Published in final edited form as:

J Dent Educ. 2018 October ; 82(10): 1043–1050. doi:10.21815/JDE.018.101.

Comparison of Dental Students' Self-Evaluation and Faculty Evaluation of Communication Skills During a Standardized Patient Exercise

R. Constance Wiener, Christopher Waters, Joan Doris, and Daniel W. McNeil

R. Constance Wiener, DMD, PhD, is Associate Professor, Department of Dental Practice and Rural Health, School of Dentistry, West Virginia University; Christopher Waters, MS, is Research Labs Director, School of Dentistry, West Virginia University; Joan Doris, DSW, LICSW, is Assistant Professor, Department of Behavioral Medicine and Psychiatry, School of Medicine, West Virginia University; and Daniel W. McNeil, PhD, is Eberly Distinguished Professor, Department of Psychology, and Clinical Professor, Department of Dental Practice and Rural Health, School of Dentistry, West Virginia University. Direct correspondence to Dr. R. Constance Wiener, Department of Dental Practice and Rural Health, School of Dentistry, West Virginia University, PO Box 9415, 104a Health Sciences Addition, Morgantown, WV 26506; 304-581-1960; rwiener2@hsc.wvu.edu.

Abstract

The aims of this retrospective study, conducted in 2017, were to explore dental students' perceptions of their first standardized patient encounters and to assess the relationship between students' self-evaluation and faculty members' evaluation of students' communication skills in those encounters. Data from a simulation training laboratory at one U.S. dental school were obtained for all 46 second-year students, who had a standardized patient communication learning session. The students had completed self-evaluations before and after a debriefing with a clinical psychology and/or social work faculty member and three or four student peers. The faculty members had also completed evaluations of the students' communications skills. The results showed that the students found the standardized patient encounter helpful. The students rated "making an introduction" most positively. Students had a weighted Kappa agreement of 0.22 ($p=0.024$) with the faculty evaluations on their post-debriefing evaluation of overall communication skills, which was within the 0.21–0.40 range of fair agreement. This study found that, during their first standardized patient simulation exercise, the second-year students rated their overall communication skills in fair agreement with the faculty member after debriefing.

Keywords

dental education; communication; patient-doctor relations; patient simulation; standardized patient; self-concept

Disclosure

The authors reported no financial, economic, or professional conflicts of interest.

Simulation-based education is an evolving alternative to the “see one, do one, teach one” method of dental education as simulation is safe and effective and addresses issues of patient care that students may not otherwise experience during their education.¹ The use of a standardized patient (a person coached to portray characteristics of a specified type of patient) is a form of simulation-based education that has been common in health professions education for decades.^{2–4} Even the recording of such interactions has existed for many years.⁵ In a ten-year review of the literature on standardized patient-based education, May et al. reported that most studies endorsed the use of standardized patients as helpful in educating students.³

There are many benefits in supplementing classroom lectures with simulations, and preclinical dental education has used various types of simulation-based education experiences^{6,7} and simulations for development of psychomotor skills needed in clinical application.⁸ Those studies found that case presentations and mechanical, high-fidelity patient simulators/mannequins were an effective and efficient use of time in improving student knowledge. However, the one-on-one experience of having a student work with a standardized patient to develop communication skills has been found to be a very effective means of teaching communication, promoting interpersonal skills, and cultural sensitivity.⁹ Although virtual reality is a promising innovation to improve problem-solving and communication skills,⁶ standardized patients currently offer greater potential for rich feedback. Additionally, Broder et al. found that dental students who had standardized patient experiences involving communication skills retained their competence and improved their data-gathering and interpersonal skills.¹⁰ In another study, dental students were more likely to include key treatment option components in their fourth year when they had had standardized patient exercises in their second year of dental education.¹¹

The first reports of the use of standardized patients in dental education were in 1989 at the University of Illinois at Chicago.² Since then, standardized patients have been used in dental curricula to address development of skills in working with tobacco cessation counseling, emergencies, and interprofessional skills.¹²

Future dental practice will require collaborations with members of other health care professions in the development of team-based care for individuals with complex and chronic medical issues.¹³ In our experience, interdepartmental collaborations, didactic education presented from faculty of other disciplines, and skills education and practices from psychologists and sociologists are common in dental curricula. Educating students to develop a wide variety of communication skills to elicit patient needs beyond dentistry is needed to improve general health outcomes, as on occasions in which dentists are expected to recognize and handle social issues such as child abuse and elder abuse. However, there is also a need to recognize, discuss, and make referrals for other social issues such as intimate partner violence and psychological distress. We found no published studies about dental students’ learning such skills with faculty members from other disciplines.

