Scientific Research Publishing

Applying *Lean* Thinking: The Assessment of Professional Conduct of Medical Students

Charlene E. Gamaldo[1], Alyssa A. Gamaldo[2], Roy E. Strowd[3], Laurence Hou[4], Aadi Kalloo[1], Mark A. Sanchez[5], Rachel E. Salas[1]

[1]Johns Hopkins School of Medicine, Neurology, Baltimore, Maryland, USA
[2]School of Aging Studies, University of South Florida, Tampa, Florida, USA
[3]Johns Hopkins School of Medicine, Neuro-Oncology, Baltimore, Maryland, USA
[4]Johns Hopkins School of Medicine, Baltimore, Maryland, USA
[5]Anne Arundel Medical Center, Annapolis, Maryland, USA
Email: cgamald1@jhmi.edu, agamaldo@gmail.com, rstrowd1@jhmi.edu, lhou6@jhu.edu,
aadi.kalloo@gmail.com, msanchez@aahs.org, rsalas3@jhmi.edu

Abstract

Methods of incorporating the evaluation of professional conduct have varied widely by specialty, institution, and level of training. Medical educators now are tasked with developing more operationally-defined evaluative tools while balancing the ongoing need to reduce administrative load and survey fatigue that have been linked with physician burnout. Our aim was to investigate the value of a single question to measure the professional conduct of a medical student. Responses to the single question, "please rate this student's potential as a resident on YOUR team", were correlated with the individual core competency domains (17 questions with six questions targeted towards professional conduct), overall clinical evaluation score, final grade, and shelf examination score. Resident and faculty ratings across the 17 questions, overall clinical evaluation scores, and final grades were significantly associated with ratings on the single "Housestaff Potential" question. While this score was also significantly associated with the shelf score for resident evaluators only, it was a significant unique predictor for the overall clinical evaluation score for both evaluators. Our findings suggest that a single question designed to rate the professional conduct of medical students can be efficiently incorporated into the clerkship clinical evaluation.

Keywords

Medical Student, Professional Conduct, Assessment, Evaluation

1. Introduction

There has been a growing push by national entities [e.g., Liaison Committee on Medical Education (LCME), and Accreditation Council for Graduate Medical Education (ACGME)] to validate and standardize the assessment methods for evaluating several critical competency domains that include medical knowledge, clinical reasoning, procedural skills and most recently professionalism across the training period (Batalden et al., 2002). Faculty and resident ratings continue to comprise the largest portion of the final clerkship grade at most institutions to assess these competency domains (Kassebaum & Eaglen, 1999). While most studies have focused on developing more structured approaches to formative assessments of medical knowledge (Merlin et al., 2014; Goldstein et al., 2014; Farrell et al., 2010) and/or data gathering (Dudas et al., 2012), few have studied standardized assessments of professionalism across the medical school curriculum (Epstein, 2007; Holmboe et al., 2010). Methods of incorporating the evaluation of professional conduct have varied widely by specialty, institution, and level of training, with some consisting of 360 evaluations (Berk, 2009; Rees & Shepherd, 2005). Previous work has shown that carefully designed comprehensive rating forms may be unable to distinguish distinct and independent clinical competencies (Silber et al., 2004) and that global rating forms in clinical assessments require high response rates in order for them to provide valid data (Littlefield et al., 2001). However, medical school curriculum leaders are tasked with developing valid and more operationally defined evaluative tools while balancing the ongoing need to reduce the administrative load and survey fatigue that have been linked with the recent rise in physician burnout rate (Jarral et al., 2015; Sigsbee & Bernat, 2014).

Lean methodology has been utilized as a "fresh approach" towards creating models that optimize productivity, cost, quality, and timely delivery of services (Kates, 2014). Recently, hospitals have begun adapting a "lean" systems-approach to streamline processes in an effort to improve costs and efficiencies, while reducing waste (Kates, 2014). While some studies have shown that ratings on any single characteristics can predict overall grade (Pulito et al., 2007), to our knowledge, a lean approach has not been studied in the context of the medical student evaluation process during a clerkship, particularly in assessing professional conduct. The business arena has demonstrated the utility of a single question as an effective evaluative tool of future success (Huhman, 2014). The primary aim of this study was to apply the aforementioned principle to investigate the value of a single question (posed to evaluators) as a potential effective tool to measure the professional conduct (and potentially replace the current six professional conduct questions in the clinical evaluation) of a medical student during a clerkship. We hypothesize that the single-question, "*please rate this student's potential as a resident on YOUR team*" will correlate with questions specifically aimed at professional conduct in a neurology clerkship and overall clinical performance.

