies in engineering had a weak, and studies in other areas had a moderate summary effect. Thus, we can say that flipped classroom is effective in increasing academic achievement, although it is partially low in engineering, in all subject areas. Nevertheless we have to note that this was just based on nine studies in engineering domain subject. As Karabulut-Ilgu et al. (2018) stated in their systematic review on flipped classroom studies in engineering, most of the studies in engineering did not report means, SDs and number of participants which necessary for a meta-analysis. In addition, the studies used different measurements (course grades, exam or quiz scores) rather than pre and post-tests that would made such a comparison difficult. While reviewing the literature, we noticed this situation in person and had to eliminate many studies because they did not meet the inclusion criteria. Otherwise, if there were more studies in the engineering, a different summary effect size value could be estimated.

It is also necessary to draw attention to the strong effect size obtained for physical education. Since physical education is designed to learn sports skills, it is usually done face-to-face. In these lessons, instructors focus on explaining the rules and showing movements that students need to repeat and imitate (Hill, 2014). However, physical education curriculums in many countries are based on the understanding that the relationship between physical activity and good health is cognitively and practically learned by students (Østerlie, 2016). This means that schools that want to fulfil the objectives of physical education curricula must embrace the fact that theoretical knowledge is part of physical education

(Solomon, 2006). Many teachers want to give students more information about physical education's basic concepts without taking the time from practical activities in physical education courses (Østerlie, 2016). However, this is not always possible because of the limited time allocated to physical education lessons and some of the time is inevitably lost in the locker rooms. As such, flipped classroom seems to have a serious potential to solve this problem in physical education. Because the research revealed that flipped classroom optimizes the time that students can be physically active (Sargent & Casey, 2020). In the classrooms where the model was applied, students' learning the necessary theoretical information before the lessons through videos may have caused them to increase their competence in basic subjects in physical education lessons and to allocate valuable time to physical activity, thus increasing their success (Østerlie, 2016). On the other hand, the low number of effect sizes included in the analysis ($k = 6$) in physical education as in engineering requires us to approach the results obtained from this moderator analysis with caution.

Although no significant difference was obtained in the remaining subgroup comparisons performed in the moderator analysis, some of the findings obtained in the analysis provided ideas for implementations would be conducted. In another comparison, it was determined that the average effect sizes of the studies with an experimental and quasi-experimental design were close to each other. This finding is in line with the results of Hew and Lo (2018) and Shi et al. (2019). It could be attributed to the inclusion of individual studies in the meta-analysis, all of which applied pre-test only prior to the experimental procedure. Based on these findings, it can be suggested that pre-test application before experimental procedures allows determining random-equivalent groups.

As a result of the subgroup comparison conducted according to the types of studies, it was determined that there was no significant difference between the summary effect sizes of the groups. This result was similar to the results of Cheng et al. (2019) and Orhan (2019), who made a similar subgroup comparison and concluded that the effect sizes did not differ significantly according to the publication types. The fact that the summary effect size values of the master's theses, doctoral dissertations and articles reviewed within the scope of meta-analysis were close to each other was another indication that the meta-analysis was not affected by publication bias.

Whether there was a difference between the summary effect sizes related to the implementers of the studies included in the meta-analysis was another possible moderator variable investigated within the scope of the research. Our analysis, coinciding with Hew and Lo (2018), Karagöl and Esen (2019) and Shi et al. (2019), also showed that the results were not moderated by whether the implementers

were different or not. The absence of a significant difference between the academic achievements of the classes in which the experimental process was implemented by the researchers or the instructors suggests that the results of the current meta-analysis do not have the Hawthorne effect that occurs in the experimental processes if the researcher exists or is aware that the participants are observed.

Apart from the possible moderator variables examined in the present meta-analysis, there may be different variables that could explain the variation in the effect sizes. For instance, the different types of videos used in flipped classroom applications may affect students' achievement (Hew & Lo, 2018). On the other hand, engagement, which is accepted as a prerequisite of learning, may vary according to the different characteristics of those videos. According to Guo et al., (2014), the most important determinant of engagement is the length of videos and that short videos increase participation more than long videos do. Similarly, Turro et al., (2016) stated that short and purposeful videos are more effective and Amresh et al. (2013) suggested that long video lectures may negatively affect students' perception of flipped classroom. In addition to length, some other variables including the timeliness of availability of videos (Wagner et al., 2013), types of videos (tutorials, lecture videos, talking-head videos, PowerPoint slides and/or code screencasts), planning phase of videos, speech rate of instructors (Guo et al., 2014), and the presence or absence of in-video questions (Cummins et al., 2015; Vigentini & Zhao, 2016) may also affect learners' academic success.

