Using "Standardized Narratives" to Explore New Ways to Represent Faculty Opinions of Resident Performance

Glenn Regehr, PhD, Shiphra Ginsburg, MD, MEd, Jodi Herold, BHSc(PT), PhD, Rose Hatala, MD, MSc, Kevin Eva, PhD, and Olga Oulanova, MA

Abstract

Purpose

Most efforts to develop reliable evaluations of clinical competence have been oriented toward deconstructing the requisite competencies into separate scales. However, many are questioning the value of this approach on theoretical and empirical bases. This study uses "standardized narratives" to explore a different approach to assessing resident performance.

Method

In 2009, based on interviews with 19 experienced clinical faculty from two institutions, 16 narrative profiles were created to represent the range of resident competence that clinical faculty might encounter during supervision. Fourteen clinicians from three institutions independently grouped the profiles into as many categories as necessary to reflect various levels of performance, described their categories, then ranked the individual profiles within each category. Then, in groups of three or four, participants negotiated a final ranking and grouping of the 16 profiles.

Results

Despite interesting idiosyncracies in the factors some participants identified as guiding their rankings, there was strong consistency across the 14 clinicians regarding the rankings (single-rater intraclass correlation [ICC] = 0.86) and groupings (single-rater ICC = 0.81) of the profiles. Similarly, across institutions, the four groups were highly consistent in their final negotiated rankings (single-group ICC = 0.91) and groupings (single-group ICC = 0.87) of the profiles.

Conclusions

Faculty showed more consistency in their decisions of what constitutes excellent, competent, and problematic performance in residents than implied by current assessment techniques that require deconstruction of resident competencies. This use of standardized narratives points to interesting opportunities for more authentically codifying faculty opinions of residents.

The evaluation of clinical competence in the practice setting continues to be a cornerstone of the process by which the health professions determine trainees' preparedness to enter into professional practice. In medical residency programs, the end-of-rotation clinical evaluation (often called the "In-training Evaluation Report," or ITER) has been one of the main mechanisms by which this clinical evaluation process is enacted.[1] In general, ITERs consist of a set of rating scales that clinical supervisors are expected to use to indicate how well a resident is meeting the expectations of the training program across multiple domains of competence. In principle, this type of clinical performance evaluation has many characteristics that should make it an excellent tool for clinical assessment: It is based on the observation of performance, it is embedded in the real practice setting, it represents an extended observation period, and it is completed by experts in the domain.[2] In practice, however, the ITER has been problematic as a mechanism to discriminate among residents,[2,3] and in particular it has been a weak tool for identifying learners who are experiencing clinical difficulties.[4,5] Despite efforts to improve the scales on which the ITERs are based,[6] and despite various efforts to train faculty to use the scales more effectively,[7] the ITER continues to be problematic as a tool to describe resident performance and discriminate among residents.

Increasingly, it is being suggested that the difficulties in developing effective scales to evaluate clinical performance in the field may have less to do with the specific details of the tools and more to do with the fundamental set of assumptions (the epistemology) that underlies the use of these tools.[8,9] That is, for the last several decades, the approach to assessment in medical education has been dominated by a psychometric epistemology in which it is presumed that psychological constructs can be deconstructed and assigned numerical values according to definable rules to obtain an accurate and concise description of an individual's ability that will be objective, replicable, easily communicable, and comparable. This model has served the field well in recent years by guiding refinements to the assessment of knowledge and spurring the development of tools such as the objective structured clinical examination.

Yet, our measurement instruments do not merely allow us to quantify a construct; they shape how we think about, evolve, and ultimately teach that construct. Thus, as Hodges[10] has warned, the psychometric construction of "competence as a reliable test score" opens the door for producing forms of "hidden incompetence." For example, our ability to measure knowledge with high reliability might lead to an overemphasis of knowledge and prevent the medical community from noticing when individual practitioners do not maintain adequate interpersonal skills. Similarly, our effort to assess each competency on a separate scale might direct our focus away from the trainee's ability to integrate these competencies into a

Please see the end of this article for information about the authors.

