Chula Med J. January - February 2027;71(1): 9 - 18 **DOI : 10.56808/2673-060X.**

Original article

Tutor and students' perceptions of tutorial assessment: A qualitative study in an Indonesian medical school

Dian Puspita Dewi[a, *], Lathifa Putry Fauzia[a], Ryan Halleyantoro[b]

[a]*Medical and Health Professions Education Development Unit, Faculty of Medicine, Universitas Diponegoro, Semarang, Central Java, Indonesia*

[b]*Departement of Parasitology, Faculty of Medicine, Universitas Diponegoro, Semarang, Central Java, Indonesia*

Abstract

Background: The tripartite framework suggests that problem-based learning (PBL) tutorial assessment should balance its summative and formative function, an ideal outcome that depends heavily on users' understanding. However, to what extent this assessment tool is actually understood remains uncertain.
Objective: This study aimed to explore tutor and students' perceptions of a PBL tutorial assessment tool and its implementation in terms of its summative and formative function.
Methods: Data were collected through focus group discussions (FGD). Four FGDs were conducted, 1 with tutors and 3 with undergraduate students from each year, recorded, and transcribed. A thematic analysis was used to identify key themes emerging from the discussions.
Results: Four main themes were defined based on qualitative data from FGDs, including acceptance and knowledge of the assessment tool, its impact on students learning (formative function), as well as concerns and challenges of its implementation (summative function). Although tutors and students generally accept the existing tutorial assessment, their conceptualization of its purpose is predominantly summative. The study identified limited assessment literacy among participants which, together with factors such as use of repeated scenarios, can influence both the summative and formative functions of tutorial assessments.
Conclusion: This study provides insights into the perceptions and experiences of tutors and medical students regarding the implementation of tutorial discussion assessments. Gaining insights into their viewpoints can lead to improvements like updating tutorial case scenarios, reforming tutorial rubrics and tutor training, and separating feedback sessions from grading.

Keywords: Assessment, problem-based learning, tutor, tutorial discussion, undergraduate medical education.

Proof Version

Problem-based learning (PBL), often called tutorial discussions, is a learning method currently applied in medical schools around the world, including Indonesia.[1] For this learning method, a small group of students engages in tutor-guided discussions, which can be triggered by a case scenario combining various disciplines about a specific learning topic.[1] Throughout the discussion, students are expected to analyze the phenomena/problems presented in the different case scenarios and express their opinions scientifically, thereby improving communication, presentation, and problem-solving skills.[2] When effective and when students actively participate in tutorial discussions, it can help students develop their mindset, facilitate the integration of basic and clinical medical knowledge, and provide insights into their future role as physicians.[3] Thus, in tutorials, the assessment is expected to act not only as a mechanism for summative assessments but as an integral component of the learning process.[4, 5]

The tripartite framework of "Assessment of Learning, Assessment for Learning, and Assessment as Learning" provides a useful lens for conceptualizing how different assessment purposes and practices can be integrated to support both accountability and learning-oriented goals in educational settings.[6–8] Assessment of Learning is typically associated with high-stakes, summative decisions about student progression or certification, whereas the other two consider feedback, learner agency, and self-regulation central to the assessment.[6, 8, 9] In principle, tutorial assessment in PBL should embody all three functions in a balanced and coherent way, with Assessment of

***Correspondence to: Dian Puspita Dewi,** Medical Education Unit, Faculty of Medicine, Universitas Diponegoro, Semarang, Central Java, Indonesia.
E-mail: dianpuspitadewi@fk.undip.ac.id
Received: March 30, 2026
Revised: May 30, 2026
Accepted: June 29, 2026

Open Access 2027 Dewi et al., published by Faculty of Medicine, Chulalongkorn University. This work is licensed under the Creative Commons Attribution 4.0 International License.

Learning providing credible summative judgments, Assessment for Learning supporting day-to-day learning, and Assessment as Learning fostering self-regulation.[6, 8]

Within this framework, the Assessment of Learning in PBL tutorials is often operationalized through global rating scales and scores that contribute to modulate or block examinations, thereby reinforcing a summative orientation.[4, 5, 10–12] In contrast, Assessment for Learning positions tutorial assessment as an ongoing, formative process whereby tutors use evidence of learning to adapt instruction and provide feedback that helps students close the gap between current and desired performance.[5, 6, 8] The last one, Assessment as Learning, extends this notion further by emphasizing students' active engagement in setting goals, monitoring their own learning, and reflecting critically on feedback, thus cultivating metacognitive skills and lifelong learning dispositions.[5, 6, 8, 9]

