Comparison of flipped learning and traditional lecture method for teaching Digestive System Diseases in undergraduate medicine: a prospective non-randomized controlled trial

[1+]Manuel Hernández-Guerra, [1+]Enrique Quintero, [1]Dalia Elena Morales-Arráez, [2]Angel Carrillo-Pallarés, [1]David Nicolás-Pérez, [1]Marta Carrillo-Palau [1]Antonio Gimeno-García, [1]Carlos González-Alayón, [1]Onofre Alarcón, [1]Elena Otón-Nieto [2]Hermógenes Díaz-Luis, [2]Norberto Hernández-Siverio, [2]Antonio Martín-Malagón, [2]Ivan Arteaga-González, [2]Alberto Bravo-Gutiérrez, [2]Maria-Nieves Lorenzo-Rocha [2]Julio Jordán-Balanza, [2]Juan Manuel Sánchez-González, [2]Manuel Barrera-Gómez, [3]Ashley Reid, [3]Nephtali Marina

1Departmento de Medicina Interna, Instituto Universitario de Tecnologías Biomédicas (ITB) & Centro de Investigación Biomédica de Canarias (CIBICAN), Universidad de La Laguna, 38320 La Laguna Tenerife. [2]Departmento de Cirugía, Universidad de La Laguna, 38320 La Laguna Tenerife, Spain. [3]CL Arena Centre for Research-Based Education, University College London. Division of Medicine, [4]U University College London, London WC1E 6BT, UK.
.

[+]Professors Hernández-Guerra and Enrique Quintero contributed equally to this article

Running Title: Flipped learning in undergraduate medical education

*** Correspondence:**
[+]Enrique Quintero equinter@ull.edu.es
https://orcid.org/0000-0002-7244-8125

Keywords: Flipped Learning, Traditional Learning, Absenteeism, Class attendance, Peer Instruction, Student engagement.

Abstract

Introduction: This study examined the effects of a large-scale flipped learning (FL) approach in an undergraduate course of Digestive System Diseases.

Methods: This prospective non-randomized trial recruited 404 students over three academic years. In 2016, the course was taught entirely in a Traditional Lecture (TL) style, in 2017 half of the course (Medical topics) was replaced by FL while the remaining half (Surgical topics) was taught by TL and in 2018, the whole course was taught entirely by FL. Academic performance, class attendance and student's satisfaction surveys were compared between cohorts.

Results: Test scores were higher in the FL module (Medical) than in the TL module (Surgical) in the 2017 cohort but were not different when both components were taught entirely by TL (2016) or by FL (2018). Also, FL increased the probability of reaching superior grades (scores >7.0) and improved class attendance and students' satisfaction.

Conclusion: The holistic FL model is more effective for teaching undergraduate clinical gastroenterology compared to traditional teaching methods and has a positive impact on classroom attendances.

Keywords: Flipped learning; absenteeism; class attendance; peer instruction; student engagement; traditional learning.

Practice points

In the vast majority of Spanish Medical Schools, undergraduate medical education is based on large master lectures (Traditional Learning).

In health care education, controlled trials have shown that Traditional Learning is associated with considerable absenteeism and suboptimal student performance when compared to active learning instruction.

Flipping educational contents with pre-class recorded lectures and transforming face-to-face meetings into problem-based discussions have shown to improve academic outcomes in pre-clinical courses of the medical curricula, but scarce evidence exists on the clinical clerkship period.

We demonstrated for the first time that a holistic Flipped Learning instruction is feasible in a large course of medical and surgical education of Digestive System Diseases, improving academic outcomes, class attendance and student´s satisfaction compared to Traditional Learning.

Introduction

Flipped learning (FL) is a strategy in which the traditional concept of classroom-based learning is inverted: students work on learning material before class and face-to-face sessions are dedicated to apply knowledge and solve problems in small groups (Tune et al. 2013). FL has been proved to encourage students to participate in a more engaging and interactive learning environment (Young et al. 2014) and to improve student performance on examination scores (Hew and Lo 2018). However, its reliance on small student to teacher ratios in a time of rising student populations and worrying teacher shortfalls and budget constraints, makes FL an impractical educational option for most institutions. Thus, we adopted an interactive instructional design known as Peer Instruction (PI) which supports active learning in large class sizes and involves students in the classroom through activities that require each student to apply core concepts to solve specific problems (Crouch and Mazur 2001). PI has been shown to have substantial pedagogical benefits in health-related subjects (Rao and DiCarlo 2000; Versteeg et al. 2019). However, controlled studies describing the teaching outcomes of PI on clinically-relevant courses are lacking. In this action research study we implemented a Holistic Flipped Learning (HFL) method based on FL, interactive PI learning and a comprehensive assessment approach which allowed to continuously collect information on students' learning performance from the pre-class learning process and throughout the face-to-face interactive sessions. Therefore, we aimed to analyze whether a HFL improves academic performance, class attendance and student´s satisfaction in an undergraduate course of Medical and Surgical management of Digestive System Diseases.

Methods

Setting and participants

The Digestive System Diseases course is taught in the 4th year of the ULL Medical School curriculum. This module consists of 24 topics that were delivered over the course of two consecutive sessions with the first session approaching medical issues (e.g., peptic ulcer as epidemiology, pathophysiology, clinical features, diagnosis and medical treatment), and the second session addressing the topic from the point of view of a surgeon (e.g., peptic ulcer complications, surgery indications, surgical techniques and prognosis) (Supplementary Table 1). Three cohorts of students were included: Cohort 2016 (n = 133) taught entirely by TL method; Cohort 2017 (n = 141) in which the Medical topics were taught by FL while the Surgical topics were taught by TL, and; Cohort 2018 (n = 130 students) taught entirely by FL (Figure 1(A)).

The HFL model consisted of three elements: (i) flipped teaching, (ii) collaborative in-class sessions with PI and, (iii) motivational formative assessment of online and in-class participation.

The flipped teaching component included PowerPoint© video lectures created in Camtasia© screen-capture (Techsmith Corporation, USA) (Supplementary Table 1). The videos included five multiple-choice questions (MCQ) with immediate feedback that were embedded approximately every five minutes in a preventing skipping mode using Edpuzzle©. Students were asked to watch video-lectures, to read supplementary book chapters and to complete the quizzes in preparation for the face-to-face sessions. Edpuzzle© platform allows teachers to assess the time each student spends in each lesson as the video and fragments can be watched ad libitum. The same teachers taught the topics on the same learning objectives all 3 years. Cohorts 2017 and 2018 undergoing FL instruction received detailed information about the rationale of this methodology and about the course evaluation prior to

the start. Post-course questionnaires were used to evaluate the effectiveness of the HFL approach in the 2017 and 2018 cohorts.

The collaborative in-class sessions employed a modified version of the previously described PI method (Crouch and Mazur 2001). Each face-to-face session had a duration of 60 minutes divided in five rounds of PI (Figure 1(B)) facilitated by two faculty members. Each round started by presentation of a clinical case study followed by a higher-order multiple-choice concept test. Students were first allowed to submit their answers individually via a classroom response system (Wooclap© platform) and were then instructed to discuss their answers with their neighbors for 5 minutes and then to vote for a second time. The instructor then facilitated a classwide discussion where students shared and justified their answers. The cycle was repeated until the question was answered correctly by more than 70% of the students (Figure 1(B)).

For the formative feedback component, students were given their scores at the end of each session to help them assess their own performance throughout the course and to keep them motivated to complete the online activities and participate in the face-to-face PI sessions.

The control groups received TL sessions (60-minute long) for both Medical and Surgical components in 2016 and for the Surgical component in 2017.

Outcomes and evaluation

The main outcome of this study was to compare the academic performance between groups educated with FL or TL, as determined by the score of an MCQ test (50 questions per component) administered at the end of the course. MCQs were prepared according to the simplified classification of Bloom's taxonomy as previously described (Rao and DiCarlo 2000) with ~80% of the questions

designed to test information recall and comprehension (Type I–II) and the remaining 20% designed to test synthesis and evaluation skills (Type III). A minimum of 65% of correct answers were required to pass each component of the course in all cohorts. Scores were graded as A (10–8.5), B (8.4–7), C (6.9–5) and D (<5). In order to ensure validity and reliability between both examination components, the questions were designed and curated by the same module lead. Assessment items were aligned to the learning objectives and approximately two questions were assigned to each topic (the first question assessing the topic from a medical perspective and the second assessing it from the surgical perspective).

