Effectiveness of flipped classrooms in Chinese baccalaureate nursing education: A meta-analysis of randomized controlled trials

Rujun Hu, Huiming Gao, Yansheng Ye, Zhihong Ni, Ning Jiang, Xiaolian Jiang*

Author information:
Rujun Hu, PhD Candidate
West China Hospital/West China School of Nursing, Sichuan University, Chengdu 610041, Sichuan, China.
Huiming Gao, M.S.
School of Nursing, Zunyi Medical College, Zunyi 563000, Guizhou, China.
Yansheng Ye, PhD Candidate
West China Hospital/West China School of Nursing, Sichuan University, Chengdu 610041, Sichuan, China.
Zhihong Ni, PhD Candidate
West China Hospital/West China School of Nursing, Sichuan University, Chengdu 610041, Sichuan, China.
Ning Jiang, PhD Candidate
West China Hospital/West China School of Nursing, Sichuan University, Chengdu 610041, Sichuan, China.
Xiaolian Jiang, PhD, Professor
West China Hospital/West China School of Nursing, Sichuan University, Chengdu 610041, Sichuan, China.

Correspondence:
Xiaolian Jiang, PhD, Professor, West China Hospital/West China School of Nursing, Sichuan University, Chengdu 610041, Sichuan, China.
Tel: +86 028 85422070; fax: +86 028 85422070.
Email address: jiangxiaolianhl@163.com; jiang_xiaolian@126.com
Abstract

Background: In recent years, the flipped classroom approach has been broadly applied to nursing courses in China. However, a systematic and quantitative assessment of the outcomes of this approach has not been conducted.

Objective: The purpose of the meta-analysis is to evaluate the effectiveness of the flipped classroom pedagogy in Chinese baccalaureate nursing education.

Design: Meta-analysis of randomized controlled studies.

Data source: All randomized controlled trials relevant to the use of flipped classrooms in Chinese nursing education were retrieved from the following databases from their date of inception through September 23, 2017: PubMed, EMBASE, the Cochrane Central Register of Controlled Trials, CINAHL, the China National Knowledge Infrastructure, the Wanfang Database, and the Chinese Scientific Journals Database. Search terms including “flipp*”, “inverted”, “classroom”, and “nurs*” were used to identify potential studies. We also manually searched the reference lists of the retrieved articles to identify potentially relevant studies.

Review Methods: Two reviewers independently assessed the eligibility of each study and extracted the data. The Cochrane risk-of-bias tool was used to evaluate the quality of the studies. RevMan (Version 5.3) was used to analyze the data. Theoretical knowledge scores and skill scores (continuous data) were synthesized using the standardized mean difference (SMD) and 95% confidence interval (CI). The statistical heterogeneity of the included studies was analyzed by calculating the I² statistic and applying a chi-square test. Publication bias was assessed by funnel plots. The quality
of the combined results was evaluated using the Grading of Recommendations Assessment, Development and Evaluation system.

**Results:** Eleven randomized controlled trials published between 2015 and 2017 were selected. All the included studies had a moderate possibility of bias due to low methodological quality. The meta-analysis indicated that the theoretical knowledge scores and skill scores were significantly higher in the flipped classroom group than in the traditional lectures group (SMD=1.06, 95% CI: 0.70-1.41, P<0.001, and SMD=1.40, 95% CI: 0.46-2.34, P<0.001). There was no significant publication bias indicated in the primary analysis. Sensitivity analysis showed that the results of our meta-analysis were reliable. The evidence grades of the results regarding the theoretical knowledge and skill scores were low and very low, respectively.

**Conclusion:** Flipped classroom pedagogy is more effective than traditional lectures at improving students’ theoretical knowledge and skill scores. Given the limitations of the included studies, more robust randomized controlled trials are warranted in a variety of educational settings to confirm our findings.

**Keywords:** flipped classroom, nursing, meta-analysis, review

1. Introduction

The flipped classroom is a term that was coined by Bergmann and Sams in 2007, and it is defined as a teaching method in which “that which is traditionally done in class is now done at home and that which is traditionally done as homework is now
completed in class” (Bergmann and Sams, 2012, Davies et al., 2013). Specifically, in the flipped classroom model, the content that is traditionally presented in the classroom setting is instead assigned as homework to be accomplished before class, while during class time, students engage in active learning, such as case studies, learning laboratories, games, and simulation-based learning under the guidance of the teacher (Herreid and Schiller, 2013). Flipped classroom pedagogy aims to help students transition from passive learning to active, self-directed learning and to improve their analytical, integration, and critical thinking skills to better address the complexity of contemporary health care (Hamer, 2000). Because of the characteristics of this approach, flipped classroom pedagogy has attracted the attention of educators and has been widely implemented and studied in many disciplines, including nursing (Galway et al., 2014, Gilboy et al., 2015, Hsu et al., 2016, Lichvar et al., 2016, Liebert et al., 2016, Mason et al., 2013, Mattis, 2015, McLaughlin et al., 2013, Mortensen and Nicholson, 2015, Park and Howell, 2015).