2. Methods

This study was approved by the Institutional Review Board at Johns Hopkins University School of Medicine (JHSOM). At the time of the study, the clinical evaluation score was worth 30% of the final grade for the clerkship. The Johns Hopkins Neurology Core Clerkship (NCC) final grade consisted of the following: 30% for inpatient clinical evaluations, 25% for the National Board of Medical Examiners (NBME) shelf examination, 25% for a neurological standardized patient encounter, 10% for an internal examination, 5% for a community outpatient clinic evaluation, and 5% for a 360 evaluation completed by non-physician healthcare staff.

Students were asked to select at least three faculty and/or resident evaluators, with at least one evaluation required from a faculty evaluator. Students chose their clinical evaluators in order to allow students the opportunity to select the faculty/residents they believe to have had the best opportunity to assess their performance including the common domains that typically encompass professional conduct (Armstrong et al., 2004). The current NCC faculty and resident evaluation consists of 29 total items including 17 that specifically evaluate clinical performance with six specifically targeted at assessing professional conduct (i.e., responsibility/reliability, compassion, respectfulness, response to feedback, rapport with patients, and rapport with colleagues). These 17 questions range from 0 (*unacceptable*) to 5 (*outstanding*) (**Table 2**). For those students with multiple faculty/resident evaluators, we took the average rating across the evaluators. Evaluators were blinded to student's performance on all other assessment measures (i.e., NBME, other clinical evaluations, etc.). On average, students spend approximately 5 - 6 hours with housestaff and 3 hours with a faculty a day during the NCC on the wards (e.g., rounding, team teaching, noon conference) which is the setting of the majority of the clinical evaluations.

To evaluate a medical student's "future housestaff potential", evaluators were asked, "please rate this student's potential as a resident on YOUR team." This "potential housestaff" rating ranged from 1 (*poor*) to 5 (*excellent*). This question was formatted as a stand-alone question with instructions to the evaluator that the response would not factor into any portion of the student's grade.

Pearson's correlations were conducted to examine potential associations between the faculty/resident rating of student housestaff potential (i.e., single question) and the following: 1) NCC final grade, 2) NBME shelf examination score and 3) faculty/resident NCC evaluation responses to each of the 17 items (including the 6 professional conduct items), and the overall clinical evaluation score. Given that the housestaff potential question was not factored into students' final grade, a subsequent regression analysis was conducted to evaluate whether the faculty and/or resident rating of housestaff potential was a significant unique predictor of the overall clinical evaluation score, NBME shelf exam score, or NCC final grade. For analysis, the following numerical assignment was used for NCC final grades: "Honors" = 5, "High Pass" = 4, "Pass" = 3, "Unsatisfactory" = 2, "Fail" = 1.

3. Results

Demographics

The sample included 193 JHSOM medical students (Mean age = 26.58, SD = 3.02; range 23 - 38) who were enrolled in the NCC from 2011-2014 (see **Table 1** on sample characteristics). Students were in their second (6%), third (72%), or fourth (22%) year of medical school. Gender was evenly divided (48% Male, 52% Female). Students rotated throughout a year's five blocks (four academic quarters plus summer) with 23% in the first block, 20% in block two, 21% in block three, 21% in block four, and 15% in block five (summer). Cronbach alphas for the 17 competency evaluation items were 0.87 (resident evaluations) and 0.93 (faculty evaluations), suggesting the scales have adequate reliability (Tavakol & Dennick, 2011; Lance et al., 2006). However, previous literature suggests that alphas greater than 90 may imply some redundancy in the scale items and potential need to shorten the scale (Tavakol & Dennick, 2011). Thus, the competency evaluations, particularly the faculty evaluations, could be shortened. On average, students selected 1.48 faculty evaluators (SD = 0.72; number of evaluations range: 1 - 5) and 2.05 resident evaluators (SD = 0.89; number of evaluations range: 1 - 7). The sample, on average, had high overall clinical evaluation scores (Mean = 4.64; SD = 0.31), with the majority of students obtaining "High Pass" for their NCC final grades (Mean = 4.21; SD = 0.63) and relatively good NBME scores (Mean = 77.80; SD = 7.60). Students had an overall clinical evaluation mean score of 4.51 (SD = 0.63) by faculty, and 4.57 (SD = 0.59) by residents with a mean of 3.47 (SD = 1.15) evaluators per student.

