



Factors That Determine the Perceived Effectiveness of Peer Feedback in Collaborative Learning: a Mixed Methods Design

Dayane Daou^{1,2} · Ramzi Sabra³ · Nathalie K. Zgheib³

Published online: 19 May 2020

© International Association of Medical Science Educators 2020

Abstract

Introduction Peer assessment has been promoted as a valuable approach to formative assessment to support learning and peer professionalism. This mixed methods study employed a conceptual framework to explore the factors that enhance the perceived effectiveness of formative peer assessment in the context of team-based learning as a form of collaborative learning.

Materials and Methods The volume and quality of written peer comments of two medical school classes at three time points were analyzed. Focus groups were then conducted to clarify issues that appeared in the quantitative data and to explore other emerging dimensions.

Results There was a notable deficiency in both the volume and quality of the comments provided, with no improvement over time. Several factors were identified, including some that are logistical and operational and can be corrected easily, such as the timing of the assignments. Others that stood out as major substantive issues and/or limitations related to the students' conceptions of the purpose of the peer assessment and to their interpersonal variables.

Discussion There were social disincentives for students to provide constructive feedback to peers with whom a continuing working relationship is necessary. There was also an inconsistency between the quality of the peer feedback being typically shallow and lacking in substance, and students considering it beneficial.

Conclusion The findings identify factors that need to be addressed in order to ensure the quality and effectiveness of formative peer assessment among medical students.

Keywords Medical students · Mixed methods · Peer assessment · Peer feedback · Collaborative learning · Team-based learning

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s40670-020-00980-7>) contains supplementary material, which is available to authorized users.

✉ Ramzi Sabra
rsabra@aub.edu.lb

✉ Nathalie K. Zgheib
nk16@aub.edu.lb

¹ Medical Education Unit, American University of Beirut, Beirut, Lebanon

² Department of Anesthesiology, Faculty of Medicine, American University of Beirut, Beirut, Lebanon

³ Program for Research and Innovation in Medical Education, Department of Pharmacology and Toxicology, Faculty of Medicine, American University of Beirut, PO. Box 11-0236, Riad El Solh, Beirut, Lebanon

Introduction

Over the past decades, peer assessment has been promoted as a valuable approach to formative assessment to support learning and peer professionalism [1–4]. During formative peer assessment of any teaching and learning method, evaluations typically include, in addition to or instead of quantitative scales or scores, qualitative comments that are referred to as “peer feedback.” Since peer feedback is a core component of peer assessment, the impact of variable contents and quality of that feedback has been evaluated in various contexts using different approaches or rubrics [1, 5–14]. Good peer feedback should enhance learning through timely and periodic checks of performance and pointing out strengths, weaknesses, and areas of improvement. Poor feedback, however, can be seen as more of an inconvenience and threat than a benefit.

Collaborative learning is an instructional approach whereby students work together in small groups towards a common goal [15]. Team-based learning (TBL) is a form of

collaborative learning and presents a distinct context for the exploration of peer feedback in the setting of collaborative learning. TBL is problem-based, interactive, and student-centered. It entails a “flipped-classroom” approach to facilitating learning whereby teams apply knowledge gained from self-study onto group application exercises [16, 17]. Similarly to the educational philosophy, theory, and evidence underlying all collaborative learning approaches, effective learning from TBL relies on two key mechanisms: positive interdependence and individual accountability [4]. With the TBL instructional design, students are accountable to their instructor and peers through the individual and group readiness assurance test and their performance on the team problems. Thus, peer assessment that entails periodically filling a peer evaluation form is at the heart of the TBL process to keep students accountable to their teammates concerning pre-class preparation and contribution to the interpersonal group dynamics and team productivity [18].

To our knowledge, only one study evaluated the impact of the quality of peer feedback on medical students [7], and no such analysis has yet been performed in collaborative learning in general, and TBL in particular [19] in the medical field. Besides, the majority of the studies that addressed this topic focused on the psychometric qualities (i.e., reliability and validity) of peer assessment, on its impact on student performance, and on the students’ attitudes towards it [20–24]; only a few assessed the quality of the peer feedback [7] or elicited factors that drive the effectiveness of peer assessment [1, 25, 26].

Several studies have reported on students’ attitudes to peer assessment and feedback, which can conceivably influence the quality of the feedback they provide. In fact, several reported on students’ reluctance to engage in this process and related this to many factors including the following: fear of negatively affecting relationships among them, not wanting to harm each other, fear of bias or lack of objectivity, lack of confidence in their abilities to provide feedback, concern that the feedback may be used for summative purposes and making judgment by the administration, concern about confidentiality and anonymity, and time constraints amid too many duties, among others [2, 6, 7, 18, 27]. It is possible to classify these factors into two main categories: one relating to the peer assessment process (i.e., in terms of logistics such as timing, format, frequency, and anonymity, as well as its purpose—i.e., formative or summative) and the other relating to the students’ interpersonal variables. Van Gennip et al. [27] developed and empirically validated the concept of peer assessment as a collaborative learning activity whereby interpersonal variables (psychological safety, trust in self and peer, value diversity, and interdependence) interact with the peer assessment (i.e., perceived purpose and characteristics of the peer assessment) for successful learning. We modified Van Gennip’s conceptual framework to examine how these factors

affect the volume and quality of written peer feedback and the perceived benefit from it (Fig. 1), and added to it the element of time (or experience), based on studies which suggested that giving effective peer feedback improved with time and experience [28–30]. We hypothesized that the characteristics of the peer assessment process, the students’ perceived conceptions of its purpose, their experience with peer assessment, and their interpersonal variables determine the volume and quality of peer feedback, which, in turn, determine its perceived effectiveness. Our specific research questions were:

1. What is the volume and quality of the written feedback given to students as part of the periodic peer assessment in TBL, and how does it change with time?
2. What are the factors that determine the volume and quality of the written feedback given to students as part of the periodic peer assessment in TBL?
3. What are the students’ perceptions of the benefits of peer feedback?

Methods

Design

We used a mixed methods approach whereby quantitative and qualitative data collection, analyses, and integrations at the level of data interpretation were carried out on the same population. The volume and quality of the peer feedback and their variability over time were collected from quantitative analysis of comments available on the peer evaluation forms (research question 1). An explanatory design model was then applied to the qualitative data collected from focus groups to elucidate factors that explain the results that appeared in the quantitative data, and to explore other emerging dimensions (research questions 2–3) [31, 32]. Themes were deductively informed using the proposed conceptual framework illustrated in Fig. 1.

