

Impact of a Faculty Development Strategy Aiming at Mitigating Erosions in a PBL-Based Medical School

*Livia Maria Moreira, Alexandre Moura, Eliza Brito, Flávia Junqueira, Ligia Ribeiro **

ABSTRACT

Problem-based learning (PBL) is an instructional method that can show erosions and failures throughout its implementation. Faculty development programs reinforcing PBL principles are essential to keep tutorial groups functioning properly. Quasi-experimental study carried out in a medical school, with a PBL curriculum. The institution has launched a faculty development program to improve tutorial performance. The program was based on the dissemination of educational material addressing five perceived erosions in tutorial groups previously identified by the tutors. Students and tutors answered a questionnaire measuring their perception on tutors' performance, before and after faculty development. The overall mean scores of tutors' performance has significantly increased among students when comparing pre- and post-program scores (0.19 + 0.06; $p < 0.001$). The study has shown that, based on students' perspective, a faculty development program focusing on the remediation of erosions identified by the tutors can help improve tutors' performance in different domains.

Keywords: PBL - Problem-based learning, medical education, faculty development

* Livia Maria Moreira, Universidade José do Rosário Vellano, Brazil
Email: liviariapm@hotmail.com
Alexandre Moura, Universidade José do Rosário Vellano, Belo Horizonte, Brazil
Email: alexandresmoura@gmail.com
Eliza Brito, Universidade José do Rosário Vellano, Belo Horizonte, Brazil
Email: elizambrito@gmail.com
Flávia Junqueira, Universidade José do Rosário Vellano, Belo Horizonte, Brazil
Email: flavia.freitas@unifenas.br
Ligia Ribeiro, Universidade José do Rosário Vellano, Belo Horizonte, Brazil
Email: ligiacayres@gmail.com

INTRODUCTION

The implementation of problem-based learning (PBL) at McMaster University in Canada, in 1969, was one of the main innovations in medical education (Bodagh et al., 2017) in the past 50 years and many medical schools worldwide adopted this instructional method since then.

PBL lies on discussing problem-situations or clinical cases in small groups, also known as tutorial groups (Bodagh et al., 2017). Although the way of conducting PBL sessions can vary among different schools, PBL principles are well-defined and must be respected, namely: activation of prior knowledge, cognitive elaboration, information structuring and restructuring, fostering intrinsic motivation and active and cooperative learning (Moust et al., 2005). PBL has been shown to be highly effective, but it is far from perfect (Hung et al., 2019). Erosions and failures can happen (Dolmans et al., 2005) leading to educational deterioration if left unmitigated (Azer et al., 2013). Medical schools must provide continuous training to teachers/tutors, as well as to find strategies capable of ensuring PBL principles (Moust et al., 2005) among faculty to prevent erosions from happening. Faculty development is defined as a wide range of activities used by institutions to help faculty members to improve their work performance. Given the new educational trends in teaching and assessment, most medical schools and educational organizations need to offer programs and activities to help faculty members improve their skills as educators (Steinert et al., 2016). Although many studies focus on describing interventions aimed at faculty development, few of them assess their effectiveness (Hewson et al., 2000; Steinert et al., 2016).

Azer (2005) has listed 12 “tips” for the successful implementation of tutorial groups in a PBL-based course, most of them focusing on what “not to do”: not criticizing; not labeling students; not adopting attitudes that can lead to distortions; not being late; not dominating group discussion; not being an information provider. A Medical School in Brazil, which has a PBL-based curriculum, has developed a faculty development program called “Wise choices in education”. This program was inspired in Azer’s (2005) recommendations for tutorial groups and aimed at strengthening the PBL principles in the institution. A workshop was conducted with tutors to identify the main perceived erosions on tutorial sessions in their current practice. An educational campaign, including educational banners and electronic messages, was then implemented at medical school to publicize good tutorial practices, focused on the workshop results, to students and teachers.

A questionnaire was designed to assess tutors’ performance related to the PBL principles covered in the educational campaign. Students and tutors responded to the questionnaire, before and after the educational campaign.

The main hypothesis of our study was that both students and tutors' perception about tutors' performance would increase after the educational campaign.

MATERIALS AND METHODS

Study design

Quasi-experimental study focused on comparing the perception about tutors' performance among tutors and students, before and after an institutional educational intervention. The study was divided into 5 phases: (1) Identification, by tutors, of perceived erosions in the tutorial groups; (2) Development of a questionnaire based on these erosions; (3) Baseline assessment of the tutors' performance by tutors and students using the questionnaire (4) Educational intervention; (5) Reassessment of tutors' performance after the educational intervention.

The comparison between tutors' performance scores before and after the educational campaign was used to measure its effectiveness.

Scenarios and participants

The study was carried out at a medical school in Brazil, from February to December 2018. The school has a PBL-based curriculum with tutorial groups as the main educational strategy from the 1st to the 8th semesters.

All students and tutors from the 1st to the 8th semesters at this medical school were invited to participate in the study.

Ethical approval for the study was given by University's ethics committee. The study was carried out in accordance with the Declaration of Helsinki.

