



Comparison of student self-assessment and teacher assessment of medical interview performance during bedside learning

Zunyi Tang, Yayoi Shikama and Koji Otani

Center for Medical Education and Career Development, Fukushima Medical University, Fukushima, Japan

(Received January 6, 2023, accepted February 8, 2023)

Abstract

Accurate and efficient self-assessment is a critical skill for medical students to develop as part of their professional development. Along with clinical training reform at Fukushima Medical University, rubric-based student self-assessment and teacher assessment of students' performance using our proposed assessment tool, which includes several aspects of clinical skills and abilities, was initiated to improve the clinical clerkship process. To investigate how students identified their weaknesses and strengths, we analyzed the results of 119 fourth-year medical students' self-assessment and corresponding teacher assessment. Our study revealed substantial consistency between student self-assessment and teacher assessment, despite some overestimation and underestimation in student self-assessments. Students who incorrectly assess themselves require a variety of feedback to increase their self-efficacy and self-confidence, as well as to identify their weaknesses.

Key words : medical education, student self-assessment, clerkship, overestimation, underestimation

Introduction

Self-assessment has been broadly defined as an essential component of professional self-regulation because it can assist in identifying not only one's weaknesses but also one's strengths for self-directed learning activities and self-improvement¹⁻⁷⁾. Understanding their weaknesses allows students to determine what they needed to learn, and knowing their strengths allows them to move forward with confidence. Medical students are expected to identify their own learning needs in order to constantly improve their performance during learning and clerkship. Thus, accurate, ongoing self-assessment becomes an important skill for medical students, especially during clerkships¹⁾. Based on this consideration and the needs of medical education reform at Fukushima Medical University, we implemented the proposed tool to facilitate student self-assessment and teacher assessment in or-

der to improve student self-assessment and the clerkship process⁸⁻¹⁰⁾. Simultaneously, we developed an assessment tool based on MoodleTM, an online learning management system, so that students and faculty could complete their assessments while also checking their learning goals or education goals⁸⁾. This change was implemented in 2018 for fourth-year medical students beginning clinical clerkships. It is necessary to investigate how students identified their strengths and weaknesses using this assessment tool.

Each student conducts a medical interview with a simulated patient during their clerkship at Fukushima Medical University's Center for Medical Education and Career Development (CMECD), and after debriefing with all participants, including peers, simulated patients, and faculty, both faculty members in charge and students themselves submit their assessments on the same day. We examined the agreement between these teacher assessments and

Corresponding author : Zunyi Tang E-mail : tangzy@fmu.ac.jp

©2023 The Fukushima Society of Medical Science. This article is licensed under a Creative Commons [Attribution-NonCommercial-ShareAlike 4.0 International] license.
<https://creativecommons.org/licenses/by-nc-sa/4.0/>

the students' self-assessments in this study to clarify the accuracy of the self-assessment and identify future issues.

Methods

Subjects

Study participants were a total of 127 fourth-year medical students in academic year 2018. They participated in this course in groups of 6-8 students from October 2018 to June 2020.

Clinical clerkships at CMECD

The CMECD clinical clerkship was a mandatory half-day course taken once during the fourth or fifth year, with 6 to 8 students, two simulated patients, two faculty advisors, and two facilitators per practice session. Each student conducted one of the following interviews: a medical interview to diagnose a new patient, a behavior change interview, and bad news telling. After each performance, students reflected on their own performance, received feedback from their peers, the simulated patient, and the faculty, and expressed their thoughts on the feedback.