Besides the variables abovementioned, some other variables have been suggested as possible sources of heterogeneity including pedagogical approaches used in classroom activities and the time allocated for them, previous experiences of teachers and students about flipped classroom, content and cognitive level of knowledge, pre-preparation of students, and the presence of pre-course quizzes (K. S. Chen et al., 2018; Gillette et al., 2018; Hew & Lo, 2018; van Alten et al., 2019). Such diversity of primary studies contributes to the heterogeneity in effect sizes. However, moderator analyses of these variables could not be performed in the present study since they were not clearly reported in the individual studies included in the meta-analysis, which could be a limitation of our research. We recommend that detailed reporting of these issues by the authors in future studies will allow further moderator analyses in the meta-analyses to be conducted on the same subject matter.

Also in our meta-analysis we reviewed nine studies which examined the effect of flipped classroom on learning retention. As far as we know, the current study is the first analysis to date to determine the overall effect size of the flipped classroom on learning retention. The mean effect size estimated in the study ($g = 0.601$) revealed that flipped classroom is more effective than traditional lecture-based

instruction in increasing learning retention. In general, it is suggested that the active learning approach increases the level of learning by making the knowledge more permanent (Collins & O'Brien, 2003; Beard & Wilson, 2005; Taylor & MacKenney, 2008), and in particular, the flipped classroom is theorized to improve learner engagement and retention (Roehl et al., 2013; Rose et al., 2016). It is possible to say that the result of the research regarding the learning retention provides an important evidence for this view.

The main effect size value obtained by synthesizing 17 study results for the attitude towards course variable was modest. According to Gall et al. (2003), for practical significance in educational research, an effect size value of at least 0.33 is considered sufficient, and according to Slavin (1996), an effect size value of 0.20–0.25 is considered pedagogically important. Taking into account these thoughts, the overall effect size value of 0.406 obtained from the research, a modest but notable effect size in educational research, provides evidence that flipped classroom is significantly effective in developing positive attitude towards course. This result supports the views that properly applied student-centered teaching will lead to more positive attitude towards course (Collins & O'Brien, 2003) and that models such as flipped classroom should be used in the learning process that will enable students to take responsibility for their own learning to develop positive attitude towards course (Demirel & Dağyar, 2016).

It was seen that the effect size of the flipped classroom on attitude towards course was not as large as the effect size on academic achievement or learning retention. However, as attitude researchers agree, some attitudes are resistant to change (Prislin & Crano, 2008) and attitude change can take time (Siegel & Ranney, 2003). Considering this feature of the attitudes, it can be said that the implication periods should be kept long enough for the approaches used in educational practices to cause students to develop positive attitude towards course. In 12 (70.58%) of the 17 individual studies analyzed within the scope of the present study, the practice time being less than 8 weeks was not sufficient for the students to develop positive attitude towards course and therefore a modest effect size value may have emerged.

The lack of moderator analyses for the attitude towards course and learning retention variables can be considered as a limitation of this review. It may be suggested that researchers can conduct more studies examining the effectiveness of the model on these variables, and as a result, implement more comprehensive meta-analyses on the subject. Another limitation of the present study was that due to the language barrier, we could not examine individual studies written in languages other than English and Turkish within the scope of our research. Moreover, further meta-analyses combining the results of individual studies in other languages will

also contribute to our understanding of the flipped classroom model.

In this meta-analysis, we tried to examine how the flipped classroom model affects learners' academic achievement, attitude towards course and learning retention. Future meta-analyses may investigate the effect of the model on other dependent variables such as self-efficacy, motivation, engagement and self-directed learning. Meta-analyses will be able to synthesize the results of studies investigating the effect of flipped classroom on preschool children and graduates. Moreover, future studies can also compare flipped classroom with other online or blended classes instead of traditional classes.

