RBEM
Revista Brasileira de Educação Médica

abem associação brasileira de educação médica

ORIGINAL ARTICLE

DOI: https://doi.org/10.1590/1981-5271v49.3-2025.0054.ING

Problem-based learning: feedbacks correlated to levels of knowledge

Aprendizagem baseada em problemas: feedbacks *correlatos a níveis de conhecimento*

Gustavo Assumpção Teixeira[1] *gassump10@gmail.com*
Gislaine Cristhina Bellusse[1] *gibellusse@gmail.com*
Reynaldo José Sant'Anna Pereira de Souza[1] *reynaldojosesantanna@gmail.com*
Sinesio Grace Duarte[1] *sinesio.duarte@unifran.edu.br*

ABSTRACT

Introduction: Problem-Based Learning (PBL) is an educational approach that fosters active and reflective learning by engaging students in meaningful problem-solving. Enhancing student performance relies on providing objective and effective feedback, supported by well-defined performance descriptors. This study aimed to validate an instrument capable of qualifying student performance while simultaneously providing feedback correlated with this assessment.

Methods: The Q-sort method was used to validate the assessment instrument in four phases: (1) evaluation by ten judges of the correlation between assessment domains and their corresponding feedback; (2) identification of outliers in the classifications given by the judges; (3) analysis of the judges' agreement index regarding the relationship between domains and feedback, and the need for adaptations using the Kappa Index; (4) submission of the domains and feedback for a new evaluation by the judges.

Results: The proposed domains and their correlated feedback were validated using the Q-sort method, with a Fleiss' Kappa coefficient of 1.0, indicating perfect agreement among the judges.

Conclusion: The validated instrument provides a reliable tool for assessing medical students' performance while offering correlated feedback that can assist tutors in delivering more effective feedback, thereby enhancing the educational process in PBL environments.

Keywords: Formative Feedback; Problem-Based Learning; Educational assessment; Academic Performances.

RESUMO

Introdução: *Aprendizagem baseada em problemas (ABP) é uma abordagem educacional que promove a aprendizagem ativa e reflexiva ao envolver os estudantes na resolução significativa de problemas. O aprimoramento da performance dos estudantes depende do fornecimento de feedback objetivo e eficaz, apoiado por descritores de desempenho bem definidos.*

Objetivo: *Este estudo teve como objetivo validar um instrumento que fosse capaz de qualificar a performance e concomitantemente prover um feedback correlato a essa avaliação.*

Método: *Para validação do instrumento de avaliação, utilizou-se o método Q-sort. O processo de validação foi realizado em quatro fases: 1. avaliação por dez juízes da correlação entre domínios de avaliação e feedbacks correlatos; 2. identificação de outlier na classificação dada pelos juízes; 3. análise do índice de concordância dos juízes na relação entre domínios e feedbacks, e necessidade de adaptações por meio do índice kappa; 4. submissão dos domínios e feedbacks para nova avaliação dos juízes.*

Resultado: *Os domínios propostos e seus feedbacks correlatos foram validados pelo método Q-sort com coeficiente kappa de Fleiss igual a 1,0 indicando uma concordância perfeita entre os juízes.*

Conclusão: *O instrumento validado fornece um meio confiável a ser utilizado nas avaliações de desempenho do estudante de Medicina, além de proporcionar feedbacks correlatos que podem auxiliar o tutor na entrega de feedbacks mais efetivos, aprimorando assim o processo educacional em ambientes de ABP.*

Palavras-chave: *Feedback Formativo; Aprendizagem Baseada em Problemas; Avaliação Educacional; Desempenho Acadêmico.*

[1] Universidade de Franca, Franca, São Paulo, Brasil.

Editora-chefe: Rosiane Viana Zuza Diniz.
Editor associado: Mario Luis Cesaretti.

Recebido em 24/02/25; Aceito em 20/05/25.

Evaluated by double blind review process.

