Original Article

Student Evaluation of Teaching. Is it valid?

Elham Alshammari

Department of Pharmacy Practice, College of Pharmacy, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Correspondence: Elham Alshammari, Department of Pharmacy Practice, College of Pharmacy, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. Email: ejalshammari @ pnu.edu.sa

ABSTRACT

Background: Many institutions of higher learning use student evaluations of teachers (SETs) as one of the requirements for employment decisions. However, the validity of SETs has been called into question using several legal measures of faculty capability. Apart from validity, SETs also tend to suffer from gender bias issues. **Purpose:** The current study sought to investigate the faculty's experience and opinions towards SETs as an assessment tool for teaching competency. **Basic procedures:** A mixed-methods approach was used to collect both qualitative and quantitative data from the study sample, which consisted of 15 pharmacy faculty. Specifically, a questionnaire was used containing close and open-ended questions. **Main findings:** In sum, the findings from the study were as follows. Most teachers believed that they received positive feedback from SET assessments. A larger percentage of the teachers also believed that the students treated them with respect. A majority of teachers encountered negative comments in their evaluation. Many teachers were not sure whether the comments made during their evaluation were honest and representative. Lastly, most pharmacy faculty thought that SET was an invalid tool and did not support its use in the evaluation of their performance. **Principle conclusion:** Student evaluation of teaching is invalid from the faculty's point of view and should be avoided. Alternatives proposed in this study is the use of teacher self-evaluation together with peer evaluation, random pre- and post-tests, ungraded pop quizzes, audio response systems before and after teaching sessions, and class participation.

Keywords: Assessment tool, Teaching evaluation, Student evaluation of teaching, University

Introduction

A majority of institutions of higher learning use student evaluations of teachers (SETs) as one of the requirements for any employment decisions, some of which include tenure and compensation [1, 2]. However, the validity of these evaluations has been tested using various legal measures of faculty capability.[3, 4] Apart from validity, past scholars have also identified various challenges associated with the use of SETs, including reliability issues, gender bias, and many other related issues.[3] Despite the problems raised by researchers, SETs are still the main assessment tool used to evaluate teaching ability, tenure, and opportunities for career advancement [3, 5]. Considering the striking lack of agreement among scholars, the current study sought to investigate the faculty's experience and opinions towards SETs as an assessment tool for teaching competency.

The following is a description of how SETs are used to evaluate faculty for promotion and tenure. The data is collected through paper-and-pencil format or electronic-based surveys.[3] These tools of data collection are administered to learners attending classes. The learners are required to rate the behavior of their instructors in the classroom on a Likert scale.[3] At the moment, the Likert scale used consists of categories, such as "unacceptable," "very poor," "poor," "satisfactory," "good," "very good," and "outstanding."[3] Because of the nature of this evaluation, statistical evaluation of SETs does not involve measures of central tendencies, such as averages or means. Based on how these tools are currently used, it is possible to legally question whether means are used in making decisions surrounding hiring, tenure, and promotion.[3] Most researchers who support the link between SET and competency among instructors do not consider the legitimacy of the data used, neither do they deliberate on whether suitable statistical measures were used.[3] Due to this shortcoming, seminal contributions have proposed the use of SETs as a separate standard that should be met.

Additionally, there is a small number of administrators that have the necessary training to decode SET data.[3] Most administrators often analyze the scores and accept that those below the mean are terrible, while those above it are decent. Such a perception is deceitful considering that the calculation of SET means is improper and pointless in the first place.[3] The reason behind SET means is based on the unlikely belief that all the members of faculty should score above average in all the categories.[3]

Another emerging concern in the use of SET relates to discriminatory practices against female instructors.[4, 6, 7] There is consensus that students use significantly different languages when evaluating male and female instructors.[4] Male instructors tend to receive higher ordinal scores than female instructors, even when administering a similar online course and the questions under evaluation are not specific to the instructors.[4] Based on this evidence, findings confirm the existence of a correlation between gender and teaching evaluations.[4] A theory related to this correlation is that the use of evaluations, such as SET in employment pronouncements discriminates against women.