Table 1. Sample characteristics.

	No. (%)	Mean [SD]
Age	-	26.6 [3.0]
Sex (n = 188)		
Male	98 (52.1)	-
Female	90 (47.9)	-
Medical School Year (n = 185)		
Second	11 (5.9)	-
Third	133 (71.9)	-
Fourth	41 (22.2)	-
Number of Evaluators		
Faculty	-	1.5 [0.7]
Resident	-	2.1 [0.9]
Overall Clinical Evaluation Score		4.6 [0.3]
NCC Final Grade	-	4.2 [0.6]
NBME Score	-	77.8 [7.6]

Individual clinical evaluation questions (*core competency items*) *and the overall clinical evaluation score*, *NCC final grade*, *and the NBME shelf examination.*

Higher (better) overall clinical evaluation scores were significantly associated with higher (better) ratings across all 17 individual competency evaluation items for the faculty (*r* range: 0.16 - 0.59, *p* range: 0.000 - 0.035) evaluators and 16 (except procedural skills) for the resident (*r* range: 0.13 - 0.57, *p* range: 0.000 - 0.089) evaluators (**Table 2**).

Resident ratings on the responsibility/reliability item had the strongest relationship with the overall clinical evaluation score (*r* = 0.57, *p* = 0.000). The faculty ratings on the clinical knowledge item had the strongest relationship (*r* = 0.59, *p* = 0.000) with the overall clinical evaluation score. Higher (better) NCC final grades were significantly associated with higher (better) ratings across the 17 competency evaluation items for the resident (*r* range: 0.27 - 0.43, *p* range: 0.000 - 0.009) and 16 (except procedural skills) for faculty (*r* range: 0.17 - 0.37, *p* range: 0.000 - 0.190) evaluators (**Table 3**).

Resident ratings on the responsibility/reliability item had the strongest relationship with the NCC final grade (*r* = 0.43, *p* = 0.000), similar to what was observed above (i.e., overall clinical evaluation score). However, faculty ratings on the "problem solving" item had the strongest relationship (*r* = 0.37, *p* = 0.000) with the NCC final grade. Lastly, higher (better) NBME scores were significantly correlated with higher (better) ratings on more evaluation items for the resident evaluators (*r* range: 0.15 - 0.30, *p* range: 0.000 - 0.039) than the faculty (*r* range: 0.03 - 0.19, *p* range: 0.009 - 0.731; **Table 4**). Resident ratings on the "problem solving" item had the strongest relationship with the NBME exam (*r* = 0.30, *p* = 0.000). However, faculty ratings on the clinical judgment item had the strongest relationship (*r* = 0.19, *p* = 0.009) with the NBME exam.

Single Question, "please rate this student's potential as a resident on YOUR team," *and the students' competency evaluation items of professionalism*, *overall clinical evaluation score*, *NCC final grade*, *and the NBME shelf score.*

Resident ratings on future housestaff potential correlated significantly with higher (better) ratings across all 6 individual competency evaluation items for professionalism (*r* range: 0.34 - 0.55, *p*-values < 0.000); the professional

Table 2. Relationships between resident/faculty competency component evaluation items and overall clinical evaluation score.

