

ONLINE FIRST – ACCEPTED ARTICLES

Accepted articles have been peer-reviewed, revised and accepted for publication by the *SMJ*.

They have not been copyedited, and are posted online in manuscript form soon after article acceptance. Each article is subsequently enhanced by mandatory copyediting, proofreading and typesetting, and will be published in a regular print and online issue of the *SMJ*.

Accepted articles are citable by their DOI upon publication.

**Enhancing the use of peer review and student feedback to
evaluate educators in early years of health professions education:
insights from a medical school**

Shuh Shing Lee¹, PhD, Inthrani Raja Indran², PhD, Celestial T Yap^{3,4,5}, MBBS, PhD,
Dujeepa D Samarasekera¹, MBBS, FRCP(Edin)¹, Sherilyn Jo-Yin Wong⁶, BSc,
Zhi Xiong Chen^{1,3,4,5,7}, PhD, MHPE

¹Centre for Medical Education, ²Department of Pharmacology, ³Department of Physiology,
⁴NUS Centre for Cancer Research, Yong Loo Lin School of Medicine, National University of
Singapore, ⁵National University Cancer Institute, Singapore, National University Health System,
⁶Agency for Science, Research and Technology (A*STAR), ⁷VIVA-KKH Paediatric Brain and
Solid Tumour Programme, KK Women's and Children's Hospital, Singapore

Correspondence: Assoc Prof Chen Zhi Xiong, Assistant Dean (Education), Department of
Physiology, Yong Loo Lin School of Medicine, National University of Singapore, MD9, 2
Medical Drive, Singapore 117593. zhixiong_chen@nus.edu.sg

Singapore Med J 2022, 1–20

<https://doi.org/10.11622/smedj.2022051>

Published ahead of print: 5 May 2022

More information, including how to cite online first accepted articles,
can be found at: <http://www.smj.org.sg/accepted-articles>

INTRODUCTION

In the last two decades, the importance of evaluating teaching performance has been greatly emphasised in the fields of health professions education.⁽¹⁾ These evaluations are often used to provide faculty feedback for curriculum development and education quality improvement as well as assist in the faculty's promotion and tenure process.⁽²⁾ Generally, medical schools use either one or mixed evaluation instruments (a combination of two or more than two evaluation instruments) to effectively and constructively assess teaching performance of the faculty members. There is a variety of instruments to evaluate teaching such as peer evaluation, self-assessment, external expert ratings, mentor ratings, teaching portfolios and students' feedback.^(1,3)

Each evaluation instrument has its own merits and demerits depending on the suitability and preference of the educational setting. Peer evaluation is considered as one of the evaluation method that provides not only formative feedback, but also serves as a reflective process for teachers.⁽⁴⁾ Self-assessment is another powerful tool to help teachers identify their strengths and weaknesses and make improvement in their teaching.⁽⁵⁾ Teaching portfolio is a collection of materials that document and reflect teaching performance of a teacher.⁽⁶⁾ External expert and mentor ratings are some alternate methods to provide feedback to improve the quality of teaching.⁽¹⁾

On the other hand, students' feedback has dominated as the key instrument to evaluate teaching and has been widely used and studied for the past few decades.⁽⁷⁻⁹⁾ Teachers' evaluation by students has a major role in providing valuable feedback to improve their teaching practices and identify areas for faculty development. The evaluation of teachers is often based on quantitative and qualitative feedback from students. A typical quantitative evaluation would use items scored using a 5-point Likert score in a questionnaire that rates the

major components of good teaching. These items include teachers' knowledge, teaching methods (innovative, interactive method), communication skills, feedback skills (direct praises, solve difficulties, give encouragement), rapport (active interaction and discussion with students) and last but not least positive personality traits of teachers (approachable, respectful, mature, observant).^(7,8,10-13)