Assessment

Two faculty advisors participated in each CMECD clinical clerkship and provided feedback to each student according to internationally accepted guidelines for behavior change interviewing and bad news telling after listening to reflections from the students, who themselves conducted the interview with peers in the same group, and with simulated patients. This way, the focus and evaluation criteria were shared among the faculty advisors. Each

faculty advisor rated each student's performance on a four-point scale based on a rubric for five items: pathophysiology understanding (knowledge), clinical reasoning, interviewing skills, communication, and learning attitude, with levels 4 (excellent as a student), 3 (desirable level), 2 (minimally acceptable level), and 1 (below acceptable level) (Appendix 1-5). The rubric for each item was displayed on the Moodle evaluation site and the evaluators entered their scores as they viewed the rubric. Students used the same rubric to rate their own performance (self-assessment) after the interview session. Upon comparing student self-assessment to teacher assessment, we defined overestimation as a score higher than the faculty assessment, and underestimation as a score lower than the teacher assessment.

Statistical analysis

The chi-squared test and the *t*-test were used. All statistical analyses were performed using IBM SPSS Statistics, version 28.0.0.0 (IBM Japan, Ltd., Japan), all the tests were two-tailed, and $p < 0.05$ was considered statistically significant.

Results

Teachers' and students' evaluations

One hundred twenty-seven students took the course and were evaluated by the faculty, and one hundred nineteen students (88 male; 31 female) completed all self-assessments. Figure 1 shows the distribution of teachers' and students' evaluations. A total of 595 teachers' evaluations were completed, with 152, 375, and 65 rated as excellent,

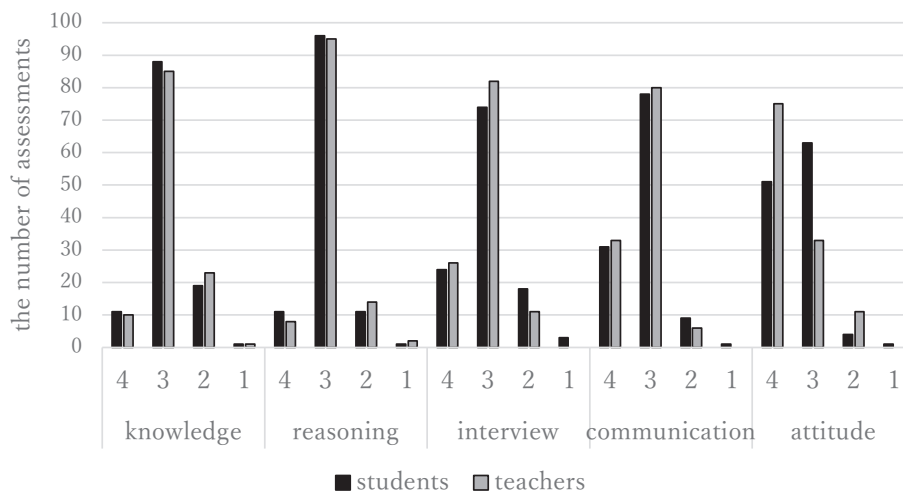


Fig. 1. Teachers' assessments and students' self-assessments.

desirable level, and minimum acceptable level, respectively. One case had knowledge that was below the minimum acceptable level, and two cases had clinical reasoning that was below the minimum acceptable level. A total of 595 students' self-assessments were completed, with 128, 399, and 61 rated as excellent, desirable, and minimum acceptable, respectively. One case fell below the minimum acceptable level in knowledge, clinical reasoning, communication, and learning attitude, and three cases fell below the minimum acceptable level in interviewing skills.

Difference between student self-assessment and teacher evaluation

When student self-assessments were compared to teacher assessments, there was 58.8% agreement, 70.6% agreement, 46.2% agreement, 58.0% agreement, and 47.9% agreement for knowledge, clinical reasoning, medical interview, communication, and learning attitude, respectively. Approximately 20% of the self-assessments were higher than the teacher's assessment in each category (Fig-

ure 2).