Resident Evaluation Items	Clinical Overall[a]	*p*-value	Faculty Evaluation Items	Clinical Overall[a]	*p*-value
Basic Science Knowledge	0.395	0.000	Basic Science Knowledge	0.297	0.000
Clinical Knowledge	0.485	0.000	Clinical Knowledge	0.589	0.000
Self-directed Learning	0.542	0.000	Self-directed Learning	0.337	0.000
Data Gathering	0.430	0.000	Data Gathering	0.556	0.000
Physical/Mental Status Exams	0.464	0.000	Physical/Mental Status Exams	0.353	0.000
Presenting Patients on Rounds	0.449	0.000	Presenting Patients on Rounds	0.531	0.000
Problem Solving	0.479	0.000	Problem Solving	0.352	0.000
Clinical Judgment	0.479	0.000	Clinical Judgment	0.438	0.000
Responsibility/Reliability	0.569	0.000	Responsibility/Reliability	0.350	0.000
Compassion	0.503	0.000	Compassion	0.291	0.000
Respectfulness	0.427	0.000	Respectfulness	0.297	0.000
Response to Feedback	0.446	0.000	Response to Feedback	0.376	0.000
Rapport with Patients	0.562	0.000	Rapport with Patients	0.324	0.000
Rapport with Colleagues	0.367	0.000	Rapport with Colleagues	0.343	0.000
Oral Patient Presentations	0.449	0.000	Oral Patient Presentations	0.416	0.000
Recording Clinical Data	0.232	0.001	Recording Clinical Data	0.218	0.003
Procedural Skills	0.125	0.089	Procedural Skills	0.155	0.035

Note: [a]Numbers in cells reflect correlation coefficients.

Table 3. Relationships between resident/faculty competency component evaluation items and NCC final grade.

Resident Evaluation Items	NCC Final Grade[a]	p-value	Faculty Evaluation Items	NCC Final Grade[a]	p-valve
Basic Science Knowledge	0.313	0.000	Basic Science Knowledge	0.245	0.001
Clinical Knowledge	0.352	0.000	Clinical Knowledge	0.321	0.000
Self-Directed Learning	0.428	0.000	Self-Directed Learning	0.260	0.000
Data Gathering	0.400	0.000	Data Gathering	0.351	0.000
Physical/Mental Status Exams	0.323	0.000	Physical/Mental Status Exams	0.289	0.000
Presenting Patients on Rounds	0.324	0.000	Presenting Patients on Rounds	0.286	0.001
Problem Solving	0.400	0.000	Problem Solving	0.371	0.000
Clinical Judgment	0.361	0.000	Clinical Judgment	0.332	0.000
Responsibility/Reliability	0.432	0.000	Responsibility/Reliability	0.334	0.000
Compassion	0.367	0.000	Compassion	0.245	0.001
Respectfulness	0.347	0.000	Respectfulness	0.232	0.002
Response to Feedback	0.335	0.000	Response to Feedback	0.210	0.005
Rapport with Patients	0.347	0.000	Rapport with Patients	0.247	0.001
Rapport with Colleagues	0.402	0.000	Rapport with Colleagues	0.284	0.000
Oral Patient Presentations	0.315	0.000	Oral Patient Presentations	0.286	0.000
Recording Clinical Data	0.358	0.000	Recording Clinical Data	0.241	0.002
Procedural Skills	0.273	0.009	Procedural Skills	0.170	0.190

Note: [a]Numbers in cells reflect correlation coefficients.

Table 4. Relationships between resident/faculty competency component evaluation items and NBME neurology examination.

Resident Evaluation Items	NBME Exam[a]	p-value	Faculty Evaluation Items	NBME Exam[a]	p-value
Basic Science Knowledge	0.256	0.000	Basic Science Knowledge	0.100	0.190
Clinical Knowledge	0.284	0.000	Clinical Knowledge	0.172	0.020
Self-Directed Learning	0.233	0.001	Self-Directed Learning	0.136	0.071
Data Gathering	0.237	0.001	Data Gathering	0.152	0.040
Physical/Mental Status Exams	0.152	0.039	Physical/Mental Status Exams	0.055	0.470
Presenting Patients on Rounds	0.206	0.013	Presenting Patients on Rounds	0.074	0.385
Problem Solving	0.303	0.000	Problem Solving	0.176	0.019
Clinical Judgment	0.280	0.000	Clinical Judgment	0.194	0.009
Responsibility/Reliability	0.227	0.002	Responsibility/Reliability	0.110	0.143
Compassion	0.187	0.011	Compassion	0.078	0.304
Respectfulness	0.203	0.006	Respectfulness	0.026	0.731
Response to Feedback	0.207	0.005	Response to Feedback	0.072	0.343
Rapport with Patients	0.187	0.011	Rapport with Patients	0.066	0.384
Rapport with Colleagues	0.259	0.000	Rapport with Colleagues	0.046	0.545
Oral Patient Presentations	0.265	0.000	Oral Patient Presentations	0.101	0.178
Recording Clinical Data	0.229	0.002	Recording Clinical Data	0.107	0.164
Procedural Skills	0.272	0.009	Procedural Skills	0.072	0.582