In medical education, ascertaining the quality of teaching has historically relied on student feedback or self-administered questionnaire.^(14,15) Numerous instruments have been designed to evaluate individual teachers, in most cases containing scaled items and open-ended questions. These instruments are mainly focused on 4 dimensions and majority are tailored to the clinical teaching context (bedside teaching): structural dimensions (the physical environment available for teaching, teaching materials and the design of a curriculum), learning process dimension (teacher-student interaction or teaching/learning atmosphere), Instructor-specific characteristics dimension (teaching skills and the level of preparation, teachers' enthusiasm as perceived by their students) and lastly, the outcome dimension (learning outcome and the development of professional attitudes).⁽¹⁴⁾

While students' feedback is a valuable source of information for evaluating teachers, there is increasing evidence indicating that this should not be the sole method of informing teaching effectiveness.^(16,17) It should be combined with other data such as peer evaluation and self-assessment to provide a more comprehensive evaluation of faculty members.⁽¹⁸⁾ Bastani et al (2017) has shown that 83.3% of the study participants chose mixed method evaluation as the best way to evaluate teachers.⁽¹⁾ Few studies have also implemented mixed method evaluation to assess teaching performance of their faculty in respective medical schools.^(5,9) These methods which offer additional perspectives on teaching performance include objective structured teaching encounters (OSTEs), teaching portfolios and patient feedback.⁽¹⁹⁾ In recent

years, peer observation of teaching or peer evaluation is increasingly gaining prominence in medical education although it has been well-established in general higher education.⁽¹⁵⁾ With the advancement of computer technology, some medical schools have implemented or are moving towards a web-based system for collection and processing of evaluation results due to its user-friendly functionality, its ability to preserve data integrity and its high compliance rates.^(2,11)

Similar to other health professions institutions, peer evaluation and students' feedback are the major instruments used to exclusively evaluate the teaching-and-learning experiences provided by the educators in the early years of health professions education (Year 1 and Year 2) in Yong Loo Lin School of Medicine (NUSMed) at the National University of Singapore (NUS) for annual appraisal, promotion and appointment since 2008^(20,21) and have not undergone many changes. Studies on these instruments on clinical years and often neglected the reconciliation of opinions and expectations of both teachers and students in determining suitable criteria, approaches and methods used to evaluate educators in early years of health professions education. Moreover, it also leaves much room for assessing whether these current instruments convey all the necessary information needed by teachers to help them develop and improve their teaching practices in a meaningful manner.

In this study, we would like to explore the strengths and weaknesses of the mixed method evaluation system (student feedback and peer evaluation), and identify approaches to enhance the system to better assess or reflect the teacher's educational intentions, capabilities and repertoire of pedagogical innovations, by examining the perspectives of students and teachers. This study might reveal some insights to institutions which implement the same programme of evaluation and to improve the existing system.

METHODS

This study was carried out at the Yong Loo Lin School of Medicine (NUSMed), National University of Singapore (NUS). Ethical approval was granted by the Institution Review Board (IRB) of the university and written consent was obtained. Although this study was conducted at NUSMed, the evaluation system has been developed and used at NUS. This study employed an exploratory qualitative method grounded in a phenomenological approach in order to gain an understanding of the current evaluation system from the experiences of the teachers and students who use it. Phenomenology approach is a qualitative research which focuses on the commonality of a lived experience within a particular group. Interviews are typically conducted to obtain first-hand knowledge of an event, situation of experience.⁽²²⁾

A purposive sample of participants was utilised to select educators in early years of health professions education (Lecturer and above in the Educator Track or Assistant Professor and above in the tenure track) who were teaching Year 1 and 2 NUSMed students (including Nursing). Year 1 and 2 NUSMed students (medical only) were selected to provide their experience in using the system as well as to triangulate with teachers' interview data. We specifically sampled Year 1 and 2 students in order to capture data specific to evaluation of educators who are mostly involved in the first and second year of the curriculum. Nursing educators were also sampled as they were teaching similar subjects in the early years of nursing education and evaluated using the same instruments, even though they were not teaching medical students.

Sample size was directed by the research questions, meaning that we recruited the educators until we attained information redundancy or saturation. The data collected from the 12 educators was triangulated with 129 Year 1 and 14 Year 2 students' open responses at which no new experiences were emerging and nothing new was being added to the data.