 DOI: https://doi.org/10.1590/1981-5271v49.3-2025.0054.ING

INTRODUCTION

Problem-based learning (PBL) has been recognised as an effective educational approach that promotes active and reflective learning[1], which encourages students to get involved in problem-solving, thus stimulating critical thinking, collaboration and self-learning[2].

In this methodology, students have the opportunity to explore complex concepts in a meaningful and contextualised manner[3].

The effectiveness of the method, however, depends not only on solving problems, but also on the objectivity and effectiveness of the feedback provided to students, which helps them to develop learning strategies.

The underlying condition for feedback to be truly effective is its correlation with the student's need.

In order to properly identify strengths and areas for improvement, the descriptors used to measure student performance must be well-founded[5].

Qualitative assessment tools can provide comprehensive insights into skills, attitudes, conceptual understanding and problem-solving ability[6].

Although the literature describes skills and attitudes, there is a shortage of effective instruments aimed at objectively measuring the knowledge demonstrated by students, restricting the development of effective feedback on this aspect.

In this context, we propose the creation of an instrument that objectively measures the knowledge presented, correlating feedback that can effectively improve student performance.

Bloom's Taxonomy was used to create the descriptors for the levels of knowledge.

Bloom's Taxonomy, originally developed by Benjamin S. Bloom and his colleagues in the 1950s and modified by Krathwohl and Anderson[7], is an organisational system widely used to categorise learning in teaching and assessment activities.

To create the feedback, strategies that have been proven to improve the performance of university students were researched in Pubmed and Medline databases.

We found that proposed assessment tools must undergo scientific validation to ensure their reliability[8] and the search for consistent results[9].

The aim of this study was to validate the proposed instrument in order to help PBL tutors provide individualised and effective feedback.

METHODOLOGY

The study was conducted at the Medical School of a private university in the city of Franca/SP. This Higher Education Institution (HEI) has 14 years of experience in active methodologies, with PBL being used from the beginning of the course as the main methodology for conducting tutorial sessions. For the face-to-face data collection, the meeting room, located inside the staff room, was used. Three researchers were responsible for providing the data collection instrument to the participants, with an estimated time of 30 minutes for completion.

The Q-sort method was used, which is based on evaluations by judges who have knowledge of the domains explored in the work, in order to offer systematised feedback on the selected scales of measurement. Agreement between the judges attests to substantive validity[10,11], i.e. agreement between the judges is evidence that the feedback is correlated to the domains assessed in the PBL process, the focus of this project.

The content evaluation process comprises four phases. Firstly, ten judges are asked to indicate the feedback that best corresponds to each domain of the learning assessment model. The second stage consists of evaluating the result of the judges' rating. The third stage is designed to analyse the disagreements pointed out by the judges and, if necessary, refine the ambiguous feedback. In the fourth stage, the feedback is submitted for further evaluation by the judges to check that the adaptations made were clear[12]. The use of multiple judges and multiple analyses is recommended to increase the reliability of the scale development process[13].

The greater the agreement between the judges' ratings of the domain and its corresponding feedback, the greater the convergent validity of the learning assessment model[12].

A questionnaire based on the Q-sort method was drawn up to validate the assessment tool. Ten judges were instructed to carefully read the feedback we have adopted for evaluating PBL and then read the domains (lines) that follow in the questionnaire. The next step was to indicate the correlated feedback that appears as an answer option (columns) for the domains presented. The domains were presented randomly in the questionnaire, disregarding the order of complexity.

- **Lack of knowledge of the content:**
  - *Feedback:* Identify the causes of lack of knowledge: assess lack of motivation[14], number of absences[15], as well as socio-economic[16],vocational[17] or emotional[18] vulnerabilities, and difficulties in relationships with peers[19]. Encourage additional assistance: use extra classes, additional bibliographies and help from peers with high proficiency levels[20]; seek psycho-pedagogical support[21].
- **Very limited knowledge, inability to identify the concepts relevant to the content:**
  - *Feedback:* Identify the causes of the lack of knowledge: assess inefficiencies in study habits, such as lack of dedication[14], commitment, learning strategies and

time management. Encourage ways of correcting inefficiencies: reflect on activities that hinder learning, such as excessive commitments unrelated to the curriculum and excessive use of social media[22]. Consider obtaining additional assistance.

