Education Tech Research Dev
https://doi.org/10.1007/s11423-021-09996-1

RESEARCH ARTICLE

A meta-analysis of relative effectiveness of flipped learning in English as second/foreign language research

Mojdeh Shahnama[1] · Behzad Ghonsooly[1] · Majid Elahi Shirvan[2]

Accepted: 23 April 2021

Abstract
Flipped learning is a pedagogical approach in which the traditional way of instruction is inverted by presenting new concepts and subject matters before class and allocating more class time to collaborative, individualized, and differentiated learning. Recently, the use of flipped learning has become popular in the context of learning English as a second/foreign language (ESL/EFL). However, no quantitative synthesis has been conducted on the results of flipped ESL/EFL studies so far. Therefore, with this gap in our mind, we conducted a meta-analysis that consisted of 69 between-subject design studies in the field of ESL/EFL, in which we compared the flipped and lectured-based classrooms in improving students' achievements. In order to interpret the results, we used the field-specific benchmark for L2 research proposed by Plonsky and Oswald (2014) in Mod Lang J 100: 538–553, 2021. We found that the influence of flipped learning on students' achievements was large and positive. Also, the findings of moderator analysis indicated that the flipped group performed better when some extra activities and exercises were accompanied by pre-class materials. We conclude that although flipped learning is new in the ESL/EFL field, it has the potential to improve students' achievements if appropriately designed and implemented. Moreover, since flipped learning is in its fledgling state in the ESL/EFL field, more empirical evidence and research is required in all educational levels to examine its effectiveness, more specifically regarding design and implementation.

Keywords Flipped learning · Flipped classroom · Meta-analysis · English as a second/foreign language · Achievements

✉ Behzad Ghonsooly
ghonsooly@um.ac.ir

Mojdeh Shahnama
shahnama.mojdeh@mail.um.ac.ir; Mojdeh.shahnama@gmail.com

Majid Elahi Shirvan
M.elahi@ub.ac.ir; Elahishmajid@gmail.com

[1] Ferdowsi University of Mashhad, Mashhad, Razavi Khorasan, Iran

[2] University of Bojnord, Bojnord, Iran

Published online: 18 May 2021

Introduction

The emerging and evolving nature of flipped learning has resulted in different definitions of this pedagogical approach over the past few years, which led to some confusion among researchers (Bishop & Verleger, 2013). Generally, flipped learning is defined as a type of pedagogical approach in which some of the contents that were traditionally taught in class by teachers are delivered to students before class. For instance, students can receive an overview of instructional contents by watching videos, listening to audio files, doing some readings and other online activities before class time. In this way, students may come to class prepared and during class time students can apply those contents in details through various pair/group works, problem-solving tasks and game or discussion-based activities (Bergmann & Sams, 2012). However, flipped learning is beyond just watching videos or using technology-based instruction. In order to engage students in flipped learning, flipped instructors must incorporate four main features or pillars of F-L-I-P into their classes: (1) flexible environment, which provides opportunities for students to choose when and where to learn the content, (2) learning culture, which means shifting focus from a teacher-centered approach toward a more learner-centered one, thereby creating an active learning environment, (3) intentional content, through which teachers and instructors can choose the contents and materials by considering students' needs and learning objectives and; finally, (4) professional educator, who has the required expertise and creativeness in generating the outline of the course, in creating videos and contents, in designing or choosing the tasks and activities that are tailored to students' needs, and in providing the immediate feedback to students on instructional videos and in-class activities (Hamdan et al., 2013).

As Bergmann and Sams (2012) and Bergmann (2017) claimed, flipped learning is a type of instructional environment that enables educators to reach every student in every class every day to enhance differentiated learning. This definition clarifies three key features of flipped learning such as maximizing student–teacher interactions, emphasizing a more student-centered instruction (Hamdan et al., 2013; Kim et al., 2014), and focusing on problems that can be solved in classrooms (Bergmann & Sams, 2012). Also, Bergmann and Sams (2012) argued that researchers should focus on talking less about the fundamentals of flipped learning and talking more about the steady stream of innovations that can be used to improve learning across the world. They also noted that researchers should also consider talking less about how to flip successfully and talking more about how to provide the intended objectives that every teacher aims to achieve through flipped leaning.

Flipped learning, as an active and student-centered learning approach, seems to have a pivotal role in transforming the teaching practice. As claimed by Bergmann and Sams (2012), certain advantages are conferred by flipped learning such as changing classroom management and instruction, increasing student–student and student–teacher interactions, and creating individualized and differentiated learning. Flipped learning has recently gained popularity in English as a second/foreign language (ESL/EFL) context and the literature investigating the effect of flipped learning in this context has grown recently. Most of the studies that investigated the effect of flipped learning on language skills (e.g., listening, speaking, reading, wiring, vocabulary, grammar) have found positive results (Aidinlou et al., 2017; Boyraz & Ocak, 2017; Ekmekci, 2017; Karimi & Hamzavi, 2017; Koroglu & Cakır, 2017; Kurt, 2017; Lin et al., 2018; Lin, & Hwang, 2018; Oraif, 2018; Saglam & Arslan, 2018; Webb & Doman, 2016; Yu & Wang, 2016;) while in some studies no significant difference has been found between flipped classes and lecture-based classes in terms of affecting ESL/EFL students' achievements; that is the results of both classroom conditions

were relatively similar in respect of improvements in students' performance (Adnan, 2017; Al-Ghamdi & Al-Bargi, 2017; Al-Harbi & Alshumaimeri, 2016; Chui, 2016; Oh, 2017) Therefore, it can be concluded that more research is required to examine this effect in more detail. However, conducting additional research may not provide the best answer regarding the effect of flipped learning on ESL/EFL skills because it might add to the current state of conundrum concerning the literature of flipped learning. Instead, it is timely to conduct a meta-analysis to assist language researchers and instructors to understand the reasons behind the ambiguous results of the previous studies on flipped learning by finding the moderating variables which can explain the inconsistent results of these studies (Stone & Rosopa, 2017). Therefore, in this study, we aimed to carry out a meta-analysis to exclusively examine the effectiveness of flipped learning with regard to English achievements of ESL/EFL students in comparison with lectured-based learning. We also conducted a comprehensive moderator analysis to examine whether different variables in terms of design (instructional design), educational context and study quality characteristics can explain the effectiveness of flipped learning in the ESL/EFL context. Moreover, we exclusively used the meta-regression model to explain the variation in results across different studies.

Previous research

The recent literature regarding the implementation of flipped learning has highlighted the potential benefits and challenges of this approach compared to the traditional lecture-based instruction (e.g., Akçayır & Akçayır, 2018; Hall & DuFrene, 2016; Lo et al., 2017; Rivera, 2016) . One of the key benefits of flipped learning is facilitating the self-paced learning or differentiation which is highly supported by Sams and Bergmann (2013). It seems that students can benefit from this feature depending on how this feature is incorporated in a flipped classroom design. For example, Hibbard et al. (2015), who used semi-self-paced flipped learning format to teach general chemistry, found that the semi-self-paced format was effective in improving the students' performance in a chemistry class. They argued that by using the semi-self-paced flipped format students could review the required materials and they could also practice and apply their knowledge over time. By doing so, they noted that learners could enhance their overall retention and mastery. The semi-flipped format implied that students had the chance to study at their own pace by using different technology-enhanced tools and assessment platforms like Moodle. This self-pacing characteristic increased students' sense of ownership, responsibility, and independence over learning which, in turn, reflected their self-regulated learning. On the other hand, Johnson (2013), who created a self-paced environment using Moodle, found mixed results regarding the self-pacing feature of flipped learning. He found that although the majority of students in his study favored the self-pacing environment, some students argued that self-pacing reduced their motivation to follow the class schedule. He also reported that some students did not like to take responsibility for their learning; they believed that the self-pacing environment decreased their motivation and they could not keep up with the class without motivation. Other students also mentioned that the self-pacing environment gave them too much freedom and that unauthorized access to the internet increased the possibility of off-task behavior.

Another benefit of flipped learning is increasing students' engagement which in turn can enhance the quality of students' performance, satisfaction, and retention (Burke & Fedorek, 2017; Chi & Wylie, 2014; Rivera, 2016). In fact, there are two types of engagement in this learning environment, each of which are associated with different learning requirements

and outcomes: (1) pre-class engagement, which requires students to be engaged in individualized, autonomous learning, and (2) in-class engagement, which requires students to be engaged in interactive, collaborative, learner-centered activities (Diemer et al., 2013; Lee et al., 2018). Moreover, Chi and Wylie (2014) distinguished four types of engagements: (1) passive (e.g., receiving knowledge, silent listening, and reading), (2) active (e.g., physical manipulation of knowledge), (3) constructive (e.g., generating knowledge, reflecting outloud and self-explaining), and finally (4) interactive (e.g., creating dialogs, debating and discussing with peers). While flipped learning is mostly associated with active, constructive and interactive modes of engagement, traditional lecture-based instructions are mostly limited to the passive type of engagement in which learners are receiving knowledge by listening to lectures (Alten et al., 2019; Chi & Wylie, 2014; Tang et al., 2017). It is worth mentioning that in flipped learning environment, the active participation and engagement is considerably increased by the use of technology (Burke & Fedorek, 2017).

Flipping a classroom does not merely mean reversing lectures and homework. Teachers may use technological tools (e.g., online resources, games, simulations, online readings) through different blogs, forums and learning management systems (LMSs) (Bergmann & Sams, 2014) to increase their students' motivation and engagement which is, in fact, a form of active learning (Bradford, 2005). However, Rivera (2016) argued that teachers' overreliance on technology could be a serious issue, especially when facing technology malfunctions (e.g., the technological tools are broken, internet connection is low or is dead). In these cases, teachers should prepare various teaching strategies and back-up plans to keep students engaged in in-class activities. Several studies have found a positive effect of flipped learning on students' engagement (e.g., Burke & Fedorek, 2017; Cronhjort et al., 2017; Elmaadaway, 2017; Johnson, 2013; Rivera, 2016). One point of caution is that since creating different tasks may require enormous time and effort; some teachers may fail to create high quality and creative content which, in turn, may decrease students' motivation and engagement (Burke & Fedorek, 2017).

Efficiency of time usage is another benefit of implementing flipped learning compared to a traditional lecture-based instruction. Since some of the instructional content is learned before class, more face to face time is dedicated to in-person interactions (e.g., student–student and student–teacher interactions). That is, teachers can use this extra time to address different misconceptions and provide more personalized guidance for each of their students (Bergmann & Sams, 2012; Hall & DuFrene, 2016). Also, in addition to receiving support from their teacher(s), students can also seek support from their peers (Alten et al., 2019). In fact, as Bergmann and Sams (2012) suggested, the essence of flipped learning is finding the best way to use in-class time to enhance the learning process. Some studies reported that flipping the classroom can increase student–teacher interactions as well as peer-interactions (e.g., Jungic et al., 2014) and, provide students with efficient time to apply their knowledge in interactive, discussion-based, and problem-solving activities (e.g., Bishop & Verleger, 2013; Farah, 2014; Rivera, 2016; Unakorn & Klongkratoke, 2015). Therefore, students have further chances for benefiting from the real-time, individualized feedback from the teacher(s), peer support, and differentiated learning in a flipped classroom, compared to the traditional lecture-based one (Alten et al., 2019; Lo et al., 2017).

Finally, flipped learning is assumed to improve the communication skills of students (Farrah & Qawasmeh, 2018; Ferreri & O'Connor, 2013; Santikarn & Wichadee, 2018; Unakorn & Klongkratoke, 2015). Rivera (2016) reported that sharing some content through online media can keep students informed of the class process in case of being absent. She also concluded that even parents can be more actively involved in their children's education by monitoring their daily progress. Moreover, she indicated that staying connected through

online media encouraged students to send their questions, exchange their ideas with each other and with the teacher(s), and receive corrective feedback when necessary. Also, several studies have revealed that different collaborative and pair/group work activities can enhance students' communication skills in flipped learning environment (Connor et al., 2014; Munir et al., 2018).