Note: [a]Numbers in cells reflect correlation coefficients.

item of responsibility/reliability had the strongest relationship with the resident ratings on the single question. In addition, the single question correlated significantly with higher overall clinical evaluation scores ($r = 0.70$, $p < 0.001$), higher final NCC grades ($r = 0.39$, $p < 0.001$) and higher NBME scores ($r = 0.29$, $p = 0.001$; **Table 5**). The magnitude of the association between the resident ratings on the housestaff potential item and overall clinical evaluation scores was greater than the association observed between the individual competency evaluation item of responsibility/reliability and overall clinical evaluation score ($r = 0.57$). However, the magnitude of the association between resident ratings on the housestaff potential item and NCC grades/NBME scores was less than the association observed between the individual competency evaluation item of responsibility/reliability and NCC final grade ($r = 0.43$) as well as the relationship between the individual competency evaluation item of problem solving and NBME (shelf) scores ($r = 0.30$).

Faculty ratings on future housestaff potential correlated significantly with higher (better) ratings across all 6 individual competency evaluation items for professionalism (r range: 0.43 - 0.57, p-values < 0.000); the professional item of rapport with colleagues had the strongest relationship with the faculty ratings on the single question. Although the single question was also significantly associated with higher overall clinical evaluation scores ($r = 0.69$, $p < 0.001$) and higher NCC final grades ($r = 0.25$, $p = 0.004$), it was not significantly associated with higher NBME shelf exam scores ($r = -0.00$, $p = 0.970$). The magnitude of the association between the faculty ratings on the housestaff potential item and overall clinical evaluation scores was greater than the association observed between the individual competency evaluation item of clinical knowledge and overall clinical evaluation score ($r = 0.59$). However, the magnitude of the association between faculty ratings on the housestaff potential item and NCC grades/NBME scores was less than the association observed between the individual competency evaluation item of problem solving and NCC final grade ($r = 0.37$) as well as the relationship between the

Table 5. Relationships between the "please rate this student's potential as a resident on YOUR team" question and performance metrics (e.g., evaluation competency components, overall clinical evaluation, NCC final grade, and NBME exam).

Resident Evaluation Items	Resident Potential[a]	p-value	Faculty Evaluation Items	Resident Potential[a]	p-value
Basic Science Knowledge	0.521	0.000	Basic Science Knowledge	0.445	0.000
Clinical Knowledge	0.607	0.000	Clinical Knowledge	0.581	0.000
Self-Directed Learning	0.550	0.000	Self-directed Learning	0.536	0.000
Data Gathering	0.472	0.000	Data Gathering	0.683	0.000
Physical/Mental Status Exams	0.511	0.000	Physical/Mental Status Exams	0.494	0.000
Presenting Patients on Rounds	0.496	0.000	Presenting Patients on Rounds	0.626	0.000
Problem Solving	0.540	0.000	Problem Solving	0.506	0.000
Clinical Judgment	0.472	0.000	Clinical Judgment	0.570	0.000
Responsibility/Reliability	0.547	0.000	Responsibility/Reliability	0.505	0.000
Compassion	0.414	0.000	Compassion	0.443	0.000
Respectfulness	0.471	0.000	Respectfulness	0.521	0.000
Response to Feedback	0.343	0.000	Response to Feedback	0.482	0.000
Rapport with Patients	0.500	0.000	Rapport with Patients	0.496	0.000
Rapport with Colleagues	0.401	0.000	Rapport with Colleagues	0.572	0.000
Oral Patient Presentations	0.533	0.000	Oral Patient Presentations	0.429	0.000
Recording Clinical Data	0.487	0.000	Recording Clinical Data	0.373	0.000
Procedural Skills	0.227	0.000	Procedural Skills	0.194	0.000
Overall Clinical Evaluation	0.704	0.000	Overall Clinical Evaluation	0.691	0.000
NCC Final Grade	0.393	0.000	NCC Final Grade	0.249	0.004
NBME Exam	0.286	0.001	NBME Exam	−0.003	0.970

NB: [a]Cells reflect correlation coefficients.

individual competency evaluation item of "clinical judgment" and NBME (shelf) scores ($r = 0.19$).