The study was approved by the Institutional Review Board (IRB) of the American University of Beirut under the expedited category.

Participants

Upon the initiation of the new Impact Curriculum at the American University of Beirut Faculty of Medicine (AUBFM) in 2013, a large number of TBL sessions were introduced. The present study included medical students of the Class of 2017 (i.e., the first Impact Curriculum class that graduated in 2017; $N = 102$) and the Class of 2019 (the class that graduated in 2019; $N = 105$) at the AUBFM. Students in both classes were randomly allocated to different 5–6

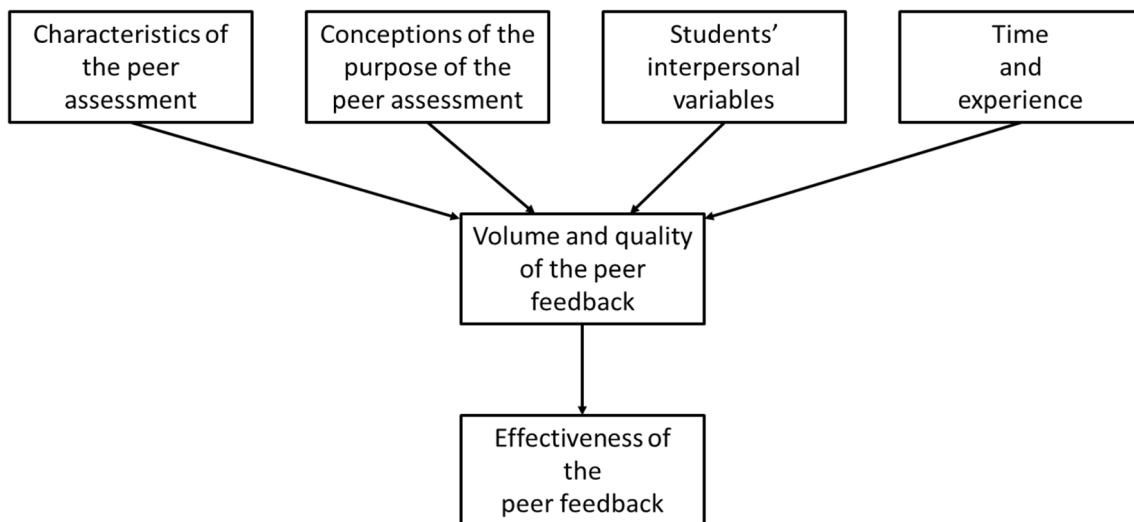


Fig. 1 Conceptual framework: factors that affect the volume and quality of the peer feedback and determine its effectiveness

membered teams on three occasions: at the beginning and middle of year 1, and at the beginning of year 2. Thus, each student belonged to 3 different teams during the two years. We refer to the peer evaluations provided within each as batches 1, 2, and 3 respectively (please refer to Table 1 and Zgheib et al. [33] for a more detailed description of the extensive implementation of TBL in the AUBFM new curriculum).

We chose the Class of 2017 because it was the first of the new Impact Curriculum and the Class of 2019 to evaluate the situation after 2 years, at a time when the curriculum had become well established.

Research Question 1: Quantitative Evaluation of the Volume and Quality of Peer Feedback over Time

Instrument

An electronic TBL peer evaluation form was periodically and anonymously filled by all medical students at different time points during years 1 and 2 of the medical program. The form was adapted from Paul Koles of the Boonshoft School of Medicine at Wright State University [18, 34] (Sup. Box 1). In addition to some objective questions that led to a summative score [33], the form included 2 open-ended questions that ask about (1) the single most valuable contribution the person makes to the team and (2) the single most important way the person could more effectively help the team. These questions were non-mandatory and used for formative purposes. In the present study, 6 peer feedback processes were used for analysis corresponding to each of the 3 team distributions for the 2 classes (Table 1).

A workshop on the method and philosophy of TBL, including the importance of peer evaluation, was delivered to all medical students at the beginning of year 1. In addition, students received education and training on providing peer

feedback during one full session of the “Learning Communities” course (see Sup. Box 2 for the learning objectives of that session). The criteria that constitute helpful feedback were introduced and clarified with several examples and exercises. The criteria stated that helpful feedback should be descriptive, specific, concerned with observed behavior, and constructive, and should include suggestions for improvement [34, 35]. Both medicine classes received the same training.

Volume

The number of comments received by each student was counted. The minimum possible number of comments received per student was zero (this is for those who did not receive any comments) and the maximum was 5 or 6 depending on the number of team members (Table 1).

Quality

The quality of each feedback was evaluated in two ways. First, words were extracted from the written comments, counted, and categorized into thematic areas. One of the investigators (DD) initially independently extracted the words and tallied them into proposed thematic areas. The three investigators then discussed and reached a consensus on the final sorting of the words into eight thematic areas. Then, each feedback was rated into a quality score that is adapted with slight modifications from Paul Koles’ [34] using the pre-specified criteria of “helpful feedback” that were shared with and explained to students as described above. Each feedback was assigned a value of 0–3 as such: (0) irrelevant comment such as “the king” or “best TBL scratcher”; (1) descriptive comment, focused on a single aspect, not specific enough, and with little practically useful information, e.g., “kind,”

Table 1 Data collected from the peer evaluation forms after three batches of TBL teams during medicine 1 and 2 of the classes (medical students of classes of 2017 and 2019 are expected to graduate in 2017 and 2019 respectively) of 2017 and 2019

	Class of 2017			Class of 2019		
	Medicine year					
	Medicine 1		Medicine 2	Medicine 1		Medicine 2
	Batches					
	Batch 1	Batch 2	Batch 3	Batch 1	Batch 2	Batch 3
Number of students	102	102	100	105	105	105
Peer feedback collection						
Batch of TBL teams	Beg. of year	Mid. year	Beg. of year	Beg. of year	Mid. year	Beg. of year
Time points analyzed	October 2013	April 2014	October 2014	October 2015	March 2016	November 2016
Number of TBLs by time point within each batch	10	11	12	11	7	15
Volume of the peer feedback						
Total number of comments	194	178	127	67	58	153
Number (%) of students who received comments	98 (96.08)	91 (89.22)	84 (84.00)	56 (53.33)	51 (48.57)	84 (80.00)
Number of comments per student: Mean \pm SEM	1.90 \pm 0.12	1.75 \pm 0.12	1.27 \pm 0.08	0.63 \pm 0.06	0.55 \pm 0.06	1.47 \pm 0.11
Quality of the peer feedback						
Distribution of words according to thematic areas (%)						
Personality traits	22.38	28.19	31.31	23.76	23.71	22.71
Contribution/participation	22.06	21.08	22.56	17.82	21.65	31.58
Cognitive abilities	29.31	20.83	20.20	20.79	18.56	18.28
Preparedness	9.02	12.25	5.72	11.39	11.34	7.76
Communication	7.73	3.92	6.40	9.90	13.92	8.03
Professionalism	4.45	6.13	7.74	4.46	7.22	3.05
Feedback	3.70	5.64	3.37	6.93	3.09	4.99
Leadership	1.45	1.96	2.69	4.95	0.52	3.60
Quality scoring						
Score ¹ : Mean \pm SEM	1.74 \pm 0.05	1.24 \pm 0.06	1.59 \pm 0.08	1.86 \pm 0.09	1.34 \pm 0.12	1.56 \pm 0.06
Intraclass correlation (95%CI) ²	0.85 (0.70–0.91)	0.80 (0.56–0.89)	0.94 (0.92–0.96)	0.85 (0.63–0.93)	0.92 (0.77–0.96)	0.88 (0.8100.92)

