Does the Flipped Classroom improve exam performance in medical education? A systematic review [version 1]

Yasmin Hughes¹, Nicholas Lyons²

¹University of Queensland
²Prince Charles Hospital

Abstract
This article was migrated. The article was marked as recommended.

Background: The 'flipped classroom' (FC) is a blended learning model in which educational material is delivered on-line prior to class, students then apply this knowledge through discussion and problem solving activities in class. Its effectiveness in medical education is debated. The aim of this systematic review is to assess the outcomes of studies which have compared the exam performance of traditional didactic teaching (DT) to the FC in medical education.

Methods: A broad based literature search was performed in accordance with PRISMA protocol. MESH terms were 'flipped classroom', 'flipped teaching', 'blended learning' and 'medical education'. The outcomes of interest were test score results following FC compared to DT methods.

Results: Eleven studies with a total of 2052 study participants were included in the review. Four studies demonstrated a significant improvement in test scores using FC compared to DT; four showed no significant difference and three demonstrated mixed results.

Discussion and Conclusions: The results of this review are equivocal. Study heterogeneity in design, participants and subject covered may account for some of this disparity. Two studies provide evidence that the FC results in improved performance on higher cognitive tasks however further robust, in depth studies are required to demonstrate this conclusively.

Keywords
Flipped classroom, Medical education, Undergraduate, Postgraduate
Corresponding author: Yasmin Hughes (yasmin.hughes@nhs.net)

Competing interests: No competing interests were disclosed.

Grant information: The author(s) declared that no grants were involved in supporting this work.

Copyright: © 2017 Hughes Y and Lyons N. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Hughes Y and Lyons N. Does the Flipped Classroom improve exam performance in medical education? A systematic review [version 1] MedEdPublish 2017, 6:100 https://doi.org/10.15694/mep.2017.000100

First published: 13 Jun 2017, 6:100 https://doi.org/10.15694/mep.2017.000100
Introduction

Medical education is faced with training medical students and doctors to become competent life-long learners at a time when medical knowledge is increasing at an exponential rate. Just as medical knowledge is advancing, so too is technology. To address the challenge faced by medical education, there has been growing interest in innovative approaches using online learning platforms. Whilst distance learning has enjoyed success in a number of formats (Kay, Reimann, & Diebold, 2013) it is unlikely that online, distance learning courses will be able to entirely replace face to face medical teaching for several reasons. Firstly, practicing medicine relies on excellent communication skills; skills that are unlikely to be developed with an online resource. Secondly, there is proposed benefit in learning collaboratively. ‘Communities of practice’ defined by organisational theorist Etienne Wenger as ‘groups of people who share a concern, a set of problems, or a passion about a topic, and who deepen their knowledge and expertise in this area by interacting on an ongoing basis,” has merits in medical education (Wenger, 1998). Thirdly, development of complex reasoning skills is at the core of medical education (Connor et al., 2016). Collaborative face to face ‘active’ learning offers the opportunity for students to engage in higher order cognitive tasks such as application, evaluation and synthesis of knowledge (Connor et al., 2016; Moffett & Mill, 2014), skills which are essential in a practicing clinician.

“Blended learning”, the combination of teaching delivered by online digital media with face to face learning potentially provides a solution to balancing the use of technology to assist in delivering large volumes of medical information with hands on clinical coaching. The ‘flipped classroom’ (FC) is a student centred, blended pedagogical model in which the homework and classroom components of a teaching session are reversed. In this type of blended learning, students work independently, prior to class, to learn basic facts and concepts through online educational content such as online videos, quizzes or podcasts. As the students are prepared prior to entering class, classroom time is devoted to the application of this knowledge through discussion, problem solving and deeper probing of concepts. This approach has been widely accepted by a number of disciplines (Adams, Garcia, & Traustadóttir, 2016; McLaughlin et al., 2013; Moffett & Mill, 2014).