However, the implementation of the flipped classroom teaching model in higher education for nursing remains controversial (Geist et al., 2015, Harrington et al., 2015, Schwartz, 2014, Simpson and Richards, 2015). Some studies have found that the flipped classroom model in nursing education presents certain advantages over traditional lecture-based learning in terms of improving students’ examination scores, grades and satisfaction (Critz and Knight, 2013, Geist et al., 2015, Missildine et al., 2013, Yacout and Shosha, 2016). Conversely, some researchers have not found the flipped classroom model to be superior to traditional models with respect to student
examination scores and satisfaction (Geist et al., 2015, Harmon and Hills, 2015, Harrington et al., 2015, Simpson and Richards, 2015, Tian et al., 2015).

In China, the flipped classroom model has been used in higher nursing education settings since 2014, and the annual number of published Chinese studies focusing on the application of flipped classrooms has increased exponentially from 6 items in 2014 to approximately 160 items in 2016. However, a systematic and quantitative assessment of the outcomes of the flipped classroom teaching model has not been conducted to date. Considering that different education systems and cultural differences among different countries may influence the effectiveness of the flipped classroom method (Frambach et al., 2012, Kim, 2008, Sharma et al., 2014), we performed a meta-analysis with the target population that was limited to Chinese baccalaureate nursing students to assess the effects of the flipped classroom model. The goals were to provide a scientific basis for evaluating the necessity and feasibility of flipped classroom learning in higher nursing education and to disseminate the findings more widely to international nursing educators and researchers. In addition, this research will provide a valuable reference for teaching-learning innovation and for researchers in different education systems and cultural backgrounds who want to assess the effectiveness of the flipped classroom model.

2. Methods

The data used in this systematic review and meta-analysis were from previously published studies; therefore, ethical approval and student consent were not necessary.

2.1 Inclusion and exclusion criteria
Two independent raters (RJH and YSY) screened and selected all the studies. The inclusion criteria were as follows: (1) types of participants: baccalaureate nursing students in China; (2) types of interventions: using flipped classrooms as the educational approach; (3) types of controls: traditional face-to-face lecture without supplementary teaching methods that could affect the results; (4) types of outcomes: theoretical knowledge scores (scores acquired by theoretical knowledge examination) and/or skill scores (scores acquired by operation skill examination); (5) types of study designs: randomized controlled trials (RCTs); and (6) language: English and Chinese. Studies were excluded if they (1) had incomplete data or (2) were conference abstracts. Any disagreements between the two raters (RJH and YSY) were resolved by discussion with a third reviewer (XLJ).

2.2 Search strategy

This study was conducted by following the guidelines in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement (PRISMA) (Moher et al., 2009). To identify the relevant studies, we searched for publications in the following databases from their date of inception through September 23, 2017: PubMed, EMBASE, the Cochrane Central Register of Controlled Trials (CENTRAL), CINAHL, the China National Knowledge Infrastructure (CNKI), the Wanfang Database, and the Chinese Scientific Journals Database (VIP). Medical subject headings (MeSH) or key words including “flipp*”, “inverted”, “classroom”, and “nurs*” were used to identify potential studies (See Supplementary File 1). We also
manually searched the reference lists of the retrieved articles to identify potentially relevant studies.

All the identified studies, including the titles and abstracts, were downloaded into Endnote X8 for review. We removed all duplicate studies. The references of the potential papers were examined to identify any additional papers that met the inclusion criteria and may have been missed by the search strategy. The titles and/or abstracts of the retrieved studies from the search strategy and from additional sources were screened independently by two raters (RJH and YSY) to identify studies that potentially met the inclusion criteria outlined above. We searched and downloaded the full papers of potentially eligible studies, which were assessed independently on the basis of inclusion criteria by the same two raters. Disagreements between the two raters were resolved by discussion with a third reviewer (XLJ).

2.3 Data extraction and quality assessment

Two raters (NJ and ZHN) used a pre-designed data collection form (Microsoft Office Excel 2013) to extract all the data independently. The following information was extracted: first author, publication year, study design, subjects, course type, participant characteristics, sample size (intervention group/control group), educational approach in the intervention group, educational approach in the control group, outcomes, and the duration of study.