SEM, standard error of the mean

¹ A quality score was assigned a value of 0–3 as such: (0) irrelevant comment; (1) descriptive comment, focused on a single aspect, not specific enough, and with little practically useful information; (2) more detailed, covers several aspects but still not very specific as to be helpful practically; and (3) very useful, multifaceted, detailed, and specific feedback

² These were calculated using the Fleiss kappa test

“helpful,” “participates”; (2) more detailed, covers several aspects but still not very specific as to be helpful practically, e.g., “Engages everyone and promotes discussion always”; and (3) very useful, multifaceted, detailed, and specific feedback, e.g., “Although [X] does a great job at making sure that we analyze and are able to explain how and why we get to a specific answer, I feel like she is not benefiting as much as she should from the TBL session. She is often going back to questions after the group has moved forward in the case and I feel like better preparation at home would allow her to follow the pace

of the group with more ease.” The three investigators initially independently scored the quality of the comments, and any discrepancies were then resolved by the three discussing the comments and reaching a consensus.

Data Analysis

Data were entered into SPSS (v. 24, IBM, Armonk, NY, USA). Means \pm standard error of the mean (SEM) were computed for the volume and quality rating of the comments. The

number of words in each thematic area was added, and the frequency (%) of words within each thematic area was computed and graphically compared among the different student batches and classes. In order to establish the baseline before the three investigators reached a consensus on the quality rating, intraclass correlations with 95% confidence intervals (CI) were calculated for each batch using the Fleiss kappa test.

Research Questions 2–3: Qualitative Focus Groups

In January 2017, an invitation email was sent to all students from the classes of 2017 and 2019. Four focus groups were conducted with 5–7 students per session to a total of 22 students. The majority of the participants were medicine 2 students of the Class of 2019. Only 2 were current medicine 4 students of the 2017 Class. Those who participated received a 25\$ voucher from a bookshop as compensation for their time.

The focus group sessions were carried out by a moderator and a note-taker. A number of probing and guiding questions were used (Supp. Box 3), and answers were tape-recorded. The sessions' duration varied between 60 and 90 min. All focus group participants signed an IRB-approved informed consent form.

Focus group conversations were transcribed verbatim. A qualitative systematic and iterative thematic analysis was then conducted as such: the transcripts were independently manually coded by the 3 investigators. The compiled codes were then compared and unified by one author. These were then transferred onto an Excel spreadsheet and consolidated into categories by the 3 authors over 3 meetings. The developed matrix allowed for the extraction of recurring themes and emerging patterns. The focus was on main themes, commonalities and conflicting views of participants, and relationships between themes [36]. Data collection stopped when it no longer generated new information (data sufficiency). Findings were summarized as text with relevant quotations. This was sent back by email to those who participated in the focus groups for feedback and confirmability.

Results

Research Question 1: Quantitative Evaluation of Volume and Quality of Peer Feedback over Time

Volume

The overall percentage of students receiving comments in both classes ranged between 49 and 96% over the six batches analyzed (Table 1). The average for the Class of 2017 was 90% while it was 61% for the Class of 2019.

The mean number \pm SEM of received comments per student in both classes was low and ranged between 0.55 and

1.90. For the Class of 2017, the number of comments per student was the highest with the first batch (1.90 ± 0.12). It was lower with the second batch (1.75 ± 0.12) and much lower with the third batch (1.27 ± 0.08). The number of comments for the Class of 2019 started as very low with the first two batches (0.63 ± 0.06 and 0.55 ± 0.06 respectively) and was the highest with the third batch (1.47 ± 0.11) (Fig. 2 and Table 1).

Quality

In terms of quality, there were eight thematic areas around which the comments revolved. These were, in decreasing order of frequency, the following: personality traits (28.0%), contribution or participation (24.1%), cognitive abilities (18.4), preparedness (9.4), communication (7.8), feedback (5.0%), professionalism (4.9%), and leadership (2.4%) (see Supp. Table 1 for examples of words extracted and categorized under each of the eight thematic areas). For the Class of 2017, cognitive abilities were mostly mentioned in the comments of the first batch (beginning of year 1) while personality traits dominated in the third batch (beginning of year 2). For the Class of 2019, personality traits dominated in the first batch, but contribution or participation was more mentioned in the third batch (Fig. 3 and Table 1).

As for the quality scores, there was a significantly high intraclass reliability for all 6 batches (Table 1). The mean quality rating of the peer feedback was overall low and ranged between 1.24 and 1.86. For both classes, the mean score \pm SEM was highest with the first batch (1.74 ± 0.05 and 1.86 ± 0.09 for classes of 2017 and 2019 respectively) and lowest with the second batch (1.24 ± 0.06 and 1.34 ± 0.12). For the

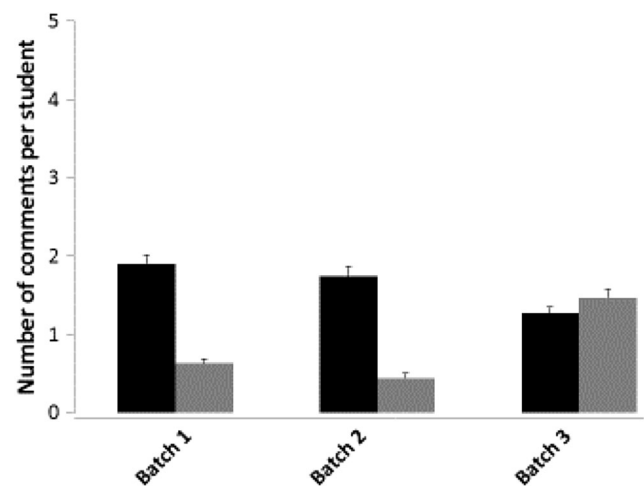


Fig. 2 Mean and standard error of the mean of the number of peer evaluation comments received per student over the 3 batches of TBL teams during medicine 1 and 2 of the classes of 2017 and 2019. The three different batches depict three different team allocations for the whole class (see Table 1). Batch 1 = beginning of medicine 1, Batch 2 = middle of medicine 1, and Batch 3 = beginning of medicine 2. Black = Class of 2017 and gray = Class of 2019

third batch, although the quality ratings for both classes (1.59 ± 0.08 and 1.56 ± 0.06) were higher when compared with the second batches, they were still lower than the very first batches (Fig. 4 and Table 1).