Conceptually, the FC is well suited to medical students who, in large, are highly motivated, independent learners. The opportunities for discussion of concepts and interactive application of knowledge to clinical problems, whilst promoting higher order cognitive thinking, simulates clinical working environments. Since its development to deliver a core biochemistry course at Stanford School of Medicine in 2012 (Prober & Heath, 2012), a number of prestigious institutions have adopted this approach to delivering medical education, including Harvard School of Medicine (Fu & Joung, 2015). This type of learning is also suited to postgraduate training doctors where teaching time needs to fit around demanding clinical commitments. Yale School of Medicine is currently using this concept to deliver anaesthesia residency training (Yale School of Medicine, 2017).

Despite its implementation, within both undergraduate and postgraduate medical education settings, little is known about the effectiveness of the FC in terms of objective quantitative measures and how these outcomes compare to traditional didactic teaching (DT) approaches. A body of research has investigated the qualitative effects of the FC and a recent systematic review addressed medical students perceptions of the FC (Ramnanan & Pound, 2017). Whilst this is of interest, it would be beneficial to evaluate learning (level two) by students rather than just reaction (level one) as per Donald Kirkpatrick’s four levels of evaluation model (Kirkpatrick, 2007). The recent review addressed undergraduate medical education, however, failed to appreciate the importance of evaluating the effect of the FC in postgraduate doctor training. At the time of writing, no systematic reviews exploring the FC and medical education have specifically considered whether the FC results in improved examination performance compared to DT methods. Moreover, no reviews to date have considered the effects of the FC in both undergraduate and postgraduate medical education settings. We aim to rectify this.

The main objective of this systematic review is to investigate and critically appraise the published data on quantitative outcomes of the use of the FC compared to DT approaches in both undergraduate and postgraduate medical education.

Materials and methods

A literature search was performed in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-analysis (PRISMA) protocol. MeSH search terms used were ‘flipped classroom’, ‘flipped learning’, ‘blended learning’ and ‘medical education’; terms were combined using ‘and/or’ as appropriate. These terms were then applied to Medline (since 1952), EMBASE (since 1980), the Cochrane library (since 1995), CINAHL (since 1982) and Google Scholar and the search conducted independently by the authors. MeSH terms were exploded to identify additional studies.

Inclusion criteria

Included studies assessed the effectiveness of the FC when used in medical education (pre or post graduate) and compared reported objective test score results for FC to scores achieved with DT. FC was defined as an educational technique that
consisted of two parts: computer-based individual instruction that occurred both prior to class and outside of the classroom, and interactive group learning activities which occurred during classroom time. All studies including this design were included irrespective of their use of the term FC. Studies were carefully evaluated for duplication or overlapping data.

**Exclusion criteria**
Studies were excluded that did not use a FC teaching modality as determined by the authors, did not specifically study medical student or doctor education, or failed to provide quantitative objective outcomes (assessment results). Medical education was determined as modern western medicine only. Chinese, Ayurveda and traditional medical practises were excluded on the basis of large heterogeneity in underlying philosophies and factual knowledge likely to result in a confounded comparison. Letters, reports, conference abstracts or abstract only reports were excluded.

**Outcome measures**
Outcomes of interest were assessment score results post a FC intervention compared to post DT.

**Study selection**
The authors independently performed the search strategy initially undertaking a title screen followed by abstract review and full text review of appropriate studies. Publications satisfying the exclusion criteria were discarded at each stage (figure 1). Publications without abstracts moved straight to full text review. Discrepancies between author searches were resolved by consensus following discussion.

**Data extraction**
Data extraction was undertaken independently by the authors using a standardised proforma for the objective scores only. The following demographic and medical education parameters were extracted from each study; study characteristics (first author, year of publication, study design, Institution), student characteristics (number and demographics), outcome measures (assessment method, quantitative results, qualitative results).

**Quality assessment**
Quality assessment was undertaken using a Modified Index for Non-Randomised Studies (MINORS) (Slim et al., 2003) modified for a non-clinical research question (see appendix). Specifically, point two (Inclusion of consecutive patients) was not relevant to any of the studies included in the final review and was thus omitted. The remaining eleven items were used as per original description. Prospective design criteria was applied to the comparison group only.

**Publication bias**
No statistical analysis assessing the risk of publication bias was undertaken as per guidance in the “Cochrane handbook of systematic reviews of interventions” (Higgins & Green, 2011). Studies of eleven or fewer articles do not require analysis for publication bias.