In order to gain experiential descriptions from the participants, we used an in-depth interview guide to ensure consistent coverage. SS and SJY from the Centre for Medical Education (CenMED) conducted the interviews which lasted 50-70 minutes. SS is an independent qualitative researcher from CenMED who has more than 10 years of experience in conducting qualitative research (which include designing qualitative research, conducting interview and analysing qualitative data) while SJY is a research assistant who has been trained by SS to conduct the interview and analyse the data. Both of them had a discussion with ZX to understand the research topic and carry out a pilot before the actual interview was taken place. This is to ensure credibility of the data collected with minimal power distance in play. The interviews were conducted in the private rooms at CenMED. All participants agreed to allow the interview to be audiotaped and transcribed verbatim. All identities of the participants were de-identified before the transcripts were sent back for member checking and analysis. Field notes were taken during the interview and the researchers (SS & SJY) created audit trails to ensure the dependability, and transferability of the study. The participants were asked broad questions and prompted by the researchers/interviewers (SS & SJY) if required. Examples of questions include: ‘What do you think of the current system in evaluating educators?’ ‘How do you think educators can be better evaluated? Why so?’ ‘How do you think an evaluation process can be successfully implemented?’

Students’ data was collected using open responses in an online survey to triangulate with the interview data. This was to ensure dependability and confirmability of the data collected. We sought student-implied consent before collecting their anonymised data. A briefing was carried out to inform the students that this data collection was voluntary and a set of 4 questions were posted to the students. They answered those four questions either using their laptop or mobile phones individually for 20-30 minutes. These 4 questions are:

- i. How do you find your experience in medical classes with respect to the teaching conducted by your teachers? Please elaborate.
- ii. Do you think the current teacher evaluation system is useful in enhancing the quality of education delivered to your future peers? Please justify your answer.
- iii. How do you think the current teacher evaluation system can be improved?
- iv. What is relevant or irrelevant in the current teacher evaluation system and why?

The above data collected was used to compare with the interview data from the teachers' perspective in order to explore the similarities and differences in using the evaluation system. An online survey was used over focus group discussion (FGD) as there was poor response to attempts to recruit students to attend FGD. Furthermore, online survey provides a safer environment for the students to voice their opinion anonymously though the data collected might not be deep enough. Nevertheless, student data collected was only used as additional evidence in addition to teachers' interview data.

The researchers (SS & SJY) from CenMED performed thematic analysis on the qualitative data. The researchers listened to the tapes, read and re-read the transcripts and immersed themselves into the text in order to acquire the actual experience the educators have gone through. The transcripts were also constantly compared with the data collected from students via the online survey to sieve out the similarities and differences from both sides. After which a discussion was carried out with all the authors to ensure the consensus was reached on the data collected before coding was initiated. Then, two of the authors (SS & SJY) generated the initial codes independently before discussions were carried out. Next, codes that appeared consistently and were associated with similar content were grouped into sub-categories, and similar sub-categories were combined into a category or forming a theme. In the event that

there were differing views on the coding/theme, they re-examined the primary data and discussed before agreeing on a code/theme.

RESULTS

In total, we gathered responses from 143 Year 1 and Year 2 students for all the questions and interviewed 12 educators from different centres/departments such as two from the Centre for Biomedical Ethics (CBmE), two from Alice Lee Centre for Nursing Studies (ALCNS) and NUSMed (eight from various biomedical science departments such as Anatomy, Biochemistry, Physiology and Pharmacology). The data collected from the educators was analysed thematically and triangulated with the data collected from the students.

Four themes emerged from the qualitative analysis of this study, namely (a) Feedback system's design, (b) Feedback Process (peer & student evaluations), (c) Operations of feedback collection, and (d) Post-feedback initiatives. Table 1 indicates the main themes with their respective sub-themes and examples of the quotes. Descriptions are provided below.

Table 1: Breakdown of themes, sub-themes and corresponding quotes.

Theme 1: Feedback system's design	Questions (scope, specificity, relevance, questions type)	E5: ... there is a shift in the type of criteria that was used years back where they focused more on soft skills of the teachers and personality-directed kind of question, I feel. For current evaluation questions, they are pretty much to the context of the teaching objectives, which is stimulation, decision-making, and reasoning skills, which I think is good.
		E1: I mean it's generic in terms of general lecture quality, it's not asking specifically for example, how knowledgeable is this person about ethical domains, how much do they help in terms of ensuring your practice has improved...
	Mode of collection (electronic feedback system vs face-to face session)	E2: The strength of it is that it's cheap and efficient I suppose. It doesn't cost them very much to do it.