Different meta-analyses and reviews, conducted in recent years, have also provided general positive views about the effectiveness of flipped classrooms regarding students' performance in different academic domains and courses (Alten et al., 2019; Cheng, et al. 2018; Karagol & Esen, 2019), in mathematics (Lo et al., 2017), in health professions education (Hew & Lo, 2018) and health science and non-health science studies (Chen et al., 2017, 2018), in nursing education (Betihavas et al., 2016; Tan et al., 2017) and pharmacy education (Gillette et al., 2018).

The Rationale for selecting the moderator variables

In this meta-analysis, the moderator variables were selected by scanning the procedures of the previous studies on flipped learning and by reading previous publications and meta-analyses, which were conducted in flipped learning environment to see which variables were frequently used in flipped learning studies. Then, based on the categorization used in Alten et al. (2019), we categorized the variables into three moderator models.

Design characteristics

In order to flip a classroom, some steps should be taken into consideration. The first step is to plan lessons by considering the students' needs and learning outcomes. Then, teachers should develop or choose the activities and resources through which the course objectives are covered. In order to have a successful flipped classroom, the activities that are assigned for pre, during, and after class should be congruous with each other. Pre-class activities are mainly designed to enhance lower levels of cognitive skills while in-class activities are associated with developing higher-level cognitive skills, based on Bloom's taxonomy of thinking skills (Bergmann, 2017). Different preparation strategies can be used to develop lower and higher levels of cognitive skills. These preparation strategies are usually divided into pre-class and in-class components of a flipped classroom. Some of the pre-class components, frequently used in primary studies are as follows: (1) the availability of online/offline quizzes or assessments (e.g.,El-Sakka, 2016; Mirshekaran, 2018; Nuon & Champakaew, 2017; Oraif, 2018), (2) providing students with extra exercises (e.g., Al-Hamdani & Al-Breiki, 2018; Farah, 2014; Gurluyer, 2019; Iyitogu & Erisen, 2017; Santosa, 2017), and (3) asking students to take notes of important or problematic parts or giving them pre-class reading assignments to help them understand the subject of lessons before class (Karakurt, 2018; Ozkal, 2019; Teng, 2017; Zou & Xie, 2018). These pre-class strategies can be created and delivered to students by using blogs, mobile applications, or LMSs (Sezer et al., 2017). The most common in-class components of flipped classes are availability of quizzes/assessments at the start of the class (e.g., Chavangklang & Suppasetseree, 2018; Li, 2015), different pair/group work activities (e.g., El-Bassuony, 2016; Oh, 2017; Zou & Xie, 2018) and problem-solving and discussion activities (e.g., Cakıroglu & Ozturk, 2017; Lin & Hwang, 2018; Yu & Wang, 2016). In fact, student preparation is an essential moderator of the effectiveness of flipped learning because the assumption is that students are prepared

for the in-class materials by completing and reviewing all assigned pre-activities (Gillette et al., 2018).

It should be mentioned that in our moderator analysis in this study, we categorized all of the above-mentioned pre-class components along with "availability of quiz in class" under instructional design moderator model. Other in-class components were not considered in our moderator analysis due to their limited information from primary studies.

"Offline/online self-assessment quizzes" can be accompanied by videos to encourage students to study the materials more carefully before coming to class. Pre-class online or in-class quizzes enable teachers to use the results of quizzes to identify the problematic areas or misconceptions (Jungic et al., 2014) and can also ensure teachers that students have mastered the subject matter before coming to class. Some reviews on flipped learning found that using quizzes moderated the findings in different academic domains and students achieved better outcomes when quizzes were added to flipped classrooms (e.g., Alten et al., 2019; Hew & Lo, 2018; Lo et al., 2017). In some studies, the pre-class activities are limited to watching videos (e.g., Alnuhayt, 2018; Hashemifardnia et al., 2018; Leis, 2016), but others have used extra activities and exercises along with instructional videos to enhance students' individualized learning and promote students' engagement and performance (e.g., Chavangklang & Suppasetseree, 2018; Farah, 2014). We expected to find similar results in the ESL/EFL field. We also analyzed the availability of "note-taking and pre-class reading assignments" as another potential moderator variable. Note-taking can be considered as a pre-class preparation method because students can come to class more prepared while knowing the general concept of the class with more organized, unresolved questions to ask. Moreover, students can use their notes to complete in-class activities or exercises. This can also ensure teachers that students watched the videos or completed the pre-class activities and came to class well-prepared (Cormier & Voisard, 2018; Jaster, 2017; McCallum et al., 2015). Besides watching videos, reading assignments is considered as the most frequently used pre-class method for flipped classes (Han & Klein, 2019) in the form of text-book reading (Saunders et al., 2017), guided reading (Lieu et al., 2017; O'Connor & Ferreri, 2013; Persky & Hogg, 2017; Persky & Pollack, 2011) and primary literature/guidelines reading (Chokshi et al., 2017; Tan et al., 2015). Finally, we were interested to see if "the use of LMSs" might account for some of the variances in effect sizes of studies because some believed that effective application of LMSs could lead to promoting students' engagement and performance (Chaubey & Bhattacharya, 2015; Swart, 2016).

Educational context characteristics

Some meta-analyses have considered "academic domains" or "subject areas" as a possible moderator variables (Alten et al., 2019; Cheng et al., 2018; Kang & Shin, 2016). However, the focus of our meta-analysis is on flipped learning studies conducted in the ESL/EFL context. Therefore, two possible moderator variables that we classified under educational context model were "duration of intervention" and "educational level".

The first moderator variable was "duration of intervention" on the basis of the inconsistent results of previous meta-analyses. Flipped classrooms have social characteristics due to their high degree of interactions and cooperative learning activities (Lag & Saele, 2019). Students may initially have negative perceptions toward flipped learning, requiring more time to be familiarized with this approach, especially if they are new to this approach. Plonsky and Oswald (2014) believed that longer interventions could lead to more substantial effects. However, some studies found no significant difference between the effectiveness

of this approach and the overall study duration (e.g., Cheng et al., 2018; Karagol & Esen, 2019).

Flipped classroom model is popular among different educational levels. Although K-12 flipped classrooms may have different meanings and challenges (e.g., student-related and operational challenges), compared to higher-education flipped classrooms, the main design characteristics of flipped classrooms seem similar in all contexts (Alten et al., 2019; Lo & Hew, 2017). Different meta-analyses have found inconsistent results regarding the effect of educational levels on students' performance in various academic domains (Alten et al., 2019; Cheng et al., 2018; Kang & Shin, 2016; Karagol & Esen, 2019). Therefore, we considered "the educational level" as a potential moderator component to see if differences between these levels could moderate the overall results in the ESL/EFL context. Since some of the primary studies, conducted in private English institutions, included the composite or mixed-age groups, we added another educational level code called "private English institute".

Study quality characteristics

In order to interpret the findings more thoroughly, Plonsky and Oswald (2014) suggested eight key criteria that can be used as guidance for interpreting the overall effect size. One of the suggested criteria is "methodological quality". Some specific features of methodological quality include "pretesting, random group allocation, delayed post-testing, control for bias, researcher experience, etc.". Adopted from Alten et al. (2019), we categorized "allocation type", "experimental/control group equivalence test", "study type" and the "type of outcome measurement (standardized vs. teacher-researcher developed tests)" under the moderator model called study quality. We assigned three codes to allocation type (no randomization, randomizing based on the individual level, and randomizing based on the pre-existing groups) to check if the possibility of selection bias may moderate the results. Since studies in which students are randomly assigned to groups have larger effect sizes than non-random allocations (Lag & Saele, 2019), we used this variable as a potential moderator to see if different groups' assignments may moderate the results (Alten et al., 2019).

Generally, "equivalence tests" are used to gain certainty about the comparability of experimental and control groups (Lo & Hew, 2017; Zhu et al., 2019). Thus, we also used "equivalence tests" as another moderator variable. Moreover, we used "study type" as another moderator variable to check the difference in effect sizes between publications in journals, conference papers, and M.A./Ph.D. theses (Alten et al., 2019; Cheng et al., 2018).

It should be mentioned that the definitions of standardized and teacher/researcher-developed tests were adopted from Mehrens and Lehmann (1978). Standardized tests are defined as the type of tests that are developed by publishing companies, formal testing agencies or universities, and skilled test-makers. These tests have specific instructions and their scoring procedures are highly valid and reliable. Also, as they are used by large sample populations, they undergo several revisions. These types of tests are highly recommended for measuring broad curriculum objectives and inter-class or school comparisons. On the other hand, teacher/researcher tests, the classroom tests that are prepared by the teacher/researcher by considering the immediate needs of students, are not applied by large populations, do not undergo revision or item analysis and are mostly appropriate for measuring the classroom objectives and intra-class comparisons (cited in Farhady et al., 2010, p.6; Mehrens & Lehmann, 1978).

Research Objectives

Our first research objective was to conduct a comprehensive meta-analysis to investigate the relative effectiveness of flipped learning with regard to students' achievements in the ESL/EFL context. Previous meta-analyses were conducted on various academic domains or courses. With the paucity of meta-analyses on flipped learning in the ESL/EFL context in mind, we only focused on the studies in which the effect of flipped learning was investigated in this context.

Our second research objective was to implement a moderator analysis to examine if the heterogeneity observed between the studies in this meta-analysis could be explained by the moderator variables. In doing so, we categorized the moderator variables under three models: design (instructional design), educational context, and study quality. Overall, the main purpose of this meta-analysis was to answer the following questions:

What is the overall effect of flipped learning on students' performances and achievements in the ESL/EFL context?
Does the effect of flipped learning on ESL/EFL students' achievements depend on design (instructional design), educational context, and study quality characteristics?

Method

Literature search

Several techniques were used to access the literature of flipped learning in the ESL/EFL context. As suggested by Polanin et al. (2016), we included unpublished studies (M.A./Ph.D. dissertations and conference papers) to reduce the potential bias from different publications. Since we only intended to consider publications in which flipped classrooms were conducted in the ESL/EFL context; first, we browsed different databases including Education Resources Information Center (ERIC), Science Direct, Web of Science, ProQuest, Google Scholar, and Chinese databases such as CNKI and Ericdata to find articles, conference articles, and dissertations. To do this, we searched keywords such as "flipped learning", "flipped classroom", "inverted classroom", "reversed classroom", "the ESL/EFL context" and "English skills. Moreover, the reference sections of the downloaded publications were also cross-checked for additional research studies. Our first search process started in October 2018. After reviewing more than 2000 documents, 49 publications, which were published between 2014 and September 2019, met our inclusion criteria. Since some of these publications included more than one single study, there were 69 independent studies in 49 publications.

Inclusion of studies

First, out of the total of 2800 retrieved publications, nearly 800 duplicates were removed. Then more than 1000 studies were excluded because their topics were not about implementing flipped learning in the ESL/EFL context, or their full texts were not available. We later excluded studies which were not written in English language, used designs other than between-subject designs with control and experimental groups (studies with no control group or within-subject designs), were not empirical studies (qualitative, action research,

case studies, etc.), and did not have enough information to calculate effect sizes. Also, three studies were excluded because flipping the classroom was not their main focus. In total, 69 independent studies were included in our meta-analysis. In should be mentioned that while conducting moderator analysis, one study (e.g., Alavi et al., 2016) had insufficient information about the procedure used in their flipped classroom. We had to exclude this study from the moderator analysis because, despite trying to contacting the author, we could not retrieve the missing information. The comprehensive set of inclusion and exclusion criteria can be seen in Table 1.