Materials and procedures

Phase 1 – Initial workshop

Phase 1 was conducted in February 2018. All 57 tutors were invited to participate in the faculty development program called “Wise choices in education” conducted by medical education specialists from the Center for Studies and Development in Medical Education at the same institution. The workshop started with the presentation of a literature review on PBL erosions to remind the tutors of the main PBL educational principles. In the workshop it was highlighted that tutors' actions that were not aligned with the best practices proposed for PBL curricula might impair students' learning. Afterwards, the tutors, in small groups, were asked to identify practices that, despite being contrary to

PBL's principles, occasionally happened at the institution. These practices were called "critical points" to be avoided.

At the end of the session, the critical points identified by the groups were shown to all participants and similar items were merged, resulting in 20 critical points associated with undesirable practices. In the following weeks, all tutors in the school, including the ones who did not attend the workshop, were asked to rank the 5 most relevant critical points from the initial 20 items. The five top ranked items were transformed into "do not" advices: (1) do not skip the activation of prior knowledge, (2) do not allow the mechanical reading of information, (3) do not forget to provide good feedback, (4) do not fear to acknowledge your own knowledge gaps, (5) do not allow the resolution map to be a non-contextualized summary of the entire subject.

Phase 2 – Development of the study questionnaire

A specific questionnaire was developed for the study due to lack of a validated instrument in the literature capable of assessing the critical points identified by the initial workshop. The initial step lied on decomposing each of the 5 critical points into questions in order to assess them based on a five-point Likert scale (1- never; 2- almost never; 3 - intermediate frequency; 4- almost always; 5- always). For example, the item "do not skip the activation of students' prior knowledge" was broken into questions such as "*how often does your tutor encourage 'brainstorming'?*" and "*how often does your tutor provide clues to activate your prior knowledge during the analysis session?*" Next, to ensure that each block of questions could reliably represent the item it was expected to assess, three PBL experts reviewed the questionnaire. These PBL experts were teachers at the institution and members of the Center for Studies and Development in Medical Education that is responsible for faculty development. All of them had more than 10 years of experience in tutoring in a PBL curriculum and in conducting tutor development programs.

Two versions of the questionnaires were developed with few adaptations to the items to make them suitable for both students (e.g., *how often does your tutor provide feedback?*) and tutors (e.g., *how often do you provide feedback?*). Face validity was established by a small group of 8 students and 1 tutor that helped to identify any differences in the interpretation of the items. This process allowed semantic adjustments before the pilot study was carried out with a group of 20 individuals, including students and tutors.

Cronbach's alpha coefficient was calculated in the pilot study to assess the reliability (internal consistency) of the set of questions covering each of the five domains. Items presenting Cronbach's alpha coefficient lower than 0.60 in the pilot test were excluded from the study; thus, the final version of the questionnaire comprised 30 items -

approximately six items for each of the evaluated domains (APPENDICES A and B). Cronbach's alpha coefficient was also calculated for data collected in phase 3.

Phase 3 – Baseline assessment

A baseline questionnaire was applied to students and tutors from the 1st to the 8th semesters of the Medical School, in June 2018. The aim of this phase was to have a baseline assessment of the quality of tutors' performance before the educational campaign, based on students and tutors' viewpoint.

Phase 4 – Educational intervention

The institution released the list with the five most relevant critical points selected by the tutors. The PBL educational campaign – supported by the institution's marketing department - was launched in September 2018, when these five items were disclosed in banners and institutional social networks to all students and tutors of the medical school. Posters and banners emphasizing the importance of each critical point were attached to walls in tutorial classrooms and in the hallways of the institution. For example: a banner said “do not skip the activation of prior knowledge”, which was followed by a brief description of the pedagogical principle that makes it an important PBL point: “*Students tend to think that they do not have relevant prior knowledge to build an initial explanation of the problem. In addition, by omitting an in-depth analysis of the problem based on their prior knowledge, students do not elaborate, which affects the restructuring of current knowledge and the acquisition of new information*”.

In addition, electronic messages, similar to the printed ones, were sent through text messaging Apps to tutors once a week throughout the campaign.

Phase 5 – Post-intervention assessment

Students and tutors from the 1st to the 8th semester answered the same pre-intervention questionnaire 12 weeks after the educational campaign.

Statistical analysis

A comparative analysis of participants' mean perception about tutors' performance before and after the educational campaign was carried out based on repeated-measures ANOVA test. The analysis was stratified by participant category (students vs. tutors) and by course stage (1-4th semester vs. 5th to 8th semester), and was conducted for each one of the evaluated domain. As each educational domain was composed of different items, domain scores were calculated as the average score of such items. All results were considered significant at probability level lower than 5% ($p < 0.05$).

RESULTS

Fifty-six percent (56%) of the 57 tutors participated in the workshop conducted in phase one of this study, 56% of them ranked the top-five critical points in phase 2 and 100% of them participated in phases 3 and 5.

There was a higher proportion of female tutors in phases 3 and 5 - 69.6% and 71.8% respectively, reflecting the composition of the institution's faculty. The median age of the tutors was 42 years (IQR: 36-49 years) in phase 3 and 43 years (IQR: 37-49 years) in phase 5. Median tutoring time of teachers in both phases was 8 semesters (IQR: 3-16 semesters in phase 3; 4-16 semesters) in phase 5.