The secondary objectives where: (1) to assess the difference of correct answers between the responses obtained individually and after PI discussion; (2) to analyze whether FL improves the student performance in higher-order concept tests; (3) to assess whether FL improves student attendance to face-to-face meetings (class attendance to FL and to TL sessions were monitored using the Wooclap© platform or by random checklist registers, respectively); and (4) to compare the student´s satisfaction to both pedagogical methodologies by an anonymous and voluntary self-reported survey.

Study design

To compare the academic performance between the FL and TL models, we applied a 'within-subjects' and 'between-subjects' design in a prospective controlled non-randomized study, following the framework of the previously described technology-enhanced learning evaluation in medical evaluation (Cook and Ellaway 2015).

‘Within-Subjects’ design

The 2017 cohort gave us a unique opportunity to compare the student performance of both interventions (FL vs TL) in the same cohort of students that served as their own control over the same period. Similarly, cohort 2018, in which both components received FL instruction served to analyze the PI process in the full course (Figure 1(A)).

‘Between-Subjects’ design

The percentage of students achieving grades A or B (score ≥7), C (score 5–6.9) and D (score <5) was compared between the three cohorts: (1) The Medical and Surgical components were assessed separately for each grade category according to the pedagogical method applied in each cohort; and (2) Grades scores from the final MCQ test were compared between FL or TL instructions, including all students registered in the three academic courses.

Student satisfaction survey

An online survey was administered at the end of 2017 and 2018 academic terms (Supplementary Table 2). Responses by students with FL and TL modalities in the three cohorts were compared. Items were related with student perception of potential gains in knowledge, encouraging potential, impact on clinical practices, academic performance and professional future and overall satisfaction regarding FL and TL modalities. Students were asked to subjectively compare their experience with both pedagogical modalities using a 5-point Likert scale (Responses, 1=’totally disagree’, 2=‘disagree’, 3 = ‘neutral’, 4= ‘agree’, 5= ‘totally agree’) (McLaughlin et al. 2014). Additionally, students could suggest changes in an open question.

Statistical analysis

We determined the effect of the FL and TL interventions on three rubrics: academic achievement, class attendance and students' satisfaction. Results for quantitative variables are expressed as

mean ± Standard Deviation and the median, whereas qualitative variables were assessed by the chi-squared test. ‘Within-subjects’ comparisons were analyzed by paired t-test and the chi-squared test as required. The medians corresponding to the number of correct answers and exam scores were compared using the Mann–Whitney U-test. ‘Between-subjects’ comparisons were assessed by the unpaired t-test or chi-squared test as required. The effect of PI sessions was assessed by the paired t-test. Binomial logistic regression was performed to calculate the probability of reaching superior grades (A or B) or failing (D) in students that followed FL instruction. To assess the effect of FL on the comprehension and resolution of higher-order MCQ, we grouped type I and type II MCQ of the Bloom’s taxonomy classification as Type 1 and Type III questions as Type 2. To assess student´s satisfaction quantitative analysis of Likert responses was performed. p-values <0.05 were considered statistically significant. We used SPSS version 25.0 for the statistical analyses.

Results

Students on FL spent in average 31.7 minutes on each video for a total of 24.8 hours on 47 videos.

Academic achievements

‘Within-Subjects’ analysis

In the 2017 cohort, the mean number of correct answers and mean exam scores were significantly higher in the FL module compared to the TL module (Table 1). Similarly, the corresponding medians were also significantly higher in the FL module versus the TL module. In contrast, there was no difference in means or median scores when both components of the module were delivered entirely by either TL (2016) or FL (2018) (Table 1).

Regarding the PI analysis (2018 cohort), answers were successfully recorded for paired analysis in 187 (79.6%) out of 235 case challenges, with a mean attendance of 127 ± 6.7 (97.6%) and 125 ± 9.0 (96.1%) students in the Medical versus Surgical components, respectively ($p = 0.29$). There was a significant increase (16.8–25.6%) of correct answers after discussion with classmates in the five case challenges assessed over the course (Figure 2 and Supplementary Table 3).

‘Between-subjects’ analysis

At the final examination, the number of correct answers and the mean exam scores were significantly higher in the FL cohort (Table 2). The median examination scores improved by 10% with FL compared to TL.

Regarding grading scores, students receiving FL sessions had better superior grades (A or B) and less fail grades (D) than those taught by TL (Figure 3). In the Medical component, the proportion of grades A or B scores were significantly higher at the 2017 and 2018 cohorts instructed by FL

compared to the 2016 cohort instructed by TL (Figure 3(A)). The proportion of students who failed the exam was significantly lower in the 2018 cohort (FL) versus de 2016 cohort (TL) with a trend that did not reach significance between cohorts 2017 (FL) and 2016 (TL). In the Surgical component, there was a trend for higher rates of grades (A or B) in the 2018 cohort (FL) as compared to the 2016 (TL) and 2017 (TL) cohorts. The percentage of students who failed the final examination was significantly lower in the 2018 cohort (FL) compared to the 2016 cohort (TL) (Figure 3(B)).

Taken into account the three courses together, students following FL instruction had significantly higher rates of superior grades (A or B) than those following TL (30.4% vs 18.7%, $p < 0.001$). Students undergoing FL instruction also had significantly less failing grades (D) compared with those receiving TL (16.0% vs 23.3%, $p < 0.01$) and the probability of reaching superior grades scores (A or B) was about twice higher with the FL module versus the TL module (OR 1.9, 95% CI 1.37–2.64, $p < 0.001$). In addition, students undergoing FL also had less probability to fail the final exam (OR 0.62, 95% CI 0.43–0.88, $p < 0.001$) (Figure 3(C)).

Regarding the categorization of the final MCQ examination by their level of complexity or specificity, the cohorts contained similar numbers of type 1 and type 2 questions, although the Surgical component had lower number of high-intellectual questions (Supplementary Table 4). As shown in Figure 4(A), FL significantly increased the rate of type 2 higher-intellectual MCC compared to TL (66.7 ± 1.0 vs. 73.4 ± 0.7 respectively; $p < 0.001$) whereas no differences were found in type 1 questions (71.6 ± 0.6 vs. 72.0 ± 0.6 respectively; $p = 0.88$). The impact of FL on high-intellectual questions was more pronounced in the surgical component compared to the medical component ($p < 0.01$; Figure 4(B)).

Class attendance

Prior to the implementation of FL (2016 cohort), class attendance to both components of the module (TL) showed the same negative trend as in previous years, with a 50% attendance rate in the first session and a progressive decline in the number of students attending the subsequent sessions (Supplementary Figure 1A). In contrast, the median of attendance for FL sessions in the 2017 Medical component was significantly higher than that found in TL sessions (Surgical component) ($n = 132$, 94.28% vs $n = 30.5$, 21.6%, $p < 0.001$). Attendance levels in the FL group were kept remarkably constant for the remaining sessions, whilst the TL group showed the same negative trend as in previous years (Supplementary Figure 1B). On the 2018 cohort, were FL instruction was extended to the full subject, the median of class meetings attendance was 125 and 126.5 students for the Medical and Surgical components, respectively (Supplementary Figure 1C).

Student satisfaction

The overall rate of response to the survey was 66.6% (94/141) and 42.3% (55/130) at the 2017 and 2018 cohorts, respectively.

Between-subjects analysis of survey responses revealed that FL modality compared to TL significantly increased student satisfaction as shown in Supplementary Table 2. Students perceived that FL might be more useful to help them achieve a better clinical practice performance (3.1 ± 1.0 in the TL 2017 group vs. 4.0 ± 0.9 in the FL 2018 group, $p < 0.001$). FL was also found to be of benefit for professional practice (3.0 ± 0.9 in the TL 2017 group vs. 4.0 ± 0.7 in the FL 2018 group, $p < 0.001$) and to encourage more in-class participation (2.5 ± 1.0 in the TL 2017 group vs. 3.7 ± 1.0 in the FL 2018 group, $p < 0.001$). Regarding the academic performance, students found FL a better modality to achieve better scores (2.8 ± 1.0 in the FL 2017 group vs. 4.2 ± 0.8 in the FL 2018 group,

$p < 0.001$). Also, FL was perceived as a better method of assimilating the content (3.1 ± 0.9 in the TL 2017 group vs. 4.2 ± 0.8 in the FL 2018 group, $p < 0.001$).

Within-group analysis showed differences between FL and TL in 2017 cohort reinforcing the perceived excellent satisfaction by students with FL. Between-subjects analysis of FL in 2017 and 2018 courses showed no differences suggesting that the performance of the modality was similar in every course.