Coding of study characteristics and moderator variables

A solid coding procedure is necessary to extract data out of the research literature (Brown et al., 2003). As suggested by Cooper (2010), we prepared a coding sheet which consisted of four parts: (1) study features including descriptive data about the study (Study ID number, author, year, publication type, funding status, peer-reviewed, sample size, first language of participants, learning context, origin, gender of participants, L2 proficiency, educational context, type of school, sampling assignment, research design, theoretical underpinning, duration of intervention, number of sessions and focus of the study), (2) instrument features and outcome measures (outcome type, outcome skill, flipped classrooms procedures including pre-class, in-class and after class strategies, measurement of outcome, instrument, test structure, piloting), (3) effect size estimate (effect size metric and statistical information required for calculating effect size), and (4) conclusion reports. We first provided a code for the control group procedure. However, we had to eliminate it because most of the studies did not report the control group procedures. We also had to eliminate the code called "the type of test" because some studies did not thoroughly explain the types of tests that were used for measuring the English outcomes. Also, some other codes such as "instructor equivalence", "types of in-class activities", and "duration of class time (hours)" were eliminated due to lack of sufficient information.

After reviewing the coding schemes, the first author independently coded all the studies and the third author also double-coded 25% of the studies randomly. They discussed and resolved any discrepancies and disagreements, which resulted in high inter-coder reliability between the two coders. The Cohen's Kappa coefficient value was found to be 0.90.

Effect size computation and data analysis

In this meta-analysis, the effect sizes were calculated by using means and standard deviations. However, when the raw data did not include means and standard deviations, other statistical values such as T value and p value were used to calculate the effect size.

In our meta-analysis, Comprehensive Meta-analysis (CMA) (Borenstein et al., 2009) (https://www.meta-analysis.com/) was used for systematic data analysis. Since different characteristics of flipped classrooms such as the frequency of flipped lessons, duration of study intervention, and types of exercises and assessments used in pre and in-class may be involved in affecting students' achievements, we assumed that there were different effect sizes underlying different studies and that the true effect size was not the same in all studies. Therefore, we used a random-effects model, which accounts for two sources of variance: within-study variance and between-study variance (Borenstein et al., 2009). We also used the standardized mean difference called *Hedges's g* that indicates the difference between the means of the experimental and the control groups as a measure of effect size to summarize the findings of

Table 1 Inclusion and exclusion criteria

	Inclusion	Exclusion
Language	Written in English	Studies written in Chinese from (CNKI), Turkish, Russian or languages other than English
Design	Empirical studies (quasi-experimental, experimental	Non-empirical studies (purely qualitative, case studies, action research)
Impact study	Studies investigating the effect of flipped learning in the ESL/EFL context	Studies investigating the effect of flipped learning on subject areas or courses other than English as a second/foreign language
Control group	Between-subject designs with real control and experimental groups	Studies with no control groups and studies with within-subject designs (not real control groups like counterbalanced designs)
Availability	Publications, the full texts of which are available	Publications, the full texts of which are not available
Statistical information	Studies that reported sufficient data for calculating the effect size like combinations of means, standard deviations, t test or ANOVA, MANOVA, MANCOVA statistics	Studies that did not report sufficient statistical information
Time period	Studies written in English and published from 2011 to September 2019 were included in the data set. Since flipped learning is a new research area in the ESL/EFL context, no study was found prior to 2011.	

each study and to estimate the effect sizes of flipped learning on English outcomes of ESL/EFL students for each study (Hedges & Olkin, 1985, p.78). The reason behind using *Hedges's g* instead of *Cohen's d* is because *d* has slight bias and overestimates the absolute value of effect size parameter in small samples (see Borenstein et al., 2009). In order to accurately interpret the magnitude of effect size, the field-specific benchmark for L2 research, proposed by Plonsky and Oswald (2014), was used, in which the small, moderate and large effects were (d = 0.40), (d = 0.70), and (d = 1.00) respectively. It should be mentioned that this benchmark is specifically designed for between-group contrasts in the field of L2 research because within-group or pre-post contrasts have larger d values due to intra-group correlations (Plonsky & Oswald, 2014).

Moreover, we reported I^2, Q, T^2, and T to reveal the statistical heterogeneity in our meta-analysis. We used a mixed-effects meta-regression to conduct the moderator analysis using a two-tailed test for p values (Two-tailed) and (95%) confidence level. In order to estimate and fully explain the amount of residual heterogeneity ($\tau 2$), which accounts for the between-study variance, the restricted maximum likelihood estimation was used. I^2 (the ratio of between-study variance to total variance) is presented as a percentage of unaccounted variation between studies due to heterogeneity rather than sampling error and R^2 is given as the proportion of variance explained by the predictors (Borenstein et al., 2009). We also used the Knapp-Hartung adjustment (Knapp & Hartung, 2003) to account for the uncertainty in the estimate of residual heterogeneity. With this adjustment, the omnibus test of the moderator analysis uses an *F-distribution* with m and $k - p$ degrees of freedom. Since using multiple (i.e., univariate) tests can cause *type I error* (falsely rejecting a true null hypothesis), and also putting all variables in one model can lead to *type II error* (failing to reject a null hypothesis that is false due to low statistical power), as suggested by Alten et al., (2019), we categorized our moderator variables under three different meta-regression models to account for the risk of *type I* and *type II* errors (Alten et al., 2019; Polanin & Pigott, 2015; Stone-Romero et al., 1994).

Publication bias

One of the most recognized sources of publication bias is the preference for publishing high-quality research i.e., studies in which only the significant results are reported, which would result in generating over-estimated overall effects (Plonsky & Oswald, 2014). Therefore, as suggested by Plonsky and Oswald (2014) as well as Polanin et al. (2016), besides the inclusion of published studies in this meta-analysis, we also included unpublished ones (M.A./Ph.D. dissertations/theses and conference papers) to reduce the potential bias from different publications. We also considered "type of publication" as a potential moderator variable to check if there was a difference in effect sizes of published and unpublished studies i.e., to check the possibility of larger effect sizes for published studies, compared to unpublished ones (Alten et al., 2019; Plonsky & Oswald, 2014). Funnel plot and classic fail-safe N, (Orwin, 1983) were also created to investigate the heterogeneity and the publication bias.

Results

Although the searching time span was from 2011 to 2019, the first included study was published in October 2014 because flipped learning is a new pedagogical approach in the ESL/EFL field. The last study that was included in the meta-analysis was published in September 2019. The descriptive information of all included studies is presented in Table 2.

Educational level[a]: *SCE* secondary education, *PEI* private English Institute, *HE* higher education, Groups assignments [b]: *RI* random on individual level, *PGA* pre-existing groups allocated, *NR* not-randomization, *g (p)*[c] standardized mean difference in Hedges g (p value of associated effect size), N^d total participants in flipped–control group.

Moreover, the funnel plot and classic fail-safe N were examined to check the publication bias. The results of Fail-Safe N indicated a z value of 32.94 and a 2-tailed p value of 0.000 with a Fail-Safe N of 9423. So, in this case, 9423 'null' studies should be found and incorporated for the combined 2-tailed p value to exceed 0.050, which seems a rather large number of studies showing lack of publication bias in this study. The results of the funnel plot under the random effects model are shown in Fig. 1.

Effect size results: effect of flipped learning on the student's outcomes in the ESL/EFL context

The result of the random-effect meta-analysis is presented in Table 3. The result of this meta-analysis indicated an overall significant effect in favor of flipped learning for English achievements of students in the ESL/EFL context (*g: 1.24, 95% CI 1.05–1.44, $p < 0.001$).* Considering the heterogeneity values in students' outcomes (*$Q = 577.05$, $df = 68$, $p = 0.000$, $I^2 = 88.22$),* the random-effects model was used and the existence of moderator variables was confirmed. Since *Q (577.05)* is much bigger than the degree of freedom *(68), T and* T^2 which represent the heterogeneity measures are big and significant, and I^2 is also big and close to one *(88.22),* given the field-specific benchmark proposed by Plonsky and Oswald (2014) in which the small, moderate and large effects were (d = 0.40), (d = 0.70), and (d = 1.00) respectively for between-subject designs. We did not use Cohen's (1988) benchmark because Plonsky and Oswald (2014) indicated that "Cohen's benchmarks generally underestimate the effects gained in L2 research" (p. 18) and it is not appropriate for interpreting the practical significance of effect sizes in field of L2 research. According to the forest plot (See Fig. 2), the smallest effect size value is −0.22, and the highest effect size value is 6.32. Moreover, the diamond shape, at the end of the plot, is located on the positive side of the plot, indicating that the students in the experimental groups had better achievements than those in the control groups. Furthermore, since 67 studies had positive effect sizes and only two studies had negative effect sizes, it could be concluded that flipped learning, which was implemented in 67 studies, had significant effects in favor of the experimental groups.

Moderator analysis

In order to conduct the meta-regression analysis, the moderators were grouped under three different models because grouping all of them under one model might lead to type I error (univariate moderator analysis) and type II error (all variables in one model). The variables were categorized as *design characteristics* (e.g., availability of quiz before class, availability of exercises and activities before class, availability of reading or note-taking before class, availability of quiz in class and availability of a LMS for delivering the materials to students and interacting with students), *educational context characteristics* (e.g., duration of flipped intervention and educational levels of students), and *study quality characteristics* (e.g., allocation type, experimental and control groups equivalence tests, study type, type of outcome measurement).

Table 2 Description of the studies included in the meta-analysis

Authors	Publication type	Educati onal level[a]	Assignments[b]	Outcome	g (p)[c]	N[d]
Abaeian and Samadi (2016)	Journal	PEI	RI	Reading	2.19 (0.00)	50/50
Abedi, Keshmirshekan, et al. (2019)	Journal	PEI	RI	Writing	1.00 (0.01)	16/16
Abedi, Namaziandost, et al. (2019)	Journal	PEI	RI	Writing	1.97 (0.00)	24/24
Afrilyasanti et al. (2016)	Journal	SCE	PGA	Writing	2.25 (0.00)	30/32
Ahmed (2016)	Journal	HE	PGA	Writing	6.32 (0.00)	30/30
Alakawi (2016)	Journal	HE	PGA	Listening	1.60 (0.00)	20/20
Alavi et al. (2016)	Journal	PEI	RI	Overall academic	1.16 (0.00)	20/20
Al-Ghamdi and Al-Bargi (2017)	Journal	HE	NR	Speaking	0.22 (0.47)	21/21
Al-Hamdani and Al Breiki (2018)	Journal	SCE	PGA	Vocabulary	0.57 (0.05)	25/25
Al-Harbi & Alshumaimeri, 2016	Journal	SCE	NR	Grammar	0.32 (0.28)	20/23
Alnuhayt (2018)	Journal	HE	NR	Vocabulary	0.99 (0.00)	24/21
Alzaytuniya (2016)	MA thesis	SCE	NR	Grammar	1.63 (0.00)	30/30
Amiryousefi (2017) (EX1) A	Journal	HE	NR	Listening	0.61 (0.04)	23/22
Amiryousefi (2017) (EX1) B	Journal	HE	NR	Listening	1.80 (0.00)	23/22
Amiryousefi (2017) (EX2) A	Journal	HE	NR	Listening	0.42 (0.16)	22/22
Amiryousefi (2017) (EX2) B	Journal	HE	NR	Listening	2.14 (0.00)	22/22
Amiryousefi (2017) (EX1)	Journal	HE	NR	Speaking	1.54 (0.00)	23/22
Amiryousefi (2017) (EX2)	Journal	HE	NR	Speaking	1.47 (0.00)	22/22
Boyraz and Ocak (2017)	Journal	HE	PGA	Grammar	1.25 (0.00)	17/23
Bulut (2018)	MA thesis	HE	NR	Grammar	0.76 (0.02)	19/20
Cavdar (2018)	MA thesis	HE	NR	overall Academic	− 0.22 (0.29)	52/44
Chavangklang and Suppasetseree (2018)	Journal	HE	PGA	Reading	1.29 (0.00)	34/37
Ekmekci (2017)	Journal	HE	RI	Writing	1.92 (0.00)	23/20
Farah (2014) A	PhD thesis	SCE	NR	Writing	1.06 (0.00)	21/18
Farah (2014) B	PhD thesis	SCE	NR	Writing	1.20 (0.00)	21/18
Farah (2014) C	PhD thesis	SCE	NR	Writing	0.88 (0.01)	21/18