Conclusion

The present meta-analysis demonstrated that the flipped classroom model significantly improves students' academic achievement, attitude towards course and learning retention compared to the traditional lecture-based instruction. However, there was large heterogeneity between the effect sizes of the primary studies carried out for academic achievement and the moderator analyses we applied could only explained some part of it. Through our meta-analysis we have learned that in terms of academic achievement the flipped classroom phenomenon; (a) has been implemented more effectively in small classes, (b) has been applied most effectively in primary schools, (c) its effectiveness has decreased as the duration of implementation extends, (d) has been implemented effectively in almost all domain subjects, (e) also in terms of attitude towards course and learning retention it has been more effective than traditional lecture-based instruction. On the other hand, the effect of the model on achievement did not differ according to variables such as publication type, implementer and research design. In order to better understand flipped classroom and use it more effectively, we invite researchers to conduct research to determine other possible sources of heterogeneity and the effect of the model on other dependent variables.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s12564-021-09706-9>.

Declarations

Conflict of interest The authors declare that they have no known conflict of interest to disclose.

References

- Ajzen, I., & Gilbert-Cote, N. (2008). Attitudes and the prediction of behavior. In W. D. Crano & R. Prislin (Eds.), *Attitudes and attitude change* (pp. 289–311). Psychology Press.
- Akçayır, G., & Akçayır, M. (2018). The flipped classroom: A review of its advantages and challenges. *Computers & Education*, *126*, 334–345. <https://doi.org/10.1016/j.compedu.2018.07.021>
- Akdeniz, E. (2019). *The effect of flipped class model on students' academic success, attitudes and retention*. Unpublished master thesis. Necmettin Erbakan University.
- Amresh, A., Carberry, A. R., & Femiani, J. (2013). Evaluating the effectiveness of flipped classrooms for teaching CS1. *IEEE Frontiers in Education Conference (FIE)*. <https://doi.org/10.1109/FIE.2013.6684923>
- Arthur, W., Jr., Bennett, W., & Huffcutt, A. I. (2001). *Conducting meta-analysis using SAS*. Lawrence Erlbaum Associates, Publishers.
- Banks, G. C., Kepes, S., & Banks, K. P. (2012). Publication bias: The antagonist of meta-analytic reviews and effective policymaking. *Educational Evaluation and Policy Analysis*, *34*(3), 259–277. <https://doi.org/10.3102/0162373712446144>
- Başol, G. (2016). Response letter to editor: In response to the letter to editor by Dinçer (2016): Concerning “a content analysis and methodological evaluation of meta-analyses on Turkish samples.” *International Journal of Human Sciences*, *13*(1), 1395–1401. <https://doi.org/10.14687/ijhs.v13i1.3694>
- Bates, J. E., Almekdash, H., & Gilchrist-Dunn, M. J. (2017). The flipped classroom: A brief, brief history. In L. Santos-Green, J. R. Banas, & R. A. Perkins (Eds.), *The flipped college classroom* (pp. 3–10). Springer.
- Beard, C., & Wilson, J. P. (2005). Ingredients for effective learning: The learning combination lock. In P. Hartley, A. Woods, & M. Pill (Eds.), *Enhancing teaching in higher education: New approaches for improving student learning* (pp. 3–16). Routledge.
- Becker, B. J. (2005). Failsafe N or file-drawer number. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 111–126). Wiley.
- Bergmann, J., & Sams, A. (2012). *Flip your classroom: Reach every student in every class every day*. International Society for Technology in Education.
- Bergmann, J., & Sams, A. (2014). *Flipped learning: Gateway to student engagement*. International Society for Technology in Education.
- Bergmann, J., & Sams, A. (2016). *Flipped learning for elementary instruction*. International Society for Technology in Education.
- Bhagat, K. K., Chang, C. N., & Chang, C. Y. (2016). The impact of the flipped classroom on mathematics concept learning in high school. *Journal of Educational Technology & Society*, *19*(3), 134–142.
- Bhatnaga, M., & Bhatnagar, P. (2020). Flipped classroom-an innovative approach. *Journal of Xi'an University of Architecture & Technology*, *12*(2), 403–413.
- Biddle, B. J. & Berliner, D. C. (2002). What research says about small classes and their effects. *Education Policy Reports Project. EPSL-0202-101-EPRP*. Arizona State University.
- Birenbaum, M. (2003). New insights into learning and teaching and their implications for assessment. In M. Segers, F. Dochy, & E. Cascallar (Eds.), *Optimising new modes of assessment: In search of qualities and standards* (pp. 13–36). Kluwer Academic Publishers.
- Bishop, J. L., & Verleger, M. A. (2013). The flipped classroom: A survey of the research. Presented at 2013 ASEE national conference proceedings, Atlanta. (Vol. 30, No. 9, pp. 6219).
- Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley & Sons.
- Bronson, D. E., & Davis, T. S. (2012). *Finding and evaluating evidence: Systematic reviews and evidence-based practice*. Oxford University Press.
- Brunton, G., Stansfield, C., & Thomas, J. (2012). Finding relevant studies. In D. Gough, S. Oliver, & J. Thomas (Eds.), *An introduction to systematic reviews* (pp. 107–134). Sage.