Research Questions 2–3: Qualitative Focus Groups

Analysis of focus group transcripts associated with research questions 2 and 3 identified themes relating to the reasons behind the low numbers and quality of comments on the peer evaluation form, and how these themes correspond to the conceptual framework we employed (Table 2). Participants in the focus groups who read the summarized qualitative results agreed with the themes identified and indicated that they “pretty much sum up [their] sentiments” and “capture the essence of what [they] were saying quite well.”

Research Question 2: Factors Behind the Low Volume and Quality of Comments on the Peer Evaluation Form

Several reasons for the low volume and quality of comments on the peer evaluation form were identified. What stood out is that, although students believe that the process ought to be perceived as constructive, many of them feel that it is a *judgment and evaluation rather than formative feedback*. In this context, they believe (and fear) that the comments may be used by the administration to grade students or to make decisions (e.g., selection for residency spots). Thus, they are not fully truthful or honest in filling the evaluations for fear of harming or “branding” their classmates.

Besides, although supposedly *anonymous*, the students indicate that frequently they can guess who wrote what about them; this could create *tension* and might drive them to

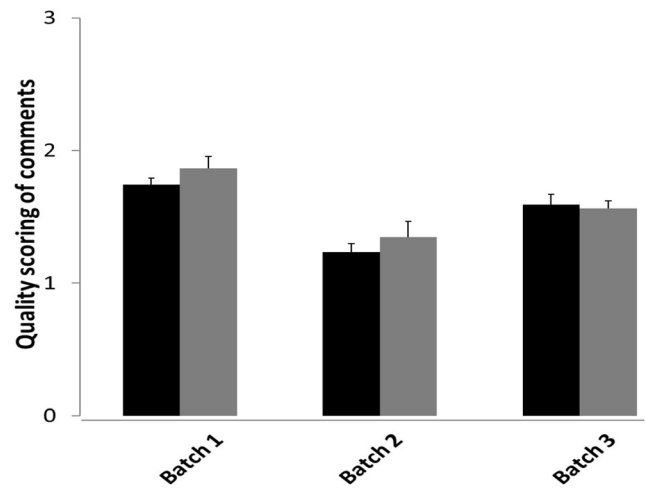


Fig. 4 Mean and standard error of the mean of the quality rating of the peer evaluation comments over the 3 batches of TBL teams during medicine 1 and 2 of the classes of 2017 and 2019. The three different batches depict three different team allocations for the whole class (see Table 1). Batch 1 = beginning of medicine 1, Batch 2 = middle of medicine 1, and Batch 3 = beginning of medicine 2. Black = Class of 2017 and gray = Class of 2019. A quality score was assigned a value of 0–3 as such: (0) irrelevant comment; (1) descriptive comment, focused on a single aspect, not specific enough, and with little practically useful information; (2) more detailed, covers several aspects but still not very specific as to be helpful practically; and (3) very useful, multifaceted, detailed, and specific feedback

reciprocate in kind, and will prevent them from writing negative comments. In order not to risk revealing their identity, students resort to writing vague rather than specific comments about each other.

Students also prefer to focus on positive aspects and comments in order to *motivate* each other. They do not want to be perceived as traitors or “rats.” This attitude increases their unity against the administration. In addition, writing positive

Fig. 3 Frequency distribution (percentages) of the words extracted for each thematic area from the peer evaluation written comments over the 3 batches of TBL teams during medicine 1 and 2 of the classes of 2017 and 2019. The three different batches depict three different team allocations for the whole class (see Table 1). Batch 1 = beginning of medicine 1, Batch 2 = middle of medicine 1, and Batch 3 = beginning of medicine 2. Black = Class of 2017 and gray = Class of 2019. Numbers depict frequency (%) of extracted words for each thematic area per batch

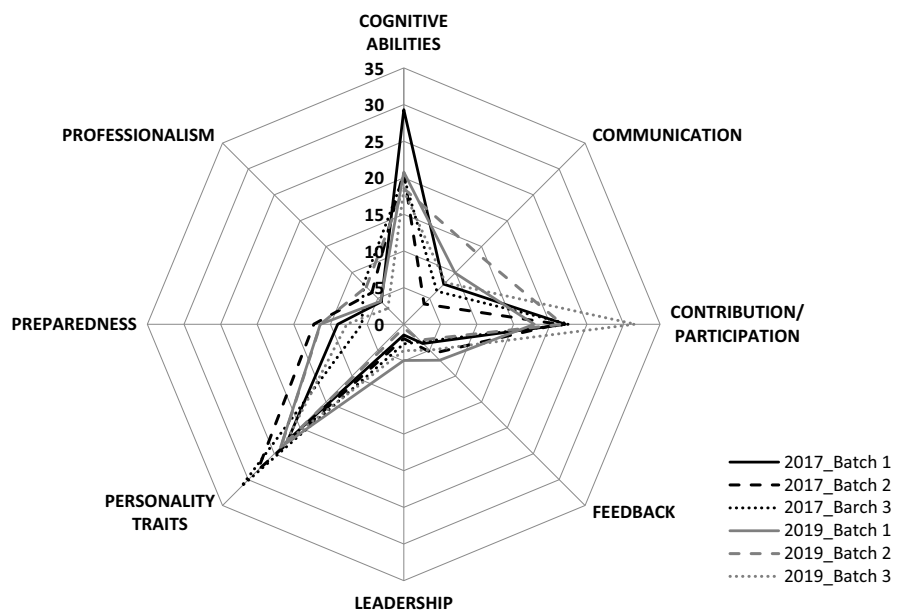


Table 2 Themes generated from the focus groups (with select representative quotes) in relation to the proposed conceptual model

