

Hybrid and team collaborative test of Medical students: case report

Prova híbrida individual e colaborativa de graduandos em Medicina: relato de experiência

Ana Carolina Aguiar Rocha¹ 

ana.aguiar@fm.usp.br

Chao Lung Wen¹ 

chao@usp.br

Dahir Ramos de Andrade Junior¹ 

dahir.amos@hc.fm.usp.br

Heraldo Possolo Souza¹ 

heraldo.possolo@fm.usp.br

Fernando Abdulkader¹ 

fkader@icb.usp.br

Izabel Cristina Rios¹ 

izabel.rios@fm.usp.br

Paulo Hilario Nascimento Saldiva¹ 

pepino@usp.br

Maria Lucia Bueno Garcia¹ 

maria.garcia@fm.usp.br

ABSTRACT

Introduction: Learning assessment is necessary to student learning. Learning should be proactive, continuous, and encompass communication and teamwork skills, while also reflecting the individual learning of each student. We present the experience of hybrid proactive individual and group-based assessment, applied to the first year of medical school, regarding its feasibility, impact on learning, and student acceptance.

Case report: During 2018, 2019 and 2022, 540 first-year medical students from the Discussão Integrada de Casos I- FMUSP- SP campus, who underwent proactive learning and hybrid assessment individually and in groups. The assessment was based on clinical cases, 25 multiple-choice questions, online, in person at the university, on institutional computers with access only to the tests. The test was first completed individually (2,2 minutes/question) and then in groups (1 minute per question, 7 students/group). Consultation was only allowed among group members during the group test. Immediately after the test, feedback was provided. Scores and a feedback instrument were presented to the students after evaluation tests regarding the learning and evaluation methodologies. Data were analyzed in SPSS23 software. The teamwork test showed engagement among students, with collaborative work and group decision-making skills within the time of the test. Teamwork tests scores were greater and presented more homogeneity than individual tests across all three analyzed years. Regarding satisfaction, students expressed high levels of satisfaction with both individual and group-based tests.

Discussion: The hybrid individual and collaborative group tests methods were feasible, with high levels of satisfaction and advantages in terms of student learning and the development of communications and decision-making skills of clinical true medical cases. Thus, it is a plausible and effective alternative to aid in solidifying the learning process and training to acquire teamwork and leadership skills, respect for peers, empathy, and problem-solving abilities in a short period.

Conclusion: It is, therefore, a method that integrates the undergraduate curriculum, enhancing the individual's potential for society while meeting the National Curriculum Guidelines.

Keywords: assessment, medical education, proactive learning, TBL.

RESUMO

Introdução: A avaliação é necessária para o aprendizado do aluno. O aprendizado deve ser proativo e contínuo, e abranger técnicas comunicacionais e de trabalho em equipe, porém é fundamental que reflita também tudo o que o aluno aprendeu. Apresentamos a experiência de uma avaliação híbrida que foi aplicada, de forma individual e em grupo, em turmas do primeiro ano de Medicina. Pretendia-se avaliar a exequibilidade dessa técnica, o impacto no aprendizado e a aceitação dos alunos.

Relato de experiência: Durante os anos de 2018, 2019, 2022, 540 alunos do primeiro ano da disciplina Discussão Integrada de Casos I da Faculdade de Medicina da Universidade de São Paulo (FMUSP), campus São Paulo, foram submetidos à aprendizagem proativa e a uma avaliação híbrida – de forma individual e em grupo – em cinco módulos/ano. Realizada presencialmente na faculdade, a avaliação, com base em casos clínicos, continha 25 perguntas fechadas de múltipla escolha. Para tanto, utilizaram-se os computadores institucionais com acesso apenas à prova. Primeiramente realizou-se a avaliação de cada aluno (2,2 minutos por questão) e, após, em grupo (sete alunos/grupo, um minuto por questão). Na prova em grupo, os poderiam conservar entre si, mas sem consulta a qualquer outro material. Em seguida, fez-se a devolutiva da prova. As notas e um instrumento de devolutiva dos alunos sobre o módulo e as provas foram analisados pelo software SPSS23. Constatamos adesão à prova em grupo pelos alunos, com trabalho colaborativo e capacidade de deliberação em grupo no tempo estipulado para a prova. Houve aumento das notas em grupo quando comparadas às individuais nos três anos analisados. As notas de grupos evidenciaram maior homogeneidade. Os alunos se mostraram satisfeitos com as provas individuais e em grupo.