- Card, N. A. (2012). *Applied meta-analysis for social science research*. The Guildford Press.
- Carlisle, C. S. (2018). *How the flipped classroom impacts students' math achievement*. Unpublished doctoral dissertation. Trevecca Nazarene University.
- Ceylan, V. K. (2015). *Effect of blended learning to academic achievement*. Unpublished master thesis. Adnan Menderes University.
- Chen, K. S., Monrouxe, L., Lu, Y. H., Jenq, C. C., Chang, Y. J., Chang, Y. C., & Chai, P. Y. C. (2018). Academic outcomes of flipped classroom learning: A meta-analysis. *Medical Education*, 52(9), 910–924. <https://doi.org/10.1111/medu.13616>.
- Chen, M., Ni, C., Hu, Y., Wang, M., Liu, L., Ji, X., Chu, H., Wu, W., Lu, C., Wang, S., Wang, S., Zhao, L., Li, Z., Zhu, H., Wang, J., Xia, Y., & Wang, X. (2018). Meta-analysis on the effectiveness of team-based learning on medical education in China. *BMC Medical Education*. <https://doi.org/10.1186/s12909-018-1179-1>.
- Cheng, L., Ritzhaupt, A. D., & Antonenko, P. (2019). Effects of the flipped classroom instructional strategy on students' learning outcomes: A meta-analysis. *Educational Technology Research and Development*, 67(4), 793–824. <https://doi.org/10.1007/s11423-018-9633-7>
- Cheung, M. W. L. (2014). Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling approach. *Psychological Methods*, 19, 211–229. <https://doi.org/10.1037/a0032968>
- Clarke, M. (2007). Overview of methods. In C. Webb & B. Roe (Eds.), *Reviewing research evidence for nursing practice: Systematic reviews* (pp. 3–8). Blackwell Publishing.
- Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education*. Routledge.
- Collins, J. W., & O'Brien, N. P. (Eds.). (2003). *The greenwood dictionary of education*. Greenwood.
- Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. Routledge.
- Cummins, S., Beresford, A. R., & Rice, A. (2015). Investigating engagement with in-video quiz questions in a programming course. *IEEE Transactions on Learning Technologies*, 9(1), 57–66. <https://doi.org/10.1109/TLT.2015.2444374>
- Dağyar, M., & Demirel, M. (2015). Effects of problem-based learning on academic achievement: A meta-analysis study. *Education and Science*, 40(181), 139–174. <https://doi.org/10.15390/EB.2015.4429>
- de Caprariis, P., Barman, C., & Magee, P. (2001). Monitoring the benefits of active learning exercises in introductory survey courses in science: An attempt to improve the education of prospective public school teachers. *Journal of the Scholarship of Teaching and Learning*, 1(2), 13–23.
- Demirel, M., & Dağyar, M. (2016). Effects of problem-based learning on attitude: A meta-analysis study. *Eurasia Journal of Mathematics, Science & Technology Education*, 12(8), 2115–2137. <https://doi.org/10.12973/eurasia.2016.1293a>
- Di Martino, P., & Zan, R. (2011). Attitude towards mathematics: A bridge between beliefs and emotions. *ZDM: the International Journal on Mathematics Education*, 43(4), 471–482. <https://doi.org/10.1007/s11858-011-0309-6>
- Diğer, S. (2014). *Applied meta-analysis in educational sciences*. Pegem Akademi.
- Divyashree, M. (2018). Flipped classroom—an innovative methodology for effective teaching learning process. *Asian Journal of Management*, 9(1), 451–456. <https://doi.org/10.5958/2321-5763.2018.00070.7>
- Dixon, K. (2017). *The effect of the flipped classroom on urban high school students' motivation and academic achievement in a high school science course*. Unpublished doctoral dissertation. Liberty University.
- Duval, S., & Tweedie, R. (2000a). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56, 455–463. <https://doi.org/10.1111/j.0006-341X.2000.00455.x>
- Duval, S., & Tweedie, R. (2000b). A nonparametric “trim and fill” method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association*, 95(449), 89–98. <https://doi.org/10.1080/01621459.2000.10473905>
- Elakovich, D. M. (2018). *Does a student's use of self-regulation change in the flipped classroom?*. Unpublished doctoral dissertation. Montana State University.
- Elian, S., & Hamaidi, D. (2018). The effect of using flipped classroom strategy on the academic achievement of fourth grade students in Jordan. *International Journal of Emerging Technologies in Learning (iJET)*, 13(2), 110–125. <https://doi.org/10.3991/ijet.v13i02.7816>
- Findlay-Thompson, S., & Mombourquette, P. (2014). Evaluation of a flipped classroom in an undergraduate business course. *Business Education & Accreditation*, 6(1), 63–71.
- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, 111(23), 8410–8415. <https://doi.org/10.1073/pnas.1319030111>
- Fulton, K. (2012). Upside down and inside out: Flip your classroom to improve student learning. *Learning & Leading with Technology*, 39(8), 12–17.
- Gall, M. D., Gall, J. P., & Borg, W. R. (2003). *Educational research: An introduction*. Allyn and Bacon.
- Gillette, C., Rudolph, M., Kimble, C., Rockich-Winston, N., Smith, L., & Broedel-Zaugg, K. (2018). A meta-analysis of outcomes comparing flipped classroom and lecture. *American Journal of Pharmaceutical Education*, 82(5), 433–440. <https://doi.org/10.5688/ajpe6898>
- Gleser, L. J., & Olkin, I. (2009). Stochastically dependent effect sizes. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 357–376). Russell Sage Foundation.
- Greenhouse, J. B., & Iyengar, S. (2009). Sensitivity analysis and diagnostics. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 417–434). Russell Sage Foundation.
- Guo, P. J., Kim, J., & Rubin, R. (2014). How video production affects student engagement: An empirical study of MOOC videos. *Proceedings of the First ACM Conference on Learning@ Scale Conference*. <https://doi.org/10.1145/2556325.2566239>
- Hanji, M. B. (2017). *Meta-analysis in psychiatry research: Fundamental and advanced methods*. CRC Press.
- Hawtrey, K. (2007). Using experiential learning techniques. *The Journal of Economic Education*, 38(2), 143–152. <https://doi.org/10.3200/JECE.38.2.143-152>
- Hedges, L. V. (1982). Estimation of effect size from a series of independent experiments. *Psychological Bulletin*, 92(2), 490–499. <https://doi.org/10.1037/0033-2909.92.2.490>
- Hedges, L. V. (2007). Meta-analysis. In C. R. Rao & S. Sinharay (Eds.), *Handbook of statistics* (Vol. 26, pp. 919–953). Elsevier.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Academic Press.
- Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods*, 3(4), 486–504.
- Hew, K. F., & Lo, C. K. (2018). Flipped classroom improves student learning in health professions education: A meta-analysis.