The aims of this retrospective study were to explore dental students’ perceptions of their first standardized patient encounters and to assess the relationship between students’ self-evaluation and faculty members’ evaluation of students’ communication skills in those

encounters. The teaching and assessment of these second-year (D2) students were done by clinical psychology and social work faculty members.

Methods

This study received “non-human subject research” acknowledgment status from the West Virginia University Institutional Review Board under protocol 1798694490. The study, conducted in 2017, was a secondary data analysis using existing data in the West Virginia University (WVU) David and Jo Ann Shaw Center for Simulation Training and Education for Patient Safety (STEPS) database (WVU IRB 130403754).

The purpose of the simulation experience from which the data were extracted was to enhance the communication skills of D2 students in the summer before their initial clinical experiences. The high-fidelity simulation was part of a dental course in which D2 students were also introduced to working with allied dental professionals. The simulation was the students’ first experience with a standardized patient. The experience was required as part of the course, but the assessment was not used in determination of final grades.

The standardized patient educational experience was explained to the students by members of the WVU STEPS team. Their presentation was followed by a two-hour didactic lecture presented by one of the coauthors (DWM) concerning the Biopsychosocial Model applied to health, health behaviors, and communication¹⁴ and Motivational Interviewing as a communication style to elicit patients’ own motivations for behavior change in the interest of their health.¹⁵ The dental curriculum has communication skills integrated into courses across all years, as part of an overall infusion model for the behavioral and social sciences. The students had previous didactic exposure and in-class activities related to working with patients throughout the life course.

During the learning session in preparation for the standardized patient experience, the D2 students demonstrated their understanding of the didactic material by selecting a classmate as a partner and role-playing the discussed skills in the following areas: introductions (fully introducing themselves and their role); expressions of empathy (using supportive remarks, appropriate tone, and continuer phrases); nonverbal aspects of communication (addressing each other at eye level, social smiling, nodding, maintaining an appropriate distance, maintaining appropriate eye contact, identifying subtleties in mannerisms indicative of need for psychological/medical/social service referral); interviewing (asking open-ended questions, waiting for responses, posing single questions at a time, avoiding assumptions); interacting (using simple reflections repeating or rephrasing comments, using complex reflections including paraphrasing, amplifying, understating, reflecting feelings, or using double-sided reflections of ambivalence, providing advice after asking permission and eliciting what the person would like to learn, checking with the person that he or she understood what was discussed); and being organized (asking about the chief concern/problem list in a logical, systematic, structured manner while allowing the other person to discuss concerns, determining if referral was needed, and leading the discussion without being domineering).

The D2 students then went to the WVU STEPS simulation educational center where they had their initial meetings with standardized patients. WVU STEPS coordinators had hired standardized patients to present one of four roles to each student: drug seeking, recent intimate partner violence, financial insecurity, and dental fear. The standardized patients had been trained by two faculty members, and extensive case histories had been developed to guide the actors to fully develop each role. The students had not been specifically briefed on the scenarios, but they had been presented information about such issues throughout their coursework. For example, the first-year, second-semester course in public health dentistry has a two-hour session (presented by DWM) that includes dental fear, health belief models, and communicating with/motivating groups and another two-hour session (presented by coauthor JD) on intimate partner violence. The students are given information about financial payment programs available to patients and other sources of care, and students become aware of drug-seeking behavior from the pharmacology sessions they attend.

The student-standardized patient encounter was video- and audio-recorded. It was viewed simultaneously in a remote debriefing room by three or four other D2 students and a male clinical psychology faculty member with a dual appointment in the Department of Psychology and the Department of Dental Practice and Rural Health (DWM) or a female social work faculty member from the School of Medicine (JD). The male and female students were equally distributed to both faculty members.