Discussion

The current study showed that the implementation of FL is an affordable and effective way of teaching Digestive System Diseases in large classes. Despite our strategy showed to impact in the MCQ test differently depending on the component taught, FL enhanced performance in high-intellectual questions which was more pronounced in the Surgical component compared to the Medical component. As a whole our HFL approach markedly improved student academic outcomes as shown by better final exam grades with significantly higher median scores, higher rates of superior grades (A or B) and less rate of fail grades (D) compared with the TL method. In addition, FL improved class attendance and student´s satisfaction versus TL.

So far, there is no robust evidence comparing learning outcomes of FL versus TL in the undergraduate education of Medical or Surgical subspecialties. The effects of FL and TL in different health professions and disciplines were assessed in a recent meta-analysis that compared 28 studies of which, only 8 were controlled trials and only 7 were performed on medical students without involving internal medicine or general surgery subspecialties (Hew and Lo 2018). The main conclusion was that FL significantly improved learning performance compared to TL, particularly when quizzes were incorporated in the face-to-face class meetings. Our study evaluated both Medical and Surgical components and included quizzes in the pre-class video-lectures and in the in-class meetings, is consistent line with these findings.

The implementation of PI in the current study was associated with a 16.8% to 25.6% improvement in correct answers following in-class discussion of case challenges, suggesting that this methodology increased the students' level of understanding and their ability to integrate and synthesize core concepts. In addition, the final MCQ test scores revealed that while FL did not affect the score in levels I-II questions of the Bloom taxonomy (fact recall), it significantly improved the score of high

difficulty (level III) questions (application and analysis). These data are consistent with previous studies that show that active learning promotes the understanding of concepts and improves problem-solving skills and academic performance (Rao and DiCarlo 2000). Interestingly, FL provided enhanced scores at complex MCQ in the Surgical component suggesting that certain topics takes more advantage from this modality.

Class absenteeism is a matter of deep concern in many medical education institutions around the world. Many factors are commonly viewed to contribute to the loss of student interest in voluntary attendance to face-to-face sessions, including the recent explosion in the availability of online medical lectures (Prober and Heath 2012; McLaughlin et al. 2014), the increasing use of tools for lecture video capture (Edwards and Clinton 2019), poor student satisfaction, and what students perceive as a lack of alignment between instructional activities and assessments (Hafeez et al. 2014). Some institutions have tried to address this issue by imposing mandatory attendance policies.

In the current study, replacement of a TL system with a holistic FL approach for continuous summative assessment of student performance produced a significant improvement in class attendance. To our knowledge, experimentally proven studies that successfully address the issue of absenteeism and offer alternatives to tackle this problem are very scarce (Wilder et al. 2001) and although an improvement in class attendance was foreseen with our intervention, the results obtained were beyond expectations. We believe that the holistic approach adopted in this approach greatly contributed to the profound effect on class attendance by providing students the opportunity of assessing continuously their efforts on online and face-to-face quizzes.

The student satisfaction surveys at the end of the 2017 and 2018 cohorts revealed a clear and robust preference for the FL instruction compared to the TL system with-in and between subject's

comparisons. Students found that using this active learning strategy encouraged them to become autonomous and independent learners and helped them take responsibility of their own learning. In addition, PI sessions were considered a good way to more deeply understand concepts and learn transferable skills. These results were in line with previous studies showing that students mostly liked the implementation of active learning methodology (Chen et al. 2017; Hew and Lo 2018).

The strengths of the current study were: first, it assessed for the first time the effect of FL on the teaching process of undergraduate Digestive System Diseases, allowing to compare under a rigorous design its feasibility on the Medical and Surgical components of the subject; second, the effect of FL versus TL on student's academic performance was compared along consecutive academic years including large size classes; third, the fact that in the 2017 cohort 50% of topics (Medical component) was taught using FL and 50% (Surgical component) using TL, gave us a unique opportunity to compare both pedagogical methods in the same students; fourth, the face-to-face PI sessions implemented in the full course at the 2018 cohort, allowed to assess its effect on 187 clinical case challenges along 47 sessions with a high student attendance per class; and fifth, teachers were the same along the three academic courses and selection and evaluation criteria for the final MCQ test were homogeneous for the three cohorts.

Our study had several limitations: first, we were not able to quantify the amount of time that students spent revising the material in TL groups; second, an increase in the final MCQ median percentage score of 10% in the FL instruction versus the TL method may seem minor but should be considered robust according to the large number of students involved and the different measurements (within and between groups). In addition, it is relevant in the framework of MCQ evaluation were one point up or down determinates whether students pass or fail the exam; third, the student performance was considered as the immediate score at the final MCQ test. Whether this effect on academic outcomes

is sustained over time cannot be ascertained by our results. However, there are positive experiences in the pre-clinical course of Anatomy (Day 2018) and in residents of Internal Medicine (Graham et al. 2019) and Anesthesiology (Martinelli et al. 2017) showing sustained long-term retention of knowledge and competences following a similar active learning approach; fourth, regarding the improved attendance to face-to-face classes in the FL instruction, it would be possible that students felt forced to attend since it counted in the evaluation of the subject if they passed the final exam. However, this was denied by a majority of students on the specific question embedded at the anonymized satisfaction survey. In addition, the possibility that higher attendance rates in the FL groups were due to a Hawthorne effect was unlikely since the course lasted for a long period of time (4 months) and altered behaviors have been shown to be normalized over time (Becker 2007); fifth, our within-subject comparison as a method to evaluate the effect of FL relative to TL may be biased by a different difficulty. However, topics were the same for medical and surgical approaches and the MCQ test was of similar difficulty to minimize these effects; Finally, the study was performed in a single institution and probably results cannot be generalizable to all Medical Schools.

In conclusion, we demonstrate that a holistic FL approach is feasible to implement in a large class setting during the clerkship of undergraduate Digestive System Diseases course, preventing the detrimental effects of class absenteeism and promoting a favorable student perception. Most importantly, FL improved the academic outcomes of the Medical and Surgical components of the subject over two consecutive terms. However, more research is warranted to determine whether this effect is sustained over time and applicable to other subspecialties in the medical clerkship education.

Acknowledgment

We thank Professor Juan Francisco Hernández Rodriguez (Colegio Hispano Ingles, Santa Cruz de Tenerife) for his inspiration on Flipped Learning and Fundación Bioavance for editorial support.

Disclosure statement

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the article.

Glossary

Flipped Learning: Is a strategy in which the traditional concept of classroom-based learning is inverted: students work on learning material before class and face-to-face sessions are dedicated to apply knowledge and solve problems in small groups.

Peer Instruction: Interactive instructional design known as which supports active learning in large class sizes and involves students in the classroom through activities that require each student to apply core concepts to solve specific problems.

Additional information

Funding

This study was supported by a grant from ULL [PITE 2020/21].

Notes on contributors

Enrique Quintero, MD, PhD, is a Full Professor at the Department of Internal Medicine and coordinator of Digestive System Diseases education in the Medical School at ULL.

Angel Carrillo-Pallarés, MD, PhD, and Norberto Hernández-Siverio are a Full Professors at the Department of Surgery in the Medical School at ULL.

Nephtali Marina, MD, PhD, is Professor at the UCL Division of Medicine, University College London

Manuel Hernández-Guerra, MD, PhD, *Antonio Gimeno-García,* MD, PhD *and Elena Otón-Nieto,* MD, PhD, are Associate Professors at the Department of Internal Medicine in the Medical School at ULL.

Hermógenes Díaz-Luis, MD, PhD, Antonio Martín-Malagón, MD, PhD, Ivan Arteaga-González, MD, PhD, Nieves Lorenzo-Rocha, MD, PhD, Alberto Bravo-Gutierrez MD, PhD, Julio Jordán-Balanza, MD, PhD, and Manuel Barrera-Gómez, MD, PhD, are Associate Professors at the Department of Surgery in the Medical School at ULL.

Dalia Elena Morales-Arráez, MD, PhD, *Marta Carrillo-Palau,* MD, PhD, *David Nicolás-Pérez,* MD, PhD, and *Carlos González-Alayón,* MD, are Clinical Assistant Professors at the Department of Internal Medicine in the Medical School at ULL.

Juan-Manuel Sánchez-González, MD, is Clinical Assistant Professor at the Department of Surgery in the Medical School at ULL.