Since the resident potential rating question was a stand-alone item and did not factor into the final grade (as the other evaluation items), three multivariate linear regression analyses were performed to explore whether the resident potential rating is a unique significant predictor of the overall clinical evaluation score, NCC final grade, or NBME (shelf) score (**Table 6**). For the overall clinical evaluation score, both the faculty rating ($\beta = 0.52$, SE = 0.03, $p < 0.001$) and the resident rating ($\beta = 0.52$, SE = 0.04, $p < 0.001$) of a students' future housestaff potential were significant unique predictors for the overall clinical evaluation score, even after adjusting for age, sex, year of medical school, and timing of rotation. The results of the regression indicated the predictors explained 69.7% of the variance in the overall clinical score ($R^2 = 0.69$, $F(6,105) = 37.94$, $p < 0.001$). However, only the resident ratings of the students' future housestaff potential served as a significant unique predictor of the NBME score ($\beta = 0.22$, SE = 2.25, $p = 0.043$) and final grade ($\beta = 0.27$, SE = 0.18, $p = 0.010$) even after adjusting for the faculty/resident potential rating (NBME: $\beta = -0.11$, SE = 1.61, $p = 0.322$; final grades: $\beta = 0.10$, SE = 0.13, $p = 0.359$), age, sex, year of medical school, and timing of rotation. The results of the regression indicated the predictors explained 6.3% of the variance in the NBME score ($R^2 = 0.06$, $F(6,105) = 1.10$, $p > 0.05$) and 9.4% of the variance in the final grade ($R^2 = 0.09$, $F(6,105)=1.71$, $p > 0.05$).

4. Discussion

Our results demonstrate that resident evaluator responses to one comprehensive question (i.e., *please rate this student's potential as a resident on YOUR team*) strongly and significantly correlates with all 17 competency questions in our clinical evaluation for residents evaluators and 16 for the faculty evaluators, and thus, as expected, the overall clinical evaluation score for both evaluators. Of interest, our findings show that this single *housestaff potential* question is strongly associated with each of the six professional conduct items of the clinical evaluation suggesting that this one question may capture this particular competency component addressed currently with six individual questions. Further, the single question was also associated with the final clinical evaluation score (from both faculty and resident evaluators) and for the NBME (shelf) score for resident evaluators. Perhaps asking a single question ("future resident potential on YOUR team") may serve as a more targeted and all-encompassing approach to verify the clinical evaluation and could serve to replace other questions aimed at assessing professional conduct in a clinical setting. While advocacy in replacing the clerkship's current clinical evaluation or clinical grade with this single question is certainly not the intent of the authors, the utility of implementing this one question in place of a set of questions may be of some benefit and is worth further consideration. In particular, the results of our study may provide strong support for the concept of having a single question assess the critical competency domain of professionalism. Moreover, the concept that this question could

Table 6. Single question predicting overall clinical score, NCC final grade, and NBME exam.

	Clinical Overall		NCC Final Grade		NBME Exam	
	Model 1 B (SE)	Model 2 B (SE)	Model 1 B (SE)	Model 2 B (SE)	Model 1 B (SE)	Model 2 B (SE)
Single Question						
Resident Housestaff Potential	0.52*** (0.04)	0.52*** (0.04)	0.26* (0.17)	0.27* (0.18)	0.22* (2.13)	0.22* (2.25)
Faculty Housestaff Potential	0.52*** (0.03)	0.52*** (0.03)	0.09 (0.12)	0.10 (0.13)	−0.13 (1.54)	−0.11 (1.61)
Covariates						
Age		0.10 (0.01)		0.07 (0.02)		−0.03 (0.25)
Sex		0.06 (0.03)		−0.02 (0.12)		−0.00 (1.51)
Year of Medical School		−0.02 (0.04)		0.04 (0.15)		0.09 (1.81)
Timing of Rotation		0.05 (0.01)		−0.03 (0.05)		−0.05 (0.63)
R^2	.68	0.69	0.08	0.09	0.04	0.06
ΔR^2		0.01		0.01		0.02

Note: *$p < 0.05$. **$p < 0.01$. ***$p < 0.001$. Model 1 represents the unadjusted regression model. Model 2 represents the regression model adjusted for the age, sex, year of medical school, and timing of rotation.