The two faculty members had previously been calibrated (by discussion) for agreement on the meaning of the assessment terms “deficient,” “needs improvement,” “meets expectations for year level,” “advanced for year level,” and “exceptional/almost or at professional level.” The faculty members developed the scenarios and had worked on a pilot of the program during the previous year in which they developed and refined their rubric for evaluation consistency.

Students are required to self-evaluate throughout the dental curriculum. They self-evaluate their clinical skills, educational presentations, and clinical experiences with patients. As part of the clinical aspect of the course in which the standardized patient experience occurred, the students received concrete criteria for self-evaluating their patient-dental assistant-student encounters.

The student-standardized patient encounter lasted up to ten minutes. Immediately afterwards, the student, faculty member, standardized patient, and D2 peers evaluated the encounter using a guided rubric with ranked responses. The rubric included items related to the student’s manner of introducing himself or herself, expression of empathy, display of non-verbal empathy, interviewing skills, interacting skills, organizing skills, expression of confidence, and an overall evaluation. The response options were 1=deficient, 2=needs improvement, 3=meets expectations for year level, 4=advanced for year level, and 5=exceptional/almost or at professional level. In addition, the students were asked to evaluate how comfortable they felt during the interview, with response options of 1=quite uncomfortable, 2=slightly uncomfortable, 3=reasonably comfortable, 4=very comfortable, and 5=quite comfortable. They were also asked if the experience was helpful, with response options of 1=poor, 2=fair, 3=satisfactory, 4=good, and 5=excellent. The evaluation

completed by the student who had the standardized patient encounter was identified as the pre-debriefing self-evaluation.

Following completion of the rubrics, the student joined his or her peers and the faculty member for a debriefing. As Quick noted, the initial moments in the debriefings can be emotionally intense and are essential to the learning process.¹⁶ After initial emotional processing, the feedback consisted of the faculty member doing the following: 1) asking the participant to describe, overall, his or her opinion of the session; 2) asking the participant to describe some of his or her good aspects of the encounter; 3) asking the participant's peers to note some good qualities exhibited by the participant during the encounter; 4) providing the faculty member's comments about good qualities shown during the encounter; 5) asking the participant to describe some aspects of the encounter that needed improvement; 6) asking the participant's peers to note some aspects of the encounter that needed improvement; 7) providing the faculty member's comments about aspects of the encounter that needed improvement; 8) asking the participant if he or she had other comments; and 9) summarizing the session.

The faculty member had the opportunity to enrich the discussion with related verbal presentation and discussion. The recording of the student-standardized patient encounter was released for the student to view several hours after the session. Each student in the student-standardized patient encounters was asked to provide a post-debriefing self-evaluation after viewing the recording, reflecting on his or her peers' and the faculty member's observations, and considering the video from a third person perspective.

The WVU STEPS database maintains faculty member, student, peer, and standardized patient responses under its WVU IRB approval. Data requests were made to access de-identified data for the D2 students' pre-debriefing and post-debriefing evaluations and the faculty members' evaluations.

The key variables of interest were the students' overall post-debriefing self-evaluation scores and the faculty members' overall evaluation scores. The data were analyzed for descriptive statistics (mean and standard deviation). The primary outcome of interest was the weighted Kappa agreement between the students' post-debriefing evaluation and the faculty members' evaluation on the "overall" item. All rubric items were evaluated to determine weighted Kappa agreement for the students' pre-debriefing self-evaluations of the eight items on the list and the faculty members' evaluations; the students' post-debriefing self-evaluations of the eight items and the faculty members' evaluations; and the students' pre- and post-debriefing self-evaluations of the eight items. A stringent criterion was used to determine "agreement," so that the student and faculty member must agree exactly on the rating. The following categories of agreement were used: 0.01–0.20 was considered a slight Kappa agreement; 0.21–0.40 was considered a fair agreement; 0.41–0.60 was considered a moderate agreement; 0.61–0.80 was considered a substantial agreement; and 0.81–0.99 was considered an almost perfect agreement.¹⁷ Analyses were conducted using SPSS version 24 (IBM Corp., Armonk, NY, USA).