Conceptual model theme	Focus group themes	Quote (focus group number)
Characteristics of the peer assessment	Anonymity of the feedback	“One can tell who wrote” (3) “A specific comment would reveal a person’s identity especially that there are only 5 members” (4)
	Face-to-face feedback	“I prefer to give negative feedback face-to-face and engage in conversation especially with close friends” (1) “Immediate face-to-face provides opportunity to understand what is going on and probably one can help the person” (2)
Conceptions of the purpose of the peer assessment	Judgment and evaluation by the administration	“Would I want to bring my classmates down?” (1) “Students mistrust the system” (3)
	Positive vs. negative feedback	“When you write good leader, you know what it means, but when you write bad leader, it needs more explanation” (2) “People know their negatives” (1)
	Importance of the feedback	“People don’t take it seriously and no one writes, so I stopped writing” (3) “A lot of people who are not serious about giving comments are also not serious about receiving them” (4)
Students’ interpersonal variables	Tension	“Is it worth giving comments and bear the tension?” (4) “Do not say things that will cause trouble” (2)
	Reciprocation	“If you don’t shoot, you are not going to be shot” (4) “Reciprocation” (3)
	Motivation	“I like to be nice and write something good” (4) “We are so stressed hence cannot take negative comments” (1)
	Vulnerability	“We are a very sensitive population” (2) “Culture of not wanting to accuse someone” (3)
	Indifference	“I am here for medicine and that’s it... if you think someone is a good leader, I don’t care. I got 90 on the test” (1) Some people don’t care about what others think of them” (3)
	Eligibility	“Cannot take comments from just anybody...we are equal” (2) “Am I eligible to give feedback to these high achievers and affect their ego” (4)
Effectiveness of the peer feedback	Beneficial	“Although the process is not optimal, it is still beneficial and people change based on feedback” (2) “Yes, I improved behavior after the feedback” (3) “It is useful if it is constructive feedback” (2)
	Essential	“What will not kill you will make you stronger” (2) “It is doing more good than harm” (4)

comments is easier than writing negative comments that have to be very well-phrased, very specific, and need more explanation.

The students explained that comments from peers are not commonly provided or when they are, they may not be easily accepted for several reasons: they could *harm* personal relationships and cause problems, they could be *biased* due to personal issues, and they may be *misunderstood*. The concept of “*culture*” was invoked: that they live in a culture that does not accept criticism easily. People, in their opinion, take things *personally* or they get emotional and oversensitive. Another aspect of culture is that people tend to *accept feedback from friends* but not from others. In addition, some people are competitive, are focused on grades, and *do not care* about feedback. Others are simply *lazy* and *do not take it seriously*. Finally, some students do not feel *qualified* to give feedback to peers who are higher achievers than they are or,

alternatively, believe that their peers are not “qualified” to criticize them, hence they would accept feedback from seniors but not peers.

The students mentioned that in many cases they fill the feedback forms haphazardly. As many of them do not receive comments, they stop writing them as well and a *vicious cycle* ensues. That is one reason why the number of comments decreased over time. In addition, with time they get to know each other better and therefore do not want to give negative comments but rather resolve issues *face-to-face*. Face-to-face communication is also believed to be more effective as it allows them to give feedback immediately after the specific incidents instead of giving feedback at the end of the course long after the events have happened. Thus, it is more specific, less formal, more constructive, and avoids misconceptions and judgments; indeed, they reported that many students already do it with their friends. They feel, however, that forcing students to

give face-to-face feedback will also not be well received, as it may create defensiveness. Notably, some students mentioned that it is *difficult* to write meaningful and constructive comments in general, as the process needs a lot of reflection.

Students provided some suggestions to improve the process of peer evaluation. They all advocated that comments become mandatory in the hope that it would then be *difficult for students to guess* who wrote what. Some suggested having the system *operated totally by students*, without the administration's involvement. All agreed that further *training and awareness* are needed. Finally, students asked for *feedback on the feedback*. They suggested the need to build a *trust circle* in their Learning Communities, to explore and potentially change behavior based on the feedback that was received.

Research Question 3: Benefits of the Peer Feedback

Despite the low number of peer feedback comments received per student and the content being typically shallow or lacking in substance, it is interesting that the process was perceived as *very beneficial for personal development* by most focus group participants. Students agreed that most of them read the comments, want to get feedback, and do not want peer feedback to be canceled. Several of them believe that feedback is helpful to improve and change behavior especially *if done in the right way*. They think the process is essential though sometimes uncomfortable. They mention that in many cases students fill them haphazardly, but they strongly believe that receiving useful comments is an *incentive for them to write good comments*.

More importantly, during the focus group discussion sessions, participants realized that the students' *conception of good or useful feedback* was different from the interviewers'. Thus, although they all agree that comments have to be specific, they believe that the details are crucial for *negative* comments when compared with *positive* ones. Although all agreed that comments on personal assets such as being intelligent are not useful, they perceived comments about good traits to be satisfying to the recipient. Most participants agreed that they need further *training* on how to write constructive feedback.

Discussion

This study showed that the volume and quality of comments provided by medical students to their peers in the context of collaborative learning are suboptimal and unlikely to produce the targeted outcomes, i.e., formative feedback that can help them improve their performance. Multiple factors were identified that could explain this finding. These were related to either the feedback process and its perceived purpose, or the interpersonal relationships among students, thus providing some validity evidence for the proposed conceptual

framework (Table 2). The lack of improvement over time is contrary to what was suggested by previous reports [29, 30]. Our results may be helpful to guide curriculum developers on how to implement peer feedback and ensure its success for the personal and professional development of medical students.

Several factors were identified to explain the decreased number and quality of the comments. Some were logistical and operational and could be corrected easily, such as the timing and frequency of the assignments, while others stood out as major substantive issues and/or limitations relating to both the purpose of the peer assessment and the students' interpersonal variables. Indeed, many students were skeptical about the purpose of peer feedback, as it was not consistently perceived as formative in nature. In addition, there were social disincentives for students to provide constructive feedback to peers with whom a continuing working relationship was necessary. There was also incongruity between the quality of the peer feedback being typically shallow and lacking in substance (as judged by the authors), and students considering it beneficial. This inconsistency necessitates further exploration and may be addressed with more training and periodic reinforcement.