Discussão: O método aplicado mostrou ser exequível, com altos níveis de satisfação e vantagens quanto ao aprendizado dos alunos com desenvolvimento de habilidades de gestão e deliberação sobre problemas em equipe. Dessa forma, mostrou-se uma alternativa plausível e eficaz para mediar o aprendizado colaborativo e um treinamento para adquirir habilidades de trabalho em equipe e liderança, respeito ao colega, empatia e resolutividade de problemas em curto espaço de tempo.

Conclusão: Trata-se, portanto, de um método que contempla o currículo de graduação que desenvolve o potencial do indivíduo para a sociedade e satisfaz as Diretrizes Curriculares Nacionais.

Palavras-chave: Avaliação; Educação Médica; Aprendizado Proativo; TBL.

¹ Universidade de São Paulo, São Paulo, São Paulo, Brazil.

Chief Editor: Rosiane Viana Zuza Diniz.

Associate Editor: Carlos Collares.

Received on 04/15/25; Accepted on 09/04/25.

Evaluated by double blind review process.

INTRODUCTION

Knowledge is dynamic, changeable and continuous, particularly in areas such as medicine, where it is based on probabilities and with important variations depending on the individual and their sociocultural environment, requiring continuous and perennial learning. Such processes and skills can be trained with good educational techniques, especially those involving proactive, critical, evidence-based and ethical work, both individually and in teams¹. Moreover, it is necessary to build skills in a solid way, aiming to facilitate future professional performance in situations of pressure that require decision-making that, despite being joint ones, must be systematic and fast. These objectives are summarized in the general competencies of the National Curriculum Guidelines (DCN, *Diretrizes Curriculares Nacionais*) of the undergraduate course in Medicine² and are: -health care, decision-making, communication, leadership, administration and management, and continuing education.

Thus, evaluation models capable of not only measuring theoretical knowledge, but also evaluating communication skills, procedural deliberations and social interaction are needed. The models of individual tests allow the evaluation of the individual's knowledge and whether they are able to be approved in view of the course objectives. Collaborative group models allow the formal development of communication skills, training and teamwork deliberations necessary in the labor market, in addition to the exercise of leadership, positive coexistence among students and the construction of other important competencies in the student's education process³. However, group evaluation models do not allow the evaluation of the individual knowledge of each student, and a student who is not able to do so may be approved in the discipline due to the greater dedication of their peers. This aspect is an important factor for the application of an assessment, as it must also include the knowledge acquired by the individual.

In medicine, there are reports of experiences of individual tests and group evaluations, both in various formats. In this context, evaluations should not only assess whether the student is able to retain information, but whether they apply it in clinical reasoning, decision-making, and resolution of problem-situations. These skills must be trained systematically so they can be performed safely even in urgent and emergency situations under pressure with limited time. However, there is great heterogeneity in the evaluation techniques and scarcity of information on individual and group hybrid tests. Having greater knowledge about proactive hybrid assessment shared between teachers and students can impact everyone's co-responsibility in learning content and skills, and can affect the student's future performance and

professional development in the job market⁴. Therefore, the search for comprehensive methods of evaluation in student performance and skill building through the evaluation process itself must be documented, systematized, trained and constantly improved.

This experience report presents a proactive didactic tool for hybrid, individual and group assessment, with a simple and accessible methodology, already tested in medical students. It proposes addressing its impact on the learning of 1st-year undergraduate medical students in individual and group hybrid formatting tests and the acceptance, feasibility, sustainability, limits, advantages and disadvantages of this methodology. Thus, it is possible to reflect on the approach, which, in addition to assessing knowledge, collaborates with the construction of some of the general competencies of the DCN², mainly communication, leadership and management among team members, health care, in an integrated way with the theoretical basis acquired with the discussions and individual study.