Results

There were 46 D2 students who had standardized patient encounters; 26 of them (56.5%) were females. Their mean evaluation of the helpfulness of the experience on the five-point scale was 3.95 (SD 0.75). One faculty member's mean evaluation of the overall experience for the students (out of a possible score of 5) was 3.05 (SD 0.38); for the second faculty member, it was 3.08 (SD 0.88). Descriptive statistics for the faculty members' evaluations are shown in Table 1.

The item that had the students' highest mean score in the pre-debriefing was "non-verbal empathy" with a mean score of 3.14. In the post-debriefing, the students' highest mean score was for "making an introduction" (3.41). The faculty members' highest mean score was for the "overall" item, which had a mean score of 3.07 (Table 2).

The difference between the students' pre-debriefing self-evaluations and the faculty evaluations for the measure "overall evaluation" was not statistically significant ($p=0.95$). The difference for "overall evaluation" was statistically significant between the students' post-debriefing self-evaluations and the faculty evaluations ($p=0.024$) with a weighted Kappa of 0.22, which was within the 0.21–0.40 range of fair agreement. Other items found to be significant in weighted Kappa were "making an introduction," "organized," and "confidence."

There was fair agreement between the students' pre- and post-debriefing self-evaluations for "making an introduction," "verbal empathy," "interviewing skills," "interaction," "organized," "confidence," and "overall." The students' mean pre-encounter level of comfort was 2.77 (SD 0.80). Immediately after completing the encounter, their mean level of comfort was 3.00 (SD 0.78). In recalling the experience later in the day, the recalled pre-encounter level of comfort was 2.87 (SD 0.75), and the recalled post-encounter level of comfort was 3.30 (SD 0.66).

The female students had a mean evaluation of the helpfulness of the experience of 4.09 (SD 0.73). Female sub-group analyses are shown in Table 3. The item that had the highest mean score in the pre-debriefing self-evaluations for the female students was "non-verbal empathy" with a mean score of 3.26. The faculty members' highest mean score for the female students was 3.13 for the items "overall," "non-verbal empathy," and "interaction."

The difference on the "overall evaluation" measure between the female students' pre-debriefing self-evaluations and the faculty evaluations was statistically significant (weighted Kappa=0.25; 95% CI: -0.03, 0.53; $p=0.036$); however, the difference between the female students' post-debriefing self-evaluations and the faculty evaluations was not significant ($p=0.081$). There were three items on which the female students' post-debriefing and the faculty evaluations agreed: "making an introduction," "verbal empathy," and "non-verbal empathy."

The male students had a mean evaluation of the helpfulness of the experience of 3.81 (SD 0.75). Male sub-group analyses are shown in Table 4. The item that had the highest mean score on the male students' pre-debriefing self-evaluation was "non-verbal empathy" (3.00).

The faculty members' highest mean score for the male students was for the "overall" item (3.05). The difference between the male students' pre-debriefing self-evaluation and the faculty evaluations on the "overall evaluation" measure was not statistically significant ($p=0.626$), nor was the difference between the male students' post-debriefing self-evaluations and the faculty evaluations statistically significant ($p=0.154$).

Discussion

In this study, the second-year dental students reported that the standardized patient experience was helpful (mean evaluation of 3.95 on a scale from 1=poor to 5=excellent). A similar result was reported by researchers who conducted literature reviews of dental simulations in which students worked with standardized patients.^{3,18} In another study, researchers reported students evaluated their standardized patient experiences as helpful immediately after the session; however, the students reported less perceived value and applicability when asked about the experience at a later time.¹⁹

In our study, the primary outcome (comparison of the overall students' post-debriefing self-evaluations with the faculty evaluation) had a significant weighted Kappa of 0.22 (95% CI: 0.01, 0.43; $p=0.024$), which was within the 0.21–0.40 range of fair agreement. The debriefing session guided by the faculty members seemed to have helped these students notice some of the factors that influenced the faculty evaluations. The students were in agreement on their post-debriefing evaluations with the manner in which they made an introduction, their organization, and their confidence. Each had an agreement between 0.21 and 0.24, which was within the 0.21–0.40 range of fair agreement. Debriefing has been found to be highly important in the learning process involving simulations, but other factors also influenced student self-evaluations.^{20,21}