In total, 564 students participated in the pre-assessment of the educational intervention: 204 students from the 1st to the 4th semester and 260 from the 5th to the 8th semester. On the other hand, 603 students participated in the post-assessment phase of the study: 346 students from the 1st to the 4th and 257 from the 5th to the 8th semester. Student participation represented approximately 88% of the target population (students enrolled from the 1st to the 8th semesters). The proportion of female students was higher in both phases of the study - 64.2%, in phase 3 and 62.7%, in phase 5 - and it reflected the composition of the institution's medical students. The median age of students was 22 years old (IQR: 20-24 years) in phase 3, and 21 years old (IQR: 20-23 years) in phase 5. Students' minimum age was 17 years old and their maximum age was 60 years old.

The internal consistency of the questionnaire was assessed through Cronbach's alpha coefficient based on the total questionnaire and after removal of each individual question. Cronbach's alpha coefficient estimated for the total questionnaire was 0.60 among tutors and 0.77 among students (APPENDIX C – SUPPLEMENTARY MATERIAL).

Global mean scores of tutors' performance pre- and post- intervention were compared to assess its effectiveness. Mean scores post-intervention were higher than those pre-intervention for both tutors (4.24 ± 0.15 vs. 4.15 ± 0.33 ; $p = 0.15$) and students (4.03 ± 0.48 vs. 3.84 ± 0.50 ; $p < 0.001$), although it was statistically significant only among the latter (Table 1).

The analysis stratified by educational domain did not show significant differences between study phases in any of the 5 domains analyzed among tutors (Table 1). The group of students has shown significant difference between study phases in all evaluated domains with significantly higher mean scores post-intervention (Table 1). Effect size (Cohen's d) per domain ranged from 0.2 to 0.5 among students (Table.1).

The global mean pre-intervention scores for tutors' self-perceived performance was higher than that of students (range 1-5, mean= 4.15 ± 0.33 vs. 3.84 ± 0.50 , respectively; p

< 0.001). A separate analyses of each of the 5 domains showed higher mean scores among tutors when compared to students ($p < 0.05$) in each of the domains (APPENDIX C – SUPPLEMENTARY MATERIAL).

Based on the analysis stratified per domain, tutors continued to show higher scores for four of the five evaluated domains after the educational campaign (phase 5). Domain “*do not allow the resolution map to be a non-contextualized summary of the entire subject*” was the only one that did not show a significant difference between students and tutors after the intervention (APPENDIX C – SUPPLEMENTARY MATERIAL).

Post-intervention mean scores for tutors’ performance was higher for tutors than for students (4.24 ± 0.39 vs. 4.03 ± 0.48 , respectively; $p < 0.001$) (APPENDIX C – SUPPLEMENTARY MATERIAL).

Comparison between mean scores for each of the 5 investigated domains of students from the 1st to the 4th ($n = 204$) and students from the 5th to the 8th semester ($n = 260$) in phase 3 showed significant differences in the following domains: “mechanical reading of information”; “feedback”; and “fear to acknowledge own knowledge gaps” - the highest means were recorded for students from 1th to 4th semester (Table 2). Scores from tutors from 1-4th vs. 5-8th semesters did not significantly differ in any of the five domains.

As shown in Table 3, the group of students from 1th to 4th semester ($n = 346$) presented significantly higher post-intervention means for the following domains than students from 5th to 8th semester ($n = 257$): “mechanical reading of information” and “feedback”. Comparing scores from tutors from 1-4th ($n = 39$) vs. 5-8th semesters ($n = 31$) tutors, “feedback” was the only domain with difference in mean scores, with higher scores observed among tutors from 1-4th (Table 4).

Domains	Students' means (standard deviation)				Tutors' means (standard deviation)		
	Before the campaign	After the campaign	p	Cohen' s d	Before the campaign	After the campaign	p
Global	3.84 (0.50)	4.03 (0.48)	< 0.001		4.15 (0.33)	4.24 (0.39)	0.15
Do not skip the activation of prior knowledge	3.92 (0.60)	4.17 (0.59)	< 0.001	0.42	4.35 (0.35)	4.42 (0.43)	0.331
Do not allow the mechanical reading of information	3.67 (0.58)	3.86 (0.60)	< 0.001	0.32	3.87 (0.48)	4.03 (0.47)	0.05
Do not forget to provide good feedback	3.08 (0.95)	3.30 (1.01)	< 0.001	0.22	3.50 (0.67)	3.66 (0.69)	0.188
Do not fear to acknowledge your own knowledge gaps	4.28 (0.69)	4.42 (0.65)	< 0.001	0.21	4.54 (0.45)	4.61 (0.46)	0.443
Do not allow the resolution map to be the summary of 'the entire' subject	4.25 (0.69)	4.42 (0.620)	< 0.001	0.26	4.49 (0.60)	4.49 (0.56)	0.997

Table 1. Descriptive and comparative measurements taken, both globally and for each of the 5 domains of interest, between phases - Group: students and tutors.