EXPERIENCE REPORT

This is an observational, retrospective, cross-sectional study with data collected from the Med USP Digital learning platform for the years 2018, 2019, and 2022. The years 2020 and 2021 were not computed due to the pandemic, when the tests were carried out with students at home and not in person at the college.

Ethics

Project approved by CAPPEsq (Ethics Committee for the Analysis of Research Projects), under CAAE number: 73025423.0.0000.0068. The waiver of the ICF was accepted because it was a retrospective case report, without identification of the participants.

Place of study

Med USP Digital learning platform of the discipline MSP4211- Integrated Discussion of Cases I- Curricular Unit 21- FMUSP (Faculty of Medicine, Universidade de São Paulo - state public institution) taught to students of the 1st year of undergraduate medical course.

Population

All 180 students of the 1st year at FMUSP of 2018, 2019 and 2022, enrolled in the discipline MSP4211- DIC1- UC21- FMUSP of the FMUSP medical course, totaling 540 students. Inclusion criteria: All students enrolled and who took the tests in the analyzed period. Exclusion criteria: enrolled students who did not take the tests in the analyzed period.

About the discipline

The discipline Integrated Case Discussion 1 (DIC1), MSP4211- Curricular Unit 21 is supra-departmental, integrative and contextualizes and integrates the concepts of the basic and clinical course in a proactive and critical way, based on real cases, using the TBL (team based learning) method and flipped class approach⁵. It includes discussion of real clinical cases based on the major clinical insufficiencies: respiratory, cardiac, renal, hepatic and metabolic. It also contains a module on digital ethics and a module on Humanization focused on Patient-Centered Medicine. All didactic material is worked on in person, or via the Med USP Digital learning platform, created especially for this discipline by the Telemedicine - FMUSP team.

Formatting of the test application

The test is organized by a multidisciplinary team consisting of teachers and employees of the informatic and learning platform and is based on a clinical case similar to the one studied in each module. It is applied to 90 students at a time divided into classes A and B, two of which are tests with different cases and questions, but with the same degree of difficulty and covering the same subject of the module. Each test consists of approximately 25 closed multiple-choice questions with 5 alternatives and only one correct alternative, performed online through the learning platform, in person at the college on institutional computers, with access only to the tests (internet blocking). It is initially applied individually, without consultations, (lasting around 2.2 minutes/question) and, soon after, the same test is applied in a group (lasting around 1 minute/question, in groups of 7 to 9 students) with consultation only among them, without access to internet material, cell phone, book, notes or any previous study

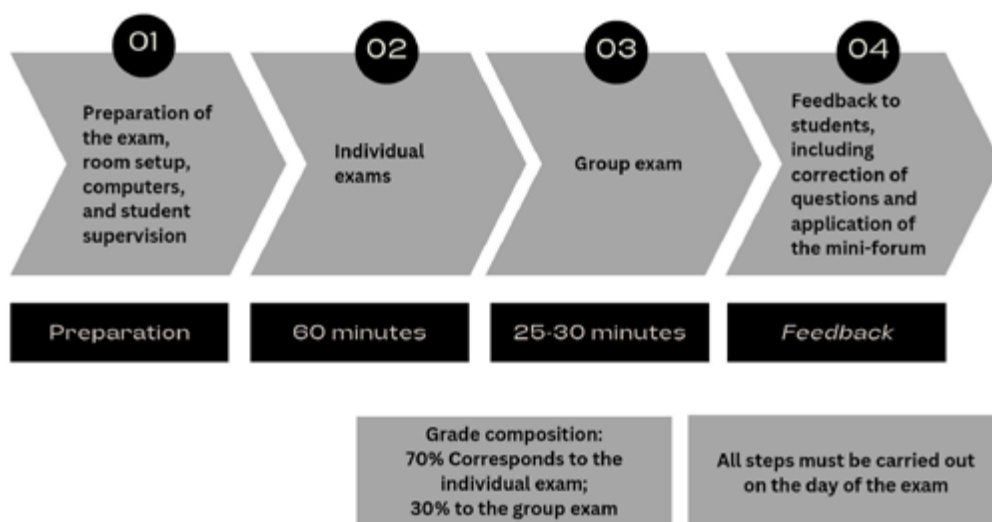
material (Figure 1). Aiming at the group's learning, the test can be submitted several times to the system until 100% of the grade is obtained, but only the first submission of the group test is considered as the students' grade. The group test is mandatory, carried out with a group password; the presence of each individual in the group is checked through a call list. The average of the tests for the composition of students' grades in the discipline is weighted with a weight of 70% for the individual and 30% for the group.