**Results**
The initial search returned 345 articles after removal of duplicates. 225 of which were excluded at title screen. 120 articles were taken forward for abstract review, 94 articles were excluded (see figure 1). The remaining 26 articles went forward to full text review, of which 15 were excluded (nine failed to compare FC and DT, five provided inadequate objective test results, one using Chinese medicine) leaving eleven studies (Belfi, Bartolotta, Giambrone, Davi, & Min, 2017; Bonnes et al., 2017; Boysen-osborn et al., 2016; Connor et al., 2016; Evans et al., 2016; Gillispie, 2016; Heitz, Prusakowski, Willis, & Franck, 2015; Liebert et al., 2017; Morton & Colbert-getz, 2016; Rui et al., 2017) which were included in quantitative analysis.

A total of 2052 students were included in the analysis, 1098 of which underwent the FC condition and 954 in the DT group. Of the DT group, six (14-16,18,19,22) used a retrospective cohort as a comparator group with the remaining five (12,13,16,17,20) studies using a contemporary DT control group.

**Allocation of groups**
Of the five Contemporary case control studies, two studies randomised students to either the intervention group (FC) or control group (DT), two studies did not state how assignment was carried out and one study assigned students according to their post-graduate year of training (PGY). In each of the six Retrospective case comparison studies, a contemporary intervention (FC) group was compared to a retrospective control (DT) group.
Contemporary case-control design
Of the five studies using a contemporary DT comparator group, three studies (Belfi et al., 2017; Bonnes et al., 2017; Gillispie, 2016) used a simple case control design with students undergoing either FC or DT teaching. The remaining two (Connor et al., 2016; Heitz et al., 2015) studies used a crossover design where students experienced both DT and FC conditions covering different teaching material for the two conditions. The study participants were either undergraduate or postgraduate medical students or postgraduate doctors. Medical students were enrolled in the first to fourth years and participating in either an Emergency medicine clerkship, Radiology clerkship or Pre-clinical ECG interpretation module. Postgraduate doctors were in either their first, second or third year of internal medicine training (see table 1).

Retrospective case-control design
Six studies (Boysen-osborn et al., 2016; Evans et al., 2016; Gillispie, 2016; Liebert et al., 2017; Morton & Colbert-getz, 2016; M. R. Sajid et al., 2016) used a retrospective comparator group, typically comparing FC test scores to those of previous years test scores following DT. Subject areas included in these six studies were; haematology undergraduate teaching, Obstetrics and gynaecology clerkship, Advanced cardiac life support teaching of medical students, Anatomy teaching and a Clinical epidemiology and biostatistics module delivered to first year medical students and a core surgery clerkship of third year medical students. Three of the retrospective case comparison studies (Evans et al., 2016; Morton & Colbert-getz, 2016; M. R. Sajid et al., 2016) used multiple choice question (MCQ) exam alone with two studies using MCQ combined with another assessment method such as an Objective structured clinical examination (OSCE), questions regarding clinical cases or ECG rhythm strip interpretation (Boysen-osborn et al., 2016; Gillispie, 2016). One study used pre and post intervention National Board of Medical Examiners (NBME) test scores as the quantitative assessment (Liebert et al., 2017).

Test scores
Contemporary case-control: There was variation in both the methodology and results of assessment within the contemporary case-control group. All studies assessed the teaching methods by examination scores; three studies used MCQ scores (Belfi et al., 2017; Connor et al., 2016; Heitz et al., 2015), one study used a pre and post course QI knowledge assessment tool (QIKAT) (Bonnes et al., 2017) and one study used an ECG interpretation examination (Rui et al., 2017). Four of the five studies demonstrated a clear, statistically significant, improvement in assessment scores for the FC groups over the DT (Belfi et al., 2017; Bonnes et al., 2017; Connor et al., 2016; Rui et al., 2017) with one study demonstrating no improvement for FC over DT (Heitz et al., 2015). Two of the studies used a cross-over design with students experiencing both DT and FC conditions in an emergency medicine clerkship and radiology clerkship (Connor et al., 2016; Heitz et al., 2015). These studies shared similar designs with two groups of students experiencing four modules, two of which were FC and two DT. Both were assessed with MCQs of 40 and 30 questions respectively. The results demonstrate no significant difference in performance for the emergency medicine clerkship whilst a significant improvement of 10.5% (p=0.013) was noted for the radiology clerkship study.