We then examined the differences between self-assessment and teacher assessment by grade level. As shown in Figure 3, the percentages of teacher assessments that were lower than the self-assessments were 100%, 80.0%, 16.8%, and 0% for items with a teacher assessment of 1, 2, 3, and 4, respectively. All three items with a teacher rating of 1 had a self-assessment that was two points higher; with a faculty rating of 2, 6.2% of self-assessments were two points higher and 73.8% were one point higher. When faculty gave students a rating of 3, 69.9% of the self-assessments agreed with the faculty rating, 16.8% of the self-assessments were higher, and 13.3% of the self-assessments were lower. When faculty gave the highest rating (4 points), 59.8% of the self-assessments were lower.

Gender differences in overestimation and underestimation

We also made a simple comparison of the differences between male and female students in overestimation and underestimation (Figure 4). Approxi-

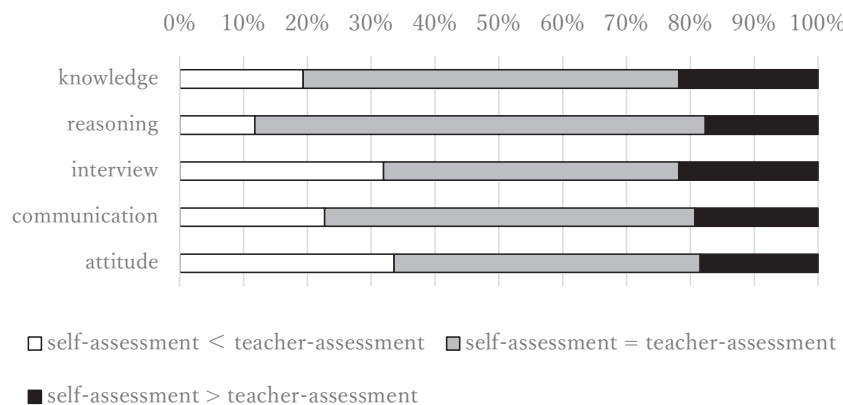


Fig. 2. Student self-assessment and teacher-assessment distributions.

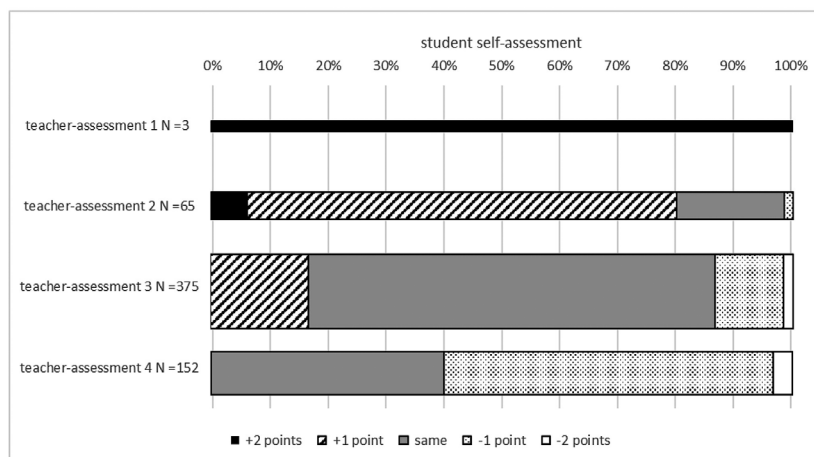


Fig. 3. Disparities in student self-assessments and teachers' assessments.

mately 50% of both male (45 of 88) and female (16 of 31) students had overestimated items (Figure 4-A). Conversely, 71.0% of the female students (22 of 31) and 64.8% of the male students (57 of 88) had underestimated items (Figure 4-A). The female students tended to underestimate themselves, however, there was no statistical gender difference in the number of students who underestimated themselves.

The average number of overestimated items per student was 1.1 for male students and 0.8 for female students (Figure 4-B). The average number of underestimated items per student was 1.1 and 1.4 for male students and female students, respectively (Figure 4-B). There were no statistical gender differences in the average number of over- and underestimated items per student.