		F1: I think face to face is ultimately the best way to get feedback, because you know whether it's genuine or not. You see the person's expression, you know it.
Theme 2: Feedback collection process (peer & student evaluation)	Selection, subjectivity and standardization of peer evaluation	E9: So I can go and invite my friend, and my friend will of course during his or her peer review invite me, so there is a mutual benefit system, which is not, I can't tell not correct, can be improved upon, that's what I feel.
	Quality of student evaluation and training	E2: ... so I think that's part of the problem – that students aren't trained on how to give effective feedback. Both to us, constructive feedback, and they're not used to giving it to each other either.
	Insight from feedback	E6: ... because you get a more accurate picture on how you did with your teaching, and then you can get real feedback in terms of where you need improvement and you can make an effort to make it better.
Theme 3: Operations of feedback collection	Time	E11: ... it's still getting collected at the end of the year, when students, if they've only seen us one time, they won't remember us
	Manpower	E4: Given actual resource constraints, where it's not feasible to have even periodic observations for each tutor, it's just not feasible. Yeah I think the idea of more periodic student focus groups is something we should seriously be looking at.
	Incentives	E7: Having more time to come up with evaluative methods to evaluate our own teaching, but we're just so time-poor, and ultimately there is so little reward for it at the end.
	Misuse of feedback system	E5: ... what I'm trying to say at the end is that the evaluation platform from the students' perspective is not being used wisely by the students, it can be a potentially abusive system to bring down the teachers or the educator's morale in trying to help the students to learn more.
Theme 4: Post-feedback initiatives	Evaluation scheme	E8: ... our department has a compulsory student feedback report evaluation and actions taken. So we have to submit at the end of the semester at least a few points that the students said or need changes, and then what changes we are implementing and when are we implementing.
	Pedagogies improvement	E10: ... they will have to meet up with the department head to discuss, so like "what's the issue?". The discussion is not more on punitive, but rather on problem solving the issue ...

Theme 1: Feedback system's design

There are two sub-themes, which fell under this theme: a) questions and b) mode of collection. Many of the interviewees commented on the scope, specificity and relevance of the questions. The questions were seen to be focused primarily on the general aspects of teaching such as the teacher's content knowledge, and ability to engage the students. Educators noted that the feedback did not attempt to capture any information on their soft skills or classroom personality. The questions in the student feedback forms were also observed to be generic and not tailored to different teaching activities. For instance, questions were not related to clinical procedure and were related to lecture quality. While educators acknowledged that the students should not be burdened with excessive feedback forms and/or questions, they emphasised the importance of ensuring that the questions are clear and relevant. While the current system of gathering student feedback using an electronic system was recognised to be more efficient and cost effective, many educators expressed that there is additional value that could be obtained via face-to-face sessions.

Theme 2: Feedback collection process (peer & student evaluation)

Objectivity of the feedback process was a very significant theme that emerged from this study. There were concerns as to whether the peer evaluation, student evaluation and insights from the feedback were obtained and/or reviewed in a fair and objective manner. With respect to peer evaluation, educators commented on the evaluator selection process, subjectivity and the lack of standardisation of the peer evaluation process. As the current system allowed some of the educators to invite their colleagues of choice to observe and evaluate their teaching, this can affect the objectivity and fairness of the peer evaluation process, where one who is providing feedback for a friend may feel compelled to leave more favourable comments as

compared to one who is evaluating independently. Moreover educators may have different repertoire of teaching activities, and differing teaching philosophies, which may further colour their perspective and evaluation of teaching quality and standards. Hence, educators shared that the peer evaluation processes should be more formalised and further standardised to ensure that the evaluations are more reliable. There is also a lack of post-feedback avenues for communication as mentioned by several interviewees to clarify or seek advice for improvement. Some educators have also suggested implementing a mentoring system which could allow for formative review processes that can help educators grow and learn, with support from trusted and capable mentors.

