How Reliable are MS-3 Clinical Clerkship Assessment Scores?


Reviewed by Madeleine W. Schrier, MD

What was the study question?
How many individual evaluations are necessary to achieve an acceptable threshold of reliability in differentiating between students’ competency levels within and across third year clerkships?

How was the study done?
Clinical competency scores for third-year medical students at the University of Michigan were examined using a generalizability analysis to determine how reliably the scores differentiate between students’ competency levels. All students who completed third-year clerkships during the 2015-2016 academic year and who had scores from at least three separate evaluators from each clerkship were included. The final data subset included 93 students (50% of the original group), and six of seven core clerkships. These data were then used to conduct a decision study to examine the effect that different numbers of assessors would have on the generalizability of the scores.

What were the results?
Within clerkships, the number of assessments required for optimal reliability ranged from 4-17 assessments per student. Interestingly, only 7% of the variance in scores was attributable to differences between students. The student-clerkship interaction (variations in students’ scores between clerkships) accounted for 10.1% of the variance. A small proportion (7.9%) was attributable to differences between the clerkships themselves. Notably, the majority of the variance (75%) was attributable to “residual error,” including variation between assessors as well as other unknown or unmeasurable variables.

What are the implications of these findings?
Clinical grades for these students had a large degree of variation both within and across clerkships, which could not be attributed solely to differences between students or clerkships. A limitation of this study design is that the degree of variation between assessors could not be specifically measured. The authors of this study have used their data to inform curriculum reform, and plan to implement changes at their university to make the grading process more reliable. This research should prompt other medical schools to closely consider the way their students are evaluated, as third year clerkship grades have a substantial impact on students’ residency placements and ultimate career paths.

Editor’s note: I think that most COMSEP members would agree that evaluating clinical competence of medical students on clerkships and assigning grades derived from those evaluations/assessments of competency is one of the more challenging responsibilities for clerkship directors. Efforts to assure the fairness and accuracy of clinical evaluations are of huge importance because clerkship grades are a major factor for selection of residents. I for one will be eager to see how the University of Michigan uses this data to make the grading process more reliable. (RR)

Reviewed by Eric Zwemer

**What was the study question?**
How do program directors interpret letters of recommendation (LORs) when reviewing applicants for residency or fellowship positions?

**How was the study done?**
Pediatric residency and fellowship Program Directors (PDs) on the APPD listserv were asked to rate 3 aspects of LORs: letter features (e.g. length of letter, academic rank of letter writer), applicant attributes (e.g. work ethic, leadership), and common phrases (“I give my highest recommendation”). Ratings were done on 5-point Likert scales assessing perceived importance or negative/positive association.

**What were the results?**
The survey had a 43% response rate (468/1079 PDs). Almost all (85%) respondents stated LORs were important for their overall impression of a candidate and that their perceptions of otherwise weaker or stronger candidates could be changed by an LOR. The highest rated letter features were depth of interaction with applicant, specific traits of applicant, and applicant’s abilities, whereas the lowest ranked features were long letters, applicant advanced degrees, and community service activities. Academic rank of letter writer was rated as important by only 33% of residency PDs and 45% of fellowship PDs. The highest rated applicant abilities were work ethic, trustworthy, team player, professional, and compassionate.

With regard to specific LOR “codes”, the phrases ranked most positively by PDs were “would like applicant to stay at our institution”, “will be an asset to any program”, “exceeded expectations”, and “I give my highest recommendation”. Notably, “I highly recommend” and “I recommend without reservation” were listed as primarily neutral phrases, while “showed improvement”, “overcame personal setbacks”, and “performed at expected level” have negative associations.

**Why is this important?**
Faculty development on writing LORs is essential for the appropriate and deliberate communication regarding applicants. In particular, faculty should be aware of how certain phrases are viewed by PDs to avoid unintended neutral or negative associations that may occur.

*Editor’s note: It has always struck me as quite peculiar that evaluating and ranking applicants often depends on decoding statements and other information that might mean one thing or might mean another, thus necessitating studies like this one. The situation is especially puzzling with letters of recommendation knowing that applicants have almost always waived their right to review the letters. Having said that, some guidance is better than no guidance and this study provides some help for writing and interpreting letters of recommendation. (RR)*
Mistreatment—we don’t necessarily know it when we see it...


Reviewed by: Janet Meller

What was the study question?
What are students’ interpretations of previously reported mistreatment incidents?

How was the study done?
Twenty-one scenarios of previously discussed instances of mistreatment were presented at a mandatory seminar for third-year medical students. Using an anonymous audience response, 127 students were asked “how likely would you label this as mistreatment?” by assigning a Likert score from 1 - strongly disagree to 5 - strongly agree for each scenario. Each scenario was discussed before moving on to the subsequent one. Students were also asked to anonymously provide demographic data and indicate whether they had experienced mistreatment. Results were tabulated using agree or strongly agree to label the scenario as mistreatment or not and the average Likert score of each was tabulated. Each student’s 21–scenario average was also calculated which was linked to gender and whether the student identified as a minority, as well as whether they had previously experienced mistreatment.

What were the results?
There was not complete agreement on any one of the scenarios as to whether there was mistreatment. 49% of students reported previously being a victim of mistreatment. In 7 scenarios the majority of students agreed that there was mistreatment. These scenarios fell into three identifiable categories: (1) faculty/resident abuse of power; (2) name-calling; and (3) inappropriate comments about gender/race. Nevertheless, although 10% of students identified more than ½ of scenarios as mistreatment, some students rarely perceived mistreatment. Finally students who were minorities or reported previous personal mistreatment were more likely to identify a scenario as mistreatment.

Why is this important?
Professionalism is a core competency for medical practitioners at all levels. It is important that all members of the learning environment recognize what constitutes mistreatment and that these behaviors cannot be tolerated. It is also important to understand and clarify the more subtle acts of mistreatment to educate and empower students and faculty to better deal with them.

Editor’s comments:
This study once again demonstrates the incredible variability in students’ perceptions of behaviours that may be labeled as mistreatment. Regardless of these different perspectives, we must continue to develop a shared understanding of mistreatment and challenge ourselves to change the culture the learning environment to one where mistreatment is not tolerated (KFo).