Relationship between clinical assessment and examination scores in determining clerkship grade
Stephen J Lurie & Christopher J Mooney
Medical Education 2010 44: 122–124

Introduction: Grades in clinical clerkships are typically based on a combination of clinical assessments from teachers, as well as results of more reliable (but perhaps less valid) scores on standardized tests of knowledge. It remains unclear how these different sources of information about students’ performance should be optimally combined in the high-stakes assignment of final grades for clinical performance. In this paper, three related questions are asked: 1) how variable is the relationship between subject examination scores and clinical assessment in a sample of students as they rotate through five clerkships?, 2) how much of the variance in clerkship grades is a reflection of subject examination scores versus measures of clinical competence, and how do these associations change as this group of students rotates through different clerkships? 3) Given these results, what are the consequences of assigning different weights to these pieces of information in terms of how they combine to determine final grades?

Methods: Participants were 83 students at the University of Rochester School of Medicine and Dentistry (URSMD) who graduated in 2008. Data from the study were obtained from five of the participants’ seven required clerkships (neurology, obstetrics and gynecology, pediatrics, psychiatric and surgery) over the past two years. Each participating clerkship provided data that contained students’ clinical evaluation scores (s), subject examination score, and final letter grade (pass, high pass or honors). After computing univariate correlations between clinical assessment scores and standardized examination scores, logistic regression analyses for each clerkship were performed to predict the final grade from these two variables. The authors then compared actual grade with predicted grade under various hypothetical policies for combining these two variables. Finally, they assessed whether some students would systematically benefit from these policies.

Results: Clerkships varied in their univariate correlations between scores on clinical assessments and scores on standardized examinations. Clerkships with the lowest correlations tended to give more weight to standardized examination scores. Grading committees adjusted a substantial minority of grades to account for factors that were not reflected in either score, indicating that they may have knowledge about individual students’ situations and contexts and factored into their decisions. There did not appear to be a systematic bias in grading committee effect across the five clerkships.

Conclusions: The study assessed patterns of grading among the same group of students as they rotated through several clerkships in the course of a single year. Students’ numerical scores failed to capture all the data that grading committees considered important in assigning final grades. However, the results give credence to the idea that the ability to perform well on standardized tests is not necessarily independent of other clinical skills – and provide an added challenge to develop and validate instruments that uniquely assess aspects of clinical performance that are not easily captured by standardized testing.

Implications for TUSM: At TUSM clinician educators are always striving to accurately assess our students’ performance during their clinical rotations. Thus, discussing grading policies and approaches, as well as unmeasured contextual information, has always been a topic of discussion at clerkship meetings. This article suggests that we should keep this discussion going and consider examining our grading patterns in an empirical fashion to refine our evaluation approaches by distilling the aspects of clinical performance that are not easily captured by standardized tests, and share best grading practices.

For questions/suggestions please contact Maria Blanco at: maria.blanco@tufts.edu