Student feedback was also noted to suffer from wide discrepancies relating to factors such as the time gap between the teaching event(s) and feedback collection exercise which could then affect recollection, the number of contact hours the teacher being evaluated had with the students providing the feedback, the overall student feedback response rate, the perceived importance the student placed on feedback exercise, the student's relationship with the teacher, training of students on how to give feedback and the voluntary nature of the feedback process. Without adequate control or review of these factors, the validity and reliability of the outcomes measured via this process could be questionable.

Despite the inherent drawbacks that the peer evaluation and student feedback system posed, most educators still found that the information collected via these means could help them reflect and build on their teaching skills. Furthermore, the feedback also provided raw insights on how their teaching performances change over the years and allowed them to benchmark against the department and faculty.

Theme 3: Operations of feedback collection

The third major theme that arose from this study was pertaining to the operations underlying the feedback collection process. First was the timing and frequency of feedback collection. When the feedback was collectively obtained at the end of the semester, the students' ability to recollect different lecturers may vary significantly and this could affect the quality of feedback provided. Feedback fatigue was yet another real issue as students have to complete the feedback for numerous teachers at the end of the semester. This was shared by both students and faculty members in the online survey and interview respectively. Both teachers and students provided suggestions to address the above issues, which include (i) reducing the number of teachers on whom each student provides feedback on, and (ii) to stagger the feedback collection periods to ensure feedback is collected in a timely manner.

Another important concern raised with respect to the operations underlying the feedback collection was manpower. For student feedback, administrators have to compile the reports and send them to respective educators. Peer evaluation involved additional manpower for class observations and filling up evaluation forms. To ensure reliability of the outcomes, two tutors acting as peer reviewers instead of one would be ideal for peer evaluation. However, human resources could be a constraint especially for small departments with limited manpower.

There were also discussions on providing incentives to increase student involvement and rewards for educators who have done well. Incentives could also be educational support such as faculty development and online teaching tools freely available for teachers so that they will be able to utilise them for teaching. Educators also raised the issue that the feedback system could be abused by the students to air their grievances over fair and necessary classroom measures adopted by the teachers. This could bring down the educators' morale instead of helping them to improve their teaching and learning. Hence, as mentioned in Theme 2 by some

educators, training the students on how to provide proper feedback is crucial, although it is lacking in the current system.

Theme 4: Post-feedback initiatives

The last theme was related to the post-feedback initiatives. For example, CBmE developed a unique set of internal evaluation questions, aimed at providing meaningful feedback to the educators, taking into account their relevant teaching domains. ALCNS provided post-feedback evaluation reports for educators, which included developing action plans to improve on potential areas of weaknesses. Most interviewees mentioned that the feedback focused on pedagogical improvement. Some of the post-feedback initiatives that have been put in place include dialogue between the heads of department and educators regarding feedback results, conducting masterclasses for educators, sponsoring educators to attend faculty development workshops and observation of individuals who demonstrate excellence in their teaching.

DISCUSSION

Although research examining the technical merits of the measures and processes used to evaluate educators in existing systems is ongoing, only a few studies have focused on the perceptions of practitioners about how to best evaluate educators in practice.⁽²³⁾ This study revealed some underlying issues within the evaluation system in the early years of education in medical courses. This is aligned with some studies highlighting that the result of evaluation procedures have not been geared towards helping educators improve their skills, but to satisfy certain requirements in institutions.^(24,25) Therefore, issues concerning the feedback system's design, processes and operations were raised.

Learner feedback is a major tool used to evaluate teachers during early and clinical years. Literature has highlighted the importance of the validity and reliability of these evaluation tools.⁽²⁶⁻²⁸⁾ However, the psychometric characteristics of these feedback tools were focusing on assessing clinical teaching instead of the tools used in the early year of medical school.^(7,29,30) Conigliaro and Stratton argued that this is because current clinical teacher evaluation often measures attributes of clinical teachers in broad terms.⁽¹⁶⁾ Unlike measurements of lectures, workshops or online educational content, which can more readily be assessed using objective criteria, clinical teacher evaluation is more subjective and susceptible to the halo effect. Although the context and questions might be different for clinical teacher and teacher evaluation in the early year of medical school, the data in this study revealed the necessity to re-look at the questions to ensure reliable, valid, and feasible evaluation of the teachers.