After the group test, it is returned in person by the teachers to the interested students with discussion of doubts about each question, and it is possible, if necessary, to correct it using the answer key (doubts regarding statements or alternatives). After the test is returned, there is a mini forum of the module for the evaluation of the discipline on the learning platform. The mini forum is a tool to evaluate the quality of the discipline in each activity applied, including the test. The answers are graded as Excellent, Good, Fair or Poor. In addition, there is a field for students to put in a cursive way, compliments, doubts, constructive criticism or suggestions about all the course activities. The instrument questions the student about the learning and retention of the content, group work skills, and its formatting.

Institutional structure and human resources for the application of the test

Technically, the test must be delivered by the teachers to the IT team at least 3 days before to be inserted and tested on the platform. The test application requires institutional structure and multidisciplinary human resources from medicine, secretariat, information technology and students to be able to be executed. Each organization team has a checklist of tasks so

Figure 1. Model for conducting Individual Tests and Group Tests: flow of steps aimed at student evaluation.



Source: prepared by the authors.

that everything goes as expected and has alternatives already tested for any major unforeseen events that may happen, such as power or internet outages, or minor ones, such as a computer crashing. The institution needs good internet, computer rooms for the application of the tests, test software with anti-cheat equipment (e.g., determined time of start and end of the test, recognition of the IP in which the test was taken - whether institutional or not - etc.). The platform has to be programmed to release the results of the tests soon after they end, in the form of individual and group numbers, graphs of the distribution of grades, questions with more correct answers and questions with more errors, so that the test can be evaluated and improved by the teachers. The student's attendance is computed twice: once individually by the individual test and once through the attendance list in each group in the group test, as the latter is done through a group password.

Data collection

The following data were collected from the learning platform: individual and group test scores for all modules and data from the mini forum on the course evaluation (scaled data). The collected data were submitted to descriptive analyses. Continuous data were submitted to comparative tests and expressed as median and interquartile data, and categorical data were submitted to associative tests. The statistical program used was SPSS, version 23, and the significance adopted was 5%.

RESULTS

The tests were feasible, applied within the determined time and the students behaved in a dedicated and resolute way, especially in the group tests. The group tests stimulated discussions among the students, especially at the time of voting on the alternative to be chosen as the group's answer. The feedback of the tests was voluntary to the students and obtained an attendance of around 50%. On average, of the 25 questions of each test, there was a change in the answer from 1 to 2 questions per post-feedback test, due to doubts regarding statements or alternatives. The most serious

unforeseen events (power and internet outages) occurred on average 1x per year.

Most of the continuous data were nonparametric.

There was no difference between the means and dispersions between classes A and B in the three years analyzed together (2028+2019+2022: individual test: $p = 0.12$; group test: $p = 0.45$, U-Mann Whitney test). When analyzed separately each year, we also did not obtain a difference between classes, except in 2018, when the group test obtained a lower score in class A than in class B (Individual tests: 2018: $p = 0.85$; 2019 $p = 0.144$; 2022 $p = 0.09$; group test: 2018: $p = 0.01$; 2019 $p = 0.25$; 2022 $p = 0.11$, Kruskal-Wallis tests and T3 Dunnet post-test).