Table 2 (continued)

Authors	Publication type	Educati onal level[a]	Assignments[b]	Outcome	g (p)[c]	N[d]
Haghighi et al. (2018)	Journal	HE	RI	Pragmatic	1.14 (0.00)	30/30
Hashemifardnia et al. (2018)	Journal	SCE	RI	Reading	0.49 (0.08)	25/25
Huang and Hong, 2015	Journal	SCE	RI	Listening	0.80 (0.00)	40/37
Hung (2015) (EX1) A	Journal	HE	PGA	Overall academic	−0.13 (0.64)	26/25
Hung (2015) (EX1) B	Journal	HE	PGA	Overall Academic	1.54 (0.00)	26/25
Hung (2015) (EX1) C	Journal	HE	PGA	Overall academic	1.52 (0.00)	26/25
Hung (2015) (EX2) A	Journal	HE	PGA	Overall academic	0.15 (0.60)	24/25
Hung (2015) (EX2) B	Journal	HE	PGA	Overall academic	0.63 (0.03)	24/25
Hung (2015) (EX2) C	Journal	HE	PGA	Overall academic	0.57 (0.05)	24/25
Hung (2017)	Journal	HE	NR	Overall academic	0.97 (0.00)	21/22
Ishikawa et.al (2015)	Conference	HE	RI	Overall academic	0.90 (0.00)	243/203
Iyitogu and Erisen (2017)	Journal	HE	PGA	Grammar	1.16 (0.00)	21/20
Iyitogu and Erisen (2017)	Journal	HE	PGA	Listening	0.35 (0.25)	21/20
Iyitogu and Erisen (2017)	Journal	HE	PGA	Overall academic	1.14 (0.00)	21/20
Iyitogu and Erisen (2017)	Journal	HE	PGA	Reading	0.96 (0.00)	21/20
Iyitogu and Erisen (2017)	Journal	HE	PGA	Vocabulary	0.87 (0.01)	21/20
Iyitogu and Erisen (2017)	Journal	HE	PGA	Writing	1.05 (0.00)	21/20
Kang (2015)	Journal	HE	NR	Grammar	0.02 (0.93)	24/42
Kang (2015)	Journal	HE	NR	Overall academic	0.18 (0.49)	24/42
Kang (2015)	Journal	HE	NR	Vocabulary	0.39 (0.13)	24/42
Karakurt (2018)	MA thesis	HE	RI	Grammar	0.63 (0.05)	20/20
Karimi and Hamzavi (2017)	Journal	PEI	RI	Reading	2.89 (0.00)	25/25
Koroglu and Cakir (2017)	Journal	HE	PGA	Speaking	2.04 (0.00)	23/25
Leis et. al (2016)	Journal	HE	NR	Writing	0.56 (0.18)	11/11
Li (2015)	MA thesis	HE	PGA	Speaking	0.75 (0.00)	46/48

Table 2 (continued)

Authors	Publication type	Educati onal level[a]	Assignments[b]	Outcome	g (p)[c]	N[d]
Lim (2015)	Journal	HE	NR	Writing	2.36 (0.00)	20/20
Lin and Hwang (2018)	Journal	HE	PGA	Speaking	3.22 (0.00)	33/16
Lin et al. (2018)	Journal	HE	PGA	Writing	1.91 (0.00)	35/33
Mirshekaran (2018)	Conference	PEI	RI	Listening	4.51 (0.00)	40/40
Nguyen et al. (2019)	Journal	HE	NR	Writing	0.68 (0.00)	40/40
Nugroho and Insana (2018)	Journal	HE	NR	Writing	0.52 (0.10)	20/20
Nuon and Champakaew (2017)	Conference	HE	NR	Grammar	0.76 (0.00)	40/41
Oraif (2018)	PhD thesis	HE	PGA	Writing	1.53 (0.00)	24/31
Ozkal (2019)	MA thesis	PEI	PGA	Vocabulary	0.90 (0.00)	52/44
Qader and Arslan (2019)	Journal	HE	RI	Writing	0.65 (0.01)	34/32
Quyen andVan Loi (2018)	Journal	HE	PGcA	Speaking	0.83 (0.00)	30/30
Santosa (2017)	Journal	HE	PGA	Writing	2.48 (0.00)	63/88
Teng (2017) (EX1)	Journal	HE	RI	Overall academic	3.50 (0.00)	30/30
Teng (2017) (EX2)	Journal	HE	RI	Overall academic	2.30 (0.00)	30/30
Tuna (2017)	MA thesis	HE	NR	Writing	1.23 (0.00)	24/24
Webb and Doman (2016)	Journal	HE	NR	Grammar	0.64 (0.01)	39/25
Yu and Wang (2016)	Journal	HE	NR	Writing	0.47 (0.05)	35/36
Zou and Xie (2018)	Journal	HE	PGA	Writing	2.81 (0.00)	34/32

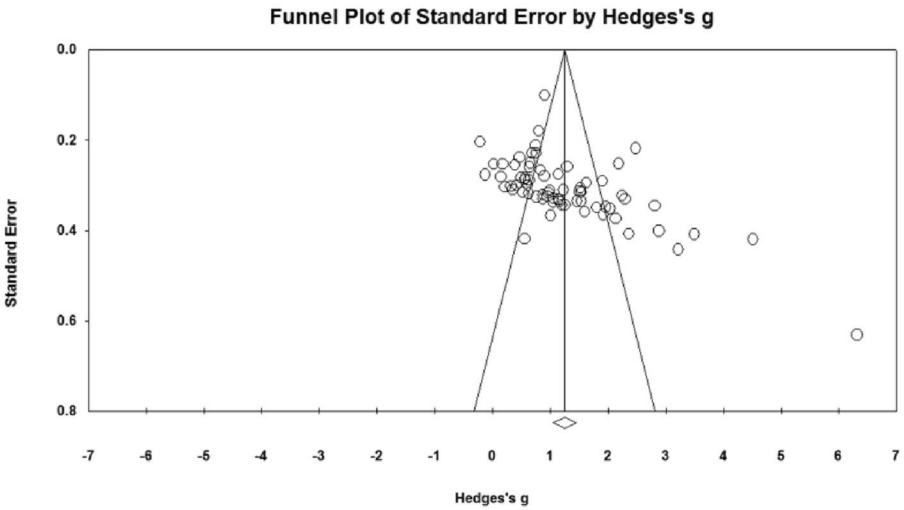

Fig. 1 Funnel plot under random-effects model

As seen in Table 4, the omnibus test of all regression coefficients in the design characteristics model (p = 0.021) was significant. Thus, it can be concluded that the design characteristics model moderated the findings as 17% of heterogeneity was explained by this model. Moreover, studies that used exercises and extra activities before class significantly moderated students' achievements in the ESL/EFL context (p = 0.003), compared to the ones in which exercises were not used before classes (with a difference of $g = -0.851$ while accounting for all of the variables in the model). It seems that other variables of the design characteristics model did not moderate the findings. As for the other two moderator models, neither educational context (p = 0.097) nor study quality (p = 0.329) moderated the findings because their omnibus tests of all regression coefficients rendered insignificant results. However, one moderator variable in the educational context model i.e., the educational level (p = 0.049) seemed to affect the results. It seemed that the studies conducted in private English institutes with multi-age groups had a significantly higher (p = 0.019) effect than the ones conducted at secondary school level (with a difference of g = − 1.06, while accounting for all of the variables in the model).

Discussion

So far all meta-analyses in the field of education, which focused on analyzing the influence of flipped learning approach on different disciplines, have used a similar benchmark i.e., Cohen's (1988), for interpreting the effect sizes, even though Cohen's (1988) is a general benchmark, mainly appropriate for the analysis of the statistical power but not suitable for interpreting the results (Plonsky & Oswald, 2014).

Moreover, Hill et al. (2008) argued that effect sizes should be interpreted with respect to the relevant context of that intervention and, in doing so, relevant empirical benchmarks should be used. Therefore, since our meta-analysis purely focused on tbecause it "underestimates the range of effects typically obtained in L2 research"

Table 3 The results of the random-effect meta-analysis

Model			Effect size and 95% confidence interval				Test of null [2-tail]-		Heterogeneity				Tau-squared			
Model	Study number	Point estimate	SE	Variance	Lower limit	Upper limit	Z value	P value	Q value	df [Q]	P value	I-squared	Tau-squared	SE	V	Tau
Fixed	69	1.06	0.03	0.00	0.99	1.13	31.36	0.00	577.05	68.00	0.00	88.22	0.60	0.15	0.02	0.77
Random	69	1.24	0.10	0.01	1.05	1.44	12.39	0.00								

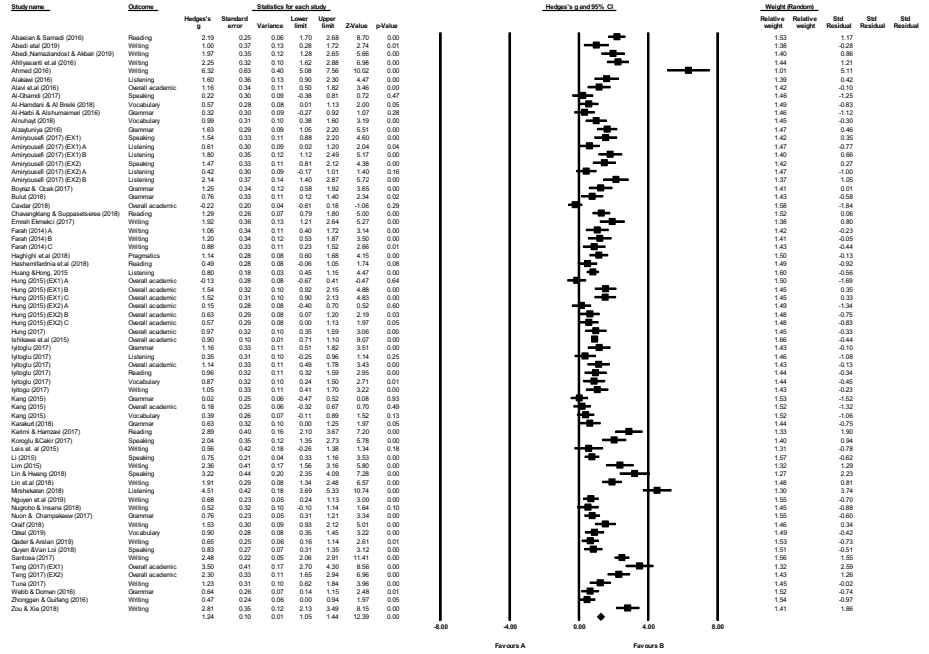

Fig. 2 Forest plot under random-effects model

(Plonsky & Oswald, 2014, p.12). Instead, we used the empirically based and field-specific benchmark proposed by Plonsky and Oswald (2014) for between-groups contrasts in an L2 context to gain more reliable and informative interpretations of the results (small effect (d = 0.40), moderate effect (d = 0.70), and large effect (d = 1.00)). However, Plonsky and Oswald (2014) suggested that their field-specific benchmark should also be used with caution and it must be considered as "very general indicators of magnitudes of mean differences and correlations typically observed in L2 research" (p. 14), and other factors must also be taken into consideration when interpreting the effect-sizes. Among the eight factors or criteria suggested by Plonsky and Oswald (2014), the first one was to compare the results with those of the previous studies, even considering the results of other meta-analyses with the same topics to explain the findings of the similar ones. However, in the absence of previous meta-analyses with similar domains, Plonsky and Oswald (2014) indicated that "when a meta-analytic replication is not available, reviews of comparable domains can still be put to good use" (p.16). Therefore, since at the time of our data analysis, we couldn't find any other meta-analysis which examined the effect of flipped learning with regard to students' achievements in the ESL/EFL context, we compared our findings with those of other meta-analyses examining the influence of flipped learning on the students' achievements in domains other than the ESL/EFL field.