Previous studies have also emphasized the significance of representation in student evaluations.[6] Representation is important since there is increasing evidence to suggest that gender and cultural background affect all aspects of employment, ranging from hiring to performance evaluation and promotion.[6] As such, predisposed performance evaluation is considered as one of the primary reasons few women achieve the upper stratum of academic hierarchy in higher education. This perspective had been tested using survey data obtained from students in a large public university located in Australia.[6] The data was used to examine the role of deliberate or unwitting prejudice from the perspective of gender and cultural background. The findings demonstrated possible unfairness towards women and teachers who had non-English speaking backgrounds.[6] These findings suggest that better representation of minority groups in the higher education setting could help reduce the occurrence of prejudice. A simple intervention in language has been proposed as a suitable approach to deal with gender bias in student evaluation.[8] The recommendation is based on a randomized experiment examining where learners evaluated the teaching process in four classes.[8] Two of the classes were taught by male instructors while the other two were taught by female instructors. The learners were provided with either a standard evaluation tool or one consisting of language designed to reduce gender bias.[8] It was clear from the findings that learners in the anti-bias language category had higher rankings for female instructors compared to those in the standard group.[8] Besides, these findings demonstrate clear evidence that evaluation tools, such as SET tend to promote unfairness towards the women instructors.

However, evidence of gender bias has been questioned in an investigation using a sample of final third-year medical student obstetrics/ gynecology.[9] The learners completed the clerkship performance evaluation in a retrospective cohort study design. An analysis of the student records revealed that females had higher mean standardized scores compared to males. [9] In other words, women reported higher scores on subject examination

and clinical performance evaluations. No evidence was found to support gender bias in clinical evaluation scores. [9] While these findings appear valid, the setting is different from the higher education sector.

Evaluation programs, such as SETs are also believed to promote unkind comments and responses directed towards female instructors. [7] What this means is that students can use course evaluations as an opportunity to write comments that female instructors consider hurtful. This hypothesis has been tested in research that sought to identify messages that women instructors receive when dealing with inappropriate course evaluation comments.[7] Some of the hurtful student comments tend to examine the appearance of the instructor and their background, instead of focusing on their learning experiences.[7] To put this into context, narrative evaluations have been examined in recent literature to characterize linguistic differences tied to gender.[10] Automated text mining was used to quantify word use during the evaluation of medical faculty. Findings showed that words such as art, trials, master, and humor were mostly associated with evaluations of male faculty. On the contrary, words like empathetic, delight, and warm were often used in the evaluation of female faculty.[10] In terms of two-word evaluations, words like run rounds, the big picture, and master clinician were common in male evaluations, while words like the model physician, just right, and attention to detail were common in female faculty evaluations.[10] Based on these findings, it is clear that there are certain linguistics differences in the evaluation of faculty, mostly favoring men over female instructors.

One possible reason for the hurtful comments is that most student evaluations allow for anonymity.[7] It is such anonymity that allows learners to feel comfortable to move beyond negative evaluation and write down hurtful comments about their instructors. Stated otherwise, learners use the opportunity to be anonymous to do more than critique the teaching methods used by their instructors.[7] However, some past literature identified no significant differences between anonymous and known student evaluations.[11] This causes many female instructors to feel devalued, frustrated, harmed, or threatened in their life and career.[7] Additionally, women faculty are likely to experience anger, aggression, and hostility from the learners, all of which can be considered to be hurtful.[7] A proposed solution for dealing with hurtful comments requires that instructors report back the outcome of the evaluation to the learners.[11] What's more, there are proposals to have instructors discuss the outcome of the evaluation openly with the learners to consider further suggestions within an open environment.[11] It is believed that such an approach will ensure that students respect and understand that instructors take their feedback seriously, and therefore consider providing honest and beneficial responses in future evaluations. While this is possible, it also shows the shortcomings of SETs and related evaluation techniques.