Two reviewers (NJ and ZHN) independently evaluated the methodological quality and risk of bias of each selected study, and disagreements were resolved by
discussion with a third reviewer (XLJ). We used the Cochrane risk-of-bias tool to assess the following items: random sequence generation (selection bias), allocation concealment (selection), blinding of participants and personnel (performance bias), blinding of outcome assessment (detection bias), incomplete outcome data (attrition bias), selective reporting (reporting bias), and other bias. Each domain was categorized as exerting a low risk of bias (unlikely to seriously alter the results), a high risk of bias (seriously weakens confidence in the results), or an unclear risk of bias. The Cochrane Collaboration risk of bias tool is available online at http://handbook.cochrane.org/.

2.4 Statistical methods

Two reviewers (RJH and HMG) separately input the data, and we used Review Manager Software 5.3, which was developed by the Cochrane Collaboration, to conduct the heterogeneity tests and meta-analysis. Before pooling the study results, the statistical heterogeneity of the included studies was analyzed by calculating the $I^2$ statistic and applying the chi-square test (Higgins et al., 2003). If $I^2<50\%$ and $P>0.10$, we considered the heterogeneity to be low, and a fixed effects model was used to pool the data. Otherwise, we used a random effects model to summarize the results. If heterogeneity was present, a sensitivity analysis was performed by excluding the study with the largest sample size and then recalculating the pooled estimates for the remaining studies to assess whether this exclusion significantly altered the meta-analysis results. Continuous data were synthesized using the standardized mean
difference (SMD) and 95% confidence interval (CI). A two-sided P<0.05 was considered to indicate a significant difference in the overall effect. We used funnel plots to assess the possible publication bias, evidence of asymmetry, and other small study effects (Hguyatt et al., 2011). In addition, we ranked the evidence quality using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) system (https://gradepro.org/).

3. Results

3.1 Literature search results

A total of 840 records from these electronic databases were identified. After the titles, the abstracts of these studies were reviewed, and 789 studies were excluded either because they were duplicated (464 studies) or they did not match the inclusion criteria (325 studies), such as not addressing undergraduate baccalaureate nursing students in China (186 studies) and non-RCT (139 studies). Then, we further reviewed the full texts of the remaining 51 studies, 40 of which were excluded because they did not address undergraduate baccalaureate nursing students in China (25 studies), they were non-RCT (8 studies), or they had unrelated outcomes (4 studies) or incomplete data (3 studies). Finally, 11 studies were included in the qualitative synthesis and meta-analysis (Deng, 2016, Ji et al., 2016, Li, 2016, Li et al., 2017, Liu, 2016, Tao et al., 2016, Tian et al., 2015, Yang, 2016, Zhang et al., 2017, Zhao et al., 2016, Zhong et al., 2016). The literature screening process and results are depicted in Fig. 1.
3.2 General study characteristics

The 11 included RCTs were all published between 2015 and 2017 in Chinese journals; all were published in Chinese and were performed in China. The studies examined the teaching methods in different nursing courses; four pertained to the fundamentals of nursing, two addressed introductions to nursing, and one of each involved preventive medicine, emergency and critical care, geriatric nursing, surgical nursing, and obstetrics and gynecological nursing. The participants in the 11 included studies were all undergraduate baccalaureate nursing students. The sample size ranged from 62 to 242 participants, and the pooled sample size was 1484 (742 in the flipped classroom group and 742 in the control group). All the studies employed the flipped classroom as the educational approach in the intervention group and traditional lectures as the approach in the control group. All the study outcomes were measured using theoretical knowledge and/or skill scores obtained after applying flipped classroom teaching. The study duration varied from one semester to two semesters.

The characteristics of the included studies are summarized in Table 1.

3.3 Risk of bias in the included studies

We used the Cochrane risk-of-bias tool to assess the risk of bias for each study. We presented the results using a “risk of bias summary” (Fig. 2). All the studies were described as “randomized”, but only 4 of them reported the randomization methods they used. None of the studies described their allocation concealment or blinding of participants, personnel and outcome data. All the studies clearly reported all the expected results to limit reporting bias. Moreover, all the studies reported that there
were no statistically significant differences in age or sex between the flipped classroom and control groups at the baseline (P>0.05). Finally, the shape of the funnel plot for the primary “theoretical knowledge scores” outcome of the 9 studies was nearly symmetrical (Fig. 3). There was no significant publication bias indicated in the primary analysis.
3.4 Meta-analysis results