Finally, we looked at the number of overestimated and underestimated items per student, by gender. As shown in Figure 5-A, approximately 50% of both male (43 of 88) and female (15 of 31) students had no overestimated items (Figure 5-A). For male students, 18.2% (16 of 88), 19.3% (17 of 88), and 13.6% (12 of 88) overestimated themselves on one, two, and three or more of the five items, respectively. For female students, 35.5% (11 of 31), 6.5% (2 of 31), and 9.7% (3 of 31) overestimated on one, two, and three or more items, respectively. Figure 5-B shows a comparison of the number of underestimates by gender. 35.2% of the male students (31 of 88) and 29.0% of the female students (9 of 31) had no underestimated items. For male students, 36.4% (32 of 88), 12.5% (11 of 88), and 5.9% (14 of 88) had one, two, and

three or more underestimates, respectively. For female students, 29.0% (9 of 31), 22.6% (7 of 31), and 19.4% (6 of 31) had one, two, and three underestimates, respectively. One male and one female student underestimated themselves on all items. Only 10 students (8 male ; 2 female) had neither over- nor underestimations. The chi-squared test showed no significant difference in the number of overestimates and underestimates by gender.

Discussion

In our department, self-assessment and faculty assessment were done on the same day, right after each clerkship, using the same rubric assessment tool. Therefore, we had anticipated that there would be minimal discrepancy between students' self-assessment and faculty assessment. However, there were obvious discrepancies between some students' self-assessments and their corresponding faculty assessments. Identifying one's own weaknesses through ongoing self-assessment is critical for learners, particularly medical students^{1,11}. The fact that some students over-assessed themselves in comparison to teachers demonstrated that they had yet to master the ability of accurate self-assessment. Students/physicians who overestimate their own performance may place patients at greater risk^{1,12}. In this case, one antidote to over-assessment is for instructors to provide high-quality feedback¹. Students who have a tendency to under estimate themselves, on the other hand, cannot fully utilize their original ability to qualify for any task or mission. Underestimation

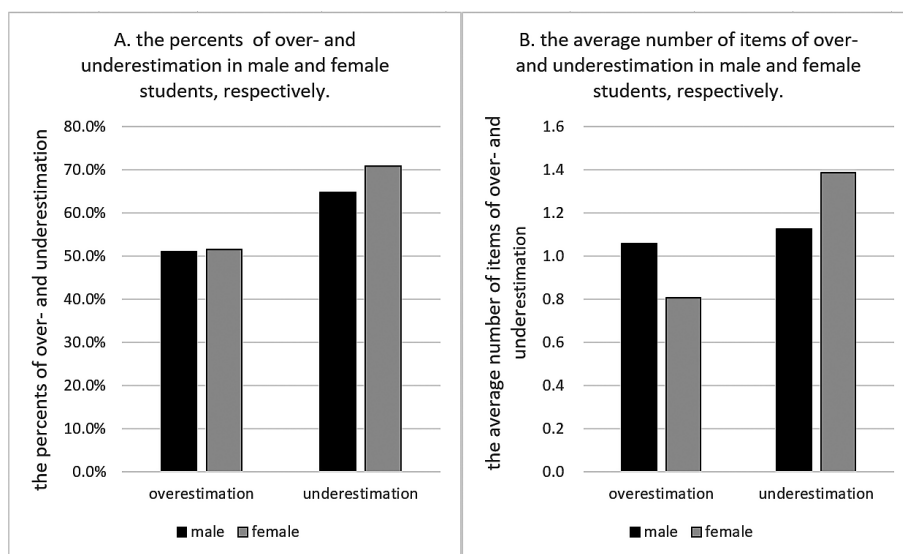


Fig. 4. Differences by gender and assessment items on over- and underestimation.