Domains	course stage (semester)	Mean (standard deviation)	p
Do not skip the activation of prior knowledge	1 st to 4 th	3.89 (0.59)	0.29
	5 th to 8 th	3.95 (0.61)	
Do not allow the mechanical reading of information	1 st to 4 th	3.75 (0.57)	< 0.001
	5 th to 8 th	3.58 (0.58)	
Do not forget to provide good feedback	1 st to 4 th	3.33 (0.92)	< 0.001
	5 th to 8 th	2.78 (0.90)	
Do not fear to acknowledge your own knowledge gaps	1 st to 4 th	4.39 (0.63)	< 0.001
	5 th to 8 th	4.16 (0.72)	
Do not allow the resolution map to be the summary of 'the entire' subject	1 st to 4 th	4.25 (0.72)	0.99
	5 th to 8 th	4.25 (0.65)	

Table 2. Analysis per course stage - descriptive and comparative measurements of each of the 5 domains of interest in the group of students - phase 3.

Domains	course stage (semester)	Mean (standard deviation)	p
Do not skip the activation of prior knowledge	1 st to 4 th	4.13 (0.56)	0.09
	5 th to 8 th	4.22 (0.62)	
Do not allow the mechanical reading of information	1 st to 4 th	3.92 (0.59)	0.004
	5 th to 8 th	3.78 (0.60)	
Do not forget to provide good feedback	1 st to 4 th	3.46 (0.96)	< 0.001
	5 th to 8 th	3.08 (1.03)	
Do not fear to acknowledge your own knowledge gaps	1 st to 4 th	4.46 (0.65)	0.09
	5 th to 8 th	4.35 (0.65)	
Do not allow the resolution map to be the summary of 'the entire' subject	1 st to 4 th	4.43 (0.63)	0.53
	5 th to 8 th	4.40 (0.61)	

Table 3. Analysis per course stage - descriptive and comparative measures of each of the 5 domains of interest in the group of students - phase 5.

Domains	course stage (semestrer)	Mean (standard deviation)	p
Do not skip the activation of prior knowledge	1 st to 4 th	4.47 (0.35)	0.24
	5 th to 8 th	4.34 (0.51)	
Do not allow the mechanical reading of information	1 st to 4 th	4.08 (0.43)	0.28
	5 th to 8 th	3.96 (0.51)	
Do not forget to provide good feedback	1 st to 4 th	3.89 (0.61)	0.001
	5 th to 8 th	3.36 (0.69)	
Do not fear to acknowledge your own knowledge gaps	1 st to 4 th	4.63 (0.38)	0.62
	5 th to 8 th	4.57 (0.54)	
Do not allow the resolution map to be the summary of 'the entire' subject	1 st to 4 th	4.41 (0.50)	0.27
	5 th to 8 th	4.56 (0.64)	

Table 4. Analysis per course stage - descriptive and comparative measurements of each of the 5 domains of interest in the group of tutors - phase 5.

DISCUSSION

We aimed to evaluate the effect of a faculty development program called “Wise choices in Education” in the perception of medical students and their tutors about tutors’ performance at a medical school. Students and tutors’ perceptions were evaluated before and after the intervention. Results have shown that the strategy had a positive impact on students’ perception of tutors’ performance. This impact was significant but had a small effect size. Tutors rated their own performance better than students, and the faculty development program did not have a significant impact on their self-perception. Students from the 1st to the 4th semester rated their tutors’ performance better than students from the 5th to the 8th semester.

Results have shown that the program was successful in revitalizing important aspects of tutorial group functioning that depend on tutors' performance. This result was expected, since the educational initiative was based on perceived erosions on PBL's principles identified in a previous workshop and focused on specific difficulties faced by this faculty group. Educational initiatives based on faculty perceived gaps (bottom-up approach), rather than on what course directors think they need (top-down approach), appear to be crucial for the design of successful interventions aimed at improving PBL functioning (Moust et al., 2005). The campaign also took advantage of experiential learning, bringing attention to good tutorial practices where they occur, while highlighting the theoretical principles underlying the learning processes fostered by PBL, two features that contribute to faculty development (Steinert et al., 2016). The intensive use of a mix of printed and electronic educational resources also appears to have been decisive for the campaign to be successful, as it allowed it to reach all tutors, even those who resist attending centralized faculty development programs - possibly those who need them most (Steinert et al., 2009).

The assessment of the intervention had a wide participation of students (90%), which indicates that most of the academic community, at least to some extent, got involved with the educational campaign. Students' involvement is very important: according to Azer (2005), PBL works best when students, and not only tutors, understand the different factors influencing the learning process. The campaign may have raised students' awareness of the importance of the tutorial group steps, making them act as practice transformation agents.

The positive effect of the campaign on tutors' performance evaluation was limited to students' perception, with a small effect-size. Nevertheless, the perception of tutors' performance by students is very important and shows how those in the center of the process feel about it. Moreover, students' scores on tutors' performance were already high before the campaign, with overall mean = 3,84, which could explain the small effect-size observed.