Based on the data collected, we would like to make several recommendations with regards to peer review and student feedback as well as propose post-feedback initiatives. For peer review, it would be further beneficial to embed the instrument within a mentee-mentor setting and serve a formative purpose in developing faculty, on top of using it as a high-stake evaluation tool for annual appraisal, promotion and appointment. Studies have suggested that the process of and results from formative evaluation of teachers' professional practice can improve instructional quality⁽³¹⁾ and student achievement,⁽³²⁾ compared to it being used for summative purpose alone.

When re-purposed in this way, it will shift the intent of peer review mainly from getting educators 'past the post' for teaching awards and promotion to getting educators 'genuinely developed'. A shift of intent will pave the way for transforming the selection process and criteria of peer reviewers. Furthermore, internal reviewers (within the department) with more

access and better knowledge of the educator can follow up with him/her over the long term. More intentional opportunities should be provided for the educator to clarify with external reviewers through post-review sessions, rather than the current single opportunity to respond to the reviews over an online system. In this way, more authentic feedback can be provided to educators with support from mentors and more time provided for educators to develop in a low-stake 'safe' setting.

One of the challenges which was widely documented in the existing literature with regards to peer review is the threat to reliability in the absence of adequate evaluator training.⁽³³⁾ This was also shared by our educators in the data collected in Theme 2. As suggested by many studies, a potential solution to this might be the use of a rubric or checklist to develop a common language among peer evaluators.⁽³⁴⁻³⁶⁾ Not only would this help in standardising the evaluation process, but would also increase opportunities for observation and feedback.⁽³⁷⁾ Other benefits include having a common ground to build constructive and collegial dialogue on the feedback.⁽³⁸⁾

For student feedback, similar issues were brought-up (questionnaire fatigue, nature of the forms, and time-consuming and error-prone) in a paper published in 2014 and has not been fully resolved hitherto.⁽²⁰⁾ In order to balance feedback fatigue and timeliness of feedback, a possible solution would be to solicit feedback after every learning event with reduced number of questions and more open-ended questions. From our study, it appears that students are not adverse to giving feedback but adverse to answering questions that they do not identify with or perceived as inadequate to address their views. When framed in the right way, they would like to have a 'larger voice' through open-ended questions. Timely feedback is important for reliability of feedback and for educators to make adjustments as close to 'real-time' as possible. Therefore, it is imperative for feedback collection periods to be free from confounding

circumstances such as just before or just after exams and to be built into a system that matches timely feedback acquisition with timely dissemination. Finally, allowing flexibility for educators with little contact time to opt in/out of the feedback process will do more to encourage rather than stifle budding educators.

Finally, instruments such as peer review and student feedback would find synergy when paired together with mentorship as part of a broader longitudinal mentoring process. Post-feedback initiatives that match individual educators to targeted faculty development opportunities such as soft skills in teaching and learning will create a holistic yet personalised evaluation-development-evaluation cycle. This will ensure the use of evaluation to drive development so as to create a continuum during the developmental journey of an educator.

This study was not without limitations. The benefits of conducting qualitative research is that the researchers will be able to provide an in-depth and rich understanding of the topic. Nonetheless, these experiences were only elicited from a small group of participants in the medical schools, ALCNS and CBmE. Therefore, the findings of this study are limited to a group of medical educators and students at one health professions institute. For further understanding, a mixed method including a quantitative component might be helpful to gather mass data on the topic. Participation was voluntary, and this may have contributed to selection bias. To mitigate this, we ensure the data collection method conformed to the trustworthiness principles of qualitative research by triangulation with other sources of data, member checking, field notes and audit trails.

Current instruments for evaluating educators in early years of health professions education are not yet optimally developed, implemented and integrated with other aspects of educator development. This prevents the full utility of these instruments in health professions education in the early years. Taking the voices of those evaluating and those evaluated, we

uncovered a set of themes that could pave the way for reforms on how educators in early years of health professions education are evaluated.