The debriefing seemed to have helped the students in our study to self-evaluate in a manner similar to the faculty evaluation of overall communication skills to a fair degree. However, the students remained consistent with their pre-debriefing evaluations. The concepts of commitment and consistency can be influencing factors. In the theory of commitment and consistency, once a stand or decision has been taken, particularly if it has been written and made public (which was the case of the students responding to the pre-debriefing survey and debriefing experience), Cialdini found there was a personal and interpersonal pressure to be consistent with that commitment.²² Typically, according to Cialdini, consistency is to one's best interests and is a motivator for behavior/habits. However, that author also reported that avoiding critical thinking (cognitive analysis utilizing new evidence, logic, and reason) for the ease of an automatic consistent response was a luxury of convenience, relative ease, and efficiency as well as an expedient that avoided the labor of thinking. Future research could explore this possibility.

In our subgroup analyses by gender, the weighted Kappa scores reported by the female students in their post-debriefing self-evaluations were higher in seven of the eight categories than the self-reported scores of the male students. Additionally, the faculty members rated the female students higher in all categories than the male students, with seven of the eight categories at or above the expected level for the female students, and two of the eight

categories at or above the expected level for the male students. The explanation for such differences requires future research.

There are limited prior studies with which to compare our outcomes of dental student communication skill development, especially since the study designs with other health professions students-standardized patient encounters were only slightly similar to our study's design or our terminology for categories. For example, the study design of a medical school communication research project with a standardized patient had a workshop setting in which self-selected medical students (n=34) completed pre and post self-ratings on competence.¹⁸ That study found that students' self-ratings increased significantly after training with a standardized patient. Similar results on the importance of training in communication skills were reported by researchers who asked medical students to evaluate an interviewing skills course being developed.²¹

Our study was conducted under stringent simulation conditions. The second-year dental students were presented with standardized patients who had practiced their assigned roles and had been employed as professional standardized patients in the past by the WVU coordinators. The environment included such items as dental instruments, a dental stool for the student, a dental chair/examination table for the patient, and a sink for hand-washing. Students arrived in scrubs and wore personal protective equipment to increase the authenticity of the experience. All of the students in the class participated.

A limitation of the study was the limited sample size (n=46). Another limitation was the potential for students to discuss the experience with students waiting for their scheduled meeting with the standardized patient; however, this was limited by the request of the faculty members to not discuss the experience. Another potential limitation was students' apprehension to openly discuss the experience with the faculty members and peers; however, this was managed by the small number of students present during the debriefing session and the objective, non-judgmental tone set by the faculty for the debriefing sessions. To address the potential for bias, the rubrics used by the faculty and students were the standardized ones used in the WVU STEPS center. Further research is warranted to better understand our findings and improve dental student education regarding communication skills.

Conclusion

The results of this study showed that the second-year dental students found the standardized patient encounter helpful, rating "making an introduction" most positively. The students had a weighted Kappa agreement of 0.22 (p=0.024) with the faculty evaluations on their post-debriefing evaluation of overall communication skills, which was within the 0.21–0.40 range of fair agreement.

Acknowledgments

Research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number U54GM104942. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The authors

would also like to thank the West Virginia University David and Jo Ann Shaw Center for Simulation Training and Education for Patient Safety Research Advisory Committee for providing access to the 2017 data.