Table 4 Results of the moderator analysis

Moderator variable	N[b]	Estimate (SE)[c]	P	95% CI[d]		F (df1, df2)[h]	p[*]	τ^{2e} res	I^{2f} res	R^2g
Design characteristics						F (5, 62) = 2.88	p = 0.021	0.74	87.07%	0.17
Intercept	68	1.890 (0.324)	0.000	(1.246, 2.539)						
Quiz before class										
No (ref)[a]	35									
Yes	33	−0.285 (0.240)	0.239	(−0.765,0.194)						
Exercise before class										
No (ref)	19									
Yes	49	−0.851 (0.270)	0.003	(−1.392, −0.311)						
Reading/note taking before class										
No (ref)	18									
Yes	50	0.351 (0.270)	0.199	(−0.189, 0.890)						
Quiz in-class										
No (ref)	52									
Yes	16	−0.497 (0.282)	0.083	(−1.061, 0.067)						
LMS										
Not used (ref)	47									
Used	21	−0.064 (0.260)	0.806	(−0.583, 0.456)						
Educational context characteristics						F (3, 64) = 2.20	p = 0.097	0.83	87.64%	0.07
Intercept	68	1.073 (0.204)	0.000	(0.665, 1.480)						
Duration of intervention										
Less or equal than 9 weeks (ref)	30	0.200 (0.250)	0.426	(−0.299, 0.699)						
More than 9 weeks	38									
Educational Level										
Higher education (ref)	53									
Private English institute	6	1.068 (0.443)	0.019	(0.182,1.954)	F = 3.16, df = 2, dfErr = 64, p = 0.049					
Secondary school	9	−0.145 (0.364)	0.693	(−0.872, 0.583)						

Table 4 (continued)

Moderator variable	N[b]	Estimate (SE)[c]	P	95% CI[d]		F (df1, df2)[h]	p*	τ^{2e} res	I^{2f} res	R^2g
Study quality characteristics						F (10, 57) = 1.17	p = 0.329	0.87	88.00%	0.03
Intercept	68	0.912 (0.384)	0.021	(0.144, 1.680)						
Allocation type										
Not randomized (ref)	27									
Individual level	14	0.706 (0.369)	0.061	(−0.034, 1.445)	F = 1.98, df = 2, dfErr = 57, p = 0.147					
Pre-existing groups	27	0.432 (0.309)	0.167	(−0.186, 1.050)						
Ex/co equivalence test										
Tested, equal (ref)	45									
Not tested at all	1	−0.135 (1.070)	0.900	(−2.277, 2.007)	F = 0.51, df = 4, df Err = 57, p = 0.729					
Tested but not reported	5	−0.709 (0.501)	0.163	(−1.713, 0.296)						
Tested on placement test	15	−0.143 (0.333)	0.669	(−0.810, 0.524)						
Tested, not equal	2	−0.320 (0.799)	0.690	(−1.920, 1.279)						
Study type										
Journal (ref)	54									
Conference	3	0.439 (0.635)	0.493	(−0.834, 1.711)	F = 0.70, df = 3, dfErr = 57, p = 0.553					
MA thesis	7	−0.498 (0.439)	0.261	(−1.377, 0.381)						
PhD thesis	4	0.145 (0.623)	0.817	(−1.103, 1.393)						
Outcome measurement										
Standardized test (ref)	15									
Teacher/researcher made	53	0.194 (0.342)	0.573	(−0.491, 0.879)						

Ref[a] reference category, *N*[b] total number of studies, estimate: effect sizes, SE[c] standard error, *CI*[d] confidence interval, τ^{2e} estimated amount of residual heterogeneity, I^{2f} percentage of unaccounted variation between studies that is due to heterogeneity rather than sampling error, R^{2g} proportion of amount of heterogeneity accounted for.

[h]Omnibus test of all regression coefficients of the moderators in the model.

*p Value of the omnibus test.

Is flipped learning effective in improving students' achievements in the ESL/EFL context?

The first research question of our meta-analysis was about the relative effectiveness of flipped learning regarding students' achievements in the ESL/EFL context. In order to interpret our results with respect to Plonsky and Oswald's (2014) benchmark, we converted our results from *g* to *d*, as a result of which, an overall significant large effect ($d = 1.26$, CI [1.06 to 1.46], $p < 0.001$) was found in favor of the flipped learning on students' achievements in the ESL/EFL context.

Most of the meta-analyses on the effectiveness of flipped learning found small to moderate positive effects on the performance, outcomes, and achievements of students in different domains and disciplines, thereby concluding that this pedagogical approach can be effective if conducted properly. The results of these meta-analyses are significant due to having a narrow range of confidence intervals (CIs). Therefore, we mainly compared the CI of our findings with those of these meta-analyses (e.g., Alten, et al., 2019; Cheng et al., 2018; Hew & Lo, 2018; Karagol & Esen, 2019; Lage & Saele, 2019; Lo et al., 2017). Tan et al. (2017) found a large effect size in favor of flipped instruction in improving students' academic performance in knowledge and skill (SMD = 1.13, CI (0.76, 1.49), $P < 0.00001$ for knowledge and SMD = 1.68, CI (1.24, 2.12), $P < 0.00001$ for skill), just like the findings of the current study in which a quite large effect was found in favor of flipped learning in improving students' achievements in the ESL/EFL context. However, it should be noted that the effect size precision of the current study was higher than that of Tan et al. (2017) due to having a narrower range of CI (SMD = 1.24, CI [1.05 to 1.44], $p < 0.001$).

It is worth noting that we did not compare our findings with those of some reviews (e.g., Kang & Shin, 2016; Zhu et al., 2019) because they suffered from some limitations. For example, although Kang and Shin (2016) found an overall moderate effect size in favor of flipped learning (SMD: 0.54, CI [0.35to 0.97], we assumed that their findings were not valid for two reasons, the most important of which was including both between-subject and pre-post contrast designs in their analysis. As Plonsky and Oswald (2014) mentioned, within-subject and pre-post contrasts may have larger effects because the same group of participants might experience both experimental and control interventions which "reduces error variance and accentuates the strength of the effect (all else being equal)" (p. 9). Moreover, both Kang and Shin (2016) and Zhu et al. (2019) examined the effect of flipped learning on English outcomes along with other disciplines and subject areas. Again, since various disciplines were evaluated in these studies, Cohen's rule of thumb did not seem to be an appropriate benchmark for interpreting the results, because, as argued by Hill et al. (2008), effect sizes should be interpreted with respect to the relevant context of that intervention. Therefore, it was more appropriate to evaluate the findings of studies in which English skills served as dependent variables with a more specific context-bound benchmark.

Do different moderator models (design characteristics (instructional design), educational context, or study quality) explain the difference in the effects of the flipped learning?

The second research question of our study was about the effect of moderator variables in terms of three models on the effectiveness of flipped learning in the ESL/EFL context. A statistically significant and large variability was also found across studies [Q

(68) = 577.05, p = 0.00, I^2 = 88.22], which indicated that the moderator variables might affect the findings.

Other meta-analyses on the effectiveness of flipped learning in different domains and disciplines have also found significant heterogeneity across studies that can be assigned to sources of variance other than sampling error. However, our moderator analysis should be interpreted with caution because, in our study, the proportion of variance explained by the three meta-regression models did not exceed 17%. We found that associating extra activities and exercises with instructional videos help students to better understand the contents and materials and improve their performance in the ESL/EFL context. Providing students with different activities as pre-class assignments can increase the accountability for in-class activities (Abeysekera & Dawson, 2015). Our results contradicted the findings of the meta-analysis conducted by Hew and Lo (2018), who found that providing students with extra activities had no significant effect on their overall performance.

As for other variables in instructional design model, no moderator effect was found. Although most of the variables led to non-significant results, some additional information is presented to interpret these non-significant results. Generally, if students are new to this approach, they may not be accustomed to doing online activities and may feel confused and not know how to follow the rules, which in turn may affect their performance (Moran, 2014). Therefore, the first step in conducting flipped learning is to fully explain and clarify this approach to learners (Oraif, 2018).

Regarding "the reading/note-taking before class" variable, we found no significant result in the ESL/EFL context. This is in agreement with Hew and Lo (2018), who also found that providing students with pre-class readings or note-taking did not affect the performance of health professions students. We expected to find a significant moderating role of reading and note-taking in improving students' performance because in the field of language learning reading is an important skill and by providing different reading activities students can nearly improve all the required skills and subskills (e.g., vocabulary, grammar, writing, and spelling). Moreover, taking notes can also be used to increase students' concentration and enhance their comprehension of different subjects because they must understand the content to decide which part of the text or the video is important, to extract the key concepts, and to identify extraneous information (Bahrami & Nosratzadeh, 2017; Dunkel & Davy, 1989). Possible reasons for finding non-significant results for this variable may be improper selection and organization of reading resources and late delivery of pre-class materials to students. In order to encourage students to take notes, teachers should provide printed and organized note-taking sheets for their students to fill out before the beginning of class time (Cormier & Voisard, 2018). In future research, teachers should select and organize the reading materials based on the learning objectives of their flipped classrooms and they should use different strategies (e.g., accompanying reading materials with guiding or reflective questions, giving annotations or highlighting some sections or key points or parts of the text, asking them to write and submit a short summary after reading the text) to encourage students to engage in assigned reading activities ("Online Activities and Assessment for the Flipped classroom," n.d.). Moreover, enough time should be allocated for practicing pre-class materials (e.g., delivering pre-class activities few days before the in-class session). Also, reading assignments should not be too long and time-consuming to discourage students from completing classroom tasks (Khanova et al., 2015). In order to deliver pre-class materials to students, different low and high-stake mediums of delivery can be used such as blogs, Google tools, screencasts, mobile applications or different LMSs.

As for "applying quizzes as pre/in-class activities", no significant result was found, in spite of the findings of other studies regarding the positive effects of using quizzes on enhancing students' understanding and performance (Alten et al., 2019; Hew & Lo, 2018). Although quizzes are mostly considered as useful activities for improving students' performance due to identifying the students' problematic areas and misconceptions (Brame, 2013), the non-significant results can be attributed to the fact that quizzes can cause anxiety and stress, especially if given every session (Alavi et al., 2016; Johnson, 2013; Tune et al., 2013). Quizzes may impose pressure and stress on some students and make them feel afraid of failing the course (e.g., Strayer, 2007). This is in line with Broman and Johnels (2019), who concluded that quizzes reduced the general perceived value of flipped learning due to their lowest educational value from students' perspectives compared to lessons, online lectures and group work activities. Broman and Johnels (2019) noted that although quizzes motivated students to watch video lectures, in practice they were not well-received by students for several reasons. First, daily quizzes put students under pressure. Second, students felt frustrated because they could not receive direct feedback based on their performance in the quizzes and they did not know how many correct responses they had in quizzes, as they were meant to be reused in the future. Some students even argued that quizzes did not help them to improve their performance, but mainly served as an incentive for them to study every day on a regular basis (e.g., Broman & Johnels, 2019). Future studies should clarify the role of quizzes for students to reduce their level of stress. Teachers should also avoid using too long and difficult quizzes to reduce the risk of overwhelming students. Finally, teachers should design or modify quizzes to sufficiently cover the required learning goals and outcomes (Oraif, 2018). This is in agreement with Wagner (2018), who concluded that modifying quizzes improved students' engagement and performance.

The findings showed that "using LMSs" was not also considered as a significant moderator in the current study. Out of 69 studies, only 21 studies used LMSs for delivering pre-class components. LMSs have several advantages for effective education. LMSs are designed in a way that enable teachers to set goals for teamwork activities, use discussion forums, give and grade online assignment and provide instant feedback. Moreover, different useful blogs or interactive video lessons (e.g., Edpuzzle) can be embedded in LMSs, enabling teachers to track students' progress and to make sure that students do not skip the instructional videos while watching these videos (Aydın & Demirer, 2016). As argued by Touchton (2015), implementing flipped learning may be costly in the initial stages, but significant results can be achieved on a long-term basis. Moreover, some LMSs (e.g., Moodle) can incorporate the mastery learning into flipped classrooms (Johnson, 2013).