- **Limited knowledge, significant conceptual errors:**
- *Feedback:* Review the literature that underpins the conceptual basis, followed by a new study of the concepts. Complement the study with other recommended basic bibliographies[20]. Consider obtaining additional assistance.

- **Limited knowledge, occasional conceptual errors:**
- *Feedback:* Re-study the concepts, correcting any inaccuracies. Assess whether occasional conceptual errors are the result of failures to study the current content or previous content that underpins the present knowledge. Complement the study with other recommended basic bibliographies[20].

- **Basic knowledge, but with gaps or inaccuracies:**
- *Feedback:* Revise the content, identifying the source of gaps and inaccuracies. Complement the study with other recommended basic bibliographies[20]. Answer basic-level multiple-choice questions. Consider participating in study groups to discuss the content[23].

• Basic knowledge of the content, but without establishing correlations:

- *Feedback:* Learn to establish correlations between concepts by creating mind maps, concept maps or illustrative diagrams[24]. Answer intermediate level multiple-choice questions. Participate in study groups to deepen the correlations between content.

• Sufficient proficiency of the content to establish specific correlations, with limited ability to apply them:

- *Feedback:* Integrate new knowledge in a critical way and correlate it with previous knowledge: explore suggested complementary bibliographies[20], answer advanced-level multiple-choice questions. Developing new maps/schemes that integrate the new knowledge can help to visualise correlations that were initially less noticeable[24].

• Sufficient proficiency of the content to establish complex correlations, with the ability to apply them in specific contexts:

- *Feedback:* Become a facilitator for the teaching-learning process of peers, a step that has been proven to improve the performance of both[20]. Creating questions for peers can lead to new perspectives for improving knowledge[25]. Explore suggested complementary bibliographies[20].

• Practically complete proficiency of the content, with the ability to establish complex correlations and the capacity to apply them in different contexts:

- *Feedback:* Encourage students to identify and share the learning strategies that have led to their outstanding performance. Consider replicating them for other learning opportunities, reading articles and encouraging scientific initiation.

• Complete proficiency of the content, with the ability to establish complex correlations and the capacity to apply them in different contexts:

- *Feedback:* Encourage the student to identify their motivations, strengths and learning strategies that have led them to excel in their performance. Consider replicating them for other learning opportunities, sharing them with peers, reading articles and encourage scientific initiation.

This study was approved by the Research Ethics Committee of the University of Franca - Unifran, in compliance with Resolution 466/2012 of the National Health Council, under CAAE number 79446624.0.0000.5495.

To validate the instrument for objectively measuring knowledge and related feedback, the judges who took part in this stage of the study were asked to sign the Free and Informed Consent Form (FICF).

Data analysis methodology:

Initially, the convergence of the judges' responses was checked to identify any possible outliers that could be excluded from the process. The benchmark for the Judge vs. Judge comparison was set at above 65%[12], although the literature also considers rates above 55% to be acceptable[26].

The second analysis corresponded to the raters' level of agreement on the relationship between proficiency and feedback. The aim was to obtain the convergence index of the learning model, which, although less rigorous than Cohen's Kappa[27], is an indication of possible problems with the feedback developed.

The third analysis corresponded to obtaining the reliability index of the judges' rating, measured using the Kappa Index[27], which is the most commonly used reliability measure in scientific literature to validate the content of the items in the evaluation model[13]. The Kappa index measures the degree of agreement between judges, where a Kappa value of 1 indicates perfect agreement between judges in all cases; and a value of 0 indicates no agreement other than that expected by chance[13,28].

The formula corresponding to the Kappa coefficient is shown in Figure 1.