The significant impact on students makes the evaluation crucial for improving/selecting appropriate assessments.[11] In addition to demonstrated validity and reliability, assessment instruments need to be underpinned by clearly defined criteria and performance standards.[12] Use of explicit and objective criteria is generally preferred over global 10-point rating scales, often perceived as less valid and transparent when assessing complex behavior.[4, 12, 13] To enhance feasibility in routine use, the number of criteria should be limited to 4–5 maximum, thereby reducing the cognitive burden on tutors during daily assessments.[12] Furthermore, each item in the instrument should be supported by a clear operational definition to minimize variation in the interpretation of score levels and to promote higher consistency and clarity in judgment.[4, 12]

Beyond these psychometric and design considerations, assessment instruments should generate rich information to support constructive and timely feedback, enabling students to recognize specific areas for improvement and to regulate their own learning.[6, 7, 14] When so designed and implemented, assessment moves beyond a narrow Assessment of Learning, more effectively supporting the other two components, becoming an integral component of the teaching–learning process that encourages deep learning and the development of professional competence rather than merely reinforcing fact memorization.[5, 7]

Existing evidence suggests that tutorial assessment is still predominantly summative-oriented (assessment of learning) in many health professional programs, with limited or inconsistent enactment of Assessment for Learning and Assessment as Learning principles in day-to-day tutorial practices.[4, 7, 8, 10] Predominant use of summative-oriented assessment within tutorials may yield several positive effects, including fostering active participation in group discussions, enhancing students' motivation for self-improvement, and increasing their receptiveness to tutor feedback.[11, 15] Simultaneously, it has potential negative consequences, particularly a heightened emphasis on grades or rankings. Such performance-oriented focus can undermine the dynamics and cohesion of the tutorial group and detract from the collaborative learning that PBL intends to promote.[4, 11, 15]

During tutorials, the discussion can be assessed based on direct observation by the tutor using an instrument developed independently by the institution or adapted/modified to suit its specific needs.[16] Instruments such as checklists, Likert-type scales, and rubrics are frequently used for this purpose.[5, 9–12, 16, 17] However, it is unclear to what extent these tools provide clear assessment criteria, can deliver high-quality feedback, and effectively support students' self-assessment and reflection **(Figure 1)**.[4, 7, 10–12]

The present study therefore adopts this tripartite framework as an overarching theoretical lens to critically examine how the existing tutorial assessment tool is perceived by both students and tutors in terms of its summative and formative functions.

Materials and methods

This study was conducted from February to October 2025 in a selected university in Central Java, Indonesia (Faculty of Medicine, Universitas Diponegoro) which implemented tutorials as a learning method and used individual assessments in the form of a checklist to assess tutorial discussions for >10 years. Ethical clearance was obtained from the Health research ethics commission (KEPK) of the Faculty of Medicine, Universitas Diponegoro no. 014/EC/KEPK/FK-UNDIP/I/2025.

The Faculty of Medicine, Universitas Diponegoro follows a competency-based curriculum. Primary teaching methods include traditional and interactive lectures, laboratory practices, student-centered tutorial

Proof Version

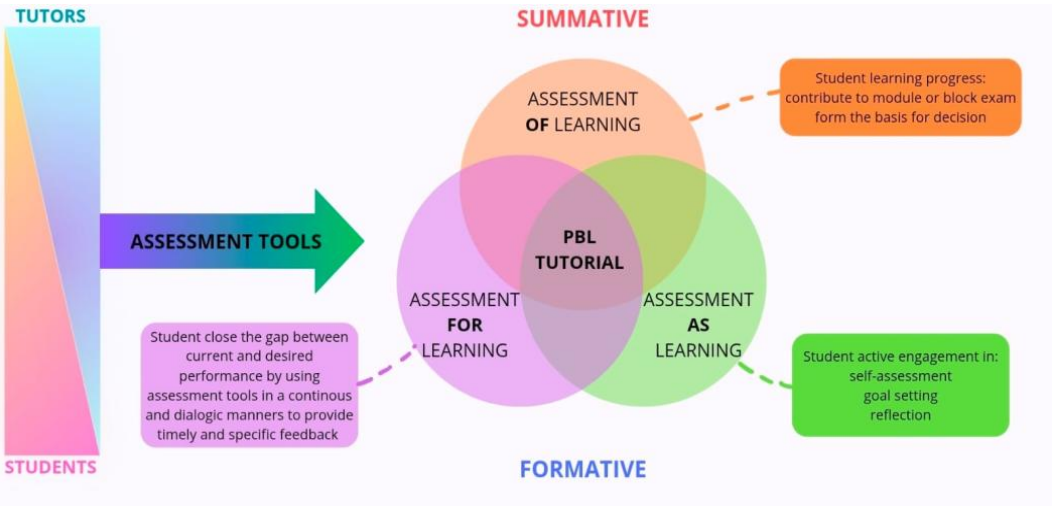


Figure 1. Tripartite framework modification (Assessment of Learning, Assessment for Learning, and Assessment as Learning) in a PBL tutorial assessment context. PBL, problem-based learning.