potentially replace the six questions to professionalism is even more appealing particularly in the current era of physician burnout. Additionally, this could potentially serve as a uniform and universal question in all of the clinical evaluations to track student growth (or deficiencies) in this domain longitudinally throughout the clinical years of medical school. This method may prove fruitful as a means of assessing Entrustable Professional Activities (EPA) (AAMC, 2014) soon to be incorporated as a graduation guideline for all U.S. medical schools. While it may be interesting that this question is predictive for both, faculty and residents, for overall clinical evaluation grade, the residents (and not the faculty) rating on this single question is also predictive of the NBME shelf and the NCC final grade. This may not be surprising since residents spend more time with the students. However, this may suggest that there may be unique benefits of asking this question to both the faculty and residents since they may glean unique insight on students' clinical ability. It is possible that residents are better able/positioned to evaluate core knowledge while faculty are more in tune with assessing more all-encompassing skills such as professional conduct and clinical reasoning. A seasoned clinician may have a better instinctual barometer for what makes a competent resident than for what defines competency in medical knowledge, clinical judgment, problem solving, or procedural skills. Educators have been focusing on defining more clearly what each rating means and maybe what we have found here is a nice example of giving the evaluators the opportunity to apply their evaluation in a more realistic and meaningful way. An important caveat to consider, in our design, is that the housestaff potential question was prefaced with the stipulation that it would not affect final clerkship grades. Given that approximately 49% of US medical schools base of final neurology clerkship grades on direct observations by faculty (e.g., clinical evaluations) and residents, the results of this study may influence clerkship directors to consider the relative weight allocated to final clinical clerkship grades and, more importantly, identify valid and efficient methods to assess specific skills and core competencies relevant to becoming a proficient clinician (Carter et al., 2014). Based on the results of this study, it appears that further investigations into the most reliable and efficient means of evaluating students are certainly warranted.

This study is not without limitation. We present a small-scale study, conducted at a single institution and in one specialty and thus, generalizability are limited. Additionally, certain data was not collected by NCC staff and is therefore, unavailable here, including ethnic distribution and age distribution. However, even after adjusting for several potential confounders (e.g., gender and medical school year), we still observed a significant relationship between the single housestaff potential ratings and standard NCC performance metrics. Students also had a variable number of evaluators (ranging from 2 - 7); however no significant findings were found between the number of faculty/resident evaluators and NBME/final grades.

5. Conclusion

In summary, a single question asking academic faculty and housestaff to rate a medical student's "future housestaff potential" may serve as a comprehensive measure of a student's current professional conduct. As academic medicine faculty strive to identify useful and cost-effective processes across all domains of healthcare and education, further studies remain needed. Implementations of strategies that can streamline, yet provide, value to the medical evaluation process are imperative. Future directions should investigate the utility of lean methodologies in medical education to improve efficiency and reduce waste in hopes of ultimately improving the educational quality, value and productivity for both medical learner and educator.

References

AAMC (2014). *Core Entrustable Professional Activities for Entering Residency: Curriculum Developers' Guide*. Association of American Medical Colleges.

Armstrong, E. G., Mackey, M., & Spear, S. J. (2004). Medical Education as a Process Management Problem. *Academic Medicine, 79,* 721-728. http://dx.doi.org/10.1097/00001888-200408000-00002

Batalden, P., Leach, D., Swing, S., Dreyfus, H., & Dreyfus, S. (2002). General Competencies and Accreditation in Graduate Medical Education. *Health Affairs, 21,* 103-111. http://dx.doi.org/10.1377/hlthaff.21.5.103

Berk, R. (2009). Using the 360˚ Multisource Feedback Model to Evaluate Teaching and Professionalism. *Medical Teacher, 31,* 1073-1080. http://dx.doi.org/10.3109/01421590802572775

Carter, J. L., Ali, I. I., Isaacson, R. S., Safdieh, J. E., Finney, G. R., Sowell, M. K., Sam, M. C., Anderson, H. S., Shin, R. K., Kraakevik, J. A. et al. (2014). Status of Neurology Medical School Education: Results of 2005 and 2012 Clerkship Director Survey. *Neurology, 83,* 1761-1766. http://dx.doi.org/10.1212/WNL.0000000000000962

Dudas, R. A., Colbert, J. M., Goldstein, S., & Barone, M. A. (2012). Validity of Faculty and Resident Global Assessment of Medical Students' Clinical Knowledge during Their Pediatrics Clerkship. *Academic Pediatrics, 12,* 138-141.