 DOI: https://doi.org/10.1590/1981-5271v49.3-2025.0054.ING

Figure 1. Kappa coefficient formula.

$$k = \frac{k1 - k2}{1 - k2}$$

$$Where: \; k1 = \sum_{i}^{r} 1 \frac{n_{ij}}{n} \; e \; k2 = \sum_{i}^{r} 1 \frac{n_i \, x \, n_j}{n^2}$$

r = assessment categories
nij = number of elements rated by judge X in category i and by judge Y in category j.
ni = number of elements rated by judge X in category i.
nj = number of elements rated by judge Y in category j.
n = total number of elements rated

Source: Derived from Cohen[27].

Table 1. Acceptable level of inter-rater agreement by the Kappa index.

Kappa Value	Interpretation
<0	No agreement
0.0 - 0.20	Slight agreement
0.21 – 0.40	Fair
0.41 – 0.60	Moderate
0.61 – 0.80	Substantial agreement
0.81 – 1.00	Perfect agreement

Source: Landis and Koch (1977).

Table 1 shows the recommended Kappa indices, as indicated in the literature by Landis and Koch[26].

To validate assessment tools, perfectly concordant coefficients are accepted as a priority.

The raters (judges) should be professionals with a degree in the health field, working as tutors in a medical course (accredited by the Ministry of Education), which used the ABP methodology in its tutorial sessions. Raters with less than 4 years' experience as a tutor were excluded.

RESULTS

Of the ten judges, nine were able to accurately correlate at least 100% of the domains to their respective feedback. The literature suggests that the accuracy level should be higher than 65%.

Judge 3 was therefore excluded from the sample because he failed to reach the minimum correct correlation level and presented a low average correlation with the other judges.

The percentage of accurately identified combinations considering the exclusion of judge 3 is shown below.

In relation to the domains and feedback, all the domains were correctly associated with their respective feedback.

Table 2. Judges' percentage of accurately identified domain/feedback combinations.

	D01	D02	D03	Accuracy
Judge01	1	1	1	100%
Judge02	1	1	1	100%
Judge03	1	0	0	33%
Judge04	1	1	1	100%
Judge05	1	1	1	100%
Judge06	1	1	1	100%
Judge07	1	1	1	100%
Judge08	1	1	1	100%
Judge09	1	1	1	100%
Judge10	1	1	1	100%
Accuracy	100%	90%	90%	

Source: Data from the study itself.

Table 3. Accuracy percentage of the domain/feedback combination after the exclusion of judge 3.

	D01	D02	D03	Accuracy
Judge01	3	2	1	100%
Judge02	3	2	1	100%
Judge04	3	2	1	100%
Judge05	3	2	1	100%
Judge06	3	2	1	100%
Judge07	3	2	1	100%
Judge08	3	2	1	100%
Judge09	3	2	1	100%
Judge10	3	2	1	100%
Acccuracy	100%	100%	100%	

Source: Data from the study itself.

 DOI: https://doi.org/10.1590/1981-5271v49.3-2025.0054.ING

Fleiss' Kappa coefficient

Fleiss' Kappa coefficient, calculated by averaging the Kappas between pairs of judges, is equal to 1.00. This means that the tool has very good degree of agreement. The literature suggests that this coefficient should be higher than 0.80 (very good agreement or perfect agreement).

DISCUSSION

PBL, which is suitable for higher education, emphasises the promotion of social skills, problem-solving and learning[2]. However, dynamic discussions with multiple concomitant perspectives present a major challenge when it comes to feedback, which consists of qualifying students' knowledge, skills and competences, as well as developing strategies for academic improvement[1].

Tutors in tutorial sessions have found it difficult to standardise, structure and deliver effective feedback, and even to find ways of sparking student interest[29]. From the students' perspective, a low percentage considers that they are evaluated fairly by their tutors, and only half of them consider the feedback to be useful and beneficial for improving their performance[30].

Although descriptions of student performance in skills and attitudes can be found in the literature, there is a shortage of effective assessment tools that involve the objective measurement of the knowledge presented by the student. Consequently, we also found a lack of correlation between the feedback strategies and the level of knowledge presented.