Retrospective case control: There was more variation in the results of the retrospective case-control group than noted in the contemporary case-control group. Of the six studies, three demonstrated no significant difference between FC and DT interventions (Evans et al., 2016; Liebert et al., 2017; M. R. Sajid et al., 2016). Three studies demonstrated mixed outcomes. One study showed significant improvement in MCQ performance but no benefit to rhythm strip interpretation with both groups scoring 100% (Boysen-osborn et al., 2016). Another showed improvement on OSCE performance with the FC but only 1/2 groups benefitting from FC on MCQ performance (Gillispie, 2016). One study implemented a FC approach to teaching Anatomy and assessed students with an MCQ exam. The MCQs were characterised according to Blooms taxonomy to assess students at different cognitive levels including knowledge, application and analysis of anatomy items. They found that students in the DT group performed better in knowledge and application anatomy items whereas students in the FC group had higher scores on analysis questions.

Statistical Analysis
Given the heterogeneity of the studies, the pooling of data for meta-analysis was deemed as inappropriate.

Quality assessment
Studies were evaluated using a modified MINORS risk of bias tool. Of a maximum score of twelve points, studies ranged from five to ten (appendix 1). One study (Bonnes et al., 2017) scored ten, four scored eight (Belfi et al., 2017; Boysen-osborn et al., 2016; Heitz et al., 2015; Rui et al., 2017), two scored seven (Connor et al., 2016; Gillispie, 2016), three scored six (Liebert et al., 2017; Morton & Colbert-getz, 2016; M. S. Sajid, Bokhari, Mallick, Cheek, & Baig, 2009) and one scored five (Evans et al., 2016). Whilst overall performance on study end point, unbiased assessment and clear aims were good, no studies undertook prospective power calculations and with the exception of two studies (Bonnes et al., 2017; Boysen-osborn et al., 2016), loss to follow up or failure to report total number of students completing examinations
was high. A prospective design was undertaken by six of the eleven studies (Belfi et al., 2017; Bonnes et al., 2017; Connor et al., 2016; Gillispie, 2016; Heitz et al., 2015; Rui et al., 2017).

Discussion
Since the widely cited article describing the use of the FC to teach a core biochemistry course at Stanford Medical School in 2012 (Prober & Heath, 2012), the FC has been introduced to deliver a wide range of medical curricula (Fu & Joung, 2015; Yale School of Medicine, 2017). Previous reviews have presented qualitative outcomes demonstrating that the FC is preferred over DT in undergraduate medical education (Ramnanan & Pound, 2017). This is the first systematic review to assess the effectiveness of the FC compared to traditional DT with objective test scores in both undergraduate and postgraduate medical education. Eleven studies were included in this review of which only four demonstrated a significant improvement for the FC over DT and a further three demonstrating what we have called a mixed outcome where more than one assessment was undertaken, some of which demonstrated benefit of FC and others not.

This variation is likely the result of several factors, the most notable of which are the variation in study design, student type and stage of learning, course material covered, study methodology and the type of assessment used (table 1). In spite of this variation there are still some points of interest. Three studies (Boysen-osborn et al., 2016; Gillispie, 2016; Morton & Colbert-getz, 2016) used more than one type of assessment in their design and reported mixed results suggesting that FC may be better for some types of learning but not others. One of the proposed benefits of the FC is that it allows students to engage in higher order cognitive tasks such as application, analysis and synthesis of knowledge (Lampinen & Arnal, 2009). In their study, Morton et al. used assessment items in the form of MCQs which tested different cognitive levels of Bloom’s taxonomy including knowledge, application and analysis of Anatomy content. This study found that students taught Anatomy by the FC outperformed students taught by a lecture based method on MCQs that tested analysis however, a DT approach resulted in higher MCQ scores based on knowledge and application (Morton & Colbert-getz, 2016). This supports the notion that, in terms of Bloom’s revised taxonomy, traditional lectures promote lower level cognitive work (gain and comprehension of factual knowledge), whereas the FC offers the opportunity for students to engage in higher order cognitive activities such as analysis, evaluation and synthesis of knowledge (Ramnanan & Pound, 2017). This may explain the mixed results from other studies.