The scores of the group tests were higher than those of the individual tests both each year and when analyzing the three years together ($p<0.001$, U- Kruskal-Wallis test and T3 Dunnet post-test). There was also less dispersion of the mean in the group scores when compared to the individual scores (Table 1, Figure 2).

The percentages of grade gain between the group and individual tests are listed in table 1. This showed an improvement in the score with the discussion and sharing of knowledge among the members, who achieved greater performance and had to apply communication, leadership, time management and basic knowledge skills.

The students' satisfaction with the mini forum was evaluated. The response to the forms varies according to the module, with the students' adherence being higher in the first modules compared to the last. In 2018, 174 students completed the mini forum for module 1, while only 81 completed the forum for module 5. For the other years analyzed, in 2019 the number decreased from 139 to 58 adherents, while in 2022 it went from 164 to 49 students.

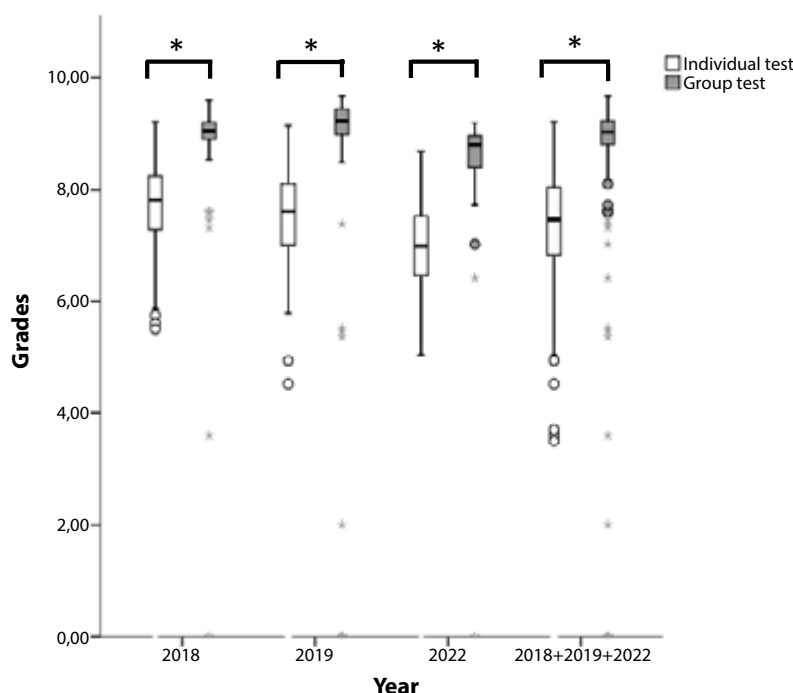
When we analyzed the answers of the first modules, most students were in favor of applying the tests in all years, and greater satisfaction was observed when asked about the group tests. For the years 2018, 2019 and 2022, 74.6%, 91.4% and 87.9% of answers between good and excellent for the individual test and 86.2%, 92.8% and 94.5% answers

Table 1. Individual and group scores in the analyzed period.

	Individual Tests	Group Tests	% of Gain (individual grade/group grade)	P
2018	7.82 (7.28-8.24)	9.04 (8.9-9.2)	13.5%	<0.001
2019	7.6 (7.0-8.1)	9.22 (8.98- 9.44)	15.6%	<0.001
2022	6.98 (6.46-7.52)	8.8 (8.4- 8.96)	20.7%	<0.001
2018+ 2019+ 2022	7.46 (6.81- 8.04)	9.02 (8.8- 9.22)	17.3%	<0.001

Values expressed as medians (interquartile); Kruskal-Wallis test, post T3 Dunnet test; significance 0.05.
Source: prepared by the authors.

Figure 2. Grades of Individual Tests and Group Tests: Data correspond to the same as in Table 1 and correspond to the scores of the set of 5 tests applied in each year.