Ashley Reid, is a student at the UCL Division of Medicine

References

Baernstein A, Oelschlager A-M, Chang TA, Wenrich MD. 2009. Learning professionalism: perspectives of preclinical medical students. Acad Med. 84(5):574–581. [Crossref], [PubMed], [Web of Science ®], [Google Scholar]

Becker HS. 2007. Sociology work: method and substance. New Brunswick (NJ): Transaction Books. [Google Scholar]

Bishop JL, Verleger MA. 2013. The flipped classroom: a survey of the research. ASEE national conference proceedings; Atlanta, GA. [Google Scholar]

Bodagh N, Bloomfield J, Birch P, Ricketts W. 2017. Problem-based learning: a review. Br J Hosp Med (Lond). 78(11):C167–C170. [Crossref], [PubMed], [Web of Science ®], [Google Scholar]

Chen F, Lui AM, Martinelli SM. 2017. A systematic review of the effectiveness of flipped classrooms in medical education. Med Educ. 51(6):585–597. [Crossref], [PubMed], [Web of Science ®], [Google Scholar]

Cook DA, Ellaway RH. 2015. Evaluating technology-enhanced learning: a comprehensive framework. Med Teach. 37(10):961–970. [Taylor & Francis Online], [Web of Science ®], [Google Scholar]

Crouch CH, Mazur E. 2001. Peer instruction: ten years of experience and results. Am J Phys. 69(9):970–977. [Crossref], [Web of Science ®], [Google Scholar]

Day LJ. 2018. A gross anatomy flipped classroom effects performance, retention, and higher-level thinking in lower performing students. Am Assoc Anatomists. 11(6):565–574. [Google Scholar]

Desalegn AA, Berhan A, Berhan Y. 2014. Absenteeism among medical and health science undergraduate students at Hawassa University, Ethiopia. BMC Med Educ. 14:81. [Crossref], [PubMed], [Web of Science ®], [Google Scholar]

Edwards MR, Clinton ME. 2019. A study exploring the impact of lecture capture availability and lecture capture usage on student attendance and attainment. High Educ. 77(3):403–421. [Crossref], [Web of Science ®], [Google Scholar]

Graham KL, Cohen A, Reynolds EE, Huang GC. 2019. Effect of a flipped classroom on knowledge acquisition and retention in an internal medicine residency program. J Grad Med Educ. 11(1):92–97. [Crossref], [PubMed], [Google Scholar]

Hafeez K, Khan M-u-Z, Jawaid M, Haroon S. 2014. Low attendance in lectures at medical colleges of Karachi – a cross sectional survey. J Postgrad Med Inst. 28(2):161–164. [Google Scholar]

Hew KF, Lo CK. 2018. Flipped classroom improves student learning in health professions education: a meta-analysis. BMC Med Educ. 18(1):38. [Crossref], [PubMed], [Google Scholar]

Ludmerer KM. 1999. Instilling professionalism in medical education. JAMA. 282(9):881–882. [Crossref], [PubMed], [Web of Science ®], [Google Scholar]

Martinelli SM, Chen F, DiLorenzo AN, Mayer DC, Fairbanks S, Moran K, Ku C, Mitchell JD, Bowe EA, Royal KD, et al. 2017. Results of a flipped classroom teaching approach in anesthesiology residents. J Grad Med Educ. 9(4):485–490. [Crossref], [PubMed], [Google Scholar]

McLaughlin JE, Roth MT, Glatt DM, Gharkholonarehe N, Davidson CA, Griffin LM, Esserman DA, Mumper RJ. 2014. The flipped classroom: a course redesign to foster learning and engagement in a health professions school. Acad Med. 89(2):236–243. [Crossref], [PubMed], [Web of Science ®], [Google Scholar]

Network FL. 2012. Flipped Learning Network. [accessed 2012]. https://flippedlearning.org/definition-of-flipped-learning/. [Google Scholar]

Prober CG, Heath C. 2012. Lecture halls without lectures – a proposal for medical education. N Engl J Med. 366(18):1657–1659. [Crossref], [PubMed], [Web of Science ®], [Google Scholar]

Rao SP, DiCarlo SE. 2000. Peer instruction improves performance on quizzes. Adv Physiol Educ. 24(1):51–55. [Crossref], [PubMed], [Web of Science ®], [Google Scholar]

Reimschisel T, Herring AL, Huang J, Minor TJ. 2017. A systematic review of the published literature on team-based learning in health professions education. Med Teach. 39(12):1227–1237. [Taylor & Francis Online], [Web of Science ®], [Google Scholar]

Rodgers JR. 2001. A panel-data study of the effect of student attendance on university performance. Austr J Educ. 45(3):284–295. [Crossref], [Web of Science ®], [Google Scholar]

Tune JD, Sturek M, Basile DP. 2013. Flipped classroom model improves graduate student performance in cardiovascular, respiratory, and renal physiology. Adv Physiol Educ. 37(4):316–320. [Crossref], [PubMed], [Web of Science ®], [Google Scholar]

Versteeg M, van Blankenstein FM, Putter H, Steendijk P. 2019. Peer instruction improves comprehension and transfer of physiological concepts: a randomized comparison with self-explanation. Adv in Health Sci Educ. 24(1):151–165. [Crossref], [PubMed], [Web of Science ®], [Google Scholar]

Wilder DA, Flood WA, Stromsnes W. 2001. The use of random extra credit quizzes to increase student attendance. J Instruct Psychol. 28(2):117–117. [Google Scholar]

Young TP, Bailey CJ, Guptill M, Thorp AW, Thomas TL. 2014. The flipped classroom: a modality for mixed asynchronous and synchronous learning in a residency program. West J Emerg Med. 15(7):938–944. [Crossref], [PubMed], [Google Scholar]

Figure Legends

Figure 1. Descriptive study algorithm. A) Description of learning methodologies implemented in the three study cohorts; B) Brief explanation of FL methodology.

Figure 2. Peer instruction results: mean percentage of correct answers before and after discussion with classmates in the five case challenges assessed over the course.

Figure 3. Comparison of grading scores at the final MCQ test between students receiving FL or TL instruction in the Medical (3 A), Surgical (3B) or independently of the components of the subjects along the three academic courses (3 C).

Figure 4. Categorization of MCQ according to Bloom´s taxonomy: A) Average percentage of correct answers according to the modified Bloom's taxonomy in FL and TL method; B) Average percentage (95% CI) of correct answers according to the modified Bloom's taxonomy and pedagogical method in Medical and Surgical components.