Gillipsie et al (Gillispie, 2016) found that the FC students outperformed the DT cohort in OSCEs however, there was a statistically significant increase in MCQ scores for some topics delivered by a DT approach. The OSCE is a well-established mode of assessment specifically designed to provide a valid and reliable measure of students clinical competence (Terry, Hing, Orr, & Milne, 2017). It requires higher order clinical reasoning skills that go beyond basic factual recall of knowledge. One could argue that, whilst the DT method may have resulted in improvement in factual recall in MCQ topics, FC resulted in improvement in higher order problem solving and clinical reasoning skills which are assessed by an OSCE. These are interesting observations that would benefit from further investigation.

The concept of the FC relies on students doing their homework. As this preparation forms the basis for basic knowledge acquisition, in its absence, the class time opportunities for application of knowledge and deeper probing of concepts become redundant. Few studies documented compliance with use of the online material. Heitz et al illustrated no significant difference between FC and DT groups during an Emergency medicine clerkship, however, almost one third (31.1%) of students stated that they were unable to adhere to the study protocol. However, when including data from only students who had followed the protocol, the authors found no statistical difference between FC and DT scores (p=0.8071). Therefore, the level of student concordance alone is unlikely to account for these results.

There are a number of limitations to this systematic review. As discussed above, the number of studies is small, only eleven in total, with large heterogeneity. Clearly there is a need for larger well designed studies to clearly demonstrate whether there should be a place for the FC in medical education and if so, where.

Conclusion
There has been extensive research demonstrating the FC is preferred by students over DT, we would argue that the evidence presented in this systematic review only demonstrates superiority of FC over DT in higher cognitive tasks and does not clearly demonstrate any benefit from widespread use of FC in medical education. More robust studies are required to clearly demonstrate the role of FC in medical education.

Take Home Messages
- The Flipped Classroom has been incorporated into a number of medical school programs and postgraduate medical training programs
• Studies of its effectiveness to date have largely been centred around qualitative student outcomes.

• A review of quantitative outcomes shows mixed outcomes of its effectiveness when applied to medical education.

• More robust studies are required to clearly demonstrate the role of FC in medical education.

Notes On Contributors

Yasmin Hughes MBiochem, MBChB is a Lecturer in Clinical Science at the University of Queensland, Brisbane.

Nicholas Lyons BSc (Hons), MBChB, MRCS is a registrar in General Surgery at Prince Charles Hospital, Brisbane.

Declarations

The author has declared that there are no conflicts of interest.

Appendices

Figure 1: PRISMA Flow chart

- 345 records identified through database searching
- 345 records (title) screened
  • 225 records excluded after title screen
  • Review = 12
  • Non-medical students = 66
  • Symposiums = 2
  • Conference proceedings = 4
  • Inappropriate = 141
- 120 abstracts screened
  • 94 abstracts excluded:
    • No quantitative data = 20
    • Non-medical students = 8
    • Review article = 11
    • Non-English = 2
    • No comparison of FC versus DT = 23
    • Not a FC design = 17
    • Editorial = 1
    • Letter to the Editor = 3
    • Author manuscript = 2
    • Perspective/Current opinion = 4
    • Abstract only = 3
- 26 full text articles assessed for eligibility
  • 15 full text articles excluded:
    • 5 studies lacked quantitative data
    • 9 studies did not compare the FC to a DT approach
    • 1 study looked at Chinese Medicine
- 11 studies included in the review
**Table 1. Summary of quantitative results**