- BMC Medical Education*, 18, 38. <https://doi.org/10.1186/s12909-018-1144-z>
- Heyborne, W. H., & Perrett, J. J. (2016). To flip or not to flip? Analysis of a flipped classroom pedagogy in a general biology course. *Journal of College Science Teaching*, 45(4), 31–37.
- Hill, K. (2014). Using mobile devices for motor-learning laboratory exercises. *Journal of Physical Education, Recreation & Dance*, 85(8), 20–26. <https://doi.org/10.1080/07303084.2014.946190>
- Hittner, J. B., & Swickert, R. (2006). Sensation seeking and alcohol use: A meta-analytic review. *Addictive Behaviors*, 31(8), 1383–1401. <https://doi.org/10.1016/j.addbeh.2005.11.004>
- Howell, D. (2013). *Effects of an inverted instructional delivery model on achievement of ninth-grade physical science honors students*. Unpublished doctoral dissertation. Gardner-Webb University.
- Huedo-Medina, T. B., Sánchez-Meca, J., Marín-Martínez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I^2 index? *Psychological Methods*, 11(2), 193–206. <https://doi.org/10.1037/1082-989X.11.2.193>
- Hughes, Y., & Lyons, N. (2017). Does the flipped classroom improve exam performance in medical education? A systematic review. *MedEdPublish*, 6(2), 38. <https://doi.org/10.15694/mep.2017.000100>
- Johnson, L., & Renner, J. (2012). *Effect of the flipped classroom model on a secondary computer applications course: Student and teacher perceptions, questions and student achievement*. Unpublished doctoral dissertation. University of Louisville.
- Kalaian, S. A., & Kasim, R. M. (2014). A meta-analytic review of studies of the effectiveness of small-group learning methods on statistics achievement. *Journal of Statistics Education*, 22(1), 1–20. <https://doi.org/10.1080/10691898.2014.11889691>
- Kan, A., & Akbaş, A. (2005). Developing attitude scale towards chemistry lesson for high school students. *Mersin University Journal of the Faculty of Education*, 1(2), 227–237.
- Karabulut-Ilgü, A., Cherrez, N. J., & Jähren, C. T. (2018). A systematic review of research on the flipped learning method in engineering education. *British Journal of Educational Technology*, 49(3), 398–411. <https://doi.org/10.1111/bjet.12548>
- Karagöl, İ., & Esen, E. (2019). The effect of flipped learning approach on academic achievement: A meta-analysis study. *Hacettepe University Journal of Education*, 34(3), 708–727. <https://doi.org/10.16986/HUJE.2018046755>
- Kennedy, E., Beaudrie, B., Ernst, D. C., & Laurent, R. (2015). Inverted pedagogy in second semester calculus. *Primus*, 25(9–10), 892–906. <https://doi.org/10.1080/10511970.2015.1031301>
- Koç-Deniz, H. (2019). *The effect of game and activity supported flipped classroom model on students' achievement in problem solving skill and reflective thinking skill based on problem solving in mathematics course*. Unpublished doctoral dissertation. Firat University.
- Konstantopoulos, S., & Hedges, L. V. (2009). Analyzing effect sizes: Fixed-effects models. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 279–294). Russell Sage Foundation.
- Låg, T., & Sæle, R. G. (2019). Does the flipped classroom improve student learning and satisfaction? A systematic review and meta-analysis. *AERA Open*, 5(3), 1–17. <https://doi.org/10.1177/2332858419870489>
- 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. <https://doi.org/10.1016/j.compedu.2016.05.006>
- Lee, N., Lee, L. W., & Kovel, J. (2016). An experimental study of instructional pedagogies to teach math-related content knowledge in construction management education. *International Journal of Construction Education and Research*, 12(4), 255–269. <https://doi.org/10.1080/15578771.2016.1141440>