Feedback that correlates with objective performance measurement makes it possible to set goals and objectives that involve desirable challenges. Desirable challenges are so named because responding to them (successfully) involves processes that support learning, understanding and remembering. However, if the student is not prepared to respond successfully, they become undesirable challenges[31].

In this context, the use of the modified Bloom's Taxonomy brings objectivity to measurement of the domains of the knowledge presented by the students and makes it possible to adequately correlate the feedback that subsidises performance improvement.

Training processes that promote continuous reflection allow students to take ownership of educational monitoring and development tools. The expected goal is for students to be co-authors and seek the leading role in their educational process[4].

CONCLUSION

This study showed that the domains of the knowledge item presented, as well as their correlated feedback, were validated by the Q-sort method with a Fleiss' Kappa coefficient equal to 1.0 (perfectly agreeable), making it a reliable tool for assessing tutorial sessions.

CONTRIBUTION OF THE AUTHORS

Gustavo Assumpção Teixeira contributed to the conceptualisation, data curation, formal analysis, research, methodology, project management, validation, writing – original draft, writing – proofreading and editing. Gislaine Cristhina Bellusse and Reynaldo José Sant'Anna Pereira de Souza contributed to the data curation, formal analysis, research, methodology, project management, writing – proofreading and editing. Sinesio Grace Duarte contributed to the formal analysis, methodology, project management, writing – proofreading and editing.

CONFLICT OF INTEREST

We declare no conflicts of interest.

FUNDING

We declare that there is no funding.

REFERENCES

1. Baden MS, Major CH. Foundations of problem based learning. New York: Open University Press; 2024. 216 p.
2. Trullàs JC, Blay C, Sarri E, Pujol R. Effectiveness of problem-based learning methodology in undergraduate medical education: a scoping review. BMC Med Educ. 2022;22(104):2-12. doi: https://doi.org/10.1186/s12909-022-03154-8.
3. Hmelo-Silver CE. Problem-based learning: what and how do students learn? Educ Psychol Rev. 2004;16:235-66.
4. Nicol DJ, Macfarlane-Dick D. Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. Stud High Educ. 2006;31(2):199-218. doi: https://doi.org/10.1080/03075070600572090.
5. Hattie J, Timperley H. The power of feedback. Rev Educ Res. 2007;77(1):81-112. doi: https://doi.org/10.3102/003465430298487.
6. Torrance H, Pryor J. Investigating formative assessment. In: Torrance H, Pryor J. Teaching, learning and assessment in the classroom. Philadelphia: Open University Press; 1998. p. 8.
7. Krathwohl DR, Anderson LW. A taxonomy for learning, teaching, and assessing: a revision of Bloom's taxonomy of educational objectives. New York: Longman; 2001. 352 p.
8. Brookhart SM. Educational assessment knowledge and skills for teachers. Educ Meas Issues Pract. 2011;30(1):3-12. doi: https://doi.org/10.1111/j.1745-3992.2010.00195.x.
9. Pelligrino J, Chudowsky N, Glaser R. Knowing what students know: the science and design of educational assessment. Washington, DC: National Academy Press; 2001. 366 p.
10. Anderson JC, Gerbing DW. Predicting the performance of measures in a confirmatory factor analysis with a pretest assessment of their substantive validities. J Appl Psychol. 1991;76(5):732-40. doi: https://doi.org/10.1037/0021-9010.76.5.732.
11. Ahmed MU, Kristal MM, Pagell M, Gattiker TF. Towards a classification of supply chain relationships: a routine-based perspective. Supply Chain Manag An Int J. 2017;22(4):341-74. doi: https://doi.org/10.1108/SCM-04-2017-0142.

