The Association of Student Examination Performance with Faculty and Resident Ratings Using a Modified RIME Process

Charles H. Griffith III, MD, MSPH^{1,2} and John F. Wilson, PhD¹

¹University of Kentucky, Lexington, KY, USA; ²K507 Kentucky Clinic, Lexington, KY, USA.

BACKGROUND: RIME is a descriptive framework in which students and their teachers can gauge progress throughout a clerkship from R (reporter) to I (interpreter) to M (manager) to E (educator). RIME, as described in the literature, is complemented by residents and attending physicians meeting with a clerkship director to discuss individual student progress, with group discussion resulting in assignment of a RIME stage.

OBJECTIVE: 1) to determine whether a student's RIME rating is associated with end-of-clerkship examination performance; and 2) to determine whose independent RIME rating is most predictive of a student's examination performance: attendings, residents, or interns.

DESIGN: Prospective cohort study.

PARTICIPANTS: Third year medical students from academic years 2004–2005 and early 2005–2006 at 1 medical school.

MEASUREMENTS AND MAIN RESULTS: Each attending, resident, and intern independently assessed the student's final RIME stage attained. For the purpose of analysis, R stage=1, I=2, M=3, and E=4. Regression analyses were performed with examination scores as dependent variables (National Board of Medical Examiners [NBME] medicine subject examination and a clinical performance examination [CPE]), with independent variables of mean attending RIME score, mean resident score, and mean intern score. For the 122 students, significant predictors of NBME subject exam score were resident RIME rating ($p=.008$) and intern RIME rating ($p=.02$). Significant predictor of CPE performance was resident RIME rating ($p=.01$).

CONCLUSION: House staff RIME ratings of students are associated with student performance on written and clinical skills examinations.

KEY WORDS: medical education; clinical evaluation; medical students.

J Gen Intern Med 23(7):1020–3

DOI: 10.1007/s11606-008-0611-3

© Society of General Internal Medicine 2008

BACKGROUND

An important goal for clinical clerkships is to provide evaluation of students that is reliable, valid, and delivers meaningful feedback.¹ The most well-studied evaluation system used in Internal Medicine clerkships is the RIME system, designed by Lou Pangaro and colleagues at the Uniformed Services University of Health Sciences (USUHS)². RIME is a synthetic descriptive framework in which students and their teachers can gauge progress throughout a clerkship from R (reporter) to I (interpreter) to M (manager) to E (educator). RIME, as described in the literature, is typically coupled with formal evaluation sessions, in which residents and faculty meet as a group with a clerkship director, during which house staff and faculty in turn discuss individual student progress, and each teacher assigns a RIME designation for each student informed by that group discussion.³ For example, at USUHS formal evaluation sessions occur every 3–4 weeks in a 12-week clerkship, with a clerkship director spending 15 minutes with each student after the evaluation session to discuss individual strengths and areas to improve. Therefore students receive formal feedback at least 3 times during the clerkship, based on the RIME taxonomy, and the evaluation sessions provide a forum for frame of reference training to ensure that house staff and faculty apply the RIME criteria appropriately when making their evaluations.^{4,5} In addition to demonstrating a high degree of reliability,² the RIME system has been shown to be able to identify students with marginal funds of knowledge,⁶ those with unprofessional behavior,⁷ and those who are more likely to receive low ratings from residency directors during internship.⁸

In a 2005 survey conducted by the Clerkship Directors of Internal Medicine, 42% of Internal Medicine clerkship directors reported using at least a form of the RIME system in their courses,⁹ and RIME has also been adopted by some Obstetrics/Gynecology clerkships.¹⁰ Nevertheless, the majority of clerkships have not been able to implement the more comprehensive RIME system (such as the formal meetings with all evaluators and the individual meetings with students), citing logistic, time and attendance constraints.⁹ One purpose of this project was to assess the feasibility and validity of using the RIME system for student evaluation outside of formal, multiple evaluation sessions and group meetings, instead having attending physicians, residents, and interns make their own individual and independent assessment of the student's progress, using the RIME taxonomy.