Based on the review of past literature, it is clear that the effectiveness of teaching can only be assessed by student knowledge and performance, mostly using pre- and post-tests.[12] Pre-and post-tests help identify the number of learners on whom the teaching has had maximum impact. Besides, these tests help

rate the extent to which the transfer of concepts fails to occur.[12] The individual readiness assurance test (iRAT) and team readiness assurance test (tRAT) may also offer better choices for student assessment.[13] iRAT requires learners to complete tests independently based on the learning session. The main purpose of the test is to have individual accountability for every student's preparation.[13] The tRAT test starts immediately after iRAT and tests learners on the same questions.[13] The goal of this test is to generate a deeper and shared understanding of course concepts. Even while using pre – and post-tests, there is a risk that the learning may have taken place elsewhere apart from the classroom setting. Nonetheless, it is flawed thinking to emphasize the use of SETs in higher education. SETs should only be used for formative assessment as a self-improvement tool instead of being used for summative purposes and not to make faculty members hesitant to try new methods of teaching.[14-18] In this regard, the goal of the current study is to investigate faculty's experience and opinions towards SETs as an assessment tool for teaching competency.

Materials and Methods

The objective of the study was to obtain pharmacy faculty's experience and opinions toward SET as an assessment tool for teaching competence. A mixed-methods approach was used to collect both qualitative and quantitative data from the study sample, which consisted of 15 pharmacy faculty. Specifically, a questionnaire was used containing close and open-ended questions. The questions were designed to explore the views of the pharmacy faculty and determine whether they support the use of this assessment tool. Participation in the study was entirely voluntary and consent was sought from all the participants. The collected data was presented and analyzed using a quantitative approach.

Results and Discussion

This section presents a summary of the descriptive statistics associated with the data obtained from the questionnaire with the pharmacy faculty. The data is presented in forms of percentages and frequencies.

Do you receive positive comments?

Pharmacy faculty were asked whether they received positive comments from the evaluation. The findings showed that the vast majority of the teachers (86.7%) believe that they received positive feedback from SET assessments. Only 13.3% believed that their assessments were negative (See table 1).

Table 1. Positive comments		
	Frequency	Percent
Yes	13	86.7
No	2	13.3
Total	15	100.0

Do you think the remarks made by students treated you with respect?

Next, the study sample was asked whether they thought the remarks made by the learners were respectful. A majority of the pharmacy faculty (66.7%) believed that the students treated them with respect, while 33% believed that they treated with disrespect (see table 2).

Table 2. Treating with respect		
	Frequency	Percent
Yes	10	66.7
No	5	33.3
Total	15	100.0

Do you ever get negative comments?

The teachers were asked if they ever experienced negative comments in their assessment outcomes. A majority of the study sample (93.3%) ascertained that they encountered negative comments, while only 6.7% did not experience any negative comments (See table 3).

Table 3. Negative comments		
	Frequency	Percent
Yes	14	93.3
No	1	6.7
Total	15	100.0

Do you think students' remarks were honest and representative of the reality?

The pharmacy faculty were asked to discern if the remarks made by the students were honest and representative of reality. Over half of the study sample (53.3%) indicated that maybe the remarks were honest and representative. At the same time, an almost equal number (46.7%) suggested that the remarks were not representative of reality (see table 4).

Table 4. Students' remarks were honest and representative of the reality		
	Frequency	Percent
No	7	46.7
Maybe	8	53.3
Total	15	100.0

Do you think SET discourage you from teaching students' new techniques or extra beneficial tasks? Afraid of their disagreement affecting SET?

The next question asked the study samples to gauge whether SET discouraged them from teaching new or extra beneficial tasks to the learners. Among the responses, 46.7% suggested that SET

did not discourage them from teaching additional beneficial tasks. An equal number of 26.7% answered either "yes" or "maybe" to this question (See table 5).