Considering the educational context model, the omnibus tests of all coefficients were not significant ($p = 0.09$). As for "the length or duration of intervention", we expected to find that longer interventions could lead to larger effects regarding students' achievements (Plonsky & Oswald, 2014). However, no significant effect was found in the current study. The findings are in agreement with those of previous meta-analyses (e.g., Cheng et al., 2018; Karagol & Esen, 2019).

Although the moderator variable called "educational level" ($p = 0.049$) seemed to moderate the findings, one of the issues of this category was the unequal distribution of data (HE: $N = 53$, Secondary level: $N = 9$, Private Institute: $N = 6$). One reason for this unequal distribution can be attributed to the omission of studies that did not meet the inclusion requirements. Other studies seem to be faced with the same problem (e.g., Alten et al., 2019). Most meta-analyses found no effect for "educational level" (Alten et al., 2019; Cheng et al., 2018; Karagol & Esen, 2019; Lag & Saele, 2019) while Zhu et al. (2019) concluded that different K-12 levels moderated the findings. Flipped learning is a new

pedagogical approach in the ESL/EFL context and more studies should be conducted on K-12 levels as well as higher education level, especially if one wants to examine the effectiveness of flipped learning on different age groups (Alten et al., 2019).

On the other hand, the variable "private institute" which consisted of studies with multi-age groups was statistically significant ($p = 0.019$). It should be noted that this educational level was only reported in our study while other meta-analyses only reported K-12 or higher educational levels. However, in a meta-analysis on the effect of mobile devices on students' performance, the highest effect on students' learning performance was attributed to primary levels ($p < 0.001$) while this type of instruction was not effective for mixed-age groups ($p = 0.615$) (e.g., Sung et al. 2016). In order to see if multi-age groups can negatively/positively affect students' achievements, more studies should be conducted in this regard in flipped learning environment.

Finally, the variables of the study quality model did not moderate the results. Our findings were in agreement with those of other meta-analyses in terms of "allocation type" (Alten et al., 2019), "EX/CO equivalence test" (e.g., Alten et al., 2019; Hew & Low, 2018; Lo et al., 2017), "study type" (Alten et al., 2019; Cheng et al., 2018), and "outcome measurement" (Alten et al., 2019). However, Lag and Saele (2019) indicated that "the allocation type" moderated the findings and studies using random-assignment had larger effect sizes, compared to the groups with unequal assignments. Zhu et al. (2019) also found that the type of publication significantly moderated the results.

Conclusion

The results of this quantitative meta-analysis, which was conducted on 69 studies, suggest that flipped learning yields a statistically significant improvement in students' achievements in the ESL/EFL context. Thus, it can be concluded that the flipped learning can be effective in improving students' achievements when it is used well. However, the satisfactory application of this approach turns out to involve the practices that require judgment from the instructors who are going to use the approach. Moreover, this approach can be more effective when students are required to practice new materials with more extra activities before entering the classroom. It somehow increases the readiness of students for face to face class time.

Implications for flipped learning research and application

The findings of our meta-analysis have some implications for future flipped learning research and application. First, our study confirms that flipped learning is more effective in improving the students' achievements and performances, compared to traditional lecture-based approach because, as claimed by Koh (2019), flipped learning supports student-learning approach through four pedagogical dimensions including personalization, higher-order thinking, self-direction, and collaboration.

Flipped classroom teachers and instructors, especially those in the ESL/EFL field, can use the finding of this study to design and implement the best possible flipped classroom. Flipping the classroom is an ongoing learning experience for both teachers and learners (Willis, 2017) and, as claimed by as Bergmann and Sams (2012), there is no single way to flip your classroom; there is no such thing as *the* flipped classroom. Rather, it is more about

a mindset which means to redirect the focus from teacher and to put more attention on the learning and learners.

Also, the findings of the moderator analysis suggest that when flipping their classrooms teachers should consider the following important factors: the age of students, students' knowledge of information technology, teachers' and students' familiarity with flipped learning, students' level of interest and engagement in doing extra activities outside the classroom, students' readiness for this type of learning, students' level of adaptability to this type of learning, students' accountability to self-study outside the classroom, availability of an appropriate platform, accessibility of students to internet and technological tools (such as computers, smart phones, tablets), and selecting or designing the learning materials that can be covered in sufficient time.

Although this study suggests that students' achievements have been improved by flipping the classrooms, research interest in different aspects of an ESL/EFL course tends to be relatively unbalanced. Most studies have focused on the writing skill while other skills (reading, speaking, and listening) and subskills (vocabulary, grammar, and pronunciation) have seemingly received less attention on the part of researchers. Therefore, more studies should be conducted on other skills (reading, speaking, and listening) and subskills (vocabulary, grammar, and pronunciation) to better understand the effect of this approach on students' achievements in the ESL/EFL context.

Moreover, in the field of L2 education, most studies have been conducted in higher education level and more studies should be conducted in K-12 to better understand the effect of this approach on different educational levels. In some countries like Iran where private language institutes, mostly consisting of classes with multi-age group, play a significant role in teaching English to EFL learners (Mohammadian Haghighi & Norton, 2016), teachers are required to be well-trained in how to design and implement flipped classrooms that are tailored to students' needs with different age groups. Accordingly, more in-depth studies are required to examine the effect of flipped learning on English outcomes of multi-age group classes.

Limitations

While conducting this study, we faced several limitations. First, we had to exclude several between-subject design studies due to the lack of sufficient statistical information for calculating the effect size. Second, as also argued by Alten et al. (2019), we were only capable of comparing flipped and lectured-based classrooms on the basis of the pre-post test results because almost all primary studies that were included in our meta-analysis failed to report the class procedure and activities of the control groups. Moreover, some other potential variables which seemed important for detecting the variation in effect sizes could not be included in this meta-analysis due to the lack of sufficient information from the majority of primary studies. These variables were as follows: (1) intensity of treatment (hours of treatment per week), (2) teacher experience about flipped learning, (3) instructor/teacher equivalence (similar or different instructors for flipped and control classes), (4) the degree of familiarity of students to flipped learning (whether it is their first time being taught by flipped approach!). In order to evaluate the effectiveness of the flipped learning in more detail, we recommend that future research fully report these types of information and also explain the activities and procedures of both research conditions in more detail. Moreover, besides gaining experience, teachers that are trained on how to implement flipped learning can assign students with more productive activities, provide them with more support and

facilitate their learning process. Third, some variables were not coded because one category was usually overrepresented (e.g., the proficiency of students: the majority of studies were conducted on students with an intermediate level of proficiency; almost all studies that met the inclusion criteria were conducted in EFL context).

Fourth, few studies have used mixed methods approach for examining the effectiveness of flipped learning in the ESL/EFL context. In order to achieve more comprehensive discussions, qualitative data should also be gathered along with the quantitative data to support and strengthen the discussed ideas.

Finally, and more importantly, despite the fact that most studies used teacher/researcher-developed tests (53 studies out of a total of 69 studies), very few ones have measured and reported the reliability estimate of the instruments used in their studies. As argued by Plonsky and Derrick (2016), reliable instrumentation is considered as one of the conditions of guaranteeing the internal validity of quantitative research studies. Therefore, we recommend future research to measure the reliability estimates of their instruments (e.g., by piloting the instruments on students other than main participants) and to use the recommendations proposed by Plonsky and Derrick (2016) in choosing, developing and interpreting their instruments.

Funding This study was not funded by any grants or contracts.

Dclarations

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

References

References that are marked with an asterisk are indicator of the studies included in the meta-analyses.

*Abaeian, H., & Samadi, L. (2016). The effect of flipped classroom on Iranian EFL learners' L2 reading comprehension: Focusing on different proficiency levels. *Journal of Applied Linguistics and Language Research, 3*(6), 295–304

*Abedi, P., Keshmirshekan, M. H., & Namaziandost, E. (2019). The comparative effect of flipped classroom instruction versus traditional instruction on Iranian intermediate EFL learners' English composition writing. *Journal of Applied Linguistics and Language Research, 6*(4), 43–56

*Abedi, P., Namaziandost, E., & Akbari, S. (2019). The impact of flipped classroom instruction on Iranian upper-intermediate EFL learners' writing skill. *English Literature and Language Review, 5*(9), 164–172

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

Adnan, M. (2017). Perceptions of senior-year ELT students for flipped classroom: A materials development course. *Computer Assisted Language Learning, 30*(3–4), 204–222

*Afrilyasanti, R., Cahyono, B., & Astuti, U. P. (2016). Effect of flipped classroom model on Indonesian EFL students' writing ability across and individual differences in learning. *International Journal of English Language and Linguistics Research, 4*(5), 65–81

*Ahmed, M. A. E. A. S. (2016). The effect of a flipping classroom on writing skill in English as a foreign language and students' attitude towards flipping. *US–China Foreign Language, 14*(2), 98–114

Aidinlou, N. A., Sharefii, S. S., & Kordabadi, F. S. (2017). Effects of flipped classroom approach on EFL Learners' reading performance with different cognitive style. *Journal of Applied Linguistics and Language Research, 4*(6), 98–104

Akçayır, G., & Akçayır, M. (2018). The flipped classroom: A review of its advantages and challenges. *Computers & Education, 126*, 334–345

*Alakawi, K. M. (2016). The impact of computer assisted language learning on Egyptian EFL learners' listening skill in the flipped classroom. *Imperial Journal of Interdisciplinary Research (IJIR), 2*(4), 1158–1161

*Alavi, S. M., Maralani, S., & K., & Salehi, Y. . (2016). A comparative study of the language learning environment of a flipped class versus a traditional class. *Journal of Applied Linguistics and Language Research, 3*(7), 207–218

*Al-Ghamdi, M., & Al-Bargi, A. (2017). Exploring the application of flipped classrooms on EFL Saudi students' speaking skill. *International Journal of Linguistics, 9*(4), 28–46

*Al-Hamdani, D., & Al-Breiki, M. (2018). The effect of flipped vocabulary learning on achievements and attitudes of ninth-grader in Oman. *International Journal of Research in Applied, Natural and Social Sciences, 6*(10), 35–44

*Al-Harbi, S. S., Alshumaimeri, U., & A. . (2016). The flipped classroom impact in grammar class on EFL Saudi secondary school students' performances and attitudes. *English Language Teaching, 9*(10), 60–80

*Alnuhayt, S. S. (2018). Investigating the use of the flipped classroom method in an EFL vocabulary course. *Journal of Language Teaching and Research, 9*(2), 236–242

Alten, D. C. D. V., Phielix, C., Janssen, J., & Kester, L. (2019). Effects of flipping the classroom on learning outcomes and satisfaction: A meta-analysis. *Educational Research Review, 28*, 1–18

*Alzaytuniya, S. H. (2016). *The effectiveness of using flipped classroom on tenth graders' grammar learning and motivation for English* (Master's thesis, the Islamic University of Gaza). Retrieved from http://hdl.handle.net/20.500.12358/20950

*Amiryousefi, M. (2017). The incorporation of flipped learning into conventional classes to enhance EFL learners' L2 speaking, L2 listening, and engagement. *Innovation in Language Learning and Teaching*. https://doi.org/10.1080/17501229.2017.1394307

Aydın, B., & Demirer, V. (2016). Flipping the drawbacks of flipped classroom: effective tools and recommendations. *Journal of Educational and Instructional Studies in the World–WJEIS, 6*(1), 33–40

Bahrami, F., & Nosratzadeh, H. (2017). The effectiveness of note-taking on reading comprehension of Iranian EFL learners. *International Journal of Applied Linguistics & English Literature, 6*(7), 308–317

Bergmann, J. (2017). *Solving the homework problem by flipping the learning*. ASCD.