3.4.1 Theoretical knowledge scores

Nine studies (Deng, 2016, Li, 2016, Li et al., 2017, Liu, 2016, Tian et al., 2015, Yang, 2016, Zhang et al., 2017, Zhao et al., 2016, Zhong et al., 2016) including 1180 students (590 in the flipped classroom group, 590 in the control group) reported their students’ theoretical knowledge scores. One study (Tian et al., 2015) showed no statistically significant difference in student theoretical knowledge scores between the flipped classroom and the control group, while the others (Deng, 2016, Li, 2016, Li et al., 2017, Liu, 2016, Yang, 2016, Zhang et al., 2017, Zhao et al., 2016, Zhong et al., 2016) showed significant differences. A high level of heterogeneity was observed between the nine studies ($I^2=87\%$, $P<0.001$), and thus, a random effects model was utilized for the meta-analysis. The pooled effect size showed a significant difference between the flipped classroom and traditional lecture groups (SMD=1.06, 95% CI:
A fixed effect model was also applied to pool the data, and the pooled effects still favored the flipped classroom group (SMD=1.04, 95% CI: 0.91-1.16, P<0.001). We observed heterogeneity between studies upon reporting theoretical knowledge scores; therefore, a sensitivity analysis was performed to verify the reliability of the results. After excluding the study with the largest sample size (Li et al., 2017) from the analyses, the pooled effect size favored the flipped classroom group (SMD=1.15, 95% CI: 0.82-1.47, P<0.001), and the effects observed in the primary analysis were not changed.

3.4.2 Skill scores

Five studies (Ji et al., 2016, Li, 2016, Tao et al., 2016, Zhang et al., 2017, Zhao et al., 2016) involving 810 students (406 in the intervention group, 404 in the control group) reported the student skill scores. One study (Zhang et al., 2017) showed no statistically significant difference in student skill scores between the flipped classroom and the control group, while the others (Ji et al., 2016, Li, 2016, Tao et al., 2016, Zhao et al., 2016) showed significant differences. There was a high degree of heterogeneity (I²=97%, P<0.001), and a random effects model was therefore used.
The pooled effect size showed a significant difference between the flipped classroom and traditional lecture (SMD=1.40, 95% CI: 0.46-2.34, P<0.001) (Fig. 5). A fixed effect model was also applied to pool the data, and the pooled effects still favored the flipped classroom group (SMD=1.56, 95% CI=1.40-1.72, P<0.001). Given the observed heterogeneity between the studies that reported skill scores, a sensitivity analysis was performed to verify the reliability of the results. After excluding the study with the largest sample size (Ji et al., 2016) from the analyses, the pooled effect size favored the flipped classroom group (SMD=1.21, 95% CI: 1.03-1.40, P<0.001), and the effects observed in the primary analysis were not changed.

3.4.3 Quality of evidence

The quality of evidence was evaluated with the GRADE system. As shown in Table 2, the evidence grade of the results regarding theoretical scores and skill scores was low and very low, respectively.

4. Discussion

4.1 Summary of main findings

To the best of our knowledge, this is the first meta-analysis to examine the effectiveness of the flipped classroom approach in Chinese baccalaureate nursing
students. A comprehensive search of the literature comparing the flipped classroom with traditional lectures was performed, and eleven studies met the predefined inclusion criteria. All the studies used a flipped classroom as the teaching approach in the intervention group and traditional lectures as the approach in the control group. The results indicated an improvement in students’ theoretical knowledge scores and skill scores through the flipped classroom approach.

Although flipped classroom pedagogy has been widely applied in nursing education around the world, the use of this teaching method in China is still in its initial stages, especially in higher nursing education. Some nursing colleges and universities have made tentative steps towards utilizing the flipped classroom approach in nursing education (especially baccalaureate nursing education) since 2014, and more and more nursing educators have realized the advantages of flipped classrooms and are trying to use the new pedagogy (Zhang et al., 2017). Because different education systems or cultural backgrounds may influence the effectiveness of the flipped classroom method (Frambach et al., 2012, Kim, 2008, Sharma et al., 2014), the target population was limited to Chinese nursing students in the Chinese education system to evaluate the potential effectiveness of the flipped classroom approach in China. In addition, we focused on two outcomes (theoretical knowledge scores and skill scores) in this meta-analysis on the basis of Kirkpatrick's framework of evaluation, which assesses outcomes on the following four different levels: level 1 (reactions), level 2 (learning), level 3 (behavior), and level 4 (results) (Ameh and Van, 2015). Among these levels, evaluation at level 2, or determining participant knowledge and skills in a test setting, is widely used (Yardley and Dornan, 2012). In addition, theoretical knowledge scores and skill scores are relatively objective and reliable for assessing the effectiveness of flipped classrooms compared with other subject outcomes.