The second aim of our project was to investigate the association of RIME ratings and student examination perfor-

Portions of this paper have been presented at the National Meeting of the Clerkship Directors of Internal Medicine, October 27, 2006, in New Orleans, LA; and at the southern regional SGIM meeting, February 10, 2007, New Orleans, LA.

mance. One should note that RIME is synthetic, incorporating the analytic model dimensions of knowledge, skills, and attitudes into each step, and therefore each stage of RIME is more than the sum of its parts, with knowledge but 1 facet. For example, a student with a strong fund of knowledge but unable to apply that knowledge into effective patient care (such as being unreliable in reporting data, or being rude or unprofessional with patients) would not be considered to attain even Reporter stage. Nevertheless, an underlying assumption is that an improved knowledge base is necessary (although not sufficient) for making the transition beyond Reporter, as suggested by the cognitivist learning theory.¹¹ An indicator of the validity of this assumption would be to demonstrate predictive ability, that is, if higher RIME stages were associated with measures of student knowledge, such as examination scores. One study has noted that "total teacher points" a student accumulated during a clerkship (as informed by the RIME scale) explained a small (1%) percent of the variance in National Board of Medical Examiners (NBME) Subject examination scores¹²; however, no study has demonstrated that higher ratings of students on the RIME scale given by individual instructors at the end of a clerkship are associated with better final examination performance. Demonstrating such an association would also suggest validity of our scaled-down RIME process, with individual ratings provided without the calibration provided by formal group evaluation sessions. Finally, a third purpose of our project was to assess whose independent RIME ratings are most predictive of a student's examination performance: attending physicians, residents, or interns?

METHODS

Our Institutional Review Board judged this project exempt from written informed consent requirements. Subjects were all third-year medical students at the University of Kentucky rotating on our 8-week inpatient Internal Medicine clerkship during academic years 2004–2005 and the first 1/3 of 2005–2006. The clerkship consists of two 4-week rotations: 1 at the University hospital and 1 at our affiliated Veteran's Affairs Medical Center. All students are on general medicine teams. During a given month, a team would consist of 2 attending physicians (each supervising for a 2-week period), a resident, 2 first year residents ("interns"), and 2 third year medical students. Therefore, students work with 4 attending physicians, 2 residents, and 4 interns during the clerkship. Each attending, resident and intern independently evaluated students at the end of the 4-week block, on a written form returned via mail to the clerkship administrative assistant. As part of the evaluation, each attending, resident and intern was asked to circle on the grade sheet the RIME descriptor that best matched his or her assessment of the student, in terms of ability achieved by the end of the rotation. The descriptors were those published in the RIME literature: a Reporter is one who is consistently good in interpersonal skills, and reliably obtains and communicates clinical findings; an Interpreter is one who is able to prioritize and analyze patient problems; a Manager is one who consistently proposes reasonable options incorporating patient preferences; an Educator is one who has consistent level of knowledge of

current medical evidence and can critically apply this knowledge to specific patients.²

Our outcomes of interest were scores on 2 end-of-clerkship examinations: the NBME Subject Examination in Medicine, and the Clinical Performance Examination (CPE), a 12-station examination of basic clinical skills, including 8 stations with standardized patients, 2 on chest x-ray interpretation, and 1 each on electrocardiogram (ECG) and arterial blood gas interpretation. The reliability of the NBME Subject examination is between 0.75 and 0.85.¹³ The reliability of the CPE over the past 12 years is 0.70 to 0.80. For purposes of analysis, each RIME score was converted to a numeric equivalent, with R=1, I=2, M=3, and E=4, similar to some studies in the RIME literature.¹⁴ Independent variables were students' mean attending RIME score, mean resident RIME score, and mean intern RIME score. Pearson's correlations were used to assess for bivariate associations between the dependent and independent variables, as well as with rotation number, to account for any differences in findings depending on the time of the academic year. Multiple regression analyses were performed with dependent variables of the final examination scores, with independent variables of mean RIME ratings by attending physician, resident, and intern.

RESULTS

Data were received regarding all 122 third year students during the study period. Mean NBME Subject Examination Score was 75.3 (SD 7.0) and Mean CPE Score was 88.3 (SD 3.6). Evaluations were received from 456 attending physicians (93% response rate), 238 residents (98% response rate), and 464 interns (95% response rate). The mean attending RIME rating was 2.86 (SD 0.8), the mean resident RIME rating was 2.87 (SD 0.9), and the mean intern RIME rating was 3.11 (SD 0.8). Attending physicians rated 2% of students at the Reporter stage by the end of the rotation, 35% Interpreter, 36% Manager, and 27% Educator. Residents rated 5% of students Reporter, 24% Interpreter, 48% Manager, and 23% Educator. Interns rated 1% of students Reporter, 20% Interpreter, 45% Manager, and 34% Educator. There were no statistically significant correlations between mean RIME rating by attending physician, resident, and intern, and no correlations between these ratings and time in the academic year. The Figure 1 depicts NBME Examination Score for the entire study sample as a function of mean RIME rating by attending physician, resident, and intern.