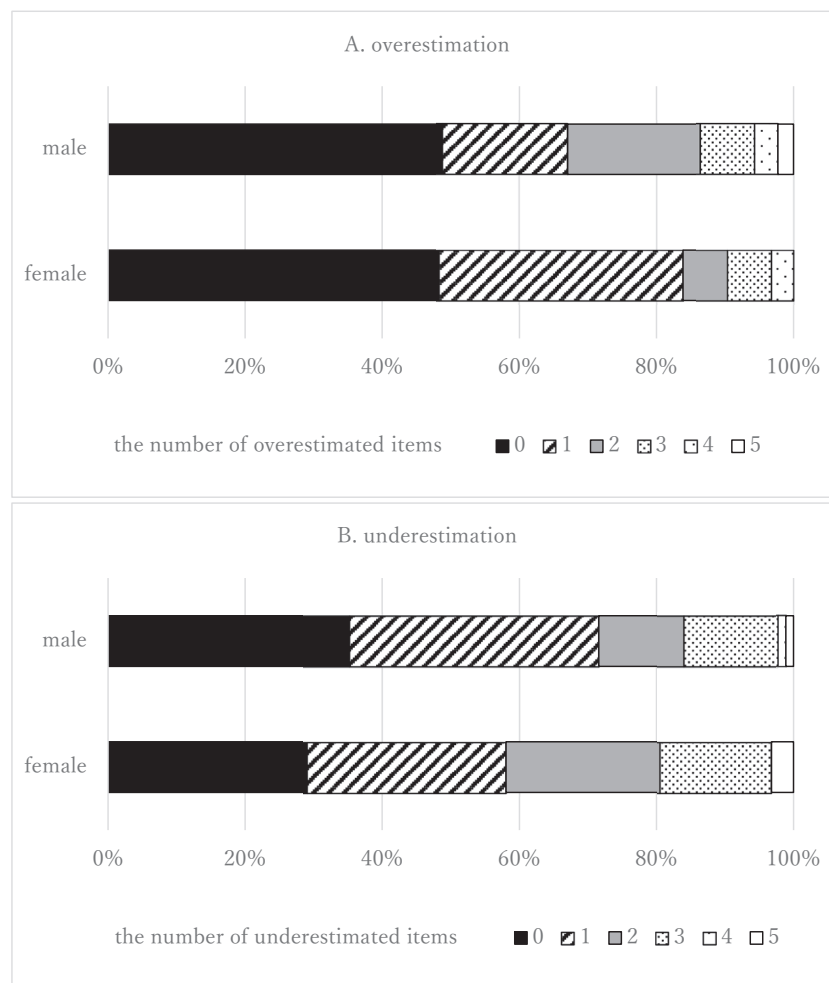


Fig. 5. Disagreements in teacher ratings per student.

of one's own abilities and strengths will not provide a medical student with the confidence to proceed with a suitable plan of action without unnecessary hesitation or trepidation¹⁾. Improving self-efficacy and experiencing success are likely to improve self-assessment¹³⁾, so medical education that encourages them to do so is critical.

Several limitations must be considered when making generalizations from this study. First, the primary learning skill for medical students at CMECD was the medical interview only, with no clinical practice. Five assessment items could not possibly represent all acceptable medical student performance¹⁵⁾. Second, at CMECD, both the faculty and students assessed performance immediately after a half-day training session, whereas in many other departments, performance over several weeks was assessed at the end. Therefore, the results of this study may not be generalizable across departments. Third, as previously reported^{1, 14)}, self-assessment is an unstable skill that varies depending

on the content, context, and perspective. As a result of situational influences, it is difficult to answer whether students are qualified self-assessors or not¹⁵⁾.

Using the evaluation developed by us, we investigated the consistency of student self-assessment and faculty assessment during clinical clerkships. This paper assumes that faculty members' qualitative evaluations are accurate and reproducible, but the reliability of qualitative evaluations needs further investigation. For example, as reported in some studies¹⁶⁻¹⁹⁾, differences in teachers' genders and ages had an effect on faculty assessed students' performance during clinical clerkships. In this study, we found no gender differences in the discrepancy between teacher ratings and students' self-ratings. Our future study will focus on discovering the causes of over- and underestimation, which will require expansion of our study participants from the current single grade (2018) to multiple grades²⁰⁾, i.e., to investigate grade differ-

ences, and from the current single clinical clerkship only involving medical interviews to other clerkships involving more clinical practices. Furthermore, in the near future, the relationships between student-assessment/teacher-assessment and a series of examinations, such as PCC-OSCE (Post Clinical Clerkship Objective Structured Clinical Examination), advancement examinations, graduation examinations, and national examinations for medical practitioners, will be investigated.