The results of the current meta-analysis revealed that the flipped classroom pedagogy had positive effects on the students’ academic performance, which is consistent with the previous research (Betihavas et al., 2016, Njie-Carr et al., 2017,
There are several possible reasons to explain the findings. First, according to Bloom’s revised taxonomy of learning, the traditional lecture has primarily involved spending classroom time (when teachers and students interact face-to-face) to promote low-level cognitive work (the acquisition and comprehension of factual knowledge). However, the flipped classroom offers the opportunity for students to engage in higher-order cognition (application, analysis, evaluation, and synthesis of knowledge) through the use of active learning strategies, such as problem-based learning, simulation, think-pair-share activities, student presentations and discussions, and others (Krathwohl, 2002, Mclaughlin et al., 2014), which can foster students’ motivation to acquire knowledge and skills (Prince, 2004). Second, the flipped classroom emphasizes individualized student-centered learning, and students can learn according to their own learning style by studying the key and difficult points repeatedly before class. In the face-to-face class phase, students use their knowledge with instructors’ direction to address challenging problems in a setting that promotes teamwork, which helps students acquire knowledge and skill better than traditional classroom lectures alone (Zhong et al., 2016). Third, Chinese students have received traditional lecture-based instruction since primary school. Instructors are usually dominant, and students are only a passive audience, which greatly affects learning enthusiasm among the students (Liu et al., 2015). However, compared with traditional learning, the flipped classroom approach is a novelty for them and has greatly motivated their learning interest (Yan and Xie, 2015).

The purpose of this meta-analysis was to assess the overall effectiveness of flipped classrooms in Chinese baccalaureate nursing education. All the studies included in this meta-analysis used the flipped classroom as the intervention and theoretical and/or skill scores as the outcome to assess the intervention effectiveness; this methodological consistency reduced the heterogeneity to some extent. However, the educational environments, courses, course contents, course duration, concentration, examination criteria, and others were different among included studies, which resulted in high heterogeneity among the included studies for theoretical
knowledge and skill scores. Therefore, a random effects model was utilized for the meta-analysis. Sensitivity analyses were performed by excluding the study with the largest sample size, while the result did not change the effects observed in the primary analysis. This result indicated that our result was stable and reliable even though there was high heterogeneity.

All the included studies had a moderate possibility of bias due to the low methodological quality, with 7 studies lacking a description of the randomization method and no studies mentioning allocation concealment. Therefore, selection bias was present. Due to the nature of the flipped classroom teaching approach, it was difficult to blind the students and faculty, but none of the studies reported whether the outcome assessors were blinded. We did not identify any attrition bias or reporting bias based on the information reported in the original studies. In addition, the funnel plot was almost symmetrical, and there was no publication bias in the meta-analysis. Based on the findings above, we did not obtain high-quality evidence in the present meta-analysis, given that the summarized evidence ranged from very low to low.

4.2 Comparisons with other published reviews

Three review articles (Betihavas et al., 2016, Njie-Carr et al., 2017, Presti, 2016) to date have summarized and analyzed studies that were previously conducted on the impact of the flipped classroom model in higher nursing education. These reviews involved only descriptive analyses and no meta-analysis. Betihavas’s (Betihavas et al., 2016) systematic review revealed that the use of the flipped classroom in higher nursing education yielded neutral or positive academic outcomes and mixed results regarding satisfaction. In that review, although some students were satisfied with the flipped model approach, whether this satisfaction translated into improved final examination scores remained unknown. Presti’s (Presti, 2016) integrative review indicated that the flipped classroom approach yielded positive outcomes, but quantifiable, significant changes in the nursing students’ knowledge, skills, and attitudes in response to the flipped learning approach were lacking. Finally, Njie-
Carr’s (Njie-Carr et al., 2017) integrative review concluded that the flipped classroom model improved student examination scores, course performance, and satisfaction. However, the results were mixed, because they were derived primarily from non-experimental studies that employed various methods in the classroom and a variety of outcomes, such as single examination scores or course grades. In summary, these findings hinted that flipped classroom yielded neutral or positive outcomes in nursing education, and our findings further confirmed previous results.

### 4.3 Strengths and limitations

#### 4.3.1 Strengths of the review

Our meta-analysis had some strengths. First, we conducted a broad search using both MeSH terms and keywords that covered flipped classroom and nursing education, and this search was conducted in multiple Chinese and English databases. In addition, the sensitivity analyses indicated that the results of the meta-analysis were credible. Finally, there was no evidence of publication bias in our meta-analysis.

#### 4.3.2 Limitations of the review

Several potential limitations of our study should be acknowledged. First, because our research focused only on undergraduates in higher nursing education in China, the results may be applicable to situations in China only. Second, our study assessed only two objective outcomes, theoretical knowledge scores and skill scores, and we did not evaluate other subjective outcomes, such as student satisfaction, critical thinking ability, and teamwork ability, etc. Third, although the search strategy was extensive and inclusive, we did not search the unpublished literature, and hence the related data in these investigations might be omitted.