Correspondence should be addressed to Dr. Ginsburg, Mount Sinai Hospital, 433-600 University Ave., Toronto, Ontario, Canada, M5G 1X5; e-mail: shiphra.ginsburg@utoronto.ca.

Acad Med. 2012;87:000–000.
First published online
doi: 10.1097/ACM.0b013e31824858a9

Supplemental digital content for this article is available at http://links.lww.com/ACADMED/A77.

coherent understanding of effective clinical practice. This warning seems particularly important as the field strives to expand the definition of competence to explicitly include the more social and relational aspects of professional expertise (such as communication, collaboration, and professionalism).[11,12] Understanding these social aspects of professional practice may require a more constructivist epistemology, which is based in the recognition that competent performance is always embedded in situated, relational contexts that are rich with information.[13] Competence, from this perspective, is recognized as being constantly constructed and reconstructed and is acknowledged as inherently subjective and integrative in nature. Although context has been recognized to be important even with scale-based evaluations (e.g., observations by multiple raters in multiple situations are known to enhance reliability and validity), variations in performance across these contexts are generally interpreted as sources of noise that hide the "true," stable score that properly represents the individual. Thus, these strategies can be seen as an attempt to "extract" the individual from the situation. In addition, scale-based strategies are explicitly designed to isolate different competencies independently, rather than asking the rater to assess the individual within a full context of performance. The clash of epistemologies that arises from the application of the psychometric approach to the complex social and relational aspects of clinical expertise was well articulated by Leach,[14] who, in describing the development and evaluation of the competencies mandated by the Accreditation Council for Graduate Medical Education, stated:

> The relevance of the work is dependent on an integrated version of the competencies, whereas measurement relies on a speciated version of the competencies. The paradox cannot be resolved easily. The more the competencies are specified the less relevant to the whole they become.

In fact, we would argue it is the very nature of paradox that it cannot be resolved *at all* using the thinking that generated the paradox in the first place. Thus, as a community, we might do well to reconsider Leach's assertion that measurement necessarily "relies on a speciated version of the competencies."[14] As Schuwirth and van der Vleuten[8] suggest, it may be worthwhile instead to develop a better understanding of how teachers process (and represent) large bodies of rich information, and it may be worthwhile developing evaluation approaches that more authentically reflect this richness of information while keeping it manageable.

One potential starting place for this process might be the recognition that narrative (i.e., the stories that individuals construct about their experiences) "is the most compelling form by which we recount our reality, understand events, and through which we make sense of our experiences and ourselves."[15] Thus, narrative (in this context, the stories supervisors tell about their residents) has the potential to authentically reflect the richness of information suggested by Schuwirth and van der Vleuten.[8] However, we would also note that unfettered narrative has the potential to violate Schuwirth and van der Vleuten's criterion of manageability. Thus, one way to address the challenge of creating evaluations that are rich, meaningful, and authentic but at the same time concise, communicable, and comparable is to formulate a set of "standardized narratives" that effectively represent the types of resident stories commonly described by experienced staff. Such an effort has been elaborated in a series of studies by Bogo and colleagues[9,16,17] in the context of evaluating social work students in the field. Although promising, their work has not been replicated, nor has it been extended into the context of residency education in medicine. Therefore, in this report, we describe the insights we gained in trying to create, rank-order, and categorize a set of resident profiles that would characterize a representative range of ways that residents present themselves that staff physicians might encounter in the clinical teaching context.

Method

The method for this research was based heavily on the work of Bogo et al[16,17] in the field of social work. It involved our interviewing a set of attending physicians to collect their stories about residents, generating from these stories a set of standardized narratives, or "profiles," then establishing a ranking and scaling of these profiles based on the collective opinions of a new set of attending physicians. Details of the method are elaborated below. For all aspects of this study, IRB approval was obtained from all institutions involved.