In conclusion, we developed a set of short-and long-term recommendations (Box 1) to refine peer review and student feedback based on our findings at NUSMed. As peer review and student feedback are common instruments used to evaluate educators in early years of health professions education in a number of medical schools, we hope our recommendations will be useful for senior education faculty and administrators to consider so as to enhance the utility of these instruments for developing their educators.

Box 1: Short-and long-term recommendations.

Short-term recommendations

1. Review selection process and criteria of peer reviewers.
2. Provide more sustained two-way communication between external reviewer and educator.
3. Solicit feedback after every learning event.
4. Reduce number of questions and create more open-ended questions.
5. Allow educators with limited contact time to opt in/out of student feedback.

Long-term recommendations

1. Embed peer review within a formative mentoring process.
2. Internal reviewer to follow up with educator and mentor over long term.
3. Use evaluation outcomes to match educators to specific faculty development opportunities.

ACKNOWLEDGEMENT

The team would like to thank and appreciate Ms Lim Yih Lin from CenMED in formatting the paper according to the requested format.

REFERENCES

1. Bastani P, Vatankhah S, TaherNejad A, Ghasemi A. Teachers evaluation methods in medical education: round views of faculty members and educational experts. *Galen Med J* 2017; 6:7.
2. Afonso NM, Cardozo LJ, Mascarenhas OAJ, Aranha ANF, Shah C. Are anonymous evaluations a better assessment of faculty teaching performance? A comparative analysis of open and anonymous evaluation processes. *Fam Med* 2005; 37:43-7.
3. Bastani P, Amini M, Taher Nejad K, Shaarbafchi Zadeh N. Faculty members' viewpoints about the present and the ideal teacher evaluation system in Tehran University of Medical Sciences. *J Adv Med Educ Prof* 2013; 1:140-7.
4. Siddiqui ZS, Jonas-Dwyer D, Carr SE. Twelve tips for peer observation of teaching. *Med Teach* 2007; 29:297-300.
5. Taheri MMH, Ryasi HR, Afshar M, Mofatteh MR. Comparison between student rating, faculty self-rating and evaluation of faculty members by heads of respective academic departments in the school of medicine in Birjand University of Medical Sciences in Iran. *J Educ Health Promot* 2014; 3:34.
6. Gelfer J, O'Hara K, Krasch D, Nguyen N. Teacher portfolios: an effective way to assess teacher performance and enhance learning. *Early Child Dev Care* 2015; 185:1495-503.
7. Fluit CRMG, Bolhuis S, Grol R, Laan R, Wensing M. Assessing the quality of clinical teachers: a systematic review of content and quality of questionnaires for assessing clinical teachers. *J Gen Intern Med* 2010; 25:1337-45.
8. Hwang JE, Kim NJ, Song M, et al. Individual class evaluation and effective teaching characteristics in integrated curricula. *BMC Med Educ* 2017; 17:252.

9. Urrutia-Aguilar ME, Sánchez-Mendiola M, Guevara-Guzmán R, Martínez-González A. Comprehensive assessment of teaching performance in medical education. *Procedia Soc Behav Sci* 2014; 141:252-9.
10. Aslam MN. Student rating as an effective tool for teacher evaluation. *J Coll Physicians Surg Pak* 2013; 23:37-41.
11. Elzubeir M, Rizk D. Evaluating the quality of teaching in medical education: are we using the evidence for both formative and summative purposes? *Med Teach* 2002; 24:313-9.
12. Lin CS. Medical students' perception of good PBL tutors in Taiwan. *Teach Learn Med* 2005; 17:179-83.
13. Pettit JE, Axelson RD, Ferguson KJ, Rosenbaum ME. Assessing effective teaching: what medical students value when developing evaluation instruments. *Acad Med* 2015; 90:94-9.
14. Schiekirka S, Feufel MA, Herrmann-Lingen C, Raupach T. Evaluation in medical education: a topical review of target parameters, data collection tools and confounding factors. *Ger Med Sci* 2015; 13:Doc15.
15. Thampy H, Kersey N. Peer observation of clinical teaching: a guide. *MedEdPublish* 2015; 5:9. <https://doi.org/10.15694/MEP.2015.005.0009>.
16. Conigliaro RL, Stratton TD. Assessing the quality of clinical teaching: a preliminary study. *Med Educ* 2010; 44:379-86.
17. Kidd RS, Latif DA. Student evaluations: are they valid measures of course effectiveness? *Am J Pharm Educ* 2004; 68:61.
18. Husain M, Khan S. Students' feedback: an effective tool in teachers' evaluation system. *Int J Appl Basic Med Res* 2016; 6:178-81.
19. Snell L, Tallett S, Haist S, et al. A review of the evaluation of clinical teaching: new perspectives and challenges. *Med Educ* 2000; 34:862-70.