Proof Version

sessions, and clinical skill practices. Student's performance in each course is evaluated through a comprehensive assessment system consisting of written block examinations, objective structured clinical examinations, laboratory practice examinations, and tutorial assessments. Tutorial assessments contribute 10%–20% to the overall course grade. During tutorial discussions, tutors evaluate students using a specific checklist with a four-point Likert-type rating scale (1, lowest and 4, highest performance level), assessing discipline, active participation, relevance of discussion to the topic, and communication skills. Both students and tutors undergo formal preparation to ensure effective implementation of student-centered learning. Tutors are required to complete a mandatory faculty development workshop focusing on small-group facilitation techniques, constructive feedback strategies, and assessment implementation. Similarly, students receive an orientation session at the beginning of the course to familiarize themselves with tutorial sessions. However, at the time of the study, there was no separate session to explain the tutorial assessment tools in detail to the students.

This qualitative study used focus group discussions (FGD) to explore different viewpoints using participants' dynamics and thought sharing to enrich the discussion. Tutor participants were recruited through intended sampling. We included individuals with relevant expertise, includ-ing practical experience, formal training, and reflective practice in using individual assessments for tutorial discussion. An invitation was sent to eight faculty staff from various backgrounds such as clinical/pre-clinical lecturers and practicing/non-practicing doctors. Faculty staff were included in this study if actively facilitating tutorial sessions during the study period. To capture a comprehensive range of perspectives, no minimum tutoring experience time was enforced as a restriction. This allowed capturing the perceptions of both newly appointed and experienced tutors. Tutors on leave or not actively teaching during the study period were excluded. Another invitation was sent to undergraduate medical students each year considering representation based on their academic achievements (highest and lowest GPA score). Student inclusion criteria required participants to be active administratively, full-time students enrolled in the program who completed ≥ 1 tutorial. First year students were not included as they just started tutorials one month before the study started. Students were excluded if they withdrew from the course or declined to provide informed consent.

The research team consisted of three medical faculty members with substantial experience in medical education. Each one played varied roles in the research process. In terms of reflexivity, they acknowledged that their background as lecturers could potentially influence their interactions with participants and data interpretation. Team members tried to remain aware of their professional experience and biases, acknowledging how these might shape both data collection and analysis. Further, they engaged in reflexive discussions throughout the process to ensure the findings were rooted in participants' experiences.

The topic guide was prepared by DPD and RH to include aspects such as participants' level of understanding of the individual assessments for tutorial discussions, experience in its use, challenges/difficulties, and suggestions for improvement. Probing

and follow-up questions were allowed to minimize subjectivity in understanding participants' responses. To ensure data accuracy and trustworthiness, credibility, dependability, confirmability, and transferability criteria were applied. Credibility was established through data source triangulation by collecting data from both students and tutors, member checking, and peer debriefing. Dependability was assured by the code–recode procedure and a clear, consistent methodology. Confirmability was strengthened by participant validation, ensuring that the findings accurately reflected their perspectives. By providing a detailed description of the study context, demographics, data collection and analysis processes, and findings, readers can assess the applicability of the results to their own contexts. This transferability is achieved.

FGD were conducted on Zoom, one with tutors and the rest with students from each year. Using the Zoom platform for data collection facilitated gathering participants at the same time after working hours. Participants were asked to turn on their cameras during the FGD so that nonverbal cues could be detected by the moderator. Each FGD was recorded with the participant's consent and subsequently transcribed verbatim to facilitate data analysis. To facilitate open and detailed discussions, these were conducted in Bahasa Indonesia. Two of the authors (DPD and LPF) acted as moderators; each discussion lasted 1.5–2.5 h. Topics were completely covered after the 4th session.

The analysis was started by transcribing audio recordings verbatim in Bahasa Indonesia. Inductive thematic analysis was used to analyze transcripts and case study notes following a six-phase approach previously outlined by Ahmed S, Mohammed R.[18] Themes and sub-themes were derived inductively from the coded data, ensuring that they directly emerged from the participants' narratives. Discrepancies in coding were discussed and resolved until consensus, ensuring intercoder reliability and transparency in the analysis. Data analysis was led by DPD, who oversaw coding and topic development, while the other authors (LPF and RH) ensured peer review and validation.