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: Maximizing face time. *T+D, 68*(2), 28–31

Betihavas, V., Bridgman, H., Kornhaber, R., & Cross, M. (2016). The evidence for 'flipping out': A systematic review of the flipped classroom in nursing education. *Nurse Education Today, 38*, 15–21

Bishop, J. L., & Verleger, M. A. (2013). *The flipped classroom: A survey of the research*. Paper presented at the 120th ASEE national conference proceedings, Atlanta, GA.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470743386

*Boyraz, S., & Ocak, G. (2017). Implementation of flipped education into Turkish EFL teaching context. *Journal of Language and Linguistic Studies, 13*(2), 426–439

Bradford, M. (2005). Motivating students through project-based service learning. *THE Journal, 32*(6), 29–31

Brame, C. (2013). *Flipping the classroom*. Retrieved from The Vanderbilt University Center for Teaching website: http://cft.vanderbilt.edu/teaching-guides/teaching-activities/flipping-the-classroom/

Broman, K., & Johnels, D. (2019). Flipping the class university chemistry students' experiences from a new teaching and learning approach. *Chemistry Teacher International, 1*(1), 20180004

Brown, S. A., Upchurch, S. L., & Acton, G. J. (2003). A framework for developing a coding scheme for meta-analysis. *Western Journal of Nursing research, 25*(2), 205–222

*Bulut, C. (2018). *Impact of flipped classroom model on EFL learners' grammar achievement: Not only inversion, but also integration* (Master's thesis YEDİTEPE University)

Burke, A., & Fedorek, B. (2017). Does "flipping" promote engagement? A comparison of a traditional, online and flipped class. *Active Learning in Higher Education, 18*(1), 11–24

Cakıroglu, U., & Ozturk, M. (2017). Flipped classroom with problem based activities: Exploring self-regulated learning in a programming language course. *Educational Technology & Society, 20*(1), 337–349

*Cavdar, O. E. (2018). *Integrating flipped classroom approach into traditional English class.* (MA thesis, Karadeniz Technical University).

Chaubey, A., & Bhattacharya, B. (2015). Learning management system in higher education. *International Journal of Science Technology & Engineering (IJSTE), 2*(3), 158–163

*Chavangklang, T., & Suppasetseree, S. (2018). Enhancing Thai EFL university students' reading comprehension through a flipped cooperative classroom. *PEOPLE: International Journal of Social Sciences, 4*(3), 238–261

Chen, F., Lui, A. M., & Martinelli, S. M. (2017). A systematic review of the effectiveness of flipped classrooms in medical education. *Medical Education, 51*, 585–597

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

Cheng, L., Ritzhaupt, A. D., & Antonenko, P. (2018–2019). Effects of the flipped classroom instructional strategy on students' learning outcomes: A meta-analysis. *Educational Technology Research & Development, 67*(4), 793–824

Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist, 49*(4), 219–243

Chokshi, B. D., Schumacher, H. K., Reese, K., Bhansali, P., Kern, J. R., Simmens, S. J., Blatt, B., & Greenberg, L. W. (2017). A "resident-as-teacher" curriculum using a flipped classroom approach: Can a model designed for efficiency also be effective? *Academic Medicine, 92*(4), 511–514

Chui, W.-L. (2016). *Effects of flipped and traditional learning on college students' English learning outcomes and motivation* (Master's thesis, Sun Yat-se University).

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Routledge Academic.

Connor, K. A., Newman, D., & Deyoe, M. M. (2014). Flipping a classroom: A continual process of refinement. *Proceedings from 121st ASEE: American Society for Engineering Education Annual Conference & Exposition*. IN: Indiana.

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

Cormier, C., & Voisard, B. (2018). Flipped classroom in organic chemistry has significant effect on students' grades. *Frontiers in ICT, 4*(30), 1–15

Cronhjort, M., Filipsson, L., & Weurlander, M. (2017). Improved engagement and learning in flipped classroom calculus. *Teaching Mathematics and its Application, 37*, 113–121

Diemer, T., Fernandez, E., & Streepey, J. (2013). Student perceptions of classroom engagement and learning using iPads. *Journal of Teaching and Learning with Technology, 1*(2), 13–25

Dunkel, P. A., & Davy, S. (1989). The heuristic of lecture note-taking: Perceptions of American and international students regarding the value and practice of note-taking. *English for Specific Purposes, 8*(1), 33–50

*Ekmekci, E. (2017). The flipped writing classroom in Turkish EFL context: A comparative study on a new model. *Turkish Online Journal of Distance Education, 18*(2), 151–167

El-Bassuony, J. M. (2016). The effectiveness of flipped learning in developing English grammatical performance of underachieving language learners at the secondary stage. *International Journal of English Language Teaching, 4*(8), 76–102

Elmaadaway, M. A. (2017). The effects of a flipped classroom approach on class engagement and skill performance in a blackboard course. *British Journal of Educational Technology, 49*(3), 479–491

El-Sakka, S. M. F. (2016). *A flipped learning-based model to develop EFL freshman university students' listening comprehension.* Presented at the 4th Scientific international Conference of Curriculum and Instruction Association, Ain Shams University, Egypt.

Farah, M. (2014). *The impact of using a flipped classroom instruction on the writing performance of twelfth grade female Emirati students in the applied technology high School (ATHS)* (Master's thesis, the British University, Dubai). Retrieved from https://bspace.buid.ac.ae/handle/1234/676.

Farhady, H., Jafarpur, A., & Birjandi, P. (2010). *Testing language skills from theory to Practice*. Iran: SAMT.

*Farrah, M., & Qawasmeh, A. (2018). English students' attitudes towards using flipped classrooms in language learning at Hebron University. *Research in English Language Pedagogy. RELP, 6*(2), 275–294

Ferreri, S. P., & O'Connor, S. K. (2013). Redesign of a large lecture course into a small group learning course. *American Journal of Pharmaceutical Education, 77*(1), 122–142

Gillette, C., Rudolph, M., Kimble, C., Rockich-Winston, N., Smith, L., & Broedel-Zaugg, K. (2018). A systematic review and meta-analysis of student pharmacist outcomes comparing flipped classroom and lecture. *American Journal of Pharmaceutical Education, 82*(5), 433–440

Gurluyer, M. (2019). *Examining EFL students' achievements and perceptions in terms of writing skill in flipped classroom environment* (Master's thesis, Kafkas University).

*Haghighi, H., Jafarigohar, M., Khoshsima, H., & Vahdany, F. (2018). Impact of flipped classroom on EFL learners' appropriate use of refusal: achievement, participation, perception. *Computer Assisted Language Learning, 32*(3), 261–293

Hall, A. A., & DuFrene, D. D. (2016). Best practices for launching a flipped classroom. *Business and Professional Communication Quarterly, 79*(2), 234–242

Hamdan, N., McKnight, P., McKnight, K. & Arfstrom, K. (2013). *A Review of Flipped Learning Flipped Learning Network*. Retrieved September 28, 2017, from http://www.flippedlearning.org/cms/lib07/VA01923112/Centricity/Domain/4 1/LitReview_FlippedLearning.pdf

Han, E., & Klein, K. C. (2019). Pre-class learning methods for flipped classrooms. *American Journal of Pharmaceutical Education*. https://doi.org/10.5688/ajpe6922

*Hashemifardnia, A., Namaziandost, E., & Shafiee, S. (2018). The effect of implementing flipped classrooms on Iranian junior high school students' reading comprehension. *Theory and Practice in Language Studies, 8*(6), 665–673

Hedges, L., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Academic Press.

Hew, K. F., & Lo, C. K. (2018). Flipped classroom improves student learning in health professions education: A meta-analysis. *BMC Medical Education, 18*(1), 1–12

Hibbard, L., Sung, S., & Wells, B. (2015). Examining the effectiveness of a semi-self paced flipped learning format in a college general chemistry sequence. *Journal of Chemical Education, 93*(1), 24–30

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives, 2*(3), 172–177

*Huang, Y. N., & Hong, Z. R. (2015). The effects of a flipped English classroom intervention on students' information and communication technology and English reading comprehension. *Educational Technology Research and Development, 64*(2), 175–193

*Hung, H. T. (2015). Flipping the classroom for English language learners to foster active learning. *Computer Assisted Language Learning, 28*(1), 81–96

*Hung, H. T. (2017). Design-based research: Redesign of an English language course using a flipped classroom approach. *TESOL Quarterly, 51*(1), 180–192

*Ishikawa, Y., Akahane-Yamada, R., Smith, C., Kondo, M., Tsubota, Y., Dantsuji, M. (2015). An EFL flipped learning course design: utilizing students' mobile online devices. In F. Helm, L. Bradley, M. Guarda, & S. Thouësny (Eds), *Critical CALL – Proceedings of the 2015 EUROCALL Conference, Padova, Italy* (pp. 261-267). Dublin: Research-publishing.net. http://dx.doi.org/https://doi.org/10.14705/rpnet.2015.000343

*Iyitogu, O., & Erisen, Y. (2017). Delving into flipping EFL classroom: A mixed method study. *European Journal of English Language Teaching, 3*(1), 120–152

Jaster, R. W. (2017). Student and instructor perceptions of a flipped college algebra classroom. *International Journal of Teaching and Learning in Higher Education, 29*(1), 1–16

Johnson, G. B. (2013). *Student perceptions of the flipped classroom* (Master's thesis, the University of British Columbia). Retrieved from https://open.library.ubc.ca/cIRcle/collections/ubctheses/24/items/1.0073641

Jungić, J., Kaur, H., Mulholland, J., & Xin, C. (2014). On flipping the classroom in large first year calculus courses. *International Journal of Mathematical Education in Science and Technology, 46*(4), 508–520

Kang, N. (2015). The comparison between regular and flipped classrooms for EFL Korean adult learners. *Multimedia-Assisted Language Learning, 18*(3), 41–72

*Kang, S., & Shin, I. (2016). The effect of flipped learning in Korea : Meta-Analysis. Retrieved from https://pdfs.semanticscholar.org/6d9d/49af1e74a7b6cf6c6f27c24274e6f645d5d8.pdf?_ga=2.90897299.885751784.1577946829-682599380.1574959810

Karagol, I., & Esen, E. (2019). The effect of flipped learning approach on academic achievement: A meta-analysis Study. *H. U. Journal of Education, 34*(3), 708–727

*Karakurt, L. (2018). *Flipped and blended grammar instruction for B1 level EFL classes at tertiary education* (Master's thesis, Hacettepe University). Retrieved from http://www.openaccess.hacettepe.edu.tr:8080/xmlui/bitstream/handle/11655/5305/10196488.pdf?sequence=1&isAllowed=y

*Karimi, M., & Hamzavi, R. (2017). The effect of flipped model of instruction on EFL learners' reading comprehension: Learners' attitudes in focus. *Advances in Language and Literary Studies, 8*(1), 95–103

Khanova, J., Roth, M. T., Rodgers, J. E., & McLaughlin, J. E. (2015). Student experiences across multiple flipped courses in a single curriculum. *Medical Education, 49*(10), 1038–1048

Kim, M. K., Kim, S. M., Khera, O., & Getman, J. (2014). The experience of three flipped classrooms in an urban university: An exploration of design principles. *Internet and Higher Education, 22*, 37–50

Knapp, G., & Hartung, J. (2003). Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine, 22*(17), 2693–2710

Koh, J. H. L. (2019). Four pedagogical dimensions for understanding flipped classroom practices in higher education: A systematic review. *Educational Sciences: Theory and Practice, 19*(4), 14–33. https://doi.org/10.12738/estp.2019.4.002

*Koroglu, Z. C., & Cakır, A. (2017). Implementation of flipped instruction in language classrooms: An alternative way to develop speaking skills of pre-service English language teachers. *International Journal of Education and Development using Information and Communication Technology (IJEDICT), 13*(2), 42–55

Kurt, G. (2017). Implementing the flipped classroom in teacher education: Evidence from Turkey. *Journal of Educational Technology & Society, 20*(1), 211–221

Lag, T., & Saele, R. G. (2019). Does the flipped classroom improve student learning and satisfaction? A systematic review and meta-Analysis. *AERA Open, 5*(3), 1–17

Lee, J., Park, T., & Davis, R. O. (2018). What affects learner engagement in flipped learning and what predicts its outcomes? *British Journal of Educational Technology, 1*, 1–18

Leis, A. (2016). Flipped learning and EFL proficiency: An empirical study. *TELES Journal, 36*, 77–90

*Leis, A., Cooke, S., & Tohei, A. (2016). The effects of flipped classrooms on English composition writing in an EFL environment. *International Journal of Computer Assisted Language Learning and Teaching, 5*(4), 37–51

*Li, S. (2015). *Enhancing EFL speaking via flipped classroom model and constructive role plays* (Master's thesis, Suranaree University of Technology).