Conceptions of the Purpose of the Peer Assessment: Judgment or Formative Assessment

The students' skeptical attitudes, and their wariness of giving peer feedback, were evident through the persistent use of the word "negative" to characterize feedback that is supposed to be constructive. In addition, it was particularly reflected in their stance "against" an administration that, in their view, may use their comments to discriminate among them or make judgments about them. Hence, it appears that the presence of a trustworthy and receptive administration designated to receive a peer report is a critical ingredient that contributes to students' engagement and buy-in of the process [37]. In agreement with these findings, Levine et al. [38] noted in their study that students made unsolicited comments, most of which were critical of the peer evaluation process as such. In addition, students were hesitant to give negative feedback about each other. The authors, however, did not explore the reasons for that beyond noting that the students believed they all contributed equally and should not be evaluated. They, however, mentioned that when the administration eliminated the requirement that students rate others in a discriminatory fashion, there was a better student response.

Students' Interpersonal Variables: the Social Disincentive to Provide Constructive Feedback to Peers with Whom a Continuing Working Relationship is Necessary

Students struggle to balance the benefits of providing constructive feedback to each other against the risks of disrupting

the social relationship that may be critical to their success. Hence, students either totally avoided giving comments or resorted to writing general or poor quality comments of little practical use or benefit to the recipient. The trend among students to give positive comments and refrain from giving “negative” ones is similar to what was reported by Bryan et al. [39], White and Sharma [8], and Cottrell et al., who reported an average of 11.7 positive vs. 1.67 negative comments per student [40]. In another study on peer assessment of professionalism, students mostly worried about bringing out anger and hurting each other’s feelings [37]. Students also expressed a preference for oral communication and indicated that they frequently give the “real” feedback to each other face-to-face, particularly to their friends whom they know will receive it positively. This is consistent with their concern about not being misunderstood and the inability to assure complete anonymity. This preference for direct communication has been voiced before as well, though it may not be accepted equally well by all students [3, 41].

Therefore, it appears that the presence of trustworthy and receptive individuals designated to receive a peer feedback report is a critical ingredient that contributes to the success of peer evaluations, as are close relationships among students and between students and faculty [37]. Importantly, the underlying “culture” may be playing a major role in the perceived lack of trust in the ability to evaluate their peers objectively. In addition, it may contribute to repressing students’ acceptance of diversity in opinion and behavior. It should be noted that, similar to what Van Gennip et al. reported [27], the concept of interdependence, whereby personal benefits and team performance are perceived to depend on the interconnections between the individual team members’ tasks and contributions, did not emerge in this study as a potential interpersonal variable that affects peer feedback.

The Inconsistency Between the Volume and Quality of the Peer Feedback and Students’ Perceptions of Its Benefits

It was surprising that students’ peer assessment was perceived to be effective despite the comments being of low volume and quality, i.e., what students consider as beneficial is not what is deemed so by the authors. There may also be an issue of social acceptability bias, where students might have felt that they need to endorse peer feedback as useful even if their experience of it has been underwhelming; the focus group questions, however, did not specifically address this possibility.

Most of the medical education literature that has reported on the benefits of peer assessment or evaluation relied on the perceptions of students rather than a detailed and objective analysis of the quality of the comments provided as we undertook in this study [8, 21, 39–43]. This distinction between students’ perceptions and external assessment of feedback is

important, as what may satisfy students (e.g., positive comments) may not necessarily provide enough information for self-improvement (which is one of the main goals for peer evaluation). As such, many studies have revealed the preponderance of “positive” over “negative” comments in peer evaluations and the reluctance of students to be critical or to give low scores [37, 39, 40]. For example, Nofziger et al. [2] showed that students had in general positive perceptions of peer evaluations, which in some cases were quite transformative for the students’ awareness, attitudes, and behaviors. As expected, they noted that such change was more likely if the feedback was specific and described areas of improvement. It was clear from their results though that only 6% of students reported having received a specific suggestion. Similarly, Levine et al. [38] undertook a qualitative assessment of peer comments to assess reasons for giving high and low scores and found that students made assessments based on 3 thematic areas that were quite similar to those observed in the current study: personal attributes, team contributions, and cognitive abilities of the students receiving the comments. They did not attempt to evaluate a quality rating for the written comments though. Nevertheless, it was interesting to note that the median length of a comment in that study was 6 words. Considering that “good” feedback should be as specific as possible, should be descriptive rather than evaluative, and should contain practical usable suggestions for improvement, an average of 6 words per comment suggests that the feedback was not uniformly very useful. Our results are also similar to Cottrell et al. [40] who addressed the different themes that emerged from peer assessment forms and found that students mainly point out whether their peers are being punctual, contribute to the group, respect others, and engage in discussion. Finally, White and Sharma [8] analyzed the written comments received by medical students on a Surgery and Anesthesiology clerkship from physicians, patients, residents, nurses, administrators, and peers. Again, while the content was analyzed in terms of themes and domains of assessment, the quality of the comments was not analyzed.

To our knowledge, only one study by Canavan et al. [7], examined the quality of feedback provided by medical students to each other, finding that most of the comments lacked characteristics of good feedback. The authors noted that most feedback was positive, was self-oriented, and lacked “actionable information” that could be useful to the recipient. To support these conclusions, Burgess et al. [21], in a systematic review of medical students as peer teachers, found that, in most of the studies that involved peer assessment, there were no objective measures to assess the quality and accuracy of peer feedback and assessment. Interestingly, although peer feedback in collaborative learning in general, and TBL in particular, is designed to specifically foster peer accountability and team learning skills [18, 19, 33], in the current

study, these themes did not emerge as perceived benefits from peer feedback, as the only perceived benefit was on the personal level.

Need for Repeated Training and Periodic Reinforcement by the Medical School Administration

Most of the studies that examined the effectiveness of peer feedback were cross-sectional in nature. None, to our knowledge, followed students over time to examine whether engagement in repeated peer feedback altered the quality of the comments. This is important considering the literature, largely from non-medical fields, suggesting that as peer evaluators gain more experience with the process, the quality of the evaluation improves. For example, Hamer et al. [30] found improved correlations between student and faculty assessment in a computer science program as the students gained more experience with peer assessment; results that were congruent with those of Calhoun et al. [28] with medical students. Furthermore, in a systematic review, Van Zundert et al. [29] showed that both the psychometric qualities of peer assessment instruments and the student attitudes to peer assessment were positively influenced by the peer assessor's experience and training. In our study, however, despite repeated engagement in peer feedback over 2 years, the quality of the students' comments did not improve. This suggests that familiarity and repeated practice per se are not enough to hone the skill and improve the outcome. One explanation for this, which may also explain the students' skepticism and possible misunderstanding of the purpose and process of feedback, may be that the single training session they received was not enough to develop habits, skills, and convictions, particularly as most students have had no previous experience in giving or receiving feedback and have not practiced it regularly during their school or college years. The same applies to the criteria concerning helpful feedback whereby, although they are a necessary ingredient for the success of peer feedback [10, 25, 44], they also ought to be reiterated and reinforced periodically.