Epstein, R. M. (2007). Assessment in Medical Education. *The New England Journal of Medicine, 356,* 387-396. http://dx.doi.org/10.1056/NEJMra054784

Farrell, T. M., Kohn, G. P., Owen, S. M., Meyers, M. O., Stewart, R. A., & Meyer, A. A. (2010). Low Correlation between Subjective and Objective Measures of Knowledge on Surgery Clerkships. *Journal of the American College of Surgeons, 210,* 680-6835. http://dx.doi.org/10.1016/j.jamcollsurg.2009.12.020

Goldstein, S. D., Lindeman, B., Colbert-Getz, J., Arbella, T., Lidor, A., & Sacks, B. (2014). Faculty and Resident Evaluations of Medical Students on a Surgery Clerkship Correlate Poorly with Standardized Exam Scores. *The American Journal of Surgery, 207,* 231-235. http://dx.doi.org/10.1016/j.amjsurg.2013.10.008

Holmboe, E. S., Sherbino, J., Long, D. M., Swing, S. R., & Frank, J. R. (2010). The Role of Assessment in Competency-Based Medical Education. *Medical Teacher, 32,* 676-682. http://dx.doi.org/10.3109/0142159X.2010.500704

Huhman, H. (2014). 9 Questions to Ask Candidates' References. http://www.entrepreneur.com/article/235359

Jarral, O. A., Baig, K., Shetty, K., & Athanasiou, T. (2015). Sleep Deprivation Leads to Burnout and Cardiothoracic Surgeons Have to Deal with Its Consequences. *International Journal of Cardiology, 179,* 70-72. http://dx.doi.org/10.1016/j.ijcard.2014.10.020

Kassebaum, D. G., & Eaglen, R. H. (1999). Shortcomings in the Evaluation of Students' Clinical Skills and Behaviors in Medical School. *Academic Medicine, 74,* 842-849. http://dx.doi.org/10.1097/00001888-199907000-00020

Kates, S. L. (2014). Lean Business Model and Implementation of a Geriatric Fracture Center. *Clinics in Geriatric Medicine, 30,* 191-205. http://dx.doi.org/10.1016/j.cger.2014.01.002

Lance, C. E., Butts, M. M., & Michels, L. C. (2006). The Sources of Four Commonly Reported Cutoff Criteria. What Did They Really Say? *Organizational Research Methods, 9,* 202-220. http://dx.doi.org/10.1177/1094428105284919

Littlefield, J., Paukert, J., & Schoolfield, J. (2001). Quality Assurance Data for Residents' Global Performance Ratings. *Academic Medicine, 76,* S102-S104. http://dx.doi.org/10.1097/00001888-200110001-00034

Merlin, L. R., Horak, H. A., Milligan, T. A., Kraakevik, J. A., & Ali, I. I. (2014). A Competency-Based Longitudinal Core Curriculum in Medical Neuroscience. *Neurology, 83,* 456-462.

Pulito, A. R., Donnelly, M. B., & Plymale, M. (2007). Factors in Faculty Evaluation of Medical Students' Performance. *Medical Education, 41,* 667-675. http://dx.doi.org/10.1111/j.1365-2923.2007.02787.x

Rees, C., & Shepherd, M. (2005). The Acceptability of 360-Degree Judgments as a Method of Assessing Undergraduate Medical Students' Personal and Professional Behaviours. *Medical Education, 39,* 49-57. http://dx.doi.org/10.1111/j.1365-2929.2004.02032.x

Sigsbee, B., & Bernat, J. L. (2014). Physician Burnout: A Neurologic Crisis. *Neurology, 83,* 2302-2306.

Silber, C. G., Nasca, T. J., Paskin, D. L., Eiger, G., Robeson, M., & Veloski, J. J. (2004). Do Global Rating Forms Enable Program Directors to Assess the ACGME Competencies? *Academic Medicine, 79,* 549-556. http://dx.doi.org/10.1097/00001888-200406000-00010

Tavakol, M., & Dennick, R. (2011). Making Sense of Cronbach's Alpha. *International Journal of Medical Education, 2,* 53-55. http://dx.doi.org/10.5116/ijme.4dfb.8dfd