Lieu, R., Wong, A., Asefirad, A., & Shaffer, J. F. (2017). Improving exam performance in introductory biology through the use of pre-class reading guides. *CBE Life Science Education, 16*(3), 1–10

*Lim, J.-Y. (2015). The effects of flipped learning on middle level of college students' essay performance. *STEM Journal, 16*(4), 197–214

*Lin, C.-J., & Hwang, G.-J. (2018). A learning analytics approach to investigating factors affecting EFL students' oral performance in a flipped classroom. *Educational Technology & Society, 21*(2), 205–219

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

Lo, C. K., & Hew, K. F. (2017). A critical review of flipped classroom challenges in K 12 education: Possible solutions and recommendations for future research. *Research and Practice in Technology Enhanced Learning, 12*(4), 1–22

Lo, C. K., Hew, K. F., & Chen, G. (2017). Toward a set of design principles for mathematics flipped classrooms: A synthesis of research in mathematics education. *Educational Research Review, 22*, 50–73

McCallum, S., Schultz, J., Sellke, K., & Spartz, J. (2015). An examination of the flipped classroom approach on college student academic involvement. *International Journal of Teaching and Learning in Higher Education, 27*(1), 42–55

Mehrens, W. A., & Lehmann, I. J. (1978). *Measurement and evaluation in education and psychology (2nd edn.)*. Library of Congress Cataloging in Publication Data.

*Mirshekaran, R. (2018). *The impact of implementing flipped classrooms of listening comprehension of Iranian EFL learners across two proficiency levels.* Conference: The First National Conference on Applied Linguistics and Material Developments At: Bojnord. Retrieved from https://www.researchgate.net/publication/328723231_The_Impact_of_Implementing_Flipped_Classrooms_on_Listening_Comprehension_of_Iranian_EFL_Learners_across_Two_Proficiency_Levels

Mohammadian Haghighi, F., & Norton, B. (2016). The role of English language institutes in Iran. *TESOL Quarterly* . https://doi.org/10.1002/tesq.338 Advance online publication.

Moran, C. M. (2014). *Changing paradigms: A mixed methods study of flipping the English language arts classroom* (Doctoral dissertation, North Carolina State University).

Munir, M. T., Baroutiana, S., Young, B. R., & Carter, S. (2018). Flipped classroom with cooperative learning as a cornerstone. *Education for Chemical Engineers, 23*, 25–33

*Nguyen, N. Q., Lee, K. W., Nguyen, D. N. P., & Naidu, S. (2019). An investigation into using flipped classroom model in an academic writing class in Vietnam. *International Journal of Computer-Assisted Language Learning and Teaching, 9*(1), 32–57

*Nugroho, N., & Insana, D. R. M. (2018). Flipped classroom, teaching solution for writing class (case study of Indraprasta EGRI university students). *Journal of English Language Teaching, 2*(2), 141–148

*Nuon, P., & Champakaew, W. (2017). Effects of ICT-aided flipped classroom on grammar achievement. *International Conference on Language, Education, Humanities and Innovation 5th & 6th August.*

O'Connor, S. K., & Ferreri, S. P. (2013). Redesign of a large lecture course into a small group learning course. *American Journal of Pharmaceutical Education*. https://doi.org/10.5688/ajpe77113

Oh, E. (2017). The effect of peer teaching via flipped vocabulary learning on class engagement and learning achievements. *Multimedia-Assisted Language Learning, 20*(3), 105–127

148. Online activities and assessment for the flipped classroom (n.d). Retrieved fromhttps://uwaterloo.ca/centre-for-teaching-excellence/teaching-resources/teaching-tips/lecturing-and-presenting/delivery/online-activities-and-assessment-flipped-classroom

*Oraif, I. M. K. (2018). *An investigation into the impact of the flipped classroom on intrinsic motivation (IM) and learning outcomes on an EFL writing course at a university in Saudi Arabia based on self-determination theory (SDT)* (Doctoral dissertation, The University of Leicester). Retrieved from https://lra.le.ac.uk/handle/2381/42165

Orwin, R. G. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics, 8*(2), 157–159

*Ozkal, C. (2019). *Flipped vocabulary learning among Turkish learners of English as a foreign language: A sequential explanatory mixed methods Study.* (Master's thesis, Middle East Technical University). Retrieved from https://etd.lib.metu.edu.tr/upload

Persky, A. M., & Hogg, A. (2017). Influence of reading material characteristics on study time for preclass quizzes in a flipped classroom. *American Journal of Pharmaceutical Education*. https://doi.org/10.5688/ajpe816103

Persky, A. M., & Pollack, G. M. (2011). A modified team-based learning physiology course. *American Journal of Pharmaceutical Education*. https://doi.org/10.5688/ajpe7510204

Plonsky, L., & Derrick, D. J. (2016). A meta-analysis of reliability coefficients in second language research. *Modern Language Journal, 100*(2), 538–553

Plonsky, L., & Oswald, F. L. (2014). How big Is "big"? Interpreting effect sizes in L2 research. *Language Learning, 64*(4), 878–912. https://doi.org/10.1111/lang.1207

Polanin, J. R., & Pigott, T. D. (2015). The use of meta-analytic statistical significance testing. *Research Synthesis Methods, 6*(1), 63–73

Polanin, J. R., Tanner-Smith, E. E., & Hennessy, E. A. (2016). Estimating the difference between published and unpublished effect sizes: A meta-review. *Review of Educational Research, 86*(1), 207–236

*Qader, R. O., & Arslan, F. Y. (2019). The effect of flipped classroom instruction in writing: A case study with Iraqi EFL learners. *Teaching English with Technology, 19*(1), 36–55

*Quyen, T. T. T., & Van Loi, N. (2018). Flipped model for improving students' English speaking performance. *Can Tho University Journal of Science, 54*(2), 90–97

Rivera, V. M. (2016). *Flipped classrooms: Advantages and disadvantages from the perspective of a practicing art teacher* (Master's thesis, State University of New York).

Saglam, D., & Arslan, A. (2016). The effect of flipped classroom on the academic achievement and attitude of higher education students. *World Journal of Education, 8*(4), 170–176

Sams, A., & Bergmann, J. (2013). Flip your students' learning. *Educational Leadership, 7*, 16–20

Santikarn, B., & Wichadee, S. (2018). Flipping the classroom for English language learners: A study of learning performance and perceptions. *International Journal of Emerging Technologies in Learning, 13*(9), 123–135

*Santosa, M. H. (2017). Learning approaches of Indonesian EFL Gen Z students in a flipped learning context. *Journal on English as a Foreign Language, 7*(2), 183–208

Saunders, A., Green, R., & Cross, M. (2017). Making the most of person-centered education by integrating flipped and simulated teaching: An exploratory study. *Nurse Education in Practice, 27*, 71–77

Sezer, B., Asli-Sezer, T., & Lampotang, S. (2017). How to flip a classroom in four simple steps. *IACB, ICE & ICTE Conference Proceedings*, New York July 30–August 3.

Stone, D. L., & Rosopa, P. J. (2017). The advantages and limitations of using meta-analysis in human resource management research. *Human Resource Management Review, 27*(1), 1–7

Stone-Romero, E. F., Alliger, G., & Aguinis, H. (1994). Type II error problems in the use of moderated multiple regression for the detection of moderating effects of dichotomous variables. *Journal of Management, 20*(1), 167–178

Strayer, J. F. (2007). *The effects of the classroom flip on the learning environment: A comparison of learning activity in a traditional classroom and a flip classroom that used an intelligent tutoring system* (Doctoral dissertation, the Ohio State University). Retrieved from http://faculty.washington.edu/rvanderp/DLData/FlippingClassDis.pdf

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

Swart, A. J. (2016). The effective use of a learning management system still promotes student engagement. *In 2016 IEEE Global Engineering Education Conference (EDUCON).*

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

Tan, E., Brainard, A., & Larkin, G. L. (2015). Acceptability of the flipped classroom approach for in-house teaching in emergency medicine. *EMA - Emerg Med Australas, 27*(5), 453–459

Tang, F., Chen, C., Zhu, Y., Zuo, C., Zhong, Y., Wang, N., Zhou, L., Zou, Y., & Liang, D. (2017). Comparison between flipped classroom and lecture-based classroom in ophthalmology clerkship. *Medical education online., 22*(1), 1395679

*Teng, M. F. (2017). Flipping the classroom and tertiary level EFL students' academic performance and satisfaction. *The Journal of Asia TEFL, 14*(4), 605–620

Touchton, M. (2015). Flipping the classroom and student performance in advanced statistics: Evidence from a quasi-experiment. *Journal of Political Science Education, 11*(1), 28–44

*Tuna, G. (2017). *An action study on college students' EFL writing skill development through flipped classroom environment.* (MA thesis, Middle East Technical University).

Tune, J. D., Sturek, M., & Basile, D. P. (2013). Flipped classroom model improves graduate student performance in cardiovascular, respiratory and renal physiology. *Advances in Physiology Education, 37*(4), 316–320

Unakorn, P., & Klongkratoke, U. (2015). *Effectiveness of flipped classroom to mathematics learning of grade 11 students.* A paper presented at the 21st & 22nd International Conference on Language, Education, and Humanities & Innovation. Retrieved from https://icsai.org/procarch/1iclehi/1iclehi-44.pdf

Wagner, B. (2018). *The effects of a flipped classroom on student comprehension and perception among second language learners in a bachelor of social work course.* (Doctoral dissertation). Retrieved from https://scholarcommons.sc.edu/etd/4620.

*Webb, M., & Doman, E. (2016). Does the flipped classroom lead to increased gains on learning outcomes in ESL/EFL contexts? *CATESOL Journal, 28*(1), 39–67

Willis, L. D. (2017). Implications for pedagogy: Flipping the classroom to engage pre-service teachers. In C. Reidsema, L. Kavanagh, R. Hadgraft, & N. Smith (Eds.), *The Flipped Classroom: Practice and practices in higher education.* (pp. 273–287). Springer.

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

*Zhu, G., Thompson, C., Suarez, M., & Peng, Z. (2019). *A meta-analysis on the effect of flipped instruction on K-12 students' academic achievement*. Paper presented at the American Educational Research Association Annual Meeting. Toronto, Canada, April 5–9.

*Zou, D., & Xie, H. (2018–2019). Flipping an English writing class with technology enhanced just-in-time teaching and peer instruction. *Interactive Learning Environments, 27*(8), 1127–1142

Mojdeh Shahnama is a Ph.D. candidate in Teaching English as a Second or Foreign Language at Ferdowsi University of Mashhad, Iran. Her research interests are mainly flipped classroom, meta-analysis and ecological research in applied linguistics.

Behzad Ghonsooly is a Full Professor of psycholinguistics and Teaching English as a Second or Foreign Language at Ferdowsi University of Mashhad. He has 20 years of experience in introspection research. He has published a book on think-aloud in reading as a second language plus a number of virtual course books for Ferdowsi University of Mashhad. Moreover, He has published a couple of articles in translation studies in JCR and Scopus journals. The total number of the articles that he has published so far in high-brow journals goes over 120 articles.

Majid Elahi Shirvan is an Assistant Professor of Teaching English as a Second or Foreign Language at University of Bojnord. His research interests are mainly meta-analysis, materials development and ecological research in applied linguistics.