REFERENCES

1. Castillo J, Kutzin J, Agoglia K, et al. An innovative clinical skills “boot camp” for dental medicine residents. *J Med Educ Curr Devel* 2017;4:1–4.
2. Johnson JA, Kopp KC, Williams RG. Standardized patients for the assessment of dental students’ clinical skills. *J Dent Educ* 1990;54(6):331–3. [PubMed: 2351773]
3. May W, Park JH, Lee JP. A ten-year review of the literature on the use of standardized patients in teaching and learning: 1996–2005. *Med Teach* 2009;31(6):487–92. [PubMed: 19811163]
4. Barrows HS, Abrahamson S. The programmed patient: a technique for appraising student performance in clinical neurology. *J Med Educ* 1964;39:802–5. [PubMed: 14180699]
5. Kopp KC. Checklist agreement between standardized patients and faculty. *J Dent Educ* 1995;59(8):824–9. [PubMed: 7499599]
6. Phillips J, Berge ZL. Second life for dental education. *J Dent Res* 2009;73:1260–4.
7. Buchanan JA. Use of simulation technology in dental education. *J Dent Res* 2001;11:1225–31.
8. Perry S, Bridges SM, Burrow MF. A review of the use of simulation in dental education. *Simulation Healthcare* 2015;10:31–7.
9. Broder HL, Janal M. Promoting interpersonal skills and cultural sensitivity among dental students. *J Dent Educ* 2006;70(4):409–16. [PubMed: 16595533]
10. Broder HL, Janal M, Mitnick DM, et al. Communication skills in dental students: new data regarding retention and generalization of training effects. *J Dent Educ* 2015;79(8):940–8. [PubMed: 26246533]
11. McKenzie CT, Tilashalski KR, Peterson DT, White ML. Effectiveness of standardized patient simulations in teaching clinical communication skills to dental students. *J Dent Educ* 2017;81(10):1179–86. [PubMed: 28966182]
12. Anders PL, Scherer YK, Hatton M, et al. Using standardized patients to teach interprofessional competencies to dental students. *J Dent Educ* 2016;80(1):65–72. [PubMed: 26729686]
13. Andrews EA. The future of interprofessional education and practice for dentists and dental education. *J Dent Educ* 2017;81(8 Suppl):eS186–92. [PubMed: 28765471]
14. Engel GL. The need for a new medical model: a challenge for biomedicine. *Science* 1977;196:129–36. [PubMed: 847460]
15. McNeil DW, Addicks SH, Randall CL. Motivational interviewing and motivational interactions for health behavior change and maintenance. New York: Oxford University Press, 2017.
16. Quick KK. The role of self- and peer assessment in dental students’ reflective practice using standardized patient encounters. *J Dent Educ* 2016;80(8):924–9. [PubMed: 27480703]
17. Viera AJ, Garrett JM. Understanding interobserver agreement: the Kappa statistic. *Fam Med* 2005;37(5):360–3. [PubMed: 15883903]
18. Engerer C, Berberat PO, Dinkel A, et al. Integrating 360 degree behavior-oriented feedback in communication skills training for medical undergraduates: concept, acceptance, and students’ self-ratings of communication competence. *BMC Med Educ* 2016;1:271.
19. Beattie BE, Kinney J, Fitzgerald M, et al. Dental and dental hygiene students’ perceptions of a standardized patient instructor conflict resolution program. *J Dent Educ* 2014;78(10):1397–404. [PubMed: 25281673]
20. Fanning RM, Gaba DM. The role of debriefing in simulation-based learning. *Simulation Healthcare* 2007;2(2): 115–25.
21. Laidlaw TS, MacLeod H, Kaufman DM, et al. Implementing a communication skills program in medical school: needs assessment and program change. *Med Educ* 2002;36(2):115–24. [PubMed: 11869438]
22. Cialdini RB. Influence: science and practice. Pymble, NSW: HarperCollins Ebooks, 2009.

Table 1.

Each faculty member's evaluations of students' performance, by mean (SD)

Item	Faculty Member 1	Faculty Member 2
Introduction	2.83 (0.49)	3.05 (0.38)
Verbal empathy	2.92 (0.78)	3.00 (0.44)
Non-verbal empathy	3.00 (0.42)	3.05 (0.38)
Interviewing skills	2.91 (0.79)	2.95 (0.21)
Interaction	3.04 (0.86)	3.05 (0.58)
Organized	2.96 (0.36)	2.95 (0.37)
Confidence	2.88 (0.54)	2.91 (0.53)
Overall	3.08 (0.88)	3.05 (0.38)

*Note:*The response options were 1=deficient, 2=needs improvement, 3=meets expectations for year level, 4=advanced for year level, and 5=exceptional/almost or at professional level.