#### 4.3.3 Limitations of the included studies

Some limitations may be present in the included studies. First, all the included studies had low methodological quality, such as inaccurate randomization methods,
allocation concealment and blinding of assessors, etc. that may have resulted in selection bias and detection bias. Second, there were no standard measurement tools used to assess the effectiveness of the flipped classroom pedagogy, and the different instruments with different validity and reliability values may have led to measurement bias. Third, the sample size in some included studies was small, which may affect the intervention effects. Lastly, the courses, contents, duration, concentration, examination criteria, and other factors vary among the included studies, which resulted in significant heterogeneity.

4.4 Implications for nursing education and future research

Our findings suggest that the flipped classroom is a promising teaching approach. However, Morton et al. (Morton and Colbert-Getz, 2017) and Tian et al. (Tian et al., 2015) found that flipped classroom students outperformed lecture classroom students on analysis items but there were no differences in performance between the two groups in terms of knowledge and application. Therefore, not all the contents are appropriate for flipping, such as courses conveying large amounts of factual content, but flipping may benefit retention when students are expected to analyze material. In addition, pre-class work prepares students for in-class work, but if students do not prepare fully pre-class, the learning effect of the in-class phase will be affected (Ramnanan and Pound, 2017). Hence, it is very important to improve students’ learning interest and compliance (Chen et al., 2017, Heitz et al., 2015). Some suggestions to achieve this improvement are as follows: refining learning materials to minimize student fatigue and distraction; and paying attention to process evaluation. For example, instructors should record and evaluate students’ pre-class performance and give quizzes to supervise and facilitate learning.

This review reveals suggestions for future research. Given the methodological drawbacks of the included studies, high-quality studies with accurate randomization and blinding to reduce the risk of bias are needed. In addition, all the included studies in this review assessed the effects of the flipped classroom across three of
Kirkpatrick’s classification measures as follows: perception, attitude and change in knowledge and skill sets, and none of the studies conducted a process evaluation of the flipped classroom intervention. Therefore, there is a need to conduct studies with evaluations of the long-term effects of flipped classrooms regarding retention and transfer of knowledge to professional practice and patient care (the highest two levels of Kirkpatrick’s measures) and to perform a process evaluation to better examine the effects of flipped classrooms. Moreover, further studies with a large sample size are warranted. Lastly, few studies applied flipped classrooms to laboratory courses. Therefore, more studies are needed to confirm the effect of flipped classrooms in laboratory courses.

5. Conclusions

This meta-analysis is the first evidence-based study to include all RCTs used to evaluate the effects of the flipped classroom approach in baccalaureate nursing students in China. The results showed that the flipped classroom pedagogy is an effective learning approach to improve students’ theoretical knowledge and skills. This finding suggests that flipped classrooms can be integrated into nursing education to improve academic performance. However, due to the limitations discussed earlier, additional studies with large samples and high methodological quality are warranted in a variety of educational settings to confirm our findings.

Conflict of interests

The authors report no conflict of interests.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author Contributions
RJH, HMG, YSY, ZHN and NJ were responsible for conceiving and designing the experiments, collecting and analyzing the data, and writing and revising the manuscript. RJH and HMG were responsible for the data interpretation. XLJ made important intellectual contributions to the research design, provided technical guidance and revised the manuscript. All the authors read and approved the final version of the manuscript.

What is already known about the topic?

- The flipped classroom approach has been widely used in higher nursing education settings in China since 2014.
- Many studies have examined the effects of the flipped classroom approach on theoretical knowledge scores and skill scores in nursing students, but the findings have been mixed.
- Previous reviews from other countries showed a positive relationship between the flipped classroom approach and improved examination scores in nursing students. These studies performed only descriptive analyses and did not involve a meta-analysis.

What this paper adds

- The flipped classroom approach was more effective than traditional lectures at improving students’ theoretical knowledge scores and skill scores.
- The studies included in this review exhibited some common weaknesses, such as low methodological quality (inaccurate randomization methods, allocation concealment and blinding, etc.), small samples, non-uniform assessment tools,
and a lack of long-term effectiveness assessments and process evaluations on flipped classrooms.

Acknowledgments

We thank American Journal Experts (AJE) for the English language editing. This manuscript was edited for English language usage by AJE.
References


Fig. 1. Flow diagram of study selection.
RCT=randomized controlled trial, CENTRAL=Cochrane Central Register of Controlled Trials, CNKI=China National Knowledge Infrastructure, Wanfang=Wanfang Database, and VIP=the Chinese Scientific Journals Database.
Fig. 2. Risk of bias summary: a review of the author evaluations of each risk of bias item for each one included in the study.
Fig. 3. Funnel plot of comparisons: flipped classroom versus traditional lecture; outcome: theoretical knowledge scores. SE=standard error, SMD=standardized mean difference.
Fig. 4. Forest plot of comparison: flipped classroom versus traditional lecture; outcome: standardized mean difference for the theoretical scores. CI=confidence interval, SD=standard deviation.
Fig. 5. Forest plot of comparisons: flipped classroom versus traditional lecture, outcome: standardized mean difference of the skill scores. CI=confidence interval, SD=standard deviation.
Table 1 Characteristics of the included studies