Ethical acceptance

The Fukushima Medical University Institutional Review Board determined that this retrospective study of students' self-assessment and teacher assessment of BSL (Bedside Learning) performance was exempt from review.

References

1. Eva KW, Regehr G. Self-assessment in the health professions: a reformulation and research agenda[J]. *Academic medicine*, **80**(10): S46-S54, 2005.
2. Colthart I, Bagnall G, Evans A, *et al.* The effectiveness of self-assessment on the identification of learner needs, learner activity, and impact on clinical practice: BEME Guide no. 10[J]. *Medical teacher*, **30**(2): 124-145, 2008.
3. Arnold L, Willoughby TL, Calkins E. Self-evaluation in undergraduate medical education: a longitudinal perspective[J]. *Journal of medical education*, 1985.
4. Gordon MJ. A review of the validity and accuracy of self-assessments in health professions training[J]. *Academic medicine*, **66**(12): 762-769, 1991.
5. Boud D. Avoiding the traps: Seeking good practice in the use of self assessment and reflection in professional courses[J]. *Social work education*, **18**(2): 121-132, 1999.
6. Ward M, Gruppen L, Regehr G. Measuring self-assessment: current state of the art[J]. *Advances in health sciences education*, **7**(1): 63-80, 2002.
7. Boud D. *Enhancing learning through self-assessment*[M]. Routledge, 2013.
8. Tang Z, Sakamoto N, Shikama Y, Otani K. A New Evaluation System on Degree of Goal Achievement for Students in Bedside Learning: Usage Status and Effect of Introduction from the Viewpoint of Faculty [In Japanese]. *Fukushima Journal of Medical Science*, **72**(1): 11-21, 2022.
9. Gärtner J, Prediger S, Harendza S. Development and pilot test of ComCare-a questionnaire for quick assessment of communicative and social competences in medical students after interviews with simulated patients[J]. *GMS journal for medical education*, **38**(3), 2021.
10. Gordon CJ, Jorm C, Shulruf B, *et al.* Development of a self-assessment teamwork tool for use by medical and nursing students[J]. *BMC Medical Education*, **16**(1): 1-7, 2016.
11. Ghosh P, Jacob J, Goldman E, *et al.* Optimizing the use of an online self-assessment exam to promote self-directed learning behaviors in medical students[J]. *Medical Science Educator*, **30**(1): 81-85, 2020.
12. Fitzgerald JT, Gruppen LD, White CB. The influence of task formats on the accuracy of medical students' self-assessments[J]. *Academic Medicine*, **75**(7): 737-741, 2000.
13. Cervone D. Thinking about self-efficacy[J]. *Behavior modification*, **24**(1): 30-56, 2000.
14. Speyer R, Pilz W, Van Der Kruijs J, *et al.* Reliability and validity of student peer assessment in medical education: a systematic review[J]. *Medical teacher*, **33**(11): e572-e585, 2011.
15. Abadel FT, Hattab AS. How does the medical graduates' self-assessment of their clinical competency differ from experts' assessment?[J]. *BMC medical education*, **13**(1): 1-9, 2013.
16. Riese A, Rappaport L, Alverson B, *et al.* Clinical performance evaluations of third-year medical students and association with student and evaluator gender[J]. *Academic Medicine*, **92**(6): 835-840, 2017.
17. Blanch-Hartigan D. Medical students' self-assessment of performance: results from three meta-analyses[J]. *Patient education and counseling*, **84**(1): 3-9, 2011.
18. Huang W, Grigoryan L. Student self-assessment versus preceptor assessment at the midpoint of a family medicine clerkship[J]. *Family Medicine and Community Health*, **5**(1): 65-70, 2017.
19. Wang-Cheng RM, Fulkerson PK, Barnas GP, *et al.* Effect of student and preceptor gender on clinical grades in an ambulatory care clerkship[J]. *Academic Medicine*, 1995.
20. Fitzgerald JT, White CB, Gruppen LD. A longitudinal study of self-assessment accuracy[J]. *Medical education*, **37**(7): 645-649, 2003.