In regression analysis, independent predictors of higher student NBME Subject Examination performance were mean RIME resident rating ($p=.008$) and mean RIME intern rating ($p=.02$); attending RIME rating did not achieve statistical significance ($p=.17$). This model explained 15% of the variance in NBME Subject examination scores, 8.5% explained by resident RIME rating and 4.5% by the intern rating (the 2% variance explained by attending rating was not statistically significant). The significant independent predictor of CPE performance was mean RIME resident rating ($p=.01$); mean RIME intern rating and mean attending RIME rating approached but did not achieve statistical significance ($p=.09$ and $.07$, respectively). This model explained 13% of the variance in CPE score, 7% explained by the resident RIME rating (the 2.5% explained by interns

and 3.5% explained by attending physicians not achieving statistical significance).

DISCUSSION

Our findings represent the first demonstration that independent evaluations of students using the RIME taxonomy are associated with end-of-clerkship examination scores, examinations of both medical knowledge (the NBME Subject examination) and clinical skills (the CPE), adding even further validity for the use of RIME. In addition, this predictive ability was demonstrated even without formal evaluation and group meetings, relying instead on an individual evaluator's independent judgment, a much less logistically challenging process, which would perhaps be more readily exportable to other institutions and other clerkships. In addition, the ratings most predictive of a student's final examination performance were those of the residents. Students often spend more time working directly with the residents than with the attending physicians, and resident teaching is an important contributor to enhanced student learning.¹⁵ Our findings lend credence for resident participation in the evaluation of students, not relying on faculty evaluations alone, perhaps even suggesting that one should weigh resident evaluations more in the student's grade.

Our study findings are limited to a single institution, in an inpatient Internal Medicine clerkship. Future studies should consider extending our findings to other sites and in other types of clerkships, including the ambulatory setting. In addition, our methodology entailed transforming the RIME descriptors to a numerical scale, similar to other studies of this instrument¹⁴; however, this makes the assumption that the "distances" between successive descriptors are equal, an assumption which could be subject to challenge. Finally, our evaluators had no formal training in applying the RIME system to their evaluations of their students. More rigorous training and orientation to RIME for the evaluators may have resulted in even stronger associations with examination performance.

An important consideration to keep in mind when interpreting our results is that attending physicians and residents differed in the length of time they worked with a student, with residents working with a student for 4 weeks, compared to 2 weeks for faculty. Our findings could be interpreted as indirect evidence that evaluation may be compromised by 2-week attending rotations.¹⁶ One should be cautious in making such an assumption, however. For one, the RIME framework is not designed to simply evaluate knowledge, and the attending ratings could have been based on other student behaviors, such as accuracy in reporting data, or issues of professionalism. Second, as can be seen in Figure 1, some of the attending associations were attenuated by what could be perceived as misclassification of 2 students as reporters who scored high on the NBME examination (one scored 78, the other 84, for the mean of 81). Declaring attending evaluation as less accurate from data skewed (rightly or wrongly) by only two evaluations would not be warranted. Nevertheless, further research should focus on the effects on student and resident evaluation with reduced attending contact from shorter attending rotations.

Although we demonstrated that resident RIME ratings were associated with student examination performance, one should note that the RIME rating assigned to students tended to be

Figure 1. NBME Subject Examination Score as a function of attending, resident, and intern rating. R=Reporter; I=Interpreter; M=Manager; E=Educator.

high, with an average rating of Manager, and only 1–5% of students receiving Reporter designation, depending on the evaluator. One could contend that as these reflected end-of-rotation ratings, that a large number of students would achieve Manager stage at that time, and are perhaps accurate assessments. However, in RIME systems in which groups meet periodically throughout a clerkship, some 14–17% of students may be targeted for further discussion,^{3,8} a much greater percentage of students “red-flagged” as perhaps at risk for marginal funds of knowledge⁶ or future internship performance⁸ than the 1–5% our end-of-rotation, independent evaluator process identified. Likely, the lack of formal evaluation sessions in our study resulted in a lack of calibration for evaluators provided by group discussion,³ with subsequent grade inflation. Course directors should consider to what purpose they are using the RIME taxonomy. If the purpose is to assign grades based on the RIME system, our findings suggest that RIME conducted by independent evaluators without formal evaluation sessions would be a valid method of assigning grades, albeit at the risk of grade inflation. However, our modified RIME process may be insufficiently sensitive to detect as many students at risk for future poor performance as RIME conducted in conjunction with formal group evaluation sessions. Further research should compare these two ways of conducting evaluation through the RIME system.