Systematic reviews conducted by Bilal, Guraya and Chen (2019) and Leslie et al. (2013) showed that faculty development programs can have positive impact in medical education practices, enriching faculty's knowledge and skills. Bilal, Guraya and Chen's review found programs to have effect sizes that ranged from small to large (Bilal et al., 2019). The nature, purposes and outcome measurements of the programs, however, vary a lot, making it difficult to compare them with our intervention. Most interventions are based on workshops, short courses and seminars, and those which assess behavioral change as the outcome measurement usually do it from teachers'/tutors' own perspectives only. Those which assess students' perspective about their tutors' behavior are scarce. Our

findings adds to the literature showing that well-structured faculty development programs in healthcare might be effective from students' point of view.

As expected, the global mean scores recorded for tutors' assessment was significantly higher than that of students' assessment, both before and after the campaign. Hewson, Copeland and Fishleder (2000) have also observed that teachers rated themselves as very competent in all teaching skills before the educational development program. In addition, the difference in perception between tutors and students had already been observed by Zanolli, Boshuizen and Grave (2002). Students appeared to be more "pessimistic" than their tutors about how the method works, likely because of tutors' overestimated view about their own performance and/or of students' underestimated view about how the method works. Tutors' high scores on their own performance before the campaign might also have caused a "ceiling effect", hindering a positive effect of the campaign.

The comparison of evaluations by students enrolled from the 1st to 4th semesters, and by those enrolled from the 5th to 8th semesters, has suggested that tutors' adherence to good PBL practices decreased in three of the evaluated domains as the course progressed, namely: allowing mechanical reading of information, providing good feedback and acknowledging knowledge gaps. Zanolli, Boshuizen and Grave (2002) have also shown significant differences in several issues found in tutorial groups between second-year students (pre-clinical phase) and third-year students (clinical phase): third-year students had more pessimistic perceptions about feedback issues than second-year students. Based on the authors' perspective, this finding may reflect the fact that the more students are experienced and adapted to the method, the more critical they tend to be towards tutors' skills.

Based on the analysis of each separate domain, acknowledging knowledge gaps had the highest mean scores among tutors and students, before and after the campaign. Tutors are not responsible for providing all content to students in PBL. In fact, they must play active roles in students' learning process so they can build their own knowledge. Therefore, acknowledging knowledge gaps does not represent a failure in the tutorial group (Chng et al., 2011). On the contrary: tutors' open attitude to admit that they may not know the answer to every question reinforces the need for lifelong learning, as long as tutors commit to remedy the situation (Pazin Filho, 2017). The current study has shown that this adult learning feature appears to be preserved in the medical school where this research was made, both in tutors and students' perspectives. This outcome may be explained by tutors' experience (8 semesters, or 4 years, on average) and by faculty development programs periodically carried out at the investigated school.

On the other hand, providing good feedback recorded the lowest scores both in tutors' self-assessment and in students' assessment, in both assessment times. Ende (1983)

defines feedback in the medical education context as the information describing students' performance in a given activity aimed at guiding their future performance in that very same activity or in a related one. Although feedback is considered a fundamental step in the adult teaching-learning process (Chng et al., 2011), there is evidence that it is often omitted or treated inappropriately (Ramani & Krachov, 2012). It is important emphasizing that the perception about what feedback is can differ between tutors and students: tutors may consider that they give feedback, but students may not perceive it (Branch Jr, & Paranjape, 2002), a fact that may explain the difference in assessment between students and tutors. Finally, feedback scores decreased from pre-clinical to clinical years, which may result from the misperception that PBL-experienced students have less to learn from the feedback provided by their tutors than PBL-beginners. However, these inferences require further investigations.

The current study had some limitations. The first of them is associated with the questionnaire that was elaborated by the authors. Such development was necessary because we could not find in the literature any validated instrument to assess the PBL-erosions identified by the tutors. Indeed, this appears to be a limitation in most studies in this field. Steinert's (2016) review has shown that most studies aimed at analyzing the effectiveness of faculty development programs have also used non-validated questionnaires, which were specifically designed to evaluate a given intervention. It is noteworthy that, although the instrument was not validated, it was developed by PBL experts and had good internal consistency. Finally, although the questionnaire has reliably assessed the herein five selected domains, other aspects of tutorial group dynamics were not addressed. Zanolli, Boshuizen and Grave (2002) have shown that addressing all relevant tutorial group aspects is not an easy task and the current study did not intend to carry out such a comprehensive assessment. A further limitation of the study was the fact that our baseline assessment of tutors' performance was conducted after an educational workshop that was attended by approximately 50% of the tutors. Therefore, we were not able to assess the effect of this initial step of our educational initiative and to evaluate how it might have contributed to the high baseline scores observed for some of the domains. On the other hand, the workshop before the intervention allowed the awareness of the erosions as perceived by the tutors that carry PBL sessions in the school, at a specific period of time, which is important to the success of interventions, as discussed before (Moust et al., 2005).

CONCLUSION

The current study has shown that, based on students' perception, a faculty development program focused on PBL-erosions identified by the tutors and that used different faculty development strategies, as workshops, banners and electronic messages, can help improving tutors' performance in PBL tutorials. Additional studies should be conducted

to determine how long the positive effect of such program might last, as well as its potential effects on domains other than those covered in the evaluated educational initiative.