Appendix 1. Pathophysiology understanding (knowledge)

	Pathophysiology comprehension (knowledge)
4 excellent as a student	A) Adequate knowledge of pathophysiology and the ability to apply it clinically (diagnosis/judgment) are required B) Capable of formulating a prescription or treatment plan based on pathophysiology
3 desirable level	A) Adequate knowledge of pathophysiology and the ability to apply it clinically (diagnosis/judgment) are required
2 minimally acceptable level	A) Pathophysiology knowledge is limited, but the ability to apply it clinically (diagnosis/judgment) is strong
1 below acceptable level	A) Inadequate pathophysiology knowledge and B) Incapability to use it clinically

Appendix 2. Clinical reasoning

	Clinical reasoning
4 excellent as a student	A) Capable of extracting diagnostic and therapeutic issues B) Capable of describing the path of problem solving based on case history and physical examination C) Capable of stating the necessary inspection plan in order of priority
3 desirable level	A) Capable of extracting diagnostic and therapeutic issues B) Capable of describing the path of problem solving based on case history and physical examination
2 minimally acceptable level	A) Capable of extracting diagnostic and therapeutic issues B) Insufficient to conduct history and physical examination for problem solving
1 below acceptable level	A) Inability or inadequacy to extract diagnostic/therapeutic issues

Appendix 3. Interviewing skills

	Interviewing skills
4 excellent as a student	A) Capable of using proper interview techniques B) Capable of taking a comprehensive case history C) Capable of understanding the problems and asking focused questions D) Capable of explaining symptoms to patients and their families in an understandable manner
3 desirable level	A) Capable of using proper interview techniques B) Capable of taking a comprehensive case history C) Capable of understanding problems and asking focused questions
2 minimally acceptable level	A) Capable of using proper interview techniques B) Capable of taking a comprehensive case history
1 below acceptable level	A) Inadequate interview technique or B) Inadequate to take a comprehensive case history

Appendix 4. Communication

	Communication
4 excellent as a student	A) Maintain a proper listening attitude toward the patient and family B) Capable of communicating with medical personnel C) Capable of demonstrating empathy and respect for patients, families, and medical personnel D) Capable of collaborating with patients, families, and medical personnel for the purpose of sharing information
3 desirable level	A) Maintain a proper attitude toward the patient and family B) Capable of communicating with medical personnel C) Capable of demonstrating empathy and respect for patients, families, and medical personnel
2 minimally acceptable level	A) Maintain a proper listening attitude toward the patient and family B) Capable of communicating with medical personnel
1 below acceptable level	A) Disregard for patient's or family's needs and wishes or B). There are times when information that should be shared with medical personnel cannot be shared

Appendix 5. Learning attitude

	Learning attitude
4 excellent as a student	A) None of a-e in the list below B) Attitude of humility and sincerity, but eagerness to learn from start to finish
3 desirable level	A) None of a-e in the list below
2 minimally acceptable level	A) Apology, remorse, and progress in medical safety for tardiness or inappropriate behavior
1 below acceptable level	A) Do you have any of the following, a. without prior notice, absence or tardiness b. improper grooming and appearance c. dishonesty/abdication of responsibility d. repeated inappropriate behavior in the field of medical safety e. dissatisfaction with or resistance to advice or guidance