We speculate that the reason we observed the most numerous and most useful comments for the Class of 2017 during their first attempt (first batch) was because this was the first class of the new curriculum, which was monitored and followed very closely, with almost daily encounters with curriculum leaders and constant reminders of the spirit and values of the new curriculum; this included emphasis on all aspects of professionalism including teamwork and provision of feedback. This degree of monitoring decreased over time, which may account for the dwindling number of comments and their lesser quality over the 2 years. In contrast, by the time the Class of 2019 had matriculated, the curriculum had been well established and did not require close follow-up, and there was less oversight and less persistent monitoring of the

students by the administration; hence, the first and second batches of peer evaluations had a markedly lower number of responses. The relatively low number of comments provided by the students supports previous findings about the reluctance of students to give feedback to peers [37, 45, 46], and the decrease over time seen with the Class of 2017 is consistent with the findings of Bryan et al. [39] who demonstrated an overall decrease in the number of comments towards the end of the course studied.

We note here that after the low number of comments was noted on the first 2 batches of the Class of 2019, the administration intervened and reiterated the value of this practice and the need for students to provide comments as part of their professional duties, and asked them to review the principles of good feedback as conveyed during the training session they received during year 1. As a result, the number of comments rose dramatically in the third batch (of the Class of 2019); however, this was not matched by a similar improvement in their quality. Furthermore, in the Fall of 2018, and to test the effect of more extensive training, we introduced a TBL session for the class of 2021, specifically targeting how to provide good quality feedback and also how to assess the quality of peer feedback, using actual student comments from previous classes. This was in addition to the regular training session they shared with the previous classes. The first batch of evaluations for that class revealed a much higher number of comments per student (mean \pm SEM 4.18 \pm 0.08) when compared with the first batch of both classes 2017 and 2019 as well as a higher quality of comments compared with the two previous classes (mean \pm SEM 2.22 \pm 0.03) when compared with the first batch of both classes 2017 and 2019, hence underscoring the benefit of additional training. Therefore, we suggest that the medical school administration can play a continuing role to educate students on how to provide feedback, reemphasize its formative nature, and remind them of its value. However, this should be done judiciously so that it does not come across as coercive, and so that peer feedback is not perceived by students as a duty or requirement, which carries the risk of losing its authenticity and validity.

Limitations

This study is limited by the fact that the evidence provided is solely based on students' perceptions rather than assessment of actual change in behavior. Furthermore, the small number of students who participated in the focus group sessions may not necessarily reflect the views of the whole class. In addition, although the results are consistent with the proposed conceptual framework, we were unable to undertake more elaborate statistical analyses to test for model fit considering the nature of the study. Nevertheless, the study offers insight into the contributing factors of the low volume and quality of

individual peer feedback and its perceived effectiveness for further research to build on.

Conclusion

In this study, we employed a conceptual framework to explore the factors that enhance the perceived effectiveness of formative peer assessment in the context of TBL as a form of collaborative learning. We added time and experience and integrated volume and quality of the peer feedback into a conceptual peer assessment model and showed that the implementation of an effective peer feedback system for student improvement may be hindered by several logistical, social, and cultural challenges. These include the conflation of peer feedback with summative assessment, the social disincentives for students to provide feedback to peers with whom a continuing working relationship was necessary, and students' lack of understanding of what constitutes good-quality feedback. Therefore, the success of the peer feedback process depends on the availability of a "safe" environment that values diversity and establishes trust among students and between them and the medical school administration, a culture of openness, tolerance, and forthrightness, and extensive training with periodic reinforcement. Further research and efforts are needed to foster student interdependence for team learning and peer accountability in collaborative learning.

Acknowledgments The authors wish to thank participants of the focus groups.

Compliance with Ethical Standards

The study was approved by the Institutional Review Board (IRB) of the American University of Beirut under the expedited category. All focus group participants signed an IRB-approved informed consent form.

Conflict of Interest The authors declare that they have no competing interests.