20. Samaraskera D, Yeo SP, Tan CH. Obtaining impactful feedback from students: a continuous quality improvement approach to enhance the quality of students' feedback. *South East Asian J Med Educ* 2014; 8:2-9.
21. Samaraskera DD, Ooi S, Yeo SP, Hooi SC. Medical education in Singapore. *Med Teach* 2015; 37:1-7.
22. Creswell JW. *Qualitative Inquiry & Research Design: Choosing Among the Five Approaches*. Thousand Oaks, CA: SAGE Publications Inc, 2013.
23. Paufler NA, Clark C. Reframing conversations about teacher quality: school and district administrators' perceptions of the validity, reliability, and justifiability of a new teacher evaluation system. *Educ Assess Evaluation Account* 2019; 31:33-60.
24. Daley G, Kim L. A teacher evaluation system that works. In: National Institute for Excellence in Teaching. Available at: <https://files.eric.ed.gov/fulltext/ED533380.pdf>.
25. Derrington ML. Implementing teacher evaluation: lattice of leadership. *J Res Leadersh Educ* 2016; 11:181-99.
26. Downing SM. Validity: on meaningful interpretation of assessment data. *Med Educ* 2003; 37:830-7.
27. Crossley J, Humphris G, Jolly B. Assessing health professionals. *Med Educ* 2002; 36:800-4.
28. Beckman TJ, Ghosh AK, Cook DA, Erwin PJ, Mandrekar JN. How reliable are assessments of clinical teaching? A review of the published instruments. *J Gen Intern Med* 2004; 19:971-7.
29. Fluit C, Bolhuis S, Grol R, et al. Evaluation and feedback for effective clinical teaching in postgraduate medical education: validation of an assessment instrument incorporating the CanMEDS roles. *Med Teach* 2012; 34:893-901.

30. Beckman TJ, Lee MC, Rohren CH, Pankratz VS. Evaluating an instrument for the peer review of inpatient teaching. *Med Teach* 2003; 25:131-5.
31. Dodson RL. An analysis of principals' perceptions of the primary teaching evaluation system used in eight US states. *Int J Educ Policy Leadersh* 2017; 12:n5.
32. Adnot M, Dee T, Katz V, Wyckoff J. Teacher turnover, teacher quality, and student achievement in DCPS. *Educ Eval Policy Anal* 2017; 39:54-76.
33. Herlihy C, Karger E, Pollard C, et al. State and local efforts to investigate the validity and reliability of scores from teacher evaluation systems. *Teach Coll Rec* 2014; 116:1-28.
34. Derrington ML, Campbell JW. Teacher evaluation policy tools: principals' selective use in instructional leadership. *Leadersh Policy Sch* 2018; 17:568-90.
35. Drew S, Phelan L, Lindsay K, et al. Formative observation of teaching: focusing peer assistance on teachers' developmental goals. *Assess Eval High Educ* 2017; 42:914-29.
36. Kraft MA, Gilmour AF. Revisiting the widget effect: teacher evaluation reforms and the distribution of teacher effectiveness. *Educ Res* 2017; 46:234-49.
37. Martinez F, Taut S, Schaaf K. Classroom observation for evaluating and improving teaching: an international perspective. *Stud Educ Eval* 2016; 49:15-29.
38. Bell M. Supported reflective practice: a programme of peer observation and feedback for academic teaching development. *Int J Acad Dev* 2001; 6:29-39.