12. Moore GC, Benbasat I. Development of an instrument to measure the perceptions of adopting an information technology innovation. Inf Syst Res. 1991;2(3):173-91. doi: https://doi.org/10.1287/isre.2.3.192.
13. Perreault WD, Leigh LE. Reliability of nominal data based on qualitative judgments. J Mark Res. 1989;26(2):135-48. doi: https://doi.org/10.1177/002224378902600201.
14. Bonsaksen T, Brown T, Lim HB, Fong K. Approaches to studying predict academic performance in undergraduate occupational therapy students: a cross-cultural study. BMC Med Educ. 2017;17:1-9. doi: https://doi.org/10.1186/s12909-017-0914-3.
15. Kauffman CA, Derazin M, Asmar A, Kibble JD. Relationship between classroom attendance and examination performance in a second-year medical pathophysiology class. Adv Physiol Educ. 2018;42(4):593-8. doi: https://doi.org/10.1152/advan.00123.2018.
16. Salvage-Jones J, Hamill J, Todorovic M, Barton MJ, Johnston AN. Developing and evaluating effective bioscience learning activities for nursing students. Nurse Educ Pract. 2016;19:63-9. doi: https://doi.org/10.1016/j.nepr.2016.05.005.
17. Trimble M. Vocation in medicine. Ulster Med J. 2022;91(2):65-6.
18. Frazier P, Gabriel A, Merians A, Lust K. Understanding stress as an impediment to academic performance. J Am Coll Health. 2019;67(6):562-70. doi: https://doi.org/10.1080/07448481.2018.1499649.
19. Zeng Y, Zhang J, Wei J, Li S. The impact of undergraduates' social isolation on smartphone addiction: the roles of academic anxiety and social media use. Int J Environ Res Public Health. 2022;19(23):15903. doi: https://doi.org/10.3390/ijerph192315903.
20. Benè KL, Bergus G. When learners become teachers: a review of peer teaching in medical student education. Fam Med. 2014;46(10):783-7.
21. Church HR, Rumbold JL, Sandars J. Applying sport psychology to improve clinical performance. Med Teach. 2017;39(12):1205-13. doi: https://doi.org/10.1080/0142159X.2017.1359523.
22. Bickerdike A, O'Deasmhunaigh C, O'Flynn S, O'Tuathaigh C. Learning strategies, study habits and social networking activity of undergraduate medical students. Int J Med Educ. 2016;17(7):230-6. doi: https://doi.org/10.5116/ijme.576f.d074.
23. Keren D, Lockyer J, Ellaway RH. Social studying and learning among medical students: a scoping review. Perspect Med Educ. 2017;6(5):311-8. doi: https://doi.org/10.1007/s40037-017-0358-9.
24. Choudhari SG, Gaidhane AM, Desai P, Srivastava T, Mishra V, Zahiruddin SQ. Applying visual mapping techniques to promote learning in community-based medical education activities. BMC Med Educ. 2021;21:210. doi: https://doi.org/10.1186/s12909-021-02646-3.
25. Touissi Y, Hjiej G, Hajjioui A, Ibrahimi A, Fourtassi M. Does developing multiple-choice questions improve medical students' learning? A systematic review. Med Educ Online. 2022;27(1):2005505. doi: https://doi.org/10.1080/10872981.2021.2005505.
26. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159-74.
27. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas. 1960;20(1):37-46. doi: https://doi.org/10.1177/001316446002000104.
28. Fleiss JL, Levin B, Paik MC. Statistical methods for rates and proportions. 3rd ed. New Jersey: John Wiley & Sons; 2013. 800 p.
29. Fine P, Leung A, Tonni I, Louca C. Dental teacher feedback and student learning: a qualitative study. Dent J (Basel). 2023;11(7):164. doi: https://doi.org/10.3390/dj11070164.
30. Al-Drees AA, Khalil MS, Irshad M, Abdulghani HM. Students' perception towards the problem based learning tutorial session in a system-based hybrid curriculum. Saudi Med J. 2015;36(3):341-8. doi: https://doi.org/10.15537/smj.2015.3.10216.
31. Bjork RA, Bjork EL. Forgetting as the friend of learning: implications for teaching and self-regulated learning. Adv Physiol Educ. 2019;43(2):164-7. doi: https://doi.org/10.1152/advan.00001.2019.