Figure 1

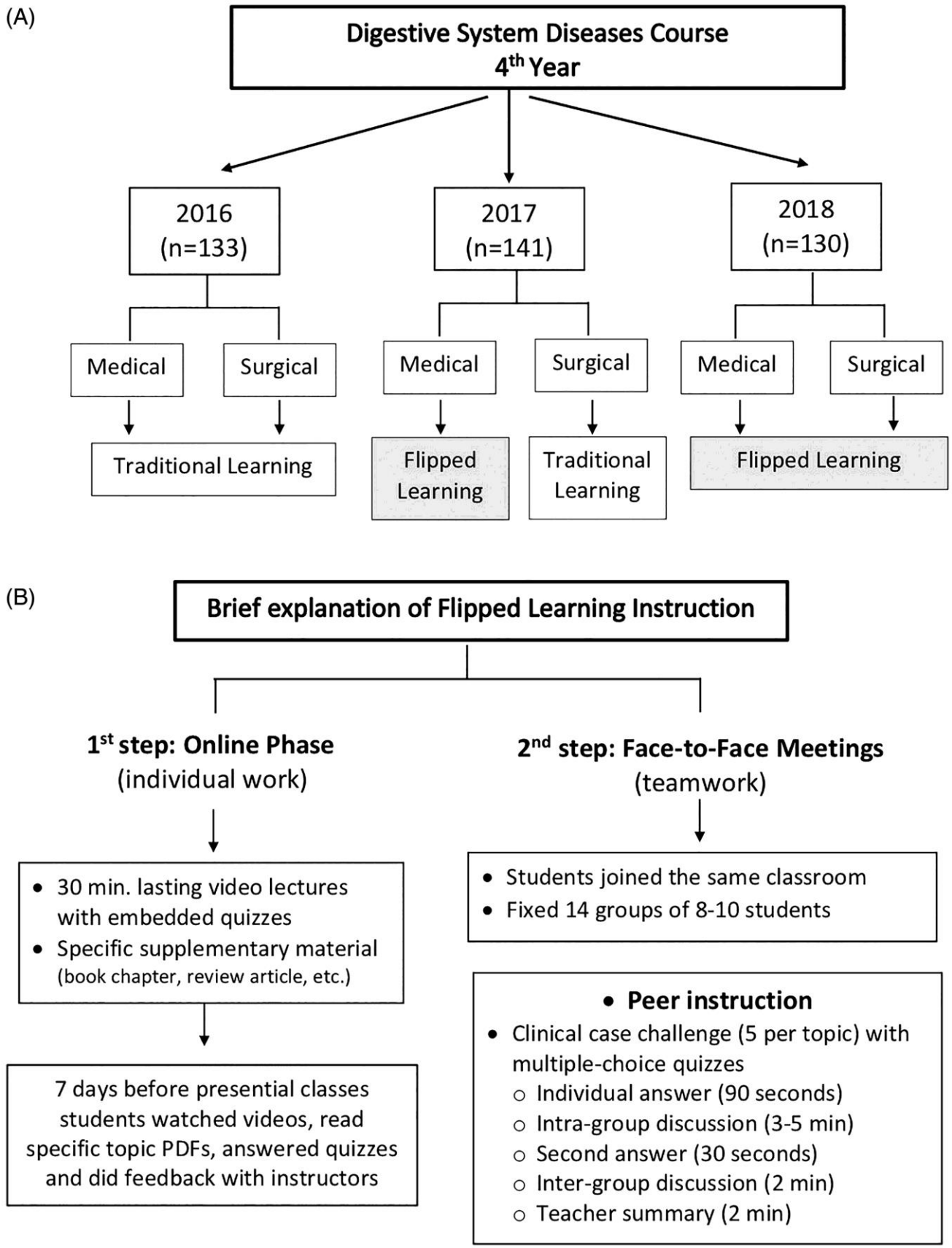

(A)
Digestive System Diseases Course
4th Year
2016
(n=133)
2017
(n=141)
2018
(n=130)
Medical
Surgical
Medical
Surgical
Medical
Surgical
Traditional Learning
Flipped Learning
Traditional Learning
Flipped Learning
(B)
Brief explanation of Flipped Learning Instruction
1st step: Online Phase
(individual work)
2nd step: Face-to-Face Meetings
(teamwork)
• 30 min. lasting video lectures with embedded quizzes
• Specific supplementary material (book chapter, review article, etc.)
• Students joined the same classroom
• Fixed 14 groups of 8-10 students
7 days before presential classes students watched videos, read specific topic PDFs, answered quizzes and did feedback with instructors
• Peer instruction
• Clinical case challenge (5 per topic) with multiple-choice quizzes
○ Individual answer (90 seconds)
○ Intra-group discussion (3-5 min)
○ Second answer (30 seconds)
○ Inter-group discussion (2 min)
○ Teacher summary (2 min)

Figure 2

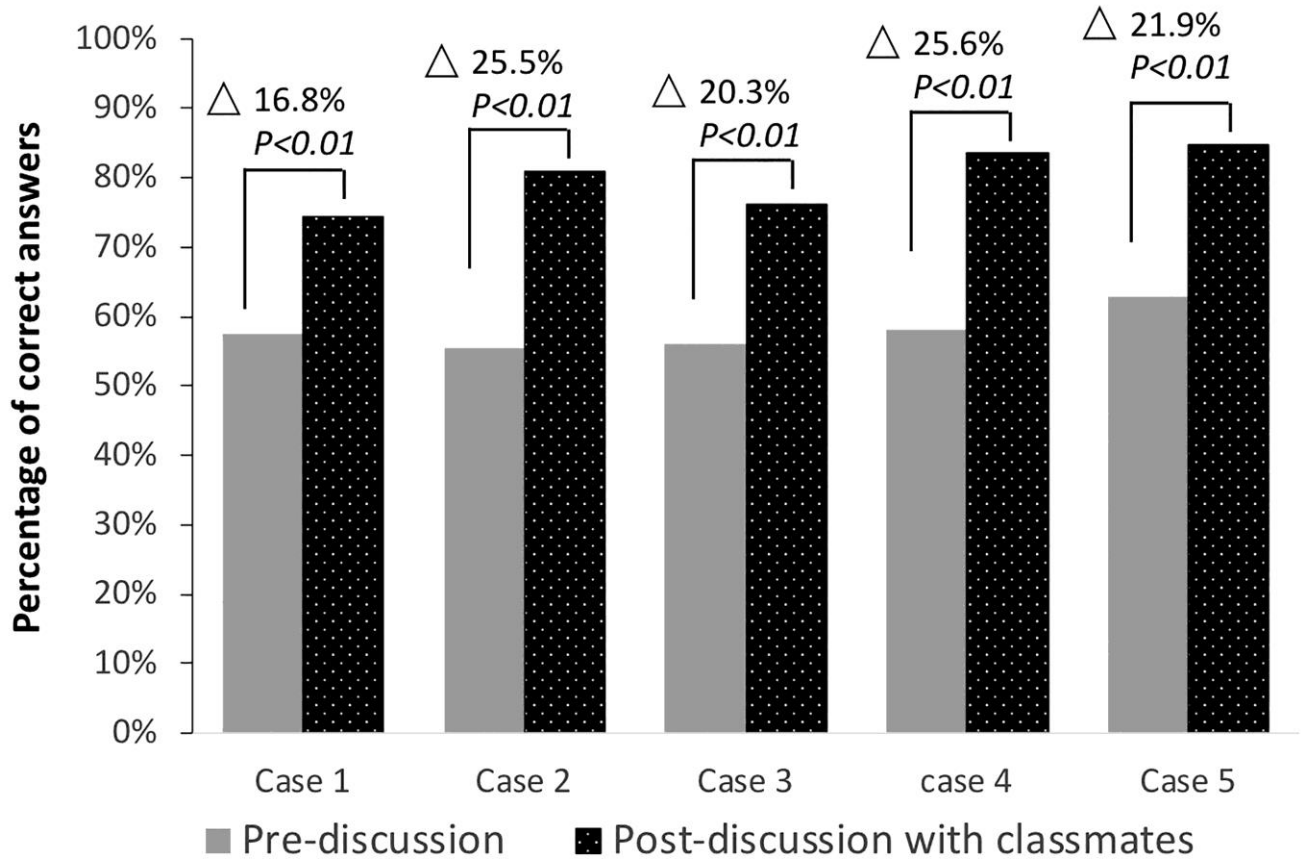

Percentage of correct answers
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
△ 16.8%
P<0.01
△ 25.5%
P<0.01
△ 20.3%
P<0.01
△ 25.6%
P<0.01
△ 21.9%
P<0.01
Case 1
Case 2
Case 3
case 4
Case 5
Pre-discussion
Post-discussion with classmates