References

- Azer, S. A. (2005). Challenges facing PBL tutors: 12 tips for successful group facilitation. *Medical Teacher*, 5;27(8):676-681. <https://doi.org/10.1080/01421590500313001>. PMID: 16451886.
- Azer, S. A., McLean, M., Onishi, H., Tagawa, M., & Scherpbier, A. (2013). Cracks in problem-based learning: What is your action plan? *Medical Teacher*, 35(10):806-814. <https://doi.org/10.3109/0142159X.2013.826792>. PMID: 23971890.
- Bilal, Guraya, S. Y., & Chen, S. (2019). The impact and effectiveness of faculty development program in fostering the faculty's knowledge, skills, and professional competence: A systematic review and meta-analysis. *Saudi Journal of Biological Sciences*, 26(4): 688-697. <https://doi.org/10.1016/j.sjbs.2017.10.024>
- Bodagh, N., Bloomfield, J., Birch, P., & Ricketts, W. (2017). Problem-based learning: a review. *British Journal of Hospital Medicine*, 78(11):167-170. <https://doi.org/10.12968/hmed.2017.78.11.C167>. PMID: 29111794.
- Branch Jr., W.T., & Paranjape A. (2002). Feedback and reflection: teaching methods for clinical settings. *Academic Medicine*, 77(12):1185-1188. <https://doi.org/10.1097/00001888-200212000-00005>. PMID: 12480619.
- Chng, E., Yew, E. H., & Schmidt, H. G. (2011). Effects of tutor-related behaviours on the process of problem-based learning. *Advances in Health Science Education: Theory and Practice*, 16:491-503. <https://doi.org/10.1007/s10459-011-9282-7>. PMID: 21547499.
- Dolmans, D. H. J. M., De Grave, W., Wolfhagen, I. H. A. P., & van der Vleuten, C. P. M. (2005). Problem-based learning: future challenges for educational practice and research. *Medical Education*, 39(7):732-741. <https://doi.org/10.1111/j.1365-2929.2005.02205.x>. PMID: 15960794.
- Ende, J. (1983). Feedback in clinical medical education. *Journal of the American Medical Association*, 250(6):777-781. <https://doi.org/10.1001/jama.1983.03340060055026> PMID: 6876333.
- Hewson, M. G., Copeland, H. L., & Fishleder, A. J. (2000). What's the Use of Faculty Development? Program Evaluation Using Retrospective Self-Assessments and Independent Performance Ratings. *Teaching and Learning in Medicine*, 13(3):153-160. https://doi.org/10.1207/S15328015TLM1303_4. PMID: 11475658.

- Hung, W., Dolmans, D. H. J. M., & van Merriënboer J. J. G. (2019). A review to identify key perspectives in PBL meta-analyses and reviews: trends, gaps and future research directions. *Advances in Health Science Education: Theory and Practice*, 24(5):943-957. <https://doi.org/10.1007/s10459-019-09945-x>.
- Leslie, K., Baker, L., Egan-Lee, E., Esdaile, M., & Reeve, S. (2013). Advancing Faculty Development in Medical Education: A Systematic Review. *Academic Medicine*, 88(7): 1038-1045. <https://doi.org/10.1097/ACM.0b013e318294fd29>
- Moust, J. H. C., Van Berkel, H. J. M., & Schmidt, H. G. (2005). Signs of erosions: Reflections on three decades of problem-based learning at Maastricht University. *Higher Education*, 50(4):665-683. <https://doi.org/10.1007/s10734-004-6371-z>.
- Pazin Filho, A. (2007). Características do aprendizado do adulto. *Medicina*, 40(1):7-16. <https://doi.org/10.11606/issn.2176-7262.v40i1p7-16> .
- Ramani, S., & Krackov, S.K. (2012). Twelve tips for giving feedback effectively in the clinical environment. *Medical Teacher*, 34:787-791. <https://doi.org/10.3109/0142159X.2012.684916>. PMID: 22730899.
- Steinert, Y., Mann, K., Anderson, B., Barnett, B. M., Centeno, A., Naismith, L., et al. (2016). A systematic review of faculty development initiatives designed to enhance teaching effectiveness: A 10-year update: BEME Guide No. 40. *Medical Teacher*, 38(8):769-786. <https://doi.org/10.1080/0142159X.2016.1181851>. PMID: 27420193.
- Steinert, Y., McLeod, P. J., Boillat, M., Meterissian, S., Elizov, M., & Macdonald, M. E. (2009). Faculty development: a 'Field of Dreams'? *Medical Education*, 43(1):42-49. <https://doi.org/10.1111/j.1365-2923.2008.03246.x>
- Zanolli, M. B., Boshuizen, H. P. A., & De Grave, W. S. (2002). Students' and Tutors' Perceptions of Problems in PBL Tutorial Groups at a Brazilian Medical School. *Education for Health*, 15(2):189-201. <https://doi.org/10.1080/13576280210136942>. PMID: 14741968.