<table>
<thead>
<tr>
<th>Study ID</th>
<th>Study design</th>
<th>Subjects</th>
<th>Course type</th>
<th>Participant characteristics</th>
<th>Sample size (IG/CG)</th>
<th>Intervention</th>
<th>Control</th>
<th>Outcomes</th>
<th>Duration of study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tao et al 2016 (Dalian, China)</td>
<td>RCT</td>
<td>Fundamentals of nursing</td>
<td>Laboratory course</td>
<td>Sophomores at a university. Mean age IG: 20.64 (SD 0.77), CG: 20.57 (SD 0.86). Gender IG: Female 28 (87.5%), CG: Female 27 (90%). Admission scores IG: 537.44 (SD 28.87), CG: 528.53 (SD 33.86).</td>
<td>62 (32/30)</td>
<td>Used flipped classrooms as the educational approach. Pre-class: interactive online modules; prerecorded mini-videos (5-8 min); textbook reading; pre-assignments. In-class: group-developed presentations; group discussion.</td>
<td>Traditional lecture</td>
<td>Skill scores (final examination)</td>
<td>One semester</td>
</tr>
<tr>
<td>Yang 2016 (Henan, China)</td>
<td>RCT</td>
<td>Preventive medicine</td>
<td>Theory course</td>
<td>Sophomores at a college. Gender IG: Male 5 (7.8%), CG: Male 4 (6.3%).</td>
<td>128 (64/64)</td>
<td>Used flipped classrooms as the educational approach. The detailed implementation process was not described.</td>
<td>Traditional lecture</td>
<td>Theoretical knowledge scores (final examination)</td>
<td>One semester</td>
</tr>
<tr>
<td>Zhong et al 2016 (Guizhou, China)</td>
<td>RCT</td>
<td>Emergency and critical care</td>
<td>Theory course</td>
<td>Juniors at a college.</td>
<td>72 (36/36)</td>
<td>Used flipped classrooms as the educational approach. Pre-class: interactive online modules; voice-over PowerPoint; prerecorded mini-videos; textbook reading; pre-assignments. In-class: PowerPoint presentations; role play; games.</td>
<td>Traditional lecture</td>
<td>Theoretical knowledge scores (unit test)</td>
<td>One semester</td>
</tr>
<tr>
<td>Tian et al 2015 (Shanxi, China)</td>
<td>RCT</td>
<td>Geriatric nursing</td>
<td>Theory course</td>
<td>Juniors at a university. Mean age IG: 21 (SD 1.14), CG: 20.72 (SD 0.85). Gender IG: Female 30 (93.8%), CG: Female 30 (93.8%)</td>
<td>64 (32/32)</td>
<td>Used flipped classrooms as the educational approach. Pre-class: prerecorded mini-videos (10-15 min); pre-assignments. In-class: discussion; presentations; role play; case studies.</td>
<td>Traditional lecture</td>
<td>Theoretical knowledge scores (final examination)</td>
<td>One semester</td>
</tr>
<tr>
<td>Deng 2016 (Hubei, China)</td>
<td>RCT</td>
<td>Introduction to nursing</td>
<td>Theory course</td>
<td>Freshmen at a university. Mean age IG: 18.91 (SD 1.03), CG: 18.69 (SD 0.90). Gender IG: Female 28 (82.4%), CG: Female 27 (84.4%). Admission scores IG: 490.88 (SD 7.23), CG: 491.97 (SD 8.26).</td>
<td>66 (34/32)</td>
<td>Used flipped classrooms as the educational approach. Pre-class: prerecorded lecture videos; PowerPoint lecture; classical cases; pre-assignments. In-class: group-developed presentations; discussion; question and answer sessions.</td>
<td>Traditional lecture</td>
<td>Theoretical knowledge scores (final examination)</td>
<td>One semester</td>
</tr>
<tr>
<td>Liu 2016 (Shaanxi, China)</td>
<td>RCT</td>
<td>Introduction to nursing</td>
<td>Theory course</td>
<td>Freshmen at a university.</td>
<td>104 (53/51)</td>
<td>Used flipped classrooms as the educational approach. Pre-class: prerecorded mini-videos; review materials; pre-assignments. In-class: group-developed presentations; discussion; question and answer sessions.</td>
<td>Traditional lecture</td>
<td>Theoretical knowledge scores (final examination)</td>
<td>One semester</td>
</tr>
<tr>
<td>Zhao et al 2016 (Hunan, China)</td>
<td>RCT</td>
<td>Fundamentals of nursing &amp; Laboratory course</td>
<td>Theory course</td>
<td>Sophomores at a university. Age IG: Ranging from 18 to 22, CG: Ranging from 19 to 22. Gender IG: Female 110 (95.7%), CG: Female 108 (94.7%).</td>