*: $p < 0.001$, Kruskal-Wallis test and post test T3 Dunnet.
Source: prepared by the authors.

between good and excellent for the group test were obtained in module 1, respectively. Over the years analyzed, there was also an increase in the number of students who considered it as excellent. Despite the smaller sample in the last module, students remained satisfied, with 92.6%, 81% and 90.2% for individual tests and 96.3%, 81% and 96% for group tests.

DISCUSSION

The hybrid individual and group test method applied in the DIC1 discipline proved to be efficient for the evaluative and formative objectives in the teaching environment in which it was implemented.

The acquisition of significant knowledge was observed by the impact of the students' performance in the test. This finding is corroborated by other studies in the literature of group evaluation regarding the acquisition of knowledge in the short and long term, showing to be a positive result of the method⁶. The individual and group tests showed similar levels of difficulty over the years between each other and between the classes, with the exception of the 2018 group test which was more difficult for class A. Most of the tests maintained the level, reflecting the acquisition of similar and comparable knowledge and skills in the period studied.

In addition to this gain in knowledge, the advantage of the hybrid individual and group test method allows students to develop proactive communication and deliberative skills with

criticism and technical grounding, individually and in groups under pressure and with limited time. These results include the DCN, reaching high levels of William's learning pyramid, especially discussing, practicing and teaching colleagues^{4,7}.

The performance of the method is plausible and feasible in a public institution of higher education – and possibly in a private institution as well – with acceptance by students submitted to the method. The easy replicability in other contexts and institutions is particularly supported by the fact that the educational design of the modules consists of elements of the PBL (Problem Based Learning), TBL (Team Based Learning) and flipped class approaches - which are frequent in several medical schools in the country and are "required" in a similar way in the cases presented in the tests^{5,7,8}. These elements are present in the problem-situations discussed during the course activities and addressed in the tests. Similar to what occurs in clinical practice, these situations need to be solved both individually and in groups, including in stressful situations, for which prior training is necessary.

This individual and group approach allowed the students to develop their communicational and deliberative potential in a group with a technical and critical basis that contemplates the DCN, with significant learning. This fact was also observed in other contexts in the group evaluation literature, where the acquisition of this knowledge in the short and long term was evaluated, which was shown to be positive⁶.

When compared with each other, the individual, group, and interclass tests showed similar levels of difficulty over the years, with the exception of the 2018 group test, which was more difficult for class A. These data suggest that, for the most part, the tests maintained the level, mirroring the acquisition of similar and comparable knowledge and skills over the years.

The method is feasible in its formal conceptual structure, and it is important to point out that it requires an institutional structure, technology and trained people. The preparation of the team that applies the test and student discipline for the proper execution of the method can cause anxiety in the team and students. In this sense, the literature mentions the students' feeling of anxiety as a discomfort in some test situations¹, while in other reports it points to a reduction in anxiety when taking the test⁷. We noticed that the anxiety, both of the team and of the students, decreased over time as they acquired experience and greater ability to deal with the unforeseen events. We detected that the main difficulty for the execution of the test was the need for a multidisciplinary team for the organization and application of the test trained to ensure the adequate functioning of the system, and the institution structure, providing material resources for its execution. For this reason, the availability of computers with access to an exam portal can represent an obstacle for institutions with fewer resources.

According to data from the mini evaluation forum of the discipline, students were satisfied with the test model, reporting gains in knowledge, proactivity and criticism, as well as communication skills and problem-solving. These are in line with what is described as the steps of Kirkpatrick's learning training already tested in other models of learning in health and in other areas of knowledge, covering the student's reaction to the proposal (satisfaction and notion of the relevance of the students' methods), understanding (understanding of the problem-situations in the test with group argumentation), behavior (communication and deliberative skills) and results (contemplated in the test scores), and, in this case, aiming at good professional performance in the future in the long term⁹.

The method also respects the 7 principles of teaching-learning of Chickering and Gamson³, reinforcing the subject's commitment to offering short and long-term learning to the student. This commitment is reinforced in relation to the second principle, which considers the development of cooperation among students a fundamental aspect for future medical practice.

