



Re: Reliability and benefits of medical student peers in rating complex clinical skills: Response to common mistake

Pamela M. Basehore & Sherry C. Pomerantz

To cite this article: Pamela M. Basehore & Sherry C. Pomerantz (2014) Re: Reliability and benefits of medical student peers in rating complex clinical skills: Response to common mistake, *Medical Teacher*, 36:11, 1008-1008, DOI: [10.3109/0142159X.2014.948836](https://doi.org/10.3109/0142159X.2014.948836)

To link to this article: <https://doi.org/10.3109/0142159X.2014.948836>



Published online: 26 Aug 2014.



Submit your article to this journal [↗](#)



Article views: 400



View related articles [↗](#)



View Crossmark data [↗](#)

with reliability and actually is one of the common mistakes in reliability analysis (Rothman et al. 2008). Reliability (repeatability or reproducibility) is often assessed by different statistical tests such as Pearson r , least square and paired t . 'Mistakes in reliability analysis are common' (Lawrence & Kuei 1989; Rothman et al. 2008).

For quantitative variables the Intra Class Correlation Coefficient (ICC) should be used. For qualitative variables the weighted kappa, which should be used with caution because kappa has its own limitation too (Lawrence & Kuei 1989; Rothman et al. 2008). It is crucial to know that there is no value of kappa that can be regarded universally as an indication good agreement. An important weakness of k value to assess agreement of a qualitative variable is that it depends upon the prevalence in each category. This means that it is possible to have a different kappa value based on the same percentage of both concordant and discordant cells.

The authors point out in their conclusion, "peer raters" of the same level of training can provide accurate ratings of complex clinical tasks and can serve as an important resource in assessing student performance in an OSCE, but have not investigated the concordance of the pass/fail decisions with respect to individual candidates

Reliability (precision) and validity (accuracy) are two completely different and important methodological issues in all fields of researches. To assess the accuracy (validity) the following tests are used:-

sensitivity (the percentage with the disease who test positive, True Positives / (True Positives + False Negative)),

specificity (the percentage of healthy who test negative, True Negatives / (True Negatives + False Positive))

positive predictive value (PPV), (percentage of positive tests who actually are diseased, True Positives / (True Positives + False Positive)),

negative predictive value (NPV) (the percentage of negative tests who are healthy, True Negatives / (True Negatives + False Negative)),

likelihood ratio positive and likelihood ratio negative as well as diagnostic accuracy [(both true positive and true negative results / total) $\times$ 100]

odds ratio (true results / false results) preferably more than 50.

These are the tests to evaluate the validity (accuracy) of a test compared to a gold standard (Rothman et al. 2008).

Therefore, the authors' conclusion is due to the confusion of reliability (precision) with validity (accuracy) and is, therefore, misleading.

Siamak Sabour, MD, MSc, DSc, PhD, Postdoc, Department of Clinical Epidemiology, Faculty of Health, Shahid Beheshti University of Medical Sciences, Tehran, I.R. Iran. Tel: +98 21 22421814. E-mail: s.sabour@sbmu.ac.ir.

Declaration of interest: The author reports no conflicts of interest.

References

Basehore PM, Pomerantz SC, Gentile M. 2014. Reliability and benefits of medical student peers in rating complex clinical skills. *Med Teach* 2014;36(5):409–414.

Lawrence I, Kuei L. 1989. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 45:255–268.

Rothman JK, Greenland S, Timothy LL. 2008. *Modern epidemiology*, 3rd ed. Baltimore, USA: Lippincott Williams & Wilkins.

Re: Reliability and benefits of medical student peers in rating complex clinical skills: Response to common mistake

Dear Sir

We want to take this opportunity to respond to the concerns raised about the reliability analysis conducted in the study. Dr Sabour has pointed out the appropriate use of intraclass correlation coefficient (ICC) as a preferred analysis to assess reliability in quantitative variables and has criticized our use of Pearson correlation coefficient.

As Dr Sabour is likely aware, the G-coefficient in generalizability analysis and ICC are both based in classical test theory and are closely related. While ICC analysis examines a single facet, generalizability analysis provides the opportunity to look at multiple facets of measurement error in a single design (Shrout & Fleiss 1979; Barch & Mathalon 2011). In our analysis, while the correlation coefficients were used to establish the relationship between peer and faculty ratings, the generalizability analysis provided the reliability measure.

We appreciate his interest in our research and the opportunity to clarify the analysis conducted.

Pamela M. Basehore and Sherry C. Pomerantz, Rowan University School of Osteopathic Medicine, Academic Affairs, 1 Medical Center Drive, Stratford, New Jersey 08084, USA. E-mail: basehore@rowan.edu

Declaration of interest: The authors report no conflicts of interest.

References

Barch DM, Mathalon DH. 2011. Using brain imaging measures in studies of precognitive pharmacological agents in schizophrenia: Psychometric and quality assurance considerations (Supplement). *Biol Psychiatry* 70(1):13–18.

Basehore PM, Pomerantz SC, Gentile M. 2014. Reliability and benefits of medical student peers in rating complex clinical skills. *Med Teach* 36(5):409–414.

Shrout PE, Fleiss JL. 1979. Intraclass correlations: Uses in assessing rater reliability. *Psych Bull* 86(2):420–428.

Better data >> Bigger data

Dear Sir

We read Ellaway et al.'s (2014) article on Big Data in health professions education with great interest. We share the authors'