Table 5. Discourage from teaching students' new techniques or extra beneficial tasks

	Frequency	Percent
Yes	4	26.7
No	7	46.7
Maybe	4	26.7
Total	15	100.0

Do you think student evaluation of teaching (SET) is a valid tool?

The teachers were asked whether they considered SET a valid tool. The highest percentage of teachers (46.7%) thought that SET was an invalid tool. Only 20% thought it as a valid tool. 33.3% were unsure. Hence, it's quite clear that teachers are more likely to consider SET as an invalid tool for evaluation.

Table 6. SET is a valid tool

	Frequency	Percent
Yes	3	20.0
No	7	46.7
Maybe	5	33.3
Total	15	100.0

Are you with or against using the SET tool?

Pharmacy faculty were also asked whether they supported the use of the SET tool. A majority of the study participants (40%) were either against or uncertain about the use of this tool. Only 20% of the teachers supported the use of the SET tool.

Table 7. with or against using the SET tool

	Frequency	Percent
With	3	20.0
Against	6	40.0
Uncertain	6	40.0
Total	15	100.0

Recommendations

The teachers also made the following recommendations for an alternative valid tool for measuring teaching competence. One of the recommendations was to match the marks received by the students (or rather student performance) with their evaluation of teaching competence. In the same light, some proposals from the teachers also advocated for measuring the amount of understanding of the material instead of focusing on the emotional point of view of the learners towards their instructors. Pharmacy faculty also suggested developing an evaluation form that is valid and reliable. This recommendation was informed by the fact that some of the points used in the evaluation are usually inappropriate. Another proposal was to conduct face to face interviews with students as a means of evaluating teaching competence. Some teachers proposed using more objective tools of assessment and avoiding the use of biased comments. Most assessment tools are subjective, thereby hindering students from providing useful comments to their teachers. For instance, one teacher noted that "I have never had any good or acceptable recommendation of improvement from my students." Next, some teachers advised focusing on teacher self-evaluation together with peer evaluation. This approach shifts the attention of assessment from the learners to the teachers and other faculty members. This recommendation is further supported by a teacher who observed the following: "Although SET is widely used across different universities in the world, my view is that it is a tool with poor validity. In my view, the surveys used to evaluate the teaching or competence of the instructors are expressing students' experience in the course. A moderately valid tool for evaluation of teaching may be peer observation of teaching (maybe two times per semester)." An alternative proposal suggested using SET together with peer evaluation techniques. Another reference was made to random pre - and post-tests, ungraded pop quizzes, audio response systems before and after teaching sessions, and class participation. Some teachers saw the opportunity in virtual learning environments and suggested that these could be used to collect student feedback before and after each teaching session.

Conclusion

The goal of the study was to examine pharmacy faculty's experience and opinion toward the use of SET as an assessment tool for teaching competence. Data was collected from a sample of 15 pharmacy faculty. A summary of the findings is as follows.

1) Most teachers believed that they received positive feedback from SET assessments.
2) Most of the teachers believed that the students treated them with respect.
3) A majority of teachers encountered negative comments in their evaluation.
4) Most of the teachers were not sure whether the comments made during their evaluation were honest and representative.
5) Many pharmacy faculty thought that SET was an invalid tool and did not support its use in the evaluation of their performance.

In sum, student evaluation of teaching is invalid from the faculty's point of view and should be avoided. A key alternative proposed in the study is the use of teacher self-evaluation together with peer evaluation. Other alternatives that can also be considered based on the study findings include random pre - and post-tests, ungraded pop quizzes, audio response systems before and after teaching sessions, and class participation.

Conflicts of interest

The author has declared that no competing interests exist

Ethics approval

Acknowledgment

References

1. Sarabi N. Nursing Students' Experiences from Two Methods of Teaching: Film Preparation and Demonstration of Physical Examinations. Int. J. Pharm. Phytopharm. Res. 2019;9(2):14-20.

2. Rakisheva G, Abdirkenova A, Abibulayeva A, Seiitkazy P, Tazhetov A. Tools of Adaptation in the Context of Globalization of Future Teachers: Transcultural Competence. Entomol. appl. sci. lett. 2018 Nov 20;5(4):49-55.