CONCLUSION

We can conclude that, despite challenges, the hybrid, individual and collaborative test method is a good method of assessment and learning with positive results in the student's

overall learning. Finally, this method can be tested in other contexts of health education, as well as in areas of exact professional sciences and other human areas, with a view to developing the intellectual and human potential of the student as a proactive and critical professional, as well as technically competent for the safe exercise of their profession in society.

AUTHOR CONTRIBUTIONS

Ana Carolina Aguiar Rocha: Conceptualization, Data Curation, Formal Analysis, Research Methodology, Project Administration, Resources, Software, Validation, Visualization, Writing - original draft, Writing - review and editing. Chao Wen: Data Curation, Writing - Original Draft, Writing - review and Editing. Dahir Ramos de Andrade Junior: Writing - original draft, Writing - review and editing. Heraldo Possolo Souza: Writing - original draft, Writing - review and editing. Fernando Abdulkader: Writing - original draft, Writing - review and editing. Izabel Cristina Rios: Writing - original draft, Writing - review and editing. Paulo Hilario Nascimento Saldiva: Writing - original draft Writing - review and editing. Maria Lucia Bueno Garcia: Conceptualization (Leadership), Data Curation (Leadership), Formal Analysis (Leadership), Investigation (Leadership), Methodology (Leadership), Project Administration (Leadership), Resources (Leadership), Software (Leadership), Supervision (Leadership), Validation (Leadership), Visualization (Leadership), Writing - Original Draft (Leadership), Writing - Review and Editing (Leadership).

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

SOURCES OF FUNDING

The authors declare no sources of funding.

DATA AVAILABILITY STATEMENT

Research data are only available upon request.

REFERÊNCIAS

1. Biesma R, Kennedy MC, Pawlikowska T, Brugha R, Conroy R, Doyle F. Peer assessment to improve medical student's contributions to team-based projects: randomized controlled trial and qualitative follow-up. *BMC Med Educ*. 2019;19(1):371. doi: <https://doi.org/10.1186/s12909-019-1783-8>
2. Brasil. Diretrizes Curriculares Nacionais do Curso de Graduação em Medicina. Brasília: Ministério da Educação; 2014.
3. Chickering AW, Gamson ZF. Seven principles for good practice in undergraduate education. *AAHE Bull*. 1987;3-7.
4. Miller GE. The assessment of clinical skills/competence/performance. *Acad Med*. 1990;65(9 Suppl):S63-7. doi: <https://doi.org/10.1097/00001888-199009000-00045>.
5. Barros EM, Carvalho G, Costa MS, Silva MM. Metodologias ativas no ensino superior. In: *Anais do XV Simpósio de Excelência em Gestão e Tecnologia (SEGeT)*; 2018 out; Resende, RJ, Brasil.

6. Vázquez-García M. Collaborative-group testing improves learning and knowledge retention of human physiology topics in second-year medical students. *Adv Physiol Educ.* 2018;42(2):232-9. doi: <https://doi.org/10.1152/advan.00113.2017>
7. Rezende-Filho FM, Nunes-Souza V, Donato FF, Gouvêa GM. Utilização da aprendizagem baseada em equipes como método de avaliação no curso de Medicina. *Rev Bras Educ Med.* 2019;43(2):85-92. doi: <https://doi.org/10.1590/1981-52712015v43n2RB20180063>.
8. Kapitanoff SH. Collaborative testing: cognitive and interpersonal processes related to enhanced test performance. *Active Learn High Educ.* 2009;10(1):56-70. doi: <https://doi.org/10.1177/1469787408100195>
9. Dorri S, Akbari M, Dorri Sebeh M. Kirkpatrick evaluation model for in-service training on cardiopulmonary resuscitation. *Iran J Nurs Midwifery Res.* 2016;21(5):493-7. doi: <https://doi.org/10.4103/1735-9066.193396>.



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.