| Study                  | Type of grouping | Assessment | Assessment method | Assessment outcome                             |
|------------------------|------------------|------------|-------------------|------------------------------------------------|
| Belfi et al            | 1 group FC only  | Pre- and Post-course MCQ (45) | FC sig better post test score vs DT (p=0.0060) |
| Heitz et al            | 2 groups FC only | Pre- and Post-course QIKAT | FC sig better performance vs DT (p < 0.0001)   |
| Heitz et al (Crossover design) | 2 groups experiencing: 2 DT modules 2 FC modules | MCQ (40) | No differences between FC and standard for MCQ scores (p=0.494) |
| O’Connor et al (Crossover design) | 2 groups experiencing: 2 DT lecture 2 FC workshops | MCQ (30) | Mean post- minus pre-test scores were 50.5% higher in the FC group than in DT group (p = 0.013) |
| Kuk et al              | 1 group FC only  | ECG interpretation exam | FC sig better post test score vs DT (p=0.001)  |

**Appendix 1 Modified MINORS risk of bias assessment tool**

| Study                  | Clear aim | Prospective design | Appropriate end point | Unbiased assessment end point | Lost to follow up (<5%) | Prospective calculation of study size | Total score |
|------------------------|-----------|--------------------|-----------------------|-------------------------------|--------------------------|--------------------------------------|-------------|
| Belfi et al            | 2         | 2                  | 2                     | 2                             | 0                        | 0                                    | 8           |
| Heitz et al            | 2         | 2                  | 2                     | 2                             | 0                        | 0                                    | 8           |
| Boysen-Osborn et al    | 2         | 0                  | 2                     | 2                             | 0                        | 0                                    | 8           |
| Evans et al            | 1         | 0                  | 2                     | 2                             | 0                        | 0                                    | 5           |
| Gillispie et al        | 1         | 2                  | 2                     | 2                             | 0                        | 0                                    | 7           |
Continued

| Study            | Clear aim | Prospective design | Appropriate end point | Unbiased assessment end point | Lost to follow up (<5%) | Prospective calculation of study size | Total score |
|------------------|-----------|--------------------|-----------------------|-------------------------------|--------------------------|---------------------------------------|-------------|
| Liebert et al.   | 2         | 0                  | 2                     | 2                             | 0                        | 0                                     | 6           |
| Morton et al.    | 2         | 0                  | 2                     | 2                             | 0                        | 0                                     | 6           |
| O’Connor et al.  | 1         | 2                  | 2                     | 2                             | 0                        | 0                                     | 7           |
| Sajid et al.     | 2         | 0                  | 2                     | 2                             | 0                        | 0                                     | 6           |
| Bonnes et al.    | 2         | 2                  | 2                     | 2                             | 2                        | 0                                     | 10          |
| Rui et al.       | 2         | 2                  | 2                     | 2                             | 0                        | 0                                     | 8           |

Bibliography/References

Adams A. E. M., Garcia J., & Traustadottir T. (2016). A quasi-experiment to determine the effectiveness of a partially flipped versus fully flipped undergraduate class in genetics and evolution. *CBE Life Sciences Education, 15*(2), 1–9.

Reference Source

Belfi L. M., Bartolotta R. J., Giambrone A. E., Davi C., & Min R. J. (2017). In *Radiology: Impact on Medical Student Performance and Perceptions. Academic Radiology, 22*(6), 794–801.

Reference Source

Bonnes S. L., Ratele J. T., Halsersen A. J., Carter K. J., Hafsdal L. T., Wang A. T., ... Wittch C. M. (2017). Flipping the Quality Improvement Classroom in Residency Education. *Journal of Educational Evaluation for Health Professions, 20*(1), 1–20.

Reference Source

Connor E. O. D., Fried J., McNulty N., Shah P., Hogg J. P., Lewis P., ... Reddy S. (2016). Flipping Radiology Education Right Side Up. *Academic Radiology, 23*(7), 810–822.

Reference Source

Evans K. H., Thompson A. C., Brien C. O., Bryant M., Basaviah P., Prober C., & Popat R. A. (2016). An Innovative Blended Preclinical Curriculum in Clinical Epidemiology and Biostatistics: Impact on Student Satisfaction and Performance. *Academic Medicine, 91*(5), 666–700.