References

- Strijbos JW, Sluijsmans D. Unravelling peer assessment: methodological, functional, and conceptual developments. *Learn Instr.* 2010;20:265–9.
- Nofziger AC, Naumburg EH, Davis BJ, Mooney CJ, Epstein RM. Impact of peer assessment on the professional development of medical students: a qualitative study. *Acad Med.* 2010;85(1):140–7.
- Arnold L. Assessing professional behavior: yesterday, today, and tomorrow. *Acad Med.* 2002;77(6):502–15.
- Prins FJ, Sluijsmans DMA, Kirschner PA, Strijbos JW. Formative peer assessment in a CSCL environment: a case study. *Assess Eval High Educ.* 2005;30(4):417–44.
- Strijbos JW, Narciss S, Dunnebie K. Peer feedback content and sender's competence level in academic writing revision tasks: are they critical for feedback perceptions and efficiency? *Learn Instr.* 2010;20:291–303.
- Gielen S, Peeters E, Dochy F, Onghena P, Struyven K. Improving the effectiveness of peer feedback for learning. *Learn Instr.* 2010;20:304–15.
- Canavan C, Holtman MC, Richmond M, Katsufarakis PJ. The quality of written comments on professional behaviors in a developmental multisource feedback program. *Acad Med.* 2010;85(10 Suppl):S106–9.
- White JS, Sharma N. Who writes what? Using written comments in team-based assessment to better understand medical student performance: a mixed-methods study. *BMC Med Educ.* 2012;12:123.
- Walker M. The quality of written peer feedback on undergraduates' draft answers to an assignment, and the use made of the feedback. *Assess Eval High Educ.* 2014;40(2):232–47.
- Ploegh K, Tillema HH, Segers MSR. In search of quality criteria in peer assessment practices. *Stud Educ Eval.* 2009;35:102–9.
- Hovardas T, Tisvitanidou OE, Zacharia ZC. Peer versus expert feedback: an investigation of the quality of peer feedback among secondary school students. *Comp Educ.* 2013;71:133–52.
- Loureiro MJ, Pombo L, Moreira A. The quality of peer assessment in a wiki-based online context: a qualitative study. *Educ Media Int.* 2012;49(2):139–49.
- Prins FJ, Sluijsmans DM, Kirschner PA. Feedback for general practitioners in training: quality, styles, and preferences. *Adv Health Sci Educ Theory Pract.* 2006;11(3):289–303.
- Burgess A, Mellis C. Feedback and assessment for clinical placements: achieving the right balance. *Adv Med Educ Pract.* 2015;6:373–81.
- Slavin R. Cooperative learning: theory, research and practice. Englewood cliffs NJ: Prentice-Hall; 1990.
- Parmelee D, Michaelsen LK, Cook S, Hudes PD. Team-based learning: a practical guide: AMEE guide no. 65. *Med Teach.* 2012;34(5):e275–87.
- Nasr R, Antoun J, Sabra R, Zgheib NK. Interactive and collaborative learning in the classroom at the medical school: automated response systems and team-based learning. *J Med Liban.* 2016;64(4):217–22.
- Levine RE. Peer evaluation in team-based learning. Team-based learning for health professions education: a guide to using small groups for improving learning. First ed. Sterling, Virginia: Stylus Publishing; 2008. p. 103–11.
- Cestone C, Levine RE, Lane DR. Peer assessment and evaluation in team-based learning. In: Michaelsen LK, Sweet M, Parmelee DX, editors. Team-based learning: small group learning's next big step. Wiley InterScience; 2008. p. 69–78.
- Speyer R, Pilz W, Van Der Kruis J, Brunings JW. Reliability and validity of student peer assessment in medical education: a systematic review. *Med Teach.* 2011;33(11):e572–85.
- Burgess A, McGregor D, Mellis C. Medical students as peer tutors: a systematic review. *BMC Med Educ.* 2014;14:115.
- Burgess AW, McGregor DM, Mellis CM. Applying established guidelines to team-based learning programs in medical schools: a systematic review. *Acad Med.* 2014;89(4):678–88.
- Han Y, James DH, McLain RM. Relationships between student peer and faculty evaluations of clinical performance: a pilot study. *J Nurs Educ Pract.* 2013;3(8):170–8.
- Reimschisel T, Herring AL, Huang J, Minor TJ. A systematic review of the published literature on team-based learning in health professions education. *Med Teach.* 2017;39(12):1227–37.
- Panadero E, Romero M, Strijbos JW. The impact of a rubric and friendship on peer assessment: effects on construct validity, performance, and perceptions of fairness and comfort. *Stud Educ Eval.* 2013;39:195–203.
- Maas MJ, Sluijsmans DM, van der Wees PJ, Heerkens YF, Nijhuis-van der Sanden MW, van der Vleuten CP. Why peer assessment

- helps to improve clinical performance in undergraduate physical therapy education: a mixed methods design. *BMC Med Educ.* 2014;14:117.
27. Van Gennip NAE, Segers MSR, Tillema HH. Peer assessment as a collaborative learning activity: the role of interpersonal variables and conceptions. *Learn Instr.* 2010;20:280–90.
 28. Calhoun JG, Ten Haken JD, Woolliscroft JO. Medical students' development of self- and peer-assessment skills: a longitudinal study. *Teach Lear Med: Int J.* 2009;2(1):25–9.
 29. van Zundert M, Sluijsmans D, van Merriënboer J. Effective peer assessment processes: research findings and future directions. *Learn Instr.* 2010;20(4):270–9.
 30. Hamer J, Purchase HC, Denny P, Luxton-Reilly A. Quality of peer assessment in CS1. In *Proceedings of the fifth international workshop on computing education research workshop.* Assoc Comput Mach. 2009:27–36.
 31. Schifferdecker KE, Reed VA. Using mixed methods research in medical education: basic guidelines for researchers. *Med Educ.* 2009;43(7):637–44.
 32. Lavelle E, Vuk J, Barber C. Twelve tips for getting started using mixed methods in medical education research. *Med Teach.* 2013;35(4):272–6.
 33. Zgheib NK, Dimassi Z, Bou AI, Badr KF, Sabra R. The long-term impact of team-based learning on medical students' team performance scores and on their peer evaluation scores. *Med Teach.* 2016;38(10):1017–24.
 34. Koles PG. Peer evaluation - examples and graded comments. 2018. https://cdn.ymaws.com/teambasedlearning.site-ym.com/resource/resmgr/Docs/Peer_Eval_examples_graded_co.pdf Accessed January 2019.
 35. Michaelsen LK, Schultheiss EE. Making feedback helpful. *Org Behav Teach Rev.* 1988;13(1):109–13.
 36. Vaismoradi M, Turunen H, Bondas T. Content analysis and thematic analysis: implications for conducting a qualitative descriptive study. *Nurs Health Sci.* 2013;15(3):398–405.
 37. Arnold L, Shue CK, Kritt B, Ginsburg S, Stern DT. Medical students' views on peer assessment of professionalism. *J Gen Intern Med.* 2005;20(9):819–24.
 38. Levine RE, Kelly PA, Karakoc T, Haidet P. Peer evaluation in a clinical clerkship: students' attitudes, experiences, and correlations with traditional assessments. *Acad Psychiatry.* 2007;31(1):19–24.
 39. Bryan RE, Krych AJ, Carmichael SW, Viggiano TR, Pawlina W. Assessing professionalism in early medical education: experience with peer evaluation and self-evaluation in the gross anatomy course. *Ann Acad Med Singap.* 2005;34(8):486–91.
 40. Cottrell S, Samora JB, Shumway J. An analysis of first-year medical student comments in a peer evaluation of professionalism. *Med Sci Educ.* 2006;17(1):27–32.
 41. Street SE, Hobbs MM. Face-to-face peer evaluation improved student perception of feedback in a preclinical course. *Med Sci Educ.* 2015;25(1):9–11.
 42. Tayem YI, James H, Al-Khaja KA, Razzak RL, Potu BK, Sequeira RP. Medical students' perceptions of peer assessment in a problem-based learning curriculum. *Sultan Qaboos Univ Med J.* 2015;15(3):e376–81.
 43. Gukas ID, Miles S, Heylings DJ, Leinster SJ. Medical students' perceptions of peer feedback on an anatomy student-selected study module. *Med Teach.* 2008;30(8):812–4.
 44. Smith H, Cooper A, Lancaster L. Improving the quality of undergraduate peer assessment: a case for student and staff development. *Innov Educ Teach Int.* 2002;39(1):71–81.
 45. Linn BS, Arostegui M, Zeppa R. Performance rating scale for peer and self assessment. *Br J Med Educ.* 1975;9(2):98–101.
 46. Freidson E. *Profession of medicine.* New York: A study of sociology of applied knowledge; 1990.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.