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical metaanalysis*. Sage.
- Littell, J. H., Corcoran, J., & Pillai, V. (2008). *Systematic reviews and meta-analysis*. Oxford University Press.
- Makinde, S. O., & Yusuf, M. O. (2017). The flipped classroom: Its effects on students' performance and retention in secondary school mathematics classroom. *International Journal of Innovative Technology Integration in Education*, 1(1), 117–126.
- Marín-Martínez, F., & Sánchez-Meca, J. (1999). Averaging dependent effect sizes in meta-analysis: A cautionary note about procedures. *The Spanish Journal of Psychology*, 2, 32–38.
- Martínez, T. S., Díaz, I. A., Rodríguez, J. M. R., & Rodríguez-García, A. M. (2019). Efficacy of the flipped classroom method at the university: Meta-analysis of impact scientific production. *REICE. Revista Iberoamericana Sobre Calidad, Eficacia y Cambio En Educación*, 17(1), 25–38. <https://doi.org/10.15366/reice2019.17.1.002>
- McLeod, B. D., & Weisz, J. R. (2004). Using dissertations to examine potential bias in child and adolescent clinical trials. *Journal of Consulting and Clinical Psychology*, 72(2), 235–251. <https://doi.org/10.1037/0022-006X.72.2.235>
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Sage.
- Mohanty, A., & Parida, D. (2016). Exploring the efficacy & suitability of flipped classroom instruction at school level in India: A pilot study. *Creative Education*, 7, 768–776. <https://doi.org/10.4236/ce.2016.75079>
- Mok, H. N. (2014). Teaching tip: The flipped classroom. *Journal of Information Systems Education*, 25(1), 7–11.
- Mosteller, F. (1995). The Tennessee study of class size in the early school grades. *The Future of Children*, 5(2), 113–127.
- Mzoughi, T. (2015). An investigation of student web activity in a “flipped” introductory physics class. *Procedia-Social and Behavioral Sciences*, 191, 235–240. <https://doi.org/10.1016/j.sbspro.2015.04.558>
- Nederveld, A., & Berge, Z. L. (2015). Flipped learning in the workplace. *Journal of Workplace Learning*, 27(2), 162–172. <https://doi.org/10.1108/JWL-06-2014-0044>
- Orhan, A. (2019). The effect of flipped learning on students' academic achievement: A meta-analysis study. *Cukurova University Faculty of Education Journal*, 48(1), 368–396. <https://doi.org/10.14812/cufej.400919>
- Orwin, R. G. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics*, 8(2), 157–159. <https://doi.org/10.3102/10769986008002157>
- Østerlie, O. (2016). Flipped learning in physical education: Why and how? In D. Novak, B. Antala, & D. Knjaz (Eds.), *Physical education and new technologies* (pp. 166–176). Croatian Kinesiology Association.
- Özdemir, A. (2016). *Flipped classroom model practice focused on blended learning in secondary mathematics education*. Unpublished doctoral dissertation. Gazi University.
- Perçin, F. (2019). *Effect of flipped learning on students' programming achievements, technology attitudes and personal innovation levels*. Unpublished master thesis. Marmara University.
- Prislin, R., & Crano, W. D. (2008). Attitudes and attitude change: The fourth peak. In W. D. Crano & R. Prislin (Eds.), *Attitudes and attitude change* (pp. 3–18). Psychology Press.
- Roehl, A., Reddy, S. L., & Shannon, G. J. (2013). The flipped classroom: An opportunity to engage millennial students through active learning strategies. *Journal of Family & Consumer Sciences*, 105(2), 44–49.
- Rose, E., Claudius, I., Tabatabai, R., Kearl, L., Behar, S., & Jhun, P. (2016). The flipped classroom in emergency medicine using online videos with interpolated questions. *The Journal of Emergency Medicine*, 51(3), 284–291. <https://doi.org/10.1016/j.jemermed.2016.05.033>

- Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin*, 86(3), 638–641. <https://doi.org/10.1037/0033-2909.86.3.638>
- Rothstein, H. R., Lortie, C. J., Stewart, G. B., Koricheva, J., & Gurevitch, J. (2013). Quality standards for research syntheses. In J. Koricheva, J. Gurevitch, & K. Mengersen (Eds.), *Handbook of meta-analysis in ecology and evolution* (pp. 323–338). Princeton University Press.
- Sargent, J., & Casey, A. (2020). Flipped learning, pedagogy and digital technology: Establishing consistent practice to optimise lesson time. *European Physical Education Review*, 26(1), 70–84. <https://doi.org/10.1177/1356336X19826603>
- Scammacca, N., Roberts, G., & Stuebing, K. K. (2014). Meta-analysis with complex research designs dealing with dependence from multiple measures and multiple group comparisons. *Review of Educational Research*, 84(3), 328–364. <https://doi.org/10.3102/0034654313500826>
- Schroll, J. B., Moustgaard, R., & Gøtzsche, P. C. (2011). Dealing with substantial heterogeneity in cochrane reviews. Cross-sectional study. *BMC Medical Research Methodology*, 11(1), 22. <https://doi.org/10.1186/1471-2288-11-22>
- Shi, Y., Ma, Y., MacLeod, J., & Yang, H. H. (2019). College students' cognitive learning outcomes in flipped classroom instruction: A meta-analysis of the empirical literature. *Journal of Computers in Education*. <https://doi.org/10.1007/s40692-019-00142-8>
- Shin, I. S., & Kim, J. H. (2013). The effect of problem-based learning in nursing education: A meta-analysis. *Advances in Health Sciences Education*, 18(5), 1103–1120. <https://doi.org/10.1007/s10459-012-9436-2>
- Siegel, M. A., & Ranney, M. A. (2003). Developing the changes in attitude about the relevance of science (CARS) questionnaire and assessing two high school science classes. *Journal of Research in Science Teaching*, 40(8), 757–775. <https://doi.org/10.1002/tea.10110>
- Şık, P. (2019). *The effect of flipped classroom on young learners' vocabulary learning in primary ELT classrooms*. Unpublished doctoral dissertation. Çanakkale Onsekiz Mart University.
- Slavin, R. E. (1996). *Education for all*. Swets & Zeitlinger.
- Solomon, M. A. (2006). Learner cognition. In D. Kirk, D. Macdonald, & M. O'Sullivan (Eds.), *The handbook of physical education* (pp. 226–241). Sage.
- Sota, M. S. (2016). Flipped learning as a path to personalization. In M. Murphy, S. Redding, & J. S. Twyman (Eds.), *Handbook on personalized learning for states, districts, and schools* (pp. 73–87). Center on Innovations in Learning.
- Spector, J. M. (2016). *Foundations of educational technology: Integrative approaches and interdisciplinary perspectives*. Routledge.
- Strayer, J. F. (2012). How learning in an inverted classroom influences cooperation, innovation and task orientation. *Learning Environments Research*, 15(2), 171–193. <https://doi.org/10.1007/s10984-012-9108-4>
- Sun, Z. (2015). *The role of self-regulation on students' learning in an undergraduate flipped math class*. Unpublished doctoral dissertation. The Ohio State University.
- Swanson, H. L., Zheng, X., & Jerman, O. (2009). Working memory, short-term memory, and reading disabilities: A selective meta-analysis of the literature. *Journal of Learning Disabilities*, 42(3), 260–287. <https://doi.org/10.1177/0022219409331958>
- Tan, C., Yue, W. G., & Fu, Y. (2017). Effectiveness of flipped classrooms in nursing education: Systematic review and meta-analysis. *Chinese Nursing Research*, 4(4), 192–200. <https://doi.org/10.1016/j.cnre.2017.10.006>