Results

Eight tutors and 17 medical students participated in this study **(Table 1)**. Four main topics were defined based on qualitative data from FGDs including general opinions such as acceptance and knowledge of assessment tools, impact on student learning (formative function), as well as concerns and challenges faced on its implementation (summative function; **Figure 2).**

General opinions: acceptance and knowledge

Participants agreed that the current assessment for tutorial activities was necessary and should continue but also welcomed improvements.

"..... the existence of this assessment is very good, it really needs to be maintained, it is very helpful even though the tagline "nothing is perfect" is true, so we will continue to improve it." (SB-U-TP)

Although all faculty demonstrated an understanding of the assessments used, the same results were not observed among student participants, particularly for second-year students. This was not only true for students with low academic achievements but also for high-achievers.

"...in general, I assume the assessment is based on performance during the presentation and on the report. But to be honest, I don't know all details yet." (A4U1L)

Formative function: impact on student learning

Participants indicated that assessments had both positive impacts on student learning and behavior if implemented correctly, but they could have negative or undesirable consequences if not.

"*That will eventually get us used to other things, like being disciplined about being on time for class... And being active itself actually forces us, forces us to practice speaking*." (A2U2L)

"*The impact related to the assessment where the lecturer is quite passive, ... the students become less enthusiastic or less, less, less giving their 100% maximum effort during the discussion so it becomes less, less useful… there is an assumption that ah this lecturer is relaxed, even if we discuss carelessly, we will get a good grade*." (A4U3L)

Proof Version

Table 1. Participant demographic data.

Characteristic	Tutors (n = 8)				Students (n = 17)			
	n (%)	Min	Max	Mean ± SD	n (%)	Min	Max	Mean ± SD
Gender								
Male	2 (25.0)				8 (47.1)			
Female	6 (75.0)				9 (52.94)			
Age (year)		36	45	41.0 ± 3.0		19	21	20 ± 0.8
Education								
High school	-				17 (100)			
Bachelor's degree	-				-			
Master's degree	5 (62.5)				-			
Doctoral degree	3 (37.5)				-			
Working years (as lecturer)		5	17			N/A	N/A	N/A
Medical practice								
None	2 (25.0)				N/A			
General practitioner	2 (25.0)				N/A			
Laboratory	2 (25.0)				N/A			
Clinical specialist	2 (25.0)				N/A			
Academic year								
Second (2024)	N/A				4 (23.53)			
Third (2023)	N/A				6 (35.3)			
Fourth (2022)	N/A				7 (41.2)			
Academic achievement								
High	N/A				10 (58.8)			
Low	N/A				7 (41.2)			

N/A not applicable. SD, standard deviation.

Proof Version

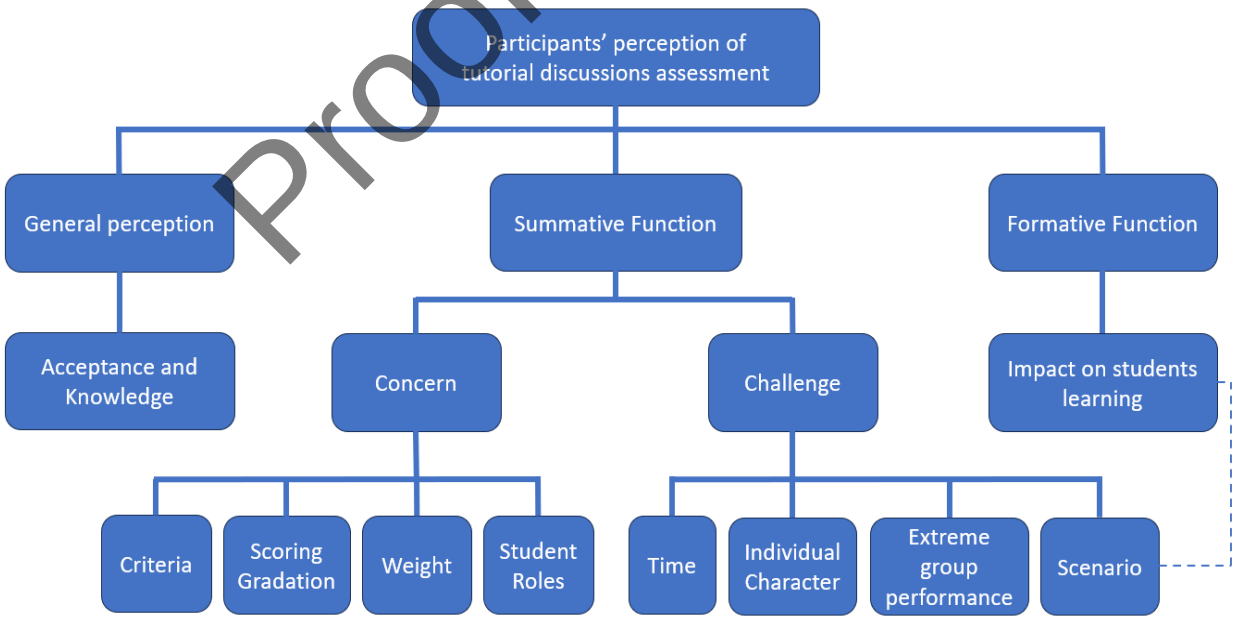


Figure 2. Thematic framework of participants' perception on tutorial assessment.