APPENDIX A - Tutorial group assessment questionnaire – students’ version

Sex: () M () F age: ____ years current course cycle: ____ period when the student started the course: ____

The items below refer to tutors’ performance during the tutorial group. Read and mark an X in each item meeting your opinion about your current tutor, based on a scale from 1 to 5, wherein:

1-never; 2-almost never; 3-intermediate frequency; 4-almost always and 5-always.

	Never	Almost never	Intermediate frequency	Almost always	Always
1. How often does your tutor encourage “brainstorming”?	1	2	3	4	5
2. How often does your tutor provide clues to activate your prior knowledge during the analysis session?	1	2	3	4	5
3. How often does your tutor address students’ previous experiences by linking them to the addressed problem?	1	2	3	4	5
4. How often does your tutor merge phases P3 (brainstorming) and P4 (analysis map)?	1	2	3	4	5
5. How often does your tutor “skip” the analysis map development stage?	1	2	3	4	5
6. How often does your tutor recover the analysis map at the beginning of the resolution session?	1	2	3	4	5
7. How often does your tutor allow you to read the studied content directly from the bibliography?	1	2	3	4	5
8. How often does your tutor encourage you to read your summary?	1	2	3	4	5

	Never	Almost never	Intermediate frequency	Almost always	Always
9. How often does your tutor encourage you to explain the problem in your own words?	1	2	3	4	5
10. How often does your tutor encourage you to summarize in your own words what you have learned?	1	2	3	4	5
11. How often does your tutor encourage knowledge application to the problem in question?	1	2	3	4	5
12. How often does your tutor encourage knowledge application to other situations or problems?	1	2	3	4	5
13. How often does your tutor encourage you to understand the concepts and mechanisms of the problem?	1	2	3	4	5
14. How often does your tutor provide feedback on the group's performance at the end of the TG?	1	2	3	4	5
15. How often does your tutor expose the TG's strengths to the group?	1	2	3	4	5
16. How often does your tutor discuss the negative aspects of TG with the group?	1	2	3	4	5
17. How often does your tutor score your participation at the end of the TG session?	1	2	3	4	5

	Never	Almost never	Intermediate	Almost always	Always
18. How often does your tutor provide individual feedback whenever needed?	1	2	3	4	5
19. How often does your tutor finish the TG session without evaluating the group's performance?	1	2	3	4	5
20. How often does your tutor ask the group for feedback on his/her performance during TG?	1	2	3	4	5
21. How often does your tutor confess to the group that he/she does not know a certain concept?	1	2	3	4	5
22. How often does your tutor tell the group that he/she will study to clarify an unresolved doubt raised by the group?	1	2	3	4	5
23. How often does your tutor ignore the doubts raised by the group?	1	2	3	4	5
24. How often does your tutor finish the tutorial group session without clarifying students' doubts?	1	2	3	4	5
25. How often does your tutor return to doubts previously raised by the group or by him/herself for clarification?	1	2	3	4	5
26. How often does your tutor encourage the recovery of the problem during the resolution session?	1	2	3	4	5

	Never	Almost never	Intermediate frequency	Almost always	Always
27. How often does your tutor encourage the application of the discussed content to solve the problem in question?	1	2	3	4	5
28. How often does your tutor ignore the analysis map at the time to build the resolution map?	1	2	3	4	5
29. How often does your tutor encourage the construction of the resolution map applied to the problem?	1	2	3	4	5
30. How often does your tutor ignore the problem in question at the time to build the resolution map?	1	2	3	4	5

APPENDIX B - Tutorial group assessment questionnaire – tutors' version

Sex: () M () F age: __ years cycle: __ Total tutoring time (in semesters): _____

The items below refer to tutors' performance during the tutorial group. Read and mark an X in each item meeting your opinion about your performance in your current group (if you were tutor for more than one period, choose only one of them for your answers). Use the following scale to score your answers: 1-never; 2-almost never; 3-intermediate frequency; 4-almost always and 5-always.

	Never	Almost never	Intermediate frequency	Almost always	Always
1. How often do you encourage "brainstorming"?	1	2	3	4	5
2. How often do you provide clues to activate students' prior knowledge during the analysis session?	1	2	3	4	5
3. How often do you stimulate students' previous experiences by linking them to the addressed problem?	1	2	3	4	5
4. How often do you merge phases P3 (brainstorming) and P4 (analysis map)?	1	2	3	4	5
5. How often do you "skip" the analysis map development stage?	1	2	3	4	5
6. How often do you recover the analysis map at the beginning of the resolution session?	1	2	3	4	5

	Never	Almost never	Intermediate frequency	Almost always	Always
7. How often do you allow students to read the studied content directly from the bibliography?	1	2	3	4	5
8. How often do you encourage students to read their own summary?	1	2	3	4	5
9. How often do you encourage students to explain the problem in their own words?	1	2	3	4	5
10. How often do you encourage students to summarize what they have learned in their own words?	1	2	3	4	5
11. How often do you encourage knowledge application to the problem in question?	1	2	3	4	5
12. How often do you encourage knowledge application to other situations or problems?	1	2	3	4	5
13. How often do you encourage students to understand the concepts and mechanisms of the problem?	1	2	3	4	5
14. How often do you provide feedback on the group's performance at the end of the tutorial session (TS)?	1	2	3	4	5
15. How often do you expose the TS's strengths to the group?	1	2	3	4	5
16. How often do you discuss the negative aspects of TS with the group?	1	2	3	4	5