Creation of the narratives

To create the standardized resident narratives, or "profiles," we interviewed 19 attending physicians from the departments of medicine at two participating institutions (the Faculty of Medicine, University of Toronto, and the Faculty of Health Sciences, McMaster University) in 30- to 60-minute interviews. As described more elaborately elsewhere,[18] each attending physician was asked to describe (without mentioning names) first a specific outstanding resident they had supervised, then a problematic resident, and finally an average resident. These descriptions could be about any aspect of performance, and there was no attempt to encourage discussion of any particular area or dimension of competence. However, descriptions had to be of actual residents rather than generalized opinions. Where needed, the research assistant probed participants to describe specific behaviors their resident(s) displayed. The interviews were audiotaped and transcribed verbatim, but with any potentially identifying features removed. We conducted a grounded theory analysis to uncover the underlying themes (e.g., knowledge base, work ethic) that the physicians appeared to be using in framing their discussions of the residents (see Ginsburg et al[18] for an elaboration of these dimensions).

From the 57 actual resident descriptions generated by the 19 physicians, we created 16 standardized profiles of residents, each about one-half to three-quarters of a page long. These profiles were designed to represent the full range of residents described by the supervisors by strategically combining various features and descriptions from different supervisors' stories while maintaining the language of the interviewed supervisors (but with any uniquely identifying information removed or altered to maintain anonymity of the residents discussed). All of the profiles were informed by the themes of performance identified in the grounded theory analysis, but no attempt was made to include each possible theme of performance in each profile. Rather, in an effort to maintain the narrative style of the 57 spontaneous descriptions offered and to authentically represent the way attendings discuss and describe residents, each profile is unique; each presents certain aspects of performance that are often different (and/or presented in a different order) from those presented in the other profiles.[18] Examples of two profiles can be seen in

Box 1. The full set of 16 profiles is in Supplemental Digital Box 1 (see http://links.lww.com/ACADMED/A77).

Ranking, sorting, and scaling the narratives

The 16 resulting profiles were read and reviewed critically by two or three attending faculty at each of three participating schools: the original two schools plus the Faculty of Medicine, University of British Columbia. This process was designed to ensure that the style and language of the narratives felt authentic to the participants at all three institutions (each of which has its own culture of residency education) and to identify any potential gaps in the profiles' ability to represent any particular residents that these attendings had interacted with. A number of minor changes in this process were made following this pilot review.

To identify the ranking and to establish a score for each of the 16 resident profiles, four groups of internal medicine (IM) attending physicians were recruited as participants. Recruitment took place via e-mail announcements sent out to all eligible IM attending faculty at the three institutions. Eligibility was based on having at least two years of attending experience requiring the evaluation of residents. We gave priority to faculty who taught on general IM teaching units, but we also included attendings from other primarily inpatient-oriented medical services. The 14 participants of this phase of our study were 2 groups of 4 faculty each from the University of Toronto, 1 group of 3 from McMaster University, and 1 group of 3 from the University of British Columbia.

All four groups followed the same procedure. The first phase of the procedure took approximately 45 minutes to complete. Following introductions and instructions, each participant was given a set of the 16 resident profiles, each profile on its own page and placed in a random order in the set. The participants read through all 16 profiles, making any notes on the pages that they wished. Highlighters were provided to allow participants to highlight relevant parts of the descriptions as they saw fit. Each then sorted the 16 profiles into as many groups or categories as he or she felt was necessary to represent the various levels of competence expressed in the profiles. Participants were asked to provide words or phrases that best described the level of performance represented by each group they created. Each participant was then asked to rank the profiles within each group from highest to lowest. Thus, each participant generated two "scores" for each profile. The first score was assigned based on its grouping, with a value of 1 assigned to profiles in the "best" group, a value of 2 assigned to profiles in the "next-best" group, and so on, for as many levels of competence as the individual produced. The second score was generated based on how each participant ranked each profile, from 1 (highest) to 16 (lowest).