Summative function: concerns

Several concerns arose during the assessment implementation:

1. Tutors' understanding of assessment criteria

Tutors revealed that some assessment criteria were subjective since they lacked clear descriptions.

"*…Personally, when I assess students, I consider some to be active and some to be less active. Actually, if you look at the frequency of their speaking during a single discussion,... So, if someone only speaks once, of course, they get a lower score than someone who repeatedly raises their hand and speaks.*" (HW-L-P)

This is also a concern for students since it affects their grades. Even though they may perform the same in each tutorial discussion, they can receive significantly different grades when assessed by different tutors.

*"Perhaps one or several tutors' assessments are slightly different from the others."(*A4U1L)

2. Unclear scoring

Tutors found it difficult to distinguish among the assessment gradations for each criterion, which lacked clear distinctions. The narrow scoring gradations (1–4) often created dilemmas for tutors, as lowering a student's score by just one point would significantly lower their total grade.

"*…if inactive/passive, it's clear but 2 (insufficient) and 3 (sufficient) are a bit ambiguous*." (SB-U-TP)

3. Equal weighting between criteria

All assessment criteria currently have equal weight, so a student who is punctual but with poor communication can receive the same final score as a student who is five minutes late with good communication. Some tutors don't see this as a problem, but others consider it unacceptable and make their own adjustments or modifications.

"So for me, the total score is checked again. Why would they have the same score? Well, as long as differences between students are really obvious, I don't think the total scores should be the same." (MD-U-TP)

4. Students' additional role in the discussion

During discussions, some students play additional roles such as moderator/chair and protocol writing. Tutors consider that the current assessment system cannot accommodate the variety of student roles, so they independently adjust or modify assessments when assigning scores for those roles.

"*For that reason, I give the moderator and note-taker higher marks than the others. I mean, the maximum marks, just like those of active students*." (HW-L-P)

"*...I personally think the moderator can get a lower mark because they don't direct the discussion content well and active discussion is lacking. Unless they can summarize what people are saying and convey it appropriately, that's probably a good activity, not just pointing fingers. That's probably the part of the activity that matters*." (SB-U-TP)

Summative function: challenges

Challenges in implementing tutorial discussion assessments can be due to tutors, students, or other factors involved in the tutorial process such as case scenario.

1. Time constraints

Time-related challenges, particularly delays in starting discussions, often occur. Like any other learning process, tutorial discussions are held on a specific schedule. Tutors and students can both arrive late to the beginning of discussions, which makes assessments difficult.

"*...But what really needs to be emphasized is when the tutor was late... So even if the student was 15 minutes late, and the lecturer 30 minutes late, they're still considered on time. How is that possible?*" (SB-U-TP)

2. Individual character

Participants reported several characteristics that could pose challenges in the assessment, such as quiet students or tutors, and the tutor's habit of multitasking.

"*...Well, in my opinion, tutors often have other jobs, so they don't really pay attention to the students. When listening, they don't really assess the students. When bored/sleepy, they don't really follow the discussion, so their assessment of the students is very biased...*" (YN-U-P)

"*...during the discussion, some friends weren't very open. So, they are shy about speaking out about what they were learning, so they were judged less active than their friends*." (A2L3P)

Proof Version

3. Group performance

Tutors reported challenges related to student group performance during assessments. Extreme variations in group performance made it difficult for tutors to assess individuals.

"*...I often find that discussions (dynamics) are flat, no one offering new input or challenging previous students' answers. So, does that mean that the discussion is inactive? (How should I assess) with no discussion taking place, and only one-way communication?*" (AR-K-P)

"*...I once found a PBL group so extreme that someone wouldn't speak at all (didn't speak from start to finish), so I had to ask. Even when I asked, they still didn't answer, which made it difficult for me to rate them, but that's rare*." (AM-K-P)

4. Scenario

Another factor reported by both tutors and students was that the discussion scenarios used were identical as those from the previous year. Use of repetitive scenarios not only creates difficulties for tutors when conducting assessments but also results in students not learning as they should.