<td>229 (115/114)</td>
<td>Used flipped classrooms as the educational approach. Pre-class: prerecorded mini-videos (approximately 10 min); interactive online modules; pre-assignments. In-class: discussion; instructors’ demonstration; standard patients.</td>
<td>Traditional lecture</td>
<td>Theoretical knowledge and skill scores (final examination)</td>
<td>Two semesters</td>
</tr>
<tr>
<td>Study</td>
<td>Design</td>
<td>Topic</td>
<td>Population</td>
<td>Participants</td>
<td>Intervention</td>
<td>Comparison</td>
<td>Assessment</td>
<td>Length</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>--------</td>
<td>-------</td>
<td>------------</td>
<td>--------------</td>
<td>-------------</td>
<td>------------</td>
<td>-----------</td>
<td>--------</td>
<td></td>
</tr>
<tr>
<td>Ji et al 2016 (Hunan, China)</td>
<td>RCT</td>
<td>Surgical nursing &amp; Laboratory course</td>
<td>Sophomores at a college.</td>
<td>242 (120/122)</td>
<td>Used flipped classrooms as the educational approach. Pre-class: review teaching materials; practice; discussion; self-developed mini-video (approximately 5 min). In-class: instructors comment on students’ self-developed mini-video and discussion.</td>
<td>Traditional lecture</td>
<td>Skill scores (final examination)</td>
<td>One semester</td>
<td></td>
</tr>
<tr>
<td>Li 2016 (Hunan, China)</td>
<td>RCT</td>
<td>Fundamentals of nursing &amp; Theory course &amp; Laboratory course</td>
<td>Sophomores at a college.</td>
<td>182 (90/92)</td>
<td>Used flipped classroom as the educational approach. Pre-class: interactive online modules; prerecorded mini-videos; PowerPoint lecture; pre-assignments. In-class: problem-based learning; case studies; discussion; instructors’ demonstration; practice.</td>
<td>Traditional lecture</td>
<td>Theoretical knowledge and skill scores (final examination)</td>
<td>One semester</td>
<td></td>
</tr>
<tr>
<td>Li et al 2017 (Hebei, China)</td>
<td>RCT</td>
<td>Obstetrics and gynecological nursing</td>
<td>Juniors at a college. Mean age IG: 21.56 (SD 1.03), CG: 21.63 (SD 1.21). Gender IG: Female 105 (89.7%), CG: Female 108 (89.3%).</td>
<td>240 (117/123)</td>
<td>Used flipped classrooms as the educational approach. Pre-class: interactive online modules; prerecorded mini-videos (8-30 min); pre-assignments. In-class: problem-based learning; group-developed presentations; discussion; question and answer sessions.</td>
<td>Traditional lecture</td>
<td>Theoretical knowledge scores (final examination)</td>
<td>One semester</td>
<td></td>
</tr>
<tr>
<td>Zhang et al 2017 (Guangdong, China)</td>
<td>RCT</td>
<td>Fundamentals of nursing &amp; Theory course &amp; Laboratory course</td>
<td>Sophomores at a university. Mean age IG: 19.63 (SD 0.76), CG: 19.54 (SD 0.72). Gender IG: Female 44 (89.8%), CG: Female 45 (97.8%).</td>
<td>95 (49/46)</td>
<td>Used flipped classroom as the educational approach. Pre-class: interactive online modules; prerecorded videos; PowerPoint lecture; pre-assignments. In-class: problem-based learning; discussion; instructors’ demonstration; practice.</td>
<td>Traditional lecture</td>
<td>Theoretical knowledge and skill scores (final examination)</td>
<td>One semester</td>
<td></td>
</tr>
</tbody>
</table>

RCT, randomized controlled trial. IG, intervention group. CG, control group. SD, standard deviation.
Table 2 Quality of evidence for the included studies.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Risk of bias</th>
<th>Inconsistency</th>
<th>Indirectness</th>
<th>Imprecision</th>
<th>Publication bias</th>
<th>Quality of evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theoretical knowledge scores</td>
<td>Serious (-1)</td>
<td>Serious (-1)</td>
<td>No</td>
<td>No</td>
<td>Undetected</td>
<td>Low</td>
</tr>
<tr>
<td>Skill scores</td>
<td>Serious (-1)</td>
<td>Serious (-1)</td>
<td>No</td>
<td>Serious (-1)</td>
<td>Undetected</td>
<td>Very low</td>
</tr>
</tbody>
</table>