Following a brief break, the participants in each group were brought together and shown each member's categorizations and

Box 1

Examples of 2 of 16 Narrative Profiles That Describe 19 Clinical Faculty Members' Views of Resident Performance, Two Canadian Medical Schools, 2009*

Profile of F

F is a resident who is highly organized, efficient, and energetic. F manages time well and is able to prioritize tasks effectively. F's efficiency and ability to prioritize improve even further during the rotation. F handles demanding situations well and does not appear stressed or fatigued on busy days. Even though F is efficient, this resident does not appear rushed. Clinically, F has a very strong knowledge base and a sound understanding of how different drugs and devices work, what their indications and contraindications are, the way in which different patients might respond to these, the side effect profile, and the expected benefits. Technically, F is very competent. F is comprehensive in the approach to obtaining a medical history and summarizing what is wrong with the patient. F is also inclusive in generating a management plan. When reporting on patients, this resident expresses ideas clearly and succinctly, both verbally and in writing, and is not overinclusive but communicates the core and essential data. As the rotation progresses, F gains more confidence, makes quicker decisions, and achieves greater finesse in his/her judgment.

F comes across as scholarly because this resident routinely looks up relevant information and is very good at critically appraising the literature and generating discussion about it. F often has a good sense of the applicability of the literature.

F has good relationships with other team members but can sometimes be demanding when he/she perceives that someone is not responding quickly enough to requests. Similarly, during the rotation, you have a chance to observe F's teaching skills and notice that F describes things well but can become frustrated with juniors who don't catch on as quickly. However, F has a very strong interpersonal approach with patients, is compassionate, and practices patient-centered care as well as family-centered care.

Profile of K

K seems to be able to take care of any patient that he/she is presented with. You notice that this resident really shines when things become chaotic on the service because K remains calm, composed, and efficient. K demonstrates an extensive knowledge base and is great at synthesizing the history and developing concise differential diagnoses. K appears to have an easy time with prioritizing patients' care. K is very adept at using the hospital system and is able to get things done for the patients. This resident has a good understanding of the hospital system and uses resources in an efficient manner.

As you observe K's performance, you notice that K seems to do a lot of direct, hands-on patient care. Although this is great for the patients, the junior members seem to become frustrated with K's tendency to micromanage. At times, K's enthusiasm and way of handling patient care can lead to tensions among the team members. The juniors sometimes feel that they are not given enough chances to make decisions on their own. When you provide feedback to K about this issue, K acknowledges the concern and adjusts his/her approach accordingly, but at times K will still get carried away and fall back into his/her routine of micromanaging.

In his/her interactions with allied health, K seems to be more directive and less collaborative. At times, K treats allied health as employees rather than as colleagues. However, K presents as respectful, caring, and compassionate in his/her communication with patients and families. K provides information to patients in a clear manner and takes time to explain things thoroughly. When interacting with patients, K consistently integrates the impact of the patient's social and ethnic background and adjusts his/her communication style accordingly. Patients tend to love K and often volunteer positive feedback about K.

* For all 16 profiles, see Supplemental Digital Box 1 at http://links.lww.com/ACADMED/A77.

rankings. They were then, as one large group, given access again to the set of 16 profiles, now laid out from highest ranked to lowest ranked based on the average of each individual's rankings. They were asked to negotiate these new rankings (moving profiles up or down the line as needed) and to collectively determine the cut points for different levels of competence by whatever criteria the group chose to use. All discussions occurring during these negotiations were audiotaped and later transcribed.

Following this process, each group was debriefed regarding their experiences of the process and their sense of the authenticity and comprehensiveness of the profiles in representing the range of residents they had encountered.