3. Hornstein HA. Student evaluations of teaching are an inadequate assessment tool for evaluating faculty performance. Cogent Education. 2017; 4(1): 1-8.

4. Mitchell KMW, Martin J. Gender bias in student evaluations. American Political Science Association. 2018; 51: 1-5.

5. Hali AU, Zhang B, Al-Qadri AH, Bakar MA. An Overview Of Science Teacher Education In Pakistan. Journal Of Organizational Behavior Research. 2020;5(1):67-74.

6. Fan Y, Shepherd LJ, Slavich E, Waters D, Stone M, Abel R, Johnston EL. Gender and cultural bias in student evaluations: Why representation matters. PLoS One. 2019; 14(2): 1-16.

7. LeFebvre LE, Carmack HJ, Pederson JR. "It's only one negative comment": women instructors' perceptions of (un)helpful support messages following hurtful course evaluations. Communication Education. 2020; 69(1): 19-47.

8. Peterson DAM, Biederman LA, Andersen D, Ditonto TM, Roe K. Mitigating gender bias in student evaluations of teaching. PLoS One. 2019; 14(5): 1-10.

9. Jacques L, Kaljo K, Treat R, Davis J, Farez R, Lund M. Intersecting gender, evaluations, and examinations: Averting gender bias in an obstetrics and gynecology clerkship in the United States. Educ Health (Abingdon). 2016; 19(1): 25-29.

10. Heath JK, Weissman GE, Clancy CB, Shou H, Farrar JT, Dine CJ. Assessment of gender-based linguistic differences in physician trainee evaluations of medical faculty using automated text mining. JAMA Netw Open. 2019; 2(5): e193520.

11. Scherer T, Straub J, Schnyder D, Schaffner N. The effects of anonymity on student ratings of teaching and course quality in a bachelor's degree Programme. GMS Z Med Ausbild. 2013; 30(3): 1-13.

12. Cariappa A, Devasundaram B. Evaluation of teaching effectiveness using student performance in pre-post tests. Medical teacher. 1988; 10(3/4): 313-321.

13. Kolluru S, Roesch DM, de la Fuente AA. A multi-instructor, team-based, active-learning exercise to integrate basic and clinical sciences content. American Journal of Pharmaceutical Education. 2012; 76(2):1-7.

14. Alshammari, E. Demonstration as an effective tool in learning: Saudi pharmacy college experience with insulin pen administration. Int. Res. J. Pharm. 2018; 9(12): 47-49.

15. Alshammari, E. Teaching Schizophrenia: 8-Minutes Video-Based Lecture Versus 1-hour Traditional Lecture. Journal of Pharmaceutical Research International. 2019; 26(5): 1-4.

16. Alshammari, E. Prescription evaluation practice by final year pharmacy students. J Adv Pharm Edu Res. 2019; 9(3): 76-79.

17. Alshammari, E. Simulated Role-playing in Pharmacy. International Journal of Pharmaceutical Quality Assurance. 2020; 11(1): 176-181.

18. Alshammari, E. Implementing Educational Game in Pharmacy. International Journal of Pharmaceutical Quality Assurance. 2020; 11(1): 148-153.
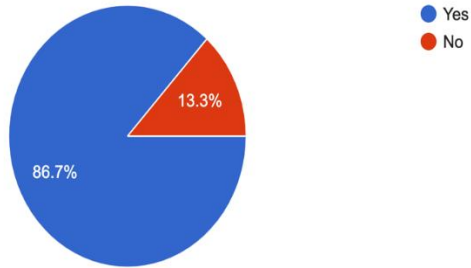
APPENDIX A
Data form:

|	
Years of teaching experience	
Do you receive positive comments?	
Do you think the remarks made by students treated you with respect?	
Do you ever get negative comments?	
Did you ever feel disappointed by getting an offensive comment?	
Do you think students' remarks were honest and representative of the reality?	
Do you think SET discourage you from teaching students new techniques or extra beneficial tasks? Afraid of their disagreement affecting SET?	
Do you think student evaluation of teaching (SET) is a valid tool?	
Are you with or against using the SET tool?	
Recommendations for alternative valid tool measuring teaching competence from your point of view.	