"*…They already know the answers to the problems, so perhaps it's necessary to create a new scenario, allowing students to have more active and critical discussions*." (YN-U-P)

"*... That's why most of the students do some preparation. Like dividing up the answers and questions so everyone can get a good score. I think that doesn't align with the goal of tutorial itself*." (A2U2L)

Discussion

This study shows that while tutors and students accept tutorial assessments, this does not inherently reflect a balanced or comprehensive conceptualization of its educational purposes. Emerging spontaneously at the outset of the FGD, this phenomenon highlights a substantial gap between the innovative aims of the curriculum and actual assessment practices, a misalignment that makes the implementation of tutorials less effective,[4, 7, 10, 15] as the effectiveness of the assessment tools depends heavily on the literacy of their users.[6, 8] Both tutors and students need a clear understanding of the aim of assessment, the processes involved, and the expected standards of performance, including timely and adequate information on how students will be evaluated.[6, 10]

Participants' conceptual understanding of tutorial discussion assessment leaned toward a summative orientation, largely perceived as a grading and ranking mechanism.[4, 8] This view is driven by limited transparency of assessment rubrics and criteria, as well as considering the tool almost exclusively as a summative instrument.[4, 8, 10] Such perception trigger performance anxiety and fear, rather than making assessment as a formative medium for Assessment for Learning or a vehicle for Assessment as Learning grounded in self-reflection and long-term professional growth.[4,10] This disproportionately summative focus underscores a critical need for regular faculty development, systematic evaluation of tutor training, and more explicit student orientation about self-regulatory assessments.[4, 8, 17, 19] Accordingly, tutorial assessments should move beyond a rigid grading ritual and reimagined as a process actively cultivating lifelong learning and metacognitive self-reflection.[4–6, 8, 9]

Variation in understanding between tutors in assessing student performance, known as interrater reliability, is a critical challenge to PBL success. This refers to the inconsistent scores granted by different tutors for students with the same performance level.[19] These variations are often influenced by the limitations in perception, mental workload, and cognitive effort of each tutor. This variation creates a feeling among students that the evaluation process is unfair and inconsistent,[4,10,12] which undermines the PBL dynamic because students become more focused on how to get a good grade from a particular tutor than on the discussion itself.[12] As a result, the entire tutorial session becomes demotivating for students.[5]

Many tutors find it difficult to establish objective criteria to accurately differentiate grades between students due to different interpretations of assessment guidelines.[4, 11] This often leads tutors to make their own subjective adjustments or modifications.[10] Clear rubrics enhance objectivity by establishing standardized performance standards, allowing both tutors and students to understand exactly what is expected during the tutorial.[16] This could minimize interrater variation by providing explicit criteria for complex behaviors, ensuring that different tutors assign consistent scores to similar levels of student performance.[10,12] Standardized rubrics also help reduce subjectivity and bias in tutor evaluations by providing a structured framework that decreases reliance on personal perceptions or mental workload.[5,12] Further, the transparency provided by

Proof Version

clear rubrics makes students perceive the evaluation process as more fair and reliable, which can reduce anxiety and encourage them to focus on the learning process rather than just the score.[4, 5, 13]

From the perspective of the tripartite framework, misalignment in weighting reflects limited conceptualization of the assessment's broader educational purposes.[4] In PBL tutorials, assessment is expected to address not only cognitive outcomes but also key processes such as self-directed learning, teamwork, and clinical reasoning.[4, 10, 11, 14, 15, 17] If an institution wishes to emphasize self-directed learning and higher-order reasoning, these domains should carry substantial weight in the overall score,[7, 13] thereby supporting formative functions and ensuring constructive alignment between curricular objectives and the assessment system, as perceived by students.[8–10] When a basic discipline such as punctuality is weighted similarly to critical reasoning, a core cognitive goal of PBL, the achievement of complex competencies can be obscured and assessment is experienced primarily as an Assessment of Learning.[4,7] This presents a practical dilemma for tutors. Since assessment significantly influences learning, generous scores for simple behaviors can inadvertently reinforce pragmatic and score-oriented strategies rather than fostering the reflective and self-regulated engagement that Assessment for Learning and Assessment as Learning aim to cultivate.[8, 11]

In tutorials, students play different roles (chair, protocol writer, member). Applying uniform criteria and weighting to all roles is often criticized because the chair's role is process management while members focus on the substance of the material. Adjusted weighting and criteria based on role would allow for a fairer, more accurate assessment of students' actual contributions to group dynamics.[10]

Tardiness, from both tutors and students, is a serious challenge in tutorials, directly impacting the quality of discussions and assessment validity. Tutors play a crucial role as facilitators and role models.[19] However, tutor tardiness is common and has negative consequences, such as suboptimal discussion time. [19] In some cases, time loss prompts tutors to assign learning objectives directly at the end of the first session rather than allowing students to formulate them independently, a practice in conflict with core PBL principles which further constrains opportunities for Assessment for Learning and Assessment as Learning.[8,19] Although discipline and punctuality are frequently listed as professional behavior criteria in rubrics, their implementation is often considered inconsistent.[5] Many students feel that their PBL assessments do not fairly consider punctuality or their actual contributions to discussions.[10]