Nevertheless, we conclude that a modified RIME system relying on individual evaluations, not coupled with formal evaluation sessions, is feasible and has demonstrated validity in the association of RIME ratings with measures of student knowledge. Our study extends the RIME literature by including the outcome of a clinical skills examination. In addition, our findings reinforce the notion that residents are accurate in their assessment of students, and should be important contributors to student evaluation.

Acknowledgments: We would especially like to offer our gratitude for the extremely helpful suggestions made by our anonymous external reviewers. Thank you.

Conflict of interest: None disclosed.

Corresponding Author: Charles H. Griffith III, MD, MSPH; K507 Kentucky Clinic, Lexington, KY 40536, USA (e-mail: cgrif00@uky.edu).

REFERENCES

1. **Pangaro LN.** Investing in descriptive evaluation: a vision for the future of assessment. *Med Teach.* 2000;22:478–81.
2. **Pangaro L.** A new vocabulary and other innovations for improving descriptive in-training evaluations. *Acad Med.* 1999;74:1203–7.
3. **Gaglione MM, Moores L, Pangaro L, Hemmer PA.** Does group discussion of student clerkship performance at an education committee affect an individual committee member's decisions. *Acad Med.* 2005;80(10 suppl):S55–8.
4. **Hemmer PA, Pangaro L.** Using formal evaluation sessions for case-based faculty development during clinical clerkships. *Acad Med.* 2000;75:1226–1.
5. **Battistone MJ, Milne C, Sande MA, Pangaro LN, Hemmer PA, Shomaker TS.** The feasibility and acceptability of implementing formal evaluation sessions and using descriptive vocabulary to assess student performance on a clinical clerkship. *Teach Learn Med.* 2002;14:5–10.
6. **Hemmer PA, Pangaro L.** The effectiveness of formal evaluation sessions during clinical clerkships in better identifying students with marginal funds of knowledge. *Acad Med.* 1997;72:641–3.
7. **Hemmer PA, Hawkins R, Jackson J, Pangaro L.** Assessing how well three evaluation methods detect deficiencies in medical students' professionalism in two settings of an internal medicine clerkship. *Acad Med.* 2000;75:167–73.
8. **Lavin B, Pangaro L.** Internship ratings as a validity outcome measure for an evaluation system to identify inadequate clerkship performance. *Acad Med.* 1998;73:998–1002.
9. **Hemmer PA, Papp KK, Mechaber AJ, Durning SJ.** Evaluation, grading, and use of a common evaluation framework by internal medicine clerkship directors: results of a national survey. *Teach Learn Med.* 2008, in press.
10. **Espey E, Ogburn T.** The R-I-M-E method for evaluation of medical students on an obstetrics and gynecology clerkship. *Am J Obstet Gynecol.* 2003;189:666–9.
11. **Torre DM, Daley BJ, Sebastian JL, Elnicki DM.** Overview of current learning theories for medical educators. *Am J Med.* 2006;119:903–7.
12. **Durning SJ, Pangaro LN, Denton GD, Hemmer PA, Wimmer A, Grau T, Gaglione MA, Moores L.** Intersite consistency as a measurement of programmatic evaluation in a medicine clerkship with multiple, geographically separated sites. *Acad Med.* 2003;78(10 suppl):S36–8.
13. Alliance for Clinical Education. Guidebook for Clerkship Directors. Chapter 6. McGaghie WC, Pangaro LN. available at: <http://familymed.uthscsa.edu/ACE/chapter6.htm#clerkship>, accessed March 3, 2008.
14. **Battistone MJ, Pendleton B, Milne C, et al.** Global descriptive evaluations are more responsive than global numeric ratings in detecting students' progress during the inpatient portion of an internal medicine clerkship. *Acad Med.* 2001;76(10 suppl.):S105–7.
15. **Griffith CH, Wilson JF, Haist SA, Ramsbottom-Lucier M.** Do students who work with better housestaff in their medicine clerkship learn more. *Acad Med.* 1998;73(10 suppl.):S57–9.
16. **Akl EA, Maroun N, Klocke RA, Schunemann HJ.** A survey of internal medicine residents and faculty about the duration of attendings' inpatient rotations. *J Gen Intern Med.* 2004;19:1133–9.