Table 2.

Comparison of total students' self-evaluations and faculty evaluations

Item	Pre-Debriefing Mean (SD); p-value [‡]		Post-Debriefing Mean (SD); p-value [‡]		Faculty Mean (SD)	Pre-Debriefing Student and Faculty Agreement [‡] [95% CI]		Post-Debriefing Student and Faculty Agreement [‡] [95% CI]		Pre-Debriefing Student and Post-Debriefing Student Agreement [95% CI]		p-value [‡]
	Mean (SD)	p-value [‡]	Mean (SD)	p-value [‡]		Agreement [‡] [95% CI]	p-value [‡]	Agreement [‡] [95% CI]	p-value [‡]	Agreement [95% CI]	p-value [‡]	
Introduction	3.00 (.81); .406	3.41 (.78); <.0005	3.41 (.78); <.0005	2.93 (.45)	2.93 (.45)	.13 [-.05, .32]	.082	.21 [.04, .40]	<.0005	.22 [.03, .41]	.013	
Verbal empathy	3.05 (.75); .579	3.20 (.72); .091	3.20 (.72); .091	2.96 (.63)	2.96 (.63)	.09 [-.14, .31]	.402	.15 [-.07, .38]	.113	.29 [.03, .54]	.004	
Non-verbal empathy	3.14 (.73); .337	3.20 (.72); .129	3.20 (.72); .129	3.02 (.39)	3.02 (.39)	.13 [-.09, .34]	.142	.09 [-.10, .28]	.256	.17 [-.06, .40]	.107	
Interviewing skills	2.61 (.69); .025	2.85 (.67); .564	2.85 (.67); .564	2.93 (.58)	2.93 (.58)	.09 [-.10, .29]	.298	.09 [-.11, .28]	.380	.35 [.15, .55]	.001	
Interaction	2.66 (.53); .002	2.96 (.67); .376	2.96 (.67); .376	3.04 (.73)	3.04 (.73)	.12 [-.06, .31]	.155	.08 [-.13, .29]	.414	.30 [.09, .50]	.005	
Organized	2.70 (.73); .034	3.04 (.84); .440	3.04 (.84); .440	2.96 (.36)	2.96 (.36)	.08 [-.11, .26]	.319	.17 [-.01, .35]	.017	.37 [.18, .56]	<.0005	
Confidence	2.91 (.78); .974	3.07 (.68); .101	3.07 (.68); .101	2.89 (.53)	2.89 (.53)	.03 [-.19, .24]	.759	.24 [.02, .46]	.007	.34 [.09, .59]	.001	
Overall	2.73 (.62); .005	3.04 (.73); .876	3.04 (.73); .876	3.07 (.68)	3.07 (.68)	.15 [-.05, .36]	.095	.22 [.01, .43]	.024	.47 [.26, .67]	<.0005	

Note: The response options were 1=deficient, 2=needs improvement, 3=meets expectations for year level, 4=advanced for year level, and 5=exceptional/almost or at professional level.

[‡] Wilcoxon signed rank test was used to compare median of differences between the students' and faculty ratings.

[‡] Agreement was determined by weighted Kappa test.

Table 3.