The character and behavior of both tutors and students profoundly influence the dynamics of discussions and the objectivity of assessments. The interaction between these cognitive and affective components forms a "social representation" that determines how assessments are perceived and implemented.[4] Within the tripartite framework, biased or inattentive tutor behavior, such as frequent distraction or repeatedly leaving the room, weakens the validity of Assessment of Learning judgments and erodes the potential of tutorials as spaces for Assessment for Learning and Assessment as Learning by reducing meaningful feedback opportunities.[6, 9] Less active or shy students are often unable to demonstrate their genuine knowledge because they feel pressured by the grading system. They often assume that good grades belong only to those who good at talking or conveying a lot of information, rather than those who provide deeper analysis, reinforcing a narrow, performance oriented summative culture and marginalizing deeper learning.[4]

Moreover, variation in group performance is a dynamic phenomenon that significantly impacts the objectivity and quality of the assessment.[14,20] Passive group members undermine reliability of assessments by limiting the observational data tutors can use to evaluate competency[4,12] and often create opportunities for other students to overly dominate discussions.[15,19] As a result, tutors tend to assign grades based on the performance of these dominant individuals, while the true contributions of the group as a whole are overlooked, resulting in assessments that do not reflect true teamwork.[11,19] Addressing this issue requires strong tutor facilitation skills to encourage participation from all members and use of assessment rubrics that emphasize the quality of analysis rather than speaking frequency.[12,15]

Finally, repeated scenarios already familiar to students from previous years can hinder knowledge acquisition. Students often simply repeat information from lecture notes or summaries from seniors rather than conducting independent information searches through in-depth study.[21] Within the tripartite framework, such practices make tutorials predictable Assessment of Learning checkpoints, rather than rich opportunities for Assessment for Learning or Assessment as Learning. In addition, the level of group engagement heavily depends on the quality of the scenarios or cases presented.[15, 20] Uninteresting or overly easy scenarios can demotivate discussion, so that group performance is low not due to student

Proof Version

incompetence, but rather due to the lack of stimulating critical thinking.[15,19] In such cases, tutorial assessment risks capturing only surface-level participation and reinforcing a summative function, while failing to elicit the kind of deep reasoning, self-regulation, and reflective dialogue that Assessment for Learning and Assessment as Learning are designed to promote.

The limitation of this study is that it was only conducted at a single institution, which may limit the generalizability of some findings to other medical schools. Therefore, future research should incorporate larger and more diverse samples from multiple institutions.

Conclusion

This study highlights that while both students and tutors highly value tutorial assessment, its practical application remains conceptually confined within a rigid Assessment of Learning framework. Systemic institutional vulnerabilities, most notably recycled case scenarios and ambiguous grading structure, actively distort the assessment's validity and overshadow its formative potential. To optimize tutorial assessments in medical education, institutions must deliberately transition from purely summative grading toward a balanced, learning-oriented culture anchored in the tripartite framework.

Recommendations for institutions include continuous updating and variation of tutorial case scenarios, transparency of rubrics and training, and separation of grades from feedback. Continuously changing tutorial case scenarios can prevent reliance on inherited knowledge from senior colleagues. This is essential for restoring the validity and reliability of tutor summative assessments. Furthermore, to unlock its Assessment for Learning potential, institutions must reform tutorial rubrics by replacing vague criteria by explicit, behaviorally-based descriptors. This transparency should be combined with long-term faculty development workshops that train tutor's to become providers of specific, actionable feedback that helps students close their learning gaps. Decoupling grades from feedback involves delivering feedback in low-stakes or non-graded contexts, so that students can reflect honestly, experiment with new strategies strategies, and develop self-regulatory and metacognitive skills, rather than focusing on protecting or maximizing their grades.

Authors contributions

DPD, RH, and LPF contributed substantially to the concept and design of this study, acquiring the data, reviewing the literature, and its analysis and interpretation. DPD and LPF contributed substantially to acquiring the data. DPD contributed to drafting the manuscript. RH and LPF edited the manuscript critically for important intellectual content. All authors approved the final version submitted for publication and take responsibility for statements made in the published article.

Acknowledgements

The authors would like to express gratitude to the assistant who helped the authors in technical aspects of data collection process until the data was ready for analysis.

Conflict of interest statement

There is no conflict of interest.