Reference Source

Fu M., & Joung J. (2015). Top Medical Schools React to Harvard’s Curriculum Change. Retrieved from Reference Source

Gillispie V. (2016). Using the Flipped Classroom to Bridge the Gap to Generation Y. *The Ochsner Journal, 16*(1), 32–6.

Reference Source

Heitz C., Prusakowski M., Willis G., & Franck C. (2015). Does the Concept of the “Flipped Classroom” Extend to the Emergency Medicine Clinical Clerkship?. *The Western Journal of Emergency Medicine, 16*(6), 851–855.

Reference Source

Higgins J., & Green S. (2011). *Cochrane handbook for systematic reviews of interventions*.

Reference Source

Kay J., Reimann P., & Diebold E. (2013). MOOCs: So Many Learners, So Much Potential. *IEEE Intelligent, 28*(3), 70–77.

Reference Source

Kirkpatrick D. L. (2007). *The Four Levels of Evaluation. In Evaluating Corporate Training: Models and Issues*. (pp. 95–112).

Reference Source

Lampinen J. M., & Arnal J. D. (2009). A Revision of Bloom’s Taxonomy: An Overview. *American Journal of Psychology, 122*(1), 39–52.

Reference Source

Liebert C. A., D. M., Lin D. T., D. M., Mazer L. M., S. M., ... S. F. A. C. (2017). Effectiveness of the Surgery Core Clerkship Flipped Classroom: a prospective cohort trial. *The American Journal of Surgery, 211*(2), 451–457.

Reference Source

McLaughlin J. E., Griffin L. M., Esserman D. A., Davidson C. A., Glatt D. M., Roth M. T., ... Mumper R. J. (2013). Pharmacy student engagement, performance, and perception in a flipped satellite classroom. *American Journal of Pharmaceutical Education, 77*(9).

Reference Source

Moffett J., & Mill A. C. (2014). Evaluation of the flipped classroom approach in a veterinary professional skills course. *Advances in Medical Education & Practice, 5*, 415–425.

Reference Source

Morton D. A., & Colbert-getz J. M. (2016). Measuring the Impact of the Flipped Anatomy Classroom: The Importance of Categorizing an Assessment by Bloom’s Taxonomy. *Journal of Medical Education and Curriculum Development, 16*(6), 1657–1659.

Reference Source

N Engl. J Med. 366(18), 1657–1659.

Reference Source

Rui Z., Lian-rui X., Rong-zheng Y., Jing Z., Xue-hong W., & Chuan Z. (2017). Flipping the Quality Improvement Classroom in Residency Education. *MedEdPublish 2017, 6:100 Last updated: 14 SEP 2023*

Reference Source

Sajid M. R., Laheji A. F., Abothenain F., Salam Y., AlJayar D., & Obeidat A. (2016). Can blended learning and the flipped classroom improve student learning and satisfaction in Saudi Arabia?. *International Journal of Medical Education, 7*, 281–285.

Reference Source

Ramnanan C., & Poud L. (2017). Advances in medical education and practice: student perceptions of the flipped classroom. *Advances in Medical Education and Practice, 8*, 63–73.

Reference Source

Sajid M. S., Bokhari S. a., Mallick A. S., Cheek E., & Baig M. K. (2009). Laparoscopic versus open repair of incisional/ventral hernia: a meta-analysis. *American Journal of Surgery, 197*(1), 64–72.

Reference Source

Slim K., Nini E., Forester D., Kwiatkowski F., Panis Y., & Chippioni J. (2003). Methodological index for non-randomized studies (Minors): Development and validation of a new instrument. *ANZ Journal of Surgery, 73*(9), 712–716.

Reference Source

Terry R., Hing W., Orr R., & Milne N. (2017). Do coursework summative assessments predict clinical performance? A systematic review. *BMJ Medical Education, 77*(1), 1657–1659.

Reference Source

Torres Alpizar J., Villalva J., & Martínez E. (2013). Proposal for a new instrument to evaluate the impact of flipped learning: A pilot study. *Medical Education and Practice, 1*, 415–425.