	Never	Almost never	Intermediate frequency	Almost always	Always
17. How often do you score the participation of each student at the end of the TS session?	1	2	3	4	5
18. How often do you provide individual feedback whenever needed?	1	2	3	4	5
19. How often do you finish the TS session without evaluating the group's performance?	1	2	3	4	5
20. How often do you ask the group for feedback on your own performance during TS?	1	2	3	4	5
21. How often do you confess to the group that you do not know a certain concept?	1	2	3	4	5
22. How often do you tell the group that you will study to clarify an unresolved doubt raised by them?	1	2	3	4	5
23. How often do you ignore the doubts raised by the group?	1	2	3	4	5
24. How often do you finish the tutorial group session without clarifying students' doubts?	1	2	3	4	5
25. How often do you return to doubts previously raised by the group or by yourself for clarification?	1	2	3	4	5
26. How often do you encourage the recovery of the problem during the resolution session?	1	2	3	4	5
27. How often do you encourage the application of the discussed content to solve the problem in question?	1	2	3	4	5

	Never	Almost never	Intermediate frequency	Almost always	Always
28. How often do you ignore the analysis map at the time to build the resolution map?	1	2	3	4	5
29. How often do you encourage the construction of the resolution map applied to the problem?	1	2	3	4	5
30. How often do you ignore the problem in question at the time to build the resolution map?	1	2	3	4	5

APPENDIX C – Supplementary Material

Question	Cronbach's alpha	Question	Cronbach's alpha	Question	Cronbach's alpha
1	0.60	11	0.51	21	0.61
2	0.58	12	0.56	22	0.62
3	0.55	13	0.57	23	0.63
4	0.61	14	0.59	24	0.62
5	0.60	15	0.57	25	0.59
6	0.61	16	0.58	26	0.59
7	0.64	17	0.57	27	0.57
8	0.59	18	0.60	28	0.65
9	0.57	19	0.64	29	0.59
10	0.55	20	0.58	30	0.64
Full questionnaire					0.60

Table 1. Analysis of the internal consistency and reliability of the questionnaire in the tutors' group, based on the removal of each indicated question and on the full questionnaire – phase 3.

Question	Cronbach's alpha	Question	Cronbach's alpha	Question	Cronbach's alpha
1	0.76	11	0.76	21	0.71
2	0.76	12	0.76	22	0.76
3	0.76	13	0.76	23	0.78
4	0.78	14	0.75	24	0.78
5	0.78	15	0.75	25	0.76
6	0.77	16	0.75	26	0.76
7	0.78	17	0.76	27	0.76
8	0.77	18	0.76	28	0.79
9	0.76	19	0.80	29	0.76
10	0.75	20	0.76	30	0.79
Full questionnaire					0.77

Table 2. Analysis of the internal consistency and reliability of the questionnaire in the students' group, based on the removal of each indicated question and on the full questionnaire – phase 3.

Domains	Group	Mean (Standard deviation)	p
Global	students	4.15 (0.33)	0.001
	tutors	3.84 (0.50)	T > A
Do not skip the activation of prior knowledge	students	4.35 (0.35)	< 0.001
	tutors	3.92 (0.60)	T > A
Do not allow the mechanical reading of information	students	3.87 (0.48)	0.014
	tutors	3.67 (0.58)	T > A
Do not forget to provide good feedback	students	3.50 (0.67)	0.001
	tutors	3.08 (0.95)	T > A
Do not fear to acknowledge your own knowledge gaps	students	4.54 (0.45)	0.005
	tutors	4.28 (0.69)	T > A
Do not allow the resolution map to be the summary of 'the entire' subject	students	4.49 (0.60)	0.014
	tutors	4.25 (0.69)	T > A

Table 3. Descriptive and comparative measures taken, both globally and for each of the 5 domains of interest, between phases - Group: students and tutors – phase 3.

Domains	Group	Mean (Standard deviation)	p
Global	students	4.24 (0.39)	< 0.001 T>A
	tutors	4.03 (0.48)	
Do not skip the activation of prior knowledge	students	4.42 (0.43)	< 0.001 T > A
	tutors	4.17 (0.59)	
Do not allow the mechanical reading of information	students	4.03 (0.47)	0.020 T > A
	tutors	3.86 (0.60)	
Do not forget to provide good feedback	students	3.66 (0.69)	0.003 T > A
	tutors	3.30 (1.01)	
Do not fear to acknowledge your own knowledge gaps	students	4.61 (0.46)	0.019 T > A
	tutors	4.42 (0.65)	
Do not allow the resolution map to be the summary of 'the entire' subject	students	4.49 (0.56)	0,374 T = A
	tutors	4.42 (0.62)	

Table 4. Descriptive and comparative measurements taken, both globally and for each of the 5 domains of interest between students and tutors – phase 5.