Figure 3

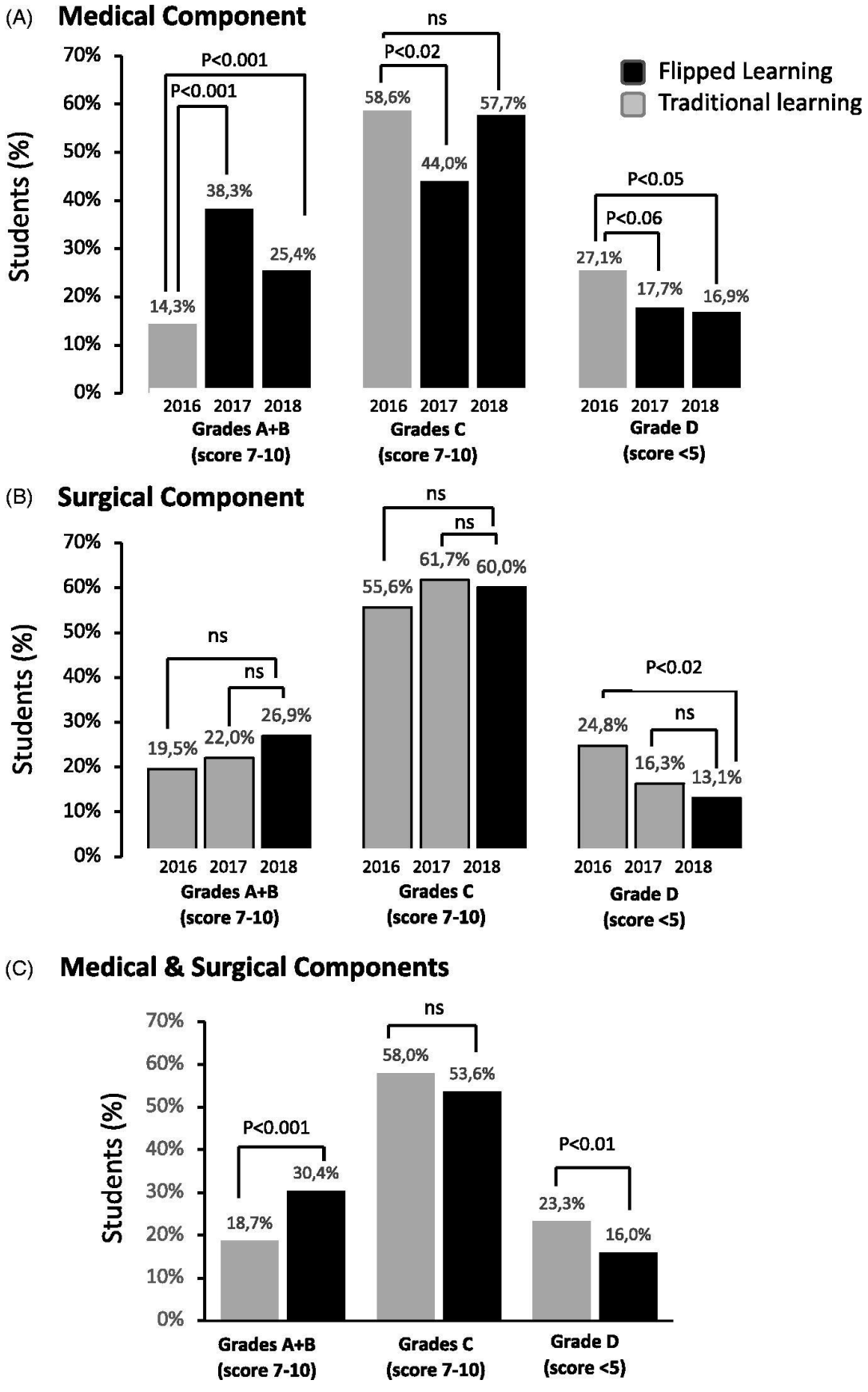

(A) Medical Component
Flipped Learning
Traditional learning
Students (%)
70%
60%
50%
40%
30%
20%
10%
0%
P<0.001
P<0.001
14,3%
38,3%
25,4%
ns
P<0.02
58,6%
44,0%
57,7%
P<0.05
P<0.06
27,1%
17,7%
16,9%
2016 2017 2018
Grades A+B
(score 7-10)
2016 2017 2018
Grades C
(score 7-10)
2016 2017 2018
Grade D
(score <5)
(B) Surgical Component
ns
ns
19,5%
22,0%
26,9%
ns
ns
55,6%
61,7%
60,0%
P<0.02
ns
24,8%
16,3%
13,1%
(C) Medical & Surgical Components
P<0.001
30,4%
18,7%
ns
58,0%
53,6%
P<0.01
23,3%
16,0%
Grades A+B
(score 7-10)
Grades C
(score 7-10)
Grade D
(score <5)

Figure 4

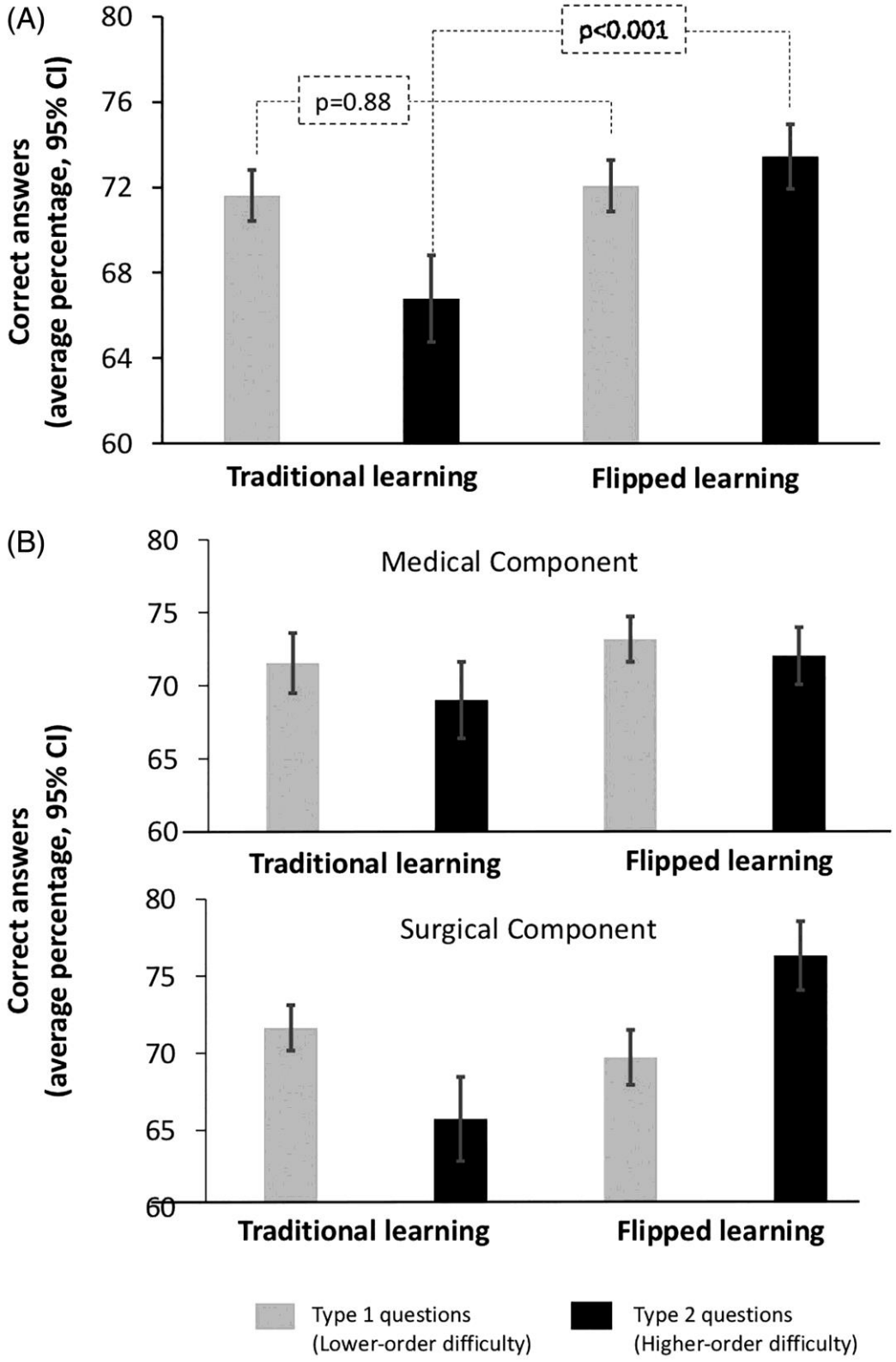

(A)
Correct answers
(average percentage, 95% CI)
80
76
72
68
64
60
p=0.88
p<0.001
Traditional learning
Flipped learning
(B)
Correct answers
(average percentage, 95% CI)
Medical Component
80
75
70
65
60
Traditional learning
Flipped learning
Surgical Component
80
75
70
65
60
Traditional learning
Flipped learning
Type 1 questions
(Lower-order difficulty)
Type 2 questions
(Higher-order difficulty)

Table 1. Multiple-Choice-Quizzes scores in the three cohorts (within-subjects analysis[1])

2016 Cohort (n=133)	**Medical component (Traditional Learning)**		**Surgical component (Traditional Learning)**		***P***
Correct answers		(95% CI[3])		(95% CI)	
Mean ± SD[4]	33.8 ± 5.6	(32.8 - 34.8)	34.5 ± 5.1	(33.6 - 35.3)	0.07
Median	35.0		35.0		0.90
Exam Scores		(95% CI)		(95% CI)	
Mean ± SD	5.7 ± 1.2	(5.4 – 5.9)	5.8 ± 1.2	(5.6 – 6.0)	0.13
Median	5.8		5.8		0.90
2017 Cohort (n=141)	**Medical component (Flipped Learning)**		**Surgical component (Traditional Learning)**		***P***
Correct answers		(95% CI[3])		(95% CI)	
Mean ± SD[4]	36.8 ± 5.7	(35.8 - 37.7)	35.3 ± 6.5	(34.2 - 36.4)	<0.001
Median	38.0		36.0		<0.05
Exam Scores		(95% CI)		(95% CI)	
Mean ± SD	6.4 ± 1.4	(6.2 - 6.6)	6.0 ± 1.4	(5.8 – 6.3)	<0.001
Median	6.7		6.1		<0.05
2018 Cohort (n=130)	**Medical component (Flipped Learning)**		**Surgical component (Flipped Learning)**		***P***
Correct answers		**(95% CI[3])**		**(95% CI)**	
Mean ± SD[4]	35.8 ± 5.8	(34.8 - 36.8)	36.1 ± 4.9	(35.3 - 37.0)	0.45
Median	37.0		36.0		0.97
Exam Scores		(95% CI)		(95% CI)	
Mean ± SD	6.1 ± 1.4	(5.9 – 6.4)	6.2 ± 1.2	(6.0 – 6.4)	0.70
Median	6.4		6.1		0.98