Reference Source

Wenger E. (1998). Communities of practice: Learning, meaning, and identity. *Cambridge Press.*

Reference Source

Yale School of Medicine. (2017). *Flipped Classroom*. Retrieved from Reference Source
Open Peer Review

Migrated Content

Version 1

Reviewer Report 15 June 2017

https://doi.org/10.21956/mep.19262.r27987

© 2017 ARUMUGAM B. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

BALAJI ARUMUGAM
TAGORE MEDICAL COLLEGE AND HOSPITAL

This review has been migrated. The reviewer awarded 3 stars out of 5

The author has made a great initiative in publishing a systematic review in an appropriate way on assessing the effectiveness of a newer and growing teaching learning method (FC). Although many institutes utilize or adapt different TL methods, generally the effectiveness or the outcome was not clearly expressed in individual studies. Or most of the time the investigator will find out that the newer TL method has always given a better result than the traditional methods. But adapting these newer methods in regular curriculum is the challenge. “The overall evidence does not clearly demonstrate any benefit from widespread use of FC in medical education”, which questions the utilization of newer TL methods.

Competing Interests: No conflicts of interest were disclosed.

Reviewer Report 13 June 2017

https://doi.org/10.21956/mep.19262.r27988

© 2017 Gibbs T. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Trevor Gibbs
AMEE

This review has been migrated. The reviewer awarded 2 stars out of 5

There is little doubt that we need to constantly evaluate the real effects of new teaching techniques upon
the students; hence this paper which was based upon a literature review covered an important area. However, I would propose that this paper addresses several issues that we still are unsure about. The flipped classroom is and was designed as a teaching methodology that moved the students from a passive learning environment to an active one. It highlighted and attempted to solve the problem of addressing key problems around didactic learning such as poor cognition from the students, a superficial learning approach whereby students were not given enough time to ask specific questions to improve their learning and an attempt to promote deeper learning by improving relevance and integration through application of new knowledge. I am not sure that the pre-learning only involved online learning—there are many ways of delivering pre-learning without using online technology. Is it enough that we use only quantitative methods to evaluate this new learning when indeed the basis of the flipped classroom is about assisting the student in their new knowledge and trying to improve their internal motivation? Should we really be basing our evaluation on level 2 Kirkpatrick when we are really hoping that the new teaching method improves longer-term retention of knowledge and application of that new knowledge? I was also worried somewhat about the search methodology—why exclude qualitative papers—why are we not searching for papers relating to higher-order cognition—why only medical education when there are so many papers delivered from other health professionals—why exclude/ not include the grey literature, particularly since this is a relatively new subject? I think that this paper, whilst useful, certainly leaves the reader with some doubts.

**Competing Interests:** No conflicts of interest were disclosed.

Reviewer Report 13 June 2017

https://doi.org/10.21956/mep.19262.r27989

© 2017 Shankar P. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

P Ravi Shankar
American International Medical University

This review has been migrated. The reviewer awarded 4 stars out of 5

Thank you for the opportunity to review this interesting manuscript. Flipped classrooms (FC) are becoming increasingly common in education and are in consonance with modern concepts of education. Obtaining objective evidence of efficacy in educational interventions is more challenging due to a variety of reasons. In the Materials and Methods section, end of paragraph 1: Can the authors clarify the sentence ‘MeSH terms were exploded to identify additional studies.’ The authors have clearly defined the criteria used during their systematic review. There may be problems in using retrospective data to compare flipped classrooms and didactic teaching methods as the two groups of students may have different demographic and other characteristics. The authors have described the strengths and
limitations of their study in detail. As medical educators we may have to discuss the importance of
different levels (hierarchy) of knowledge, skills and attitudes in medical education. The authors have
mentioned that FC is more effective in developing higher order cognitive skills. Accessing knowledge has
become easier with the advent of the internet and other technologies and critical analysis and judgment
skills are becoming increasingly important. This has influenced the use of different learning methods in
medical education. The tables and the figures are well designed and supplement the information
provided in the text. As mentioned by the authors future studies may be able to provide a more definitive
answer with regard to the influence of the flipped classroom on exam performance.

**Competing Interests:** No conflicts of interest were disclosed.