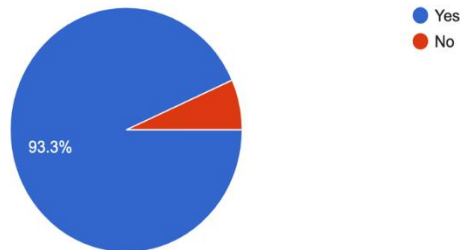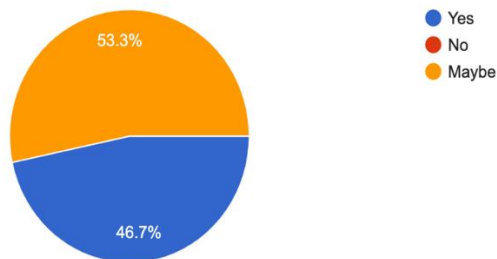**APPENDIX B**

Do you receive positive comments?
15 responses

- Yes
- No

13.3%
86.7%

Do you think the remarks made by students treated you with respect?
15 responses

- Yes
- No

33.3%
66.7%
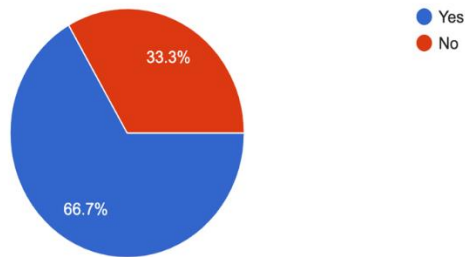
Do you ever get negative comments?
15 responses

- Yes
- No

93.3%

Did you ever feel disappointed by getting offensive comment?
15 responses

- Yes
- No
- Maybe

53.3%
46.7%

Do you think students' remarks were honest and representative to the reality?
15 responses
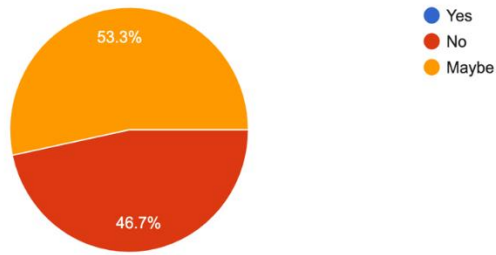
- Yes
- No
- Maybe

53.3%
46.7%

Do you think SET discourage you from teaching students new techniques or extra beneficial tasks?
Afraid of their disagreement affecting SET?
15 responses

- Yes
- No
- Maybe

26.7%
46.7%
26.7%
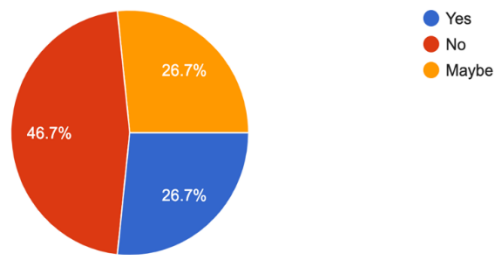
Do you think student evaluation of teaching (SET) is a valid tool?
15 responses

- Yes
- No
- Maybe

33.3%
46.7%
20%
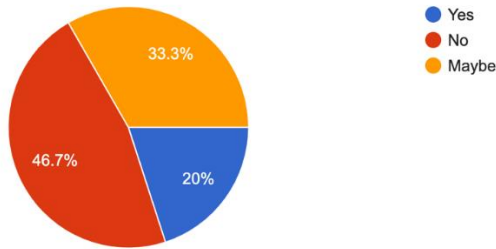
Are you with or against using SET tool?
15 responses

- With
- Against
- Uncertain

40%
40%
20%