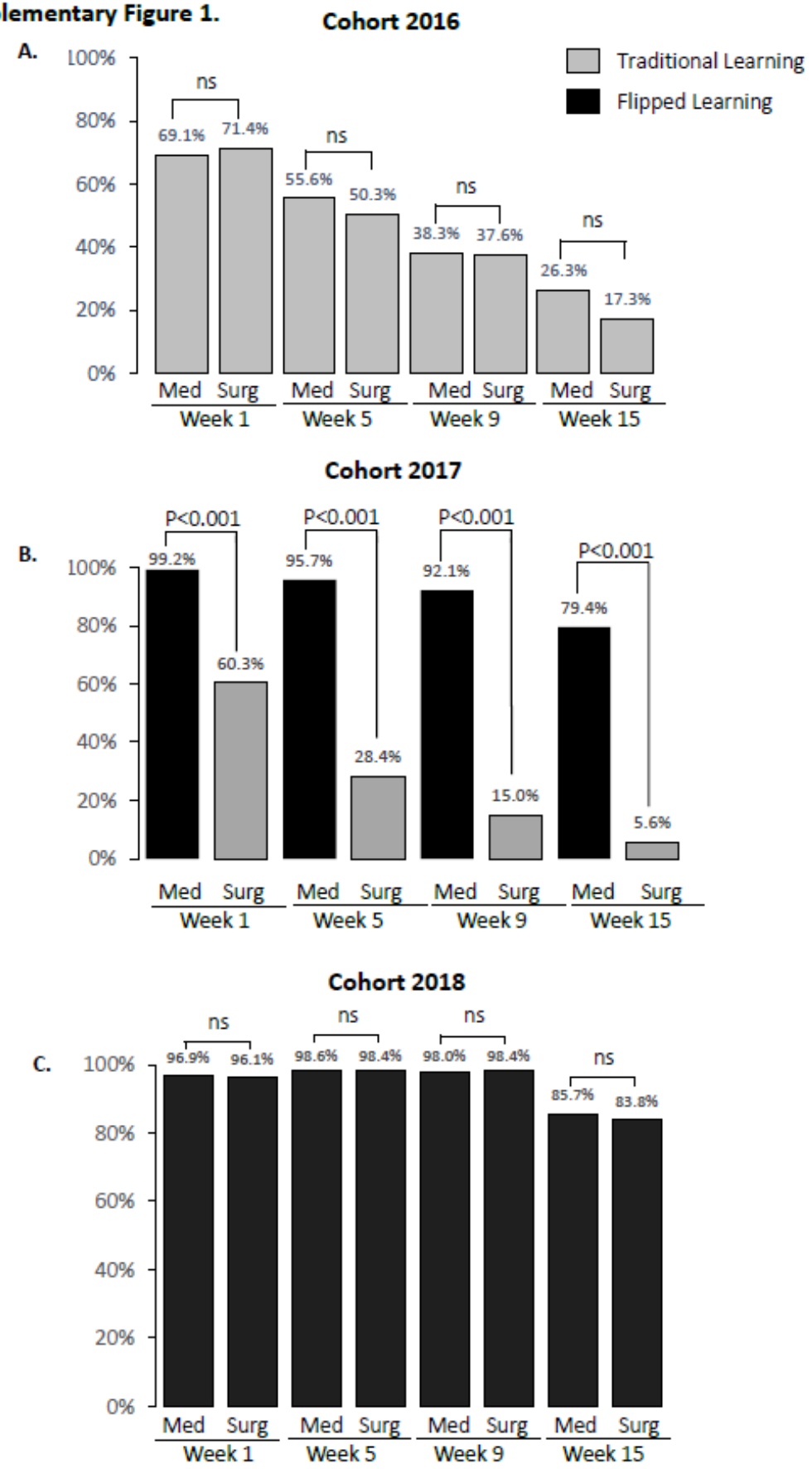
[1] Paired -t-test
[2] MCQ = multiple-choice quizzes
[3] CI = confident intervals
[4] SD= standard deviation

Table 2. Final Multiple-Choice test results (Between-subjects analysis)

2016 - 2018 Cohorts	Traditional Learning (70 Sessions, 404 students)	Flipped Learning (71 Sessions, 404 students)	*P*
Correct answers	(95% CI[1])	(95% CI)	
Mean ± SD[2]	34.5 ± 5.8 (35.1 - 34.0)	36.3 ± 5.5 (36.8 – 35.7	<0.001[3]
Median	35.0	37.0.	<0.001[4]
Exam Scores	(95% CI)	(95% CI)	
Mean ± SD	5.8 ± 1.3 (5.7 – 6.0)	6.2 ± 1.3 (6.1– 6.4)	<0.001[3]
Median	5.8	6.4	<0.001[4]

[1]CI = confident intervals
[2]SD= standard deviation
[3] Unpaired-t-test
[4] Mann-Whitney U test

Supplementary Figure 1.



Supplementary Table 1. Topics and examples of video lectures with medical and surgical approach.

Medical topics	*Medical Approach*	*Surgical Approach*
Cirrhosis and its complications, and other liver diseases	**Liver tumours** https://youtu.be/cNyrzEX_3ys	**Surgery of liver tumours**
	Portal hypertension https://youtu.be/RBOm0Z1efLk	**Hepatic cysts and liver abscess**
	Complications of cirrhosis https://youtu.be/ymXowo5vn70	
	Autoimmune hepatitis	
Diseases of the gallbladder and bile ducts	**Jaundice and cholestasis**	**Biliary colic and cholecystitis**
	Obstructive jaundice	
Diseases of the pancreas	**Acute pancreatitis**	**Surgery of the pancreas**
	Chronic pancreatitis	**Pancreatic tumours**
Metabolic diseases of the liver	**Wilson disease, hemochromatosis and NASH** *https://youtu.be/BvJO_NPFNKA*	
Chronic hepatitis	**Chronic hepatitis** *https://youtu.be/eDtMjp3ktWw*	
Diseases of the esophagus	**Dysphagia and diseases of the esophagus**	**Surgery of diseases of the esophagus**
		Tumours of the esophagus *https://youtu.be/XXleeTRuF88*
	Gastroesophageal reflux disease	**Surgery of hiatal hernia**
Peptic ulcer disease and related disorders	**Peptic ulcer disease and dyspepsia**	**Surgery of peptic ulcer disease**
	Upper gastrointestinal bleeding	**Gastric cancer** *https://youtu.be/PcMnBfQGA7Y*
	Nonsteroidal anti-inflammatory drug gastropathy	

Approach to the patient with gastrointestinal disease	**Diagnostic tests in digestive diseases** *https://youtu.be/pTquZdqaVXs*	
Irritable bowel syndrome	**Irritable bowel syndrome**	
Inflammatory bowel disease	**Ulcerative colitis**	**Surgery of inflammatory bowel disease**
	Crohn disease	
Disorders of absorption	**Chronic diarrhea**	
	Approach to acute diarrhea	
Diverticular disease and common anorectal disorders		**Diverticular disease** *https://youtu.be/RQ5fh4q1ILI*
		Proctology I ***https://youtu.be/03LqXWXZvBY***
		Proctology II ***https://youtu.be/0sje4-Si15Y***
Colorectal cancer	**Sporadic colorectal cancer**	**Colorectal cancer surgery**
	Hereditary colorrectal cancer	**Surgery of advanced polyps and polyposis**
Acute viral hepatitis	**Acute viral hepatitis**	
Mesenteric vascular insufficiency		**Intestinal ischemia** *https://youtu.be/JnOo_K9v-RI*
		Surgery of large and small bowel diseases
		Intestinal vascular syndrome
Acute appendicitis and peritonitis		**Acute appendicitis and peritonitis I**
		Acute appendicitis and peritonitis II
Acute intestinal obstruction		**Acute intestinal obstruction I**
		Acute intestinal obstruction II