Data sharing statement

All data generated or analyzed during the present study are included in this published article. Further details are available for non-commercial purposes from the corresponding author on reasonable request.

ORCID

Dian Puspita Dewi https://orcid.org/0000-0002-5699-1152
Lathifa Putry Fauzia https://orcid.org/0009-0002-9113-0881
Ryan Halleyantoro https://orcid.org/0000-0003-2270-8299

References

1. Fitri AD. Penerapan problem based learning (PBL) dalam Kurikulum Berbasis Kompetensi. Jambi Med J Jur Ked dan Kes 2016;4.
2. Sari S. Validitas dan reabilitas metode penilaian students oral case analysis (SOCA) pada mahasiswa tahap sarjana kedokteran. J Pendidik Kedokt Indones 2013;2:97.
3. Ding X, Zhao L, Chu H, Tong N, Ni C, Hu Z, et al. Assessing the effectiveness of problem-based learning of preventive medicine education in China. Sci Rep 2014;4;5126.
4. Bollela VR, Gabarra MH, da Costa C, Lima RC. Students and tutors' social representations of assessment in problem-based learning tutorials supporting change. BMC Med Educ 2009;9:30.

Proof Version

5. Elizondo-Montemayor L. Formative and Summative Assessment of the Problem Based Learning Tutorial Session Using a Criterion Referenced System. JIAMSE 2004;8–14.
6. Schellekens LH, Bok HGJ, de Jong LH, van der Schaaf MF, Kremer WDJ, van der Vleuten CPM. A scoping review on the notions of Assessment as Learning (AaL), Assessment for Learning (AfL), and Assessment of Learning (AoL). Stud Educ Eval 2021;71:101094.
7. van der Vleuten CPM, Schuwirth LWT. Assessment in the context of problem-based learning. Adv Health Sci Educ Theory Pract 2019;24:903–14.
8. Zeng W, Huang F, Yu L, Chen S. Towards a learning-oriented assessment to improve students' learning—a critical review of literature. Educ Asse Eval Acc 2018;30:211–50.
9. Lakhtakia R, Otaki F, Alsuwaidi L, Zary N. Assessment as learning in medical education: feasibility and perceived impact of student-generated formative assessments. JMIR Med Educ 2022;8:e35820.
10. Animaw Z, Asaminew T. Medical students' perspective on assessment mechanism during problem-based learning at Debre Tabor University: an explanatory mixed study. Adv Med Educ Pract 2023;14:859–73.
11. Khadafianto F, Rahayu G, Suryadi E. Rancangan perbaikan penilaian tutorial berdasarkan sudut pandang tutor dan mahasiswa serta literatur. J Pendidik Kedokt Indones 2016;5:101.
12. Khawaji B, Masuadi E, Alraddadi A, Khan MA, Aga SS, Al-Jifree H, et al. Tutor assessment of medical students in problem-based learning sessions. J Educ Health Promot 2024;13:237.
13. Piryani RM, Piryani S, Ansari N, Hussain A. Assessment of problem-based learning (PBL): poster presentation as a tool of assessment. Indones J Educ Res Rev 2022;5:394–403.
14. Bhandary S. Problem based learning curriculum and process assessment system for the undergraduate competency based medical education/ : experiences from nepal. Arch Med Health Sci 2021;9:331–6.
15. Khoiriyah U. Implementation of summative assessment in PBL tutorials: cultural perspectives. J Pendidik Kedokt Indones 2018;7:64.
16. Akbar S, Claramita M, Kristina T. Pengembangan kuesioner penilaian proses belajar problem-based learning dengan model SPICES. J Pendidik Kedokt Indones 2014;3:137.
17. Cordero MA, Abdur Rashid N, El Hasnaoui R, Ganesh K, Khired Z. Tutors and self-assessment of medical students' performance in pre-clinical problem-based learning tutorial sessions. MedEdPublish (2016) 2018;7:265.
18. Ahmed S, Mohammed R, Nashwan A, Ibrahim R, Qadir A, Ameen B, et al. Using thematic analysis in qualitative research. J Med Surg Public Health 2025;6:100198.
19. Oktafany. Hubungan kinerja tutor dengan kegiatan belajar mandiri dan pelaporan hasil belajar mandiri dalam diskusi problem based learning di Fakultas Kedokteran Universitas Lampung [thesis]. Jakarta: Universitas Indonesia; 2016.
20. Kassab SE, El-Baz A, Hassan N, Hamdy H, Mamede S, Schmidt HG. Construct validity of a questionnaire for measuring student engagement in problem-based learning tutorials. BMC Med Educ 2023;23:844.
21. Lim WK. Problem based learning in medical education: handling objections and sustainable implementation. Adv Med Educ Pract 2023;14:1460.

Proof Version