- Taylor, G. R., & MacKenney, L. (2008). *Improving human learning in the classroom: Theories and teaching practices*. Rowman & Littlefield Education.
- Toraman, Ç., & Ulubey, Ö. (2016). The effect of creative drama method on the attitude towards course: A meta-analysis study. *Journal of Educational Sciences Research*, 6(1), 87–115. <https://doi.org/10.12973/jesr.2016.61.5>
- Torgerson, C. (2003). *Systematic reviews*. Continuum.
- Tutal, Ö. (2019). *The effect of active learning approach on students' academic achievement, attitudes towards the course and retention: A meta-analysis study*. Unpublished doctoral dissertation. Dicle University.
- Turro, C., Mengod, R., Morales, J. C., & Busquets, J. (2016). Video is key for flipped learning: An experience at Universitat Politècnica de Valencia In: Giannakos MN, Sampson DG, Kidziński Ł, Pardo A (eds) Proceedings of the LAK 2016 workshop on smart environments and analytics in video-based learning (SE@VBL 2016). Edinburgh, Scotland, April 26, 2016.
- van Alten, D. C., Phielix, C., Janssen, J., & Kester, L. (2019). Effects of flipping the classroom on learning outcomes and satisfaction: A meta-analysis. *Educational Research Review*. <https://doi.org/10.1016/j.edurev.2019.05.003>
- Vevea, J. L., & Hedges, L. V. (1995). A general linear model for estimating effect size in the presence of publication bias. *Psychometrika*, 60(3), 419–435.
- Vevea, J. L., & Woods, C. M. (2005). Publication bias in research synthesis: Sensitivity analysis using a priori weight functions. *Psychological Methods*, 10(4), 428. <https://doi.org/10.1037/1082-989X.10.4.428>
- Vigentini, L., & Zhao, C. (2016). Evaluating the 'Student' experience in MOOCs. *Proceedings of the Third (2016) ACM Conference on Learning@ Scale*. <https://doi.org/10.1145/2876034.2893469>
- Wagner, D., Laforge, P., & Cripps, D. (2013). Lecture material retention: A first trial report on flipped classroom strategies in electronic systems engineering at the University of Regina. *Proceedings of the Canadian Engineering Education Association (CEEA)*. <https://doi.org/10.24908/pceea.v0i0.4804>
- Wigfield, A., Klauda, S. L., & Cambria, J. (2011). Influences on the development of academic self-regulatory processes. In D. H. S. Barry & J. Zimmerman (Eds.), *Handbook of self-regulation of learning and performance* (pp. 33–48). Routledge.
- Wiginton, B. (2013). *Flipped instruction: An investigation into the effect of learning environment on student self-efficacy, learning style, and academic achievement in an algebra I classroom*. Unpublished doctoral dissertation. The University of Alabama.
- Yestrebtsky, C. L. (2016). Direct comparison of flipping in the large lecture environment. In J. L. Muzyka & C. S. Luker (Eds.), *The flipped classroom: Results from practice* (Vol. 2, pp. 1–18). American Chemical Society.
- Zainuddin, Z., & Halili, S. H. (2016). Flipped classroom research and trends from different fields of study. *The International Review of Research in Open and Distributed Learning*, 17(3), 313–340. <https://doi.org/10.19173/irrodl.v17i3.2274>
- Zimmerman, B. J., & Martinez-Pons, M. (1990). Student differences in self-regulated learning: Relating grade, sex, and giftedness to self-efficacy and strategy use. *Journal of Educational Psychology*, 82(1), 51.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.