Supplementary Table 2. Survey analysis on the FL and TL methods

		2017 course N=94 (Score 1-5)						2018 course N=54 (Score 1-5)						P-value
Items		**1**	**2**	**3**	**4**	**5**	**Mean±SD**	**1**	**2**	**3**	**4**	**5**	**Mean±SD**	
The teaching method employed in the medical sessions stimulated content understanding	**%**	1,1	2,1	4,3	29,8	62,8	4.5±0.7	0,0	1,8	9,1	36,4	50,9	4.3±0.7	.350
	n	1	2	4	28	59		0	1	5	20	28		
The teaching method employed in the surgical sessions stimulated content understanding	**%**	5,3	15,9	45,7	27,6	5,3	3.1±0.9#	0,0	3,6	14,5	36,3	43,6	4.2±0.8	**<0.001**
	n	5	15	43	26	5		0	2	8	20	24		
I think the teaching method implemented in the medical theoretical sessions facilitated the delivery of clinical practicals	**%**	1,0	1,0	24,4	32,9	40,4	4.1±0.8	0,00	7,2	10,9	38,1	41,8	4.1±0.9	.693
	n	1	1	23	31	38		0	4	6	21	23		
I think the teaching method implemented in the surgical theoretical sessions facilitated the delivery of clinical practicals	**%**	8,5	12,7	41,4	28,7	8,5	3.1±1.0#	0,0	9,0	14,5	34,5	40,0	4.0±0.9	**<0.001**
	n	8	12	39	27	8		0	5	8	19	22		
Attendance marks motivated me to attend face-to-face medical sessions	**%**	7,4	13,8	18,0	38,3	19,1	3.4±1.1	9,0	12,7	34,5	18,1	21,8	3.3±1.2	.406
	n	7	13	17	36	18		5	7	19	10	12		
Active participation in the face-to-face medical sessions helped me understand the content	**%**	1,0	4,2	30,8	39,3	24,4	3.8±0.8	1,8	7,2	29,0	30,9	29,0	3.8±1.0	.887
	n	1	4	29	37	23		1	4	16	17	16		
Active participation in the face-to-face surgical sessions helped me understand the content	**%**	24,4	19,1	38,3	15,9	2,1	2.5±1.0#	1,8	9,0	27,2	34,5	25,4	3.7±1.0	**<0.001**
	n	23	18	36	15	2		1	5	15	19	14		
I think the teaching method employed in the medical sessions will be helpful for my future clinical practice	**%**	1,0	4,2	23,4	35,1	35,1	3.9±1.0	1,8	0,0	16,3	52,7	27,2	3.8±1.1	.713
	n	1	4	22	33	33		1	0	9	29	15		
I think the teaching method employed in the surgical sessions will be helpful for my future clinical practice	**%**	10,6	11,7	48,9	23,4	5,3	3.0±0.9	0,0	3,6	20,0	43,6	30,9	4.0±0.7	**<0.001**
	n	10	11	46	22	5		0	2	11	24	17		
	%	3,1	2,1	11,7	29,7	52,1	4.2±0.9	0,0	1,8	9,0	47,2	40,0	4.2±.7	.953

The teaching method employed in the medical sessions helped me to prepare the theoretical exam	n	3	2	11	28	49		0	1	5	26	22		
I think the teaching method employed in the surgical sessions helped me to prepare the theoretical exam	**%**	11,7	21,2	40,4	20,2	6,3	2.8±1.0#	0,0	5,4	9,0	38,1	45,4	4.2±0.8	**<0.001**
	n	11	20	38	19	6		0	3	5	21	25		
The teaching method employed in the medical sessions helped me enjoy the course	**%**	3,1	6,3	18,0	39,3	32,9	3.9±1.0	1,8	7,2	20,0	34,5	32,7	3.9±1.0	.995
	n	3	6	17	37	31		1	4	11	19	18		
The teaching method employed in the surgical sessions helped me enjoy the course	**%**	11,7	17,0	43,6	20,2	7,4	2.9±1.0#	3,6	5,4	18,1	40,0	30,9	3.9±1.0	**<0.001**
	n	11	16	41	19	7		2	3	10	22	17		
Attending face-to-face medical sessions helped me improve my academic performance	**%**	4,2	2,1	13,8	40,4	39,3	4.9±1.0	1,8	7,2	12,7	38,1	36,3	4.0±0.9	.783
	n	4	2	13	38	37		1	4	7	21	20		
Attending face-to-face surgical sessions helped me improve my academic performance	**%**	12,7	18,0	38,3	20,2	9,5	2.9±1.1#	1,8	5,4	14,5	43,6	29,0	3.9±0.9	**<0.001**
	n	12	17	36	19	9		1	3	8	24	16		
I prefer the flipped learning method	**%**	4,2	3,1	10,6	26,6	55,3	4.2±1.0	1,8	1,8	5,4	36,3	52,7	4.3±0.8	.427
	n	4	3	10	25	52		1	1	3	20	29		
I think the whole course should be delivered with the flipped classroom method	**%**	5,3	5,3	11,7	18,0	59,5	4.2±1.1	1,8	3,6	7,2	23,6	61,8	4.4±0.9	.253
	n	5	5	11	17	56		1	2	4	13	34		

# p<0.001 intragroups

Supplementary Table 3. Results of case challenge quizzes in face-to-face Peer Instruction sessions.

Compo-nents	Case challenge 1		Case challenge 2		Case challenge 3		Case challenge 4		Case challenge 5	
Medical	**Q1a** (19 sessions) (%)	**Q1b** (19 sessions) (%)	**Q2a** (23 sessions) (%)	**Q2b** (23 sessions) (%)	**Q3a** (18 sessions) (%)	**Q3b** (18 sessions) (%)	**Q4a** (17 sessions) (%)	**Q4b** (17 sessions) (%)	**Q5a** (15 sessions) (%)	**Q5b** (15 sessions) (%)
Mean$\pm SD$	50,07 ± 20,3	70,13 ± 33,06	49,21 ± 15,65	79,62 ± 23,0	49,95 ± 17,3	74,67 ± 25,0	54,37 ± 18,02	83,13 ± 24,41	53,73 ± 23,8	77,21 ± 24,3
CI 95%	40,25 - 59,90	54,65-85,61	42,44 - 55,98	69,66 – 89,59	41,31 – 58,58	62,21 – 87,12	45,10 – 63,64	70,58 – 95,68	40,54 – 66,92	63,75 – 90,67
Median	46,9	87,3	52,6	86,9	53,4	79,9	49,5	94,2	45,0	86,6
Surgical	**Q1a** (20 sessions)	**Q1b** (20 sessions)	**Q2a** (20 sessions)	**Q2b** (20 sessions)	**Q3a** (19 sessions)	**Q3b** (19 sessions)	**Q4a** (19 sessions)	**Q4b** (19 sessions)	**Q5a** (17 sessions)	**Q5b** (17 sessions)
Mean$\pm SD$	64,69 ± 21,7	86,76 ± 26,2	62,89 ± 19,64	82,59 ± 23,7	61,46 ± 25,1	77,88 ± 23,3	61,38 ± 22,3	84,15 ± 22,5	71,25 ± 17,75	91,66 ± 15,0
CI 95%	54,52 - 74,8	74,47 - 99,06	53,69 – 72,02	71,69 – 93,48	49,34 -73,58	66,62 – 89,14	50,62 – 72,13	73,29 -95,00	62,12 – 80,38	83,94 – 99,39
Median	67,6	98,7	62,0	92,9	60,98	85,4	60,6	95,1	76,3	97,7
Total	**Q1a** (39 sessions)	**Q1b** (39 sessions)	**Q2a** (43 sessions)	**Q2b** (43 sessions)	**Q3a** (37 sessions)	**Q3b** (37 sessions)	**Q4a** (36 sessions)	**Q4b** (36 sessions)	**Q5a** (32 sessions)	**Q5b** (32 sessions)
Mean$\pm SD$	57,57 ± 20,3	74,45 ± 30,6	55,57 ± 18,72	81,00 ± 22,9	56,01 ± 21,9	76,32 ± 23,9	58,07 ± 20,4	83,67 ± 23,0	63,04 ± 22,3	84,89 ± 20,9
CI 95%	50,41 – 64,73	68,64 – 88,25	49,81 – 61,33	73,94 – 88,06	48,81 – 63,21	68,35 – 84,29	51,16 – 64,98	75,85 – 91,48	54,99 – 71,08	77,35 – 92,43
Median	55,43	94,10	56,36	88,55	54,99	84,55	57,81	94,70	70,12	93,72

Q = case challenge quiz
a = average rate of individual responses
b = average rate following intra-groups discussion

Supplementary table 4. Percentage of Type 1 and Type 2 questions between the medical and Surgical components of the subject

	2016 Cohort		2017 cohort		2018 Cohort	
	Medical component (%)	Surgical component (%)	Medical component (%)	Surgical component (%)	Medical component (%)	Surgical component (%)
Type 1[1]	78	88	74	92	78	78
Type 2[2]	22	12	26	8	22	22

[1] Type 1 questions includes Type I and Type II questions of the Bloom´s Taxonomy.

[2] Type 2 questions refers to type III (higher-grade difficulty) questions of the Bloom´s Taxonomy.