Comparison of female students' self-evaluation with faculty evaluations

Item	Pre-Debriefing Mean (SD);		Post-Debriefing Mean (SD);		Faculty Mean (SD)		Pre-Debriefing Student and Faculty Agreement [‡] [95% CI]		Post-Debriefing Student and Faculty Agreement [‡] [95% CI]		Pre-Debriefing Student and Post-Debriefing Student Agreement [95% CI]		p-value [†]
	p-value [‡]		p-value [‡]										
Introduction	3.09 (.67); .572		3.50 (.83); .003		3.00 (.42)		.26 [-.12, .63]		.27 [.03, .50]		.28 [.02, .54]		.027
Verbal empathy	3.17 (.78); .218		3.33 (.70); .013		2.96 (.62)		.23 [-.07, .53]		.36 [.07, .65]		.15 [-.17, .47]		.310
Non-verbal empathy	3.26 (.26); .454		3.29 (.75); .248		3.13 (.45)		.21 [-.13, .56]		.32 [.01, .63]		.22 [-.08, .52]		.146
Interviewing skills	2.61 (.58); .008		2.96 (.62); .439		3.08 (.65)		.19 [-.08, .46]		.10 [-.19, .39]		.24 [-.04, .51]		.105
Interaction	2.70 (.56); .012		3.13 (.63); .741		3.13 (.80)		.22 [-.03, .47]		.08 [-.24, .40]		-.03 [-.25, .20]		.834
Organization	2.83 (.65); .157		3.13 (.80); .417		3.00 (.42)		.35 [.03, .67]		.19 [-.12, .49]		.42 [.12, .72]		.002
Confidence	3.00 (.62); .627		3.08 (.58); .248		2.92 (.58)		.02 [-.27, .30]		.20 [-.12, .51]		.49 [.16, .83]		.001
Overall	2.74 (.54); .007		3.00 (.78); .439		3.13 (.62)		.25 [-.03, .53]		.24 [-.06, .53]		.43 [.15, .71]		.004

Note: The response options were 1=deficient, 2=needs improvement, 3=meets expectations for year level, 4=advanced for year level, and 5=exceptional/almost or at professional level.

[†] Wilcoxon signed rank test was used to compare median of differences between the students' and faculty ratings.

[‡] Agreement was determined by weighted Kappa test.

Table 4.

Comparison of male students' self-evaluation with faculty evaluations

Item	Pre-Debriefing Mean (SD); p-value [‡]		Post-Debriefing Mean (SD); p-value [‡]		Faculty Mean (SD)		Pre-Debriefing Student and Faculty Agreement [‡] [95% CI]		Post-Debriefing Student and Faculty Agreement [‡] [95% CI]		Pre-Debriefing Student and Post-Debriefing Student Agreement [95% CI]		p-value [‡]
	Mean (SD)	p-value [‡]	Mean (SD)	p-value [‡]	Mean (SD)		Agreement [‡] [95% CI]	p-value [‡]	Agreement [‡] [95% CI]	p-value [‡]	Agreement [95% CI]	p-value [‡]	
Introduction	2.90 (.94); .577		3.32 (.72); .004		2.86 (.48)		.04 [-.14, .22]	.639	.14 [-.08, .37]	.002	.16 [-.08, .40]	.184	
Verbal empathy	2.90 (.70); .851		3.05 (.72); .597		2.91 (.68)		-.11 [-.40, .18]	.442	-.11 [-.34, .13]	.457	.40 [-.003, .81]	.006	
Non-verbal empathy	3.00 (.71); 1.00		3.09 (.68); .593		3.00 (.31)		-.07 [-.15, .01]	.518	-.08 [-.18, .02]	.433	.08 [-.24, .39]	.604	
Interviewing skills	2.62 (.81); .378		2.73 (.70); .617		2.86 (.66)		-.03 [-.25, .20]	.842	.03 [-.23, .28]	.851	.44 [.17, .71]	.005	
Interaction	2.62 (.50); .071		2.77 (.69); .317		2.95 (.65)		-.02 [-.26, .22]	.854	.07 [-.22, .35]	.607	.61 [.36, .86]	<.0005	
Organization	2.57 (.81); .108		2.95 (.90); .796		2.91 (.29)		-.16 [-.37, .06]	.086	.16 [-.03, .36]	.050	.29 [.02, .55]	.045	
Confidence	2.81 (.93); .617		3.05 (.79); .248		2.86 (.47)		.02 [-.30, .34]	.882	.27 [-.03, .57]	.033	.21 [-.14, .57]	.124	
Overall	2.71 (.72); .108		3.09 (.68); .796		3.05 (.84)		.06 [-.21, .34]	.626	.19 [-.08, .45]	.154	.51 [.24, .78]	<.0005	

Note: The response options were 1=deficient, 2=needs improvement, 3=meets expectations for year level, 4=advanced for year level, and 5=exceptional/almost or at professional level.

[‡] Wilcoxon signed rank test was used to compare median of differences between the students' and faculty ratings.

[‡] Agreement was determined by weighted Kappa test.