Analyses

The interrater reliability of the 14 participants across the four groups was calculated as both an average-rater intraclass correlation coefficient (Cronbach alpha) and a single-rater intraclass correlation (ICC) for both the categories generated and also the rankings (1–16). The intergroup reliability was also calculated as both an average-group and single-group ICC to determine whether there was evidence of differences in institutional culture.

Because participants were not limited in the number of categories they could create, the "average" group assignment for each profile across the 14 participants was generated using latent partition analysis[19] as enacted by Miller et al.[20] As described by Wiley,[19] latent partition analysis is a statistical procedure designed to combine several participants' "partitions" of a set of items into categories to generate a description of the underlying (or "latent") category structure that is common across participants. In short, by applying latent partition analysis to the categorical decisions made by each of our participants, we can estimate the "average," or common, categorical structure that is latent in those collective partitions.

Finally, discussion transcripts and field notes were analyzed by two of us (S.G., O.O.). A formal thematic analysis was not undertaken because one group's tape was lost and only field notes were available. We did, however, use the discussion notes to help understand and explore each group's process during the exercises, and we have included quotations where appropriate to support these explanations.

Results

Table 1 presents the data for the unnegotiated categories and ranking within each category made by all 14 participants (the profiles are sorted from highest average rank to lowest average

Table 1

Category Assignments and Rankings Given to Each of 16 Resident Profiles by Each of 14 Faculty Participants in Four Groups, Three Canadian Medical Schools, 2009*

	Participant assignments of category by number (and overall rank from 1 to 16) for each profile													
	Group One				Group Two				Group Three			Group Four		
	1	2	3	4	1	2	3	4	1	2	3	1	2	3
Profile identifier														
H	1 (1)	1 (2)	1 (2)	1 (1)	1 (3)	1 (2)	1 (2)	1 (2)	2 (2)	1 (4)	2 (4)	1 (3)	1 (1)	1 (2)
A	1 (3)	1 (1)	1 (1)	1 (2)	2 (5)	1 (4)	1 (4)	1 (1)	1 (1)	1 (1)	1 (1)	2 (5)	1 (3)	1 (1)
R	2 (6)	2 (3)	3 (5)	5 (10)	1 (1)	1 (1)	1 (1)	2 (5)	2 (4)	1 (2)	1 (3)	1 (1)	1 (4)	1 (6)
F	2 (4)	2 (4)	3 (7)	5 (9)	1 (2)	1 (3)	1 (3)	2 (7)	3 (7)	1 (3)	1 (2)	1 (2)	1 (6)	1 (4)
J	1 (2)	3 (6)	2 (4)	2 (4)	2 (6)	2 (6)	2 (7)	4 (11)	2 (3)	2 (3)	2 (5)	3 (6)	1 (2)	1 (5)
K	2 (5)	3 (5)	3 (6)	2 (3)	1 (4)	2 (5)	2 (5)	2 (6)	4 (10)	3 (7)	6 (14)	2 (4)	1 (5)	1 (3)
N	2 (7)	3 (7)	1 (3)	3 (5)	4 (10)	3 (7)	2 (6)	1 (3)	3 (6)	3 (9)	3 (7)	3 (7)	2 (7)	2 (7)
C	3 (11)	4 (9)	4 (10)	4 (6)	3 (7)	3 (9)	3 (11)	4 (12)	3 (5)	2 (10)	3 (6)	4 (9)	2 (11)	2 (10)
D	3 (9)	5 (12)	4 (11)	4 (7)	3 (8)	3 (8)	3 (12)	3 (9)	3 (8)	2 (6)	6 (11)	4 (13)	2 (9)	2 (12)
P	3 (8)	4 (8)	4 (8)	5 (11)	4 (11)	4 (11)	3 (9)	3 (10)	4 (11)	3 (5)	6 (15)	3 (8)	2 (10)	2 (8)
M	3 (10)	5 (11)	4 (9)	6 (14)	4 (12)	4 (12)	3 (10)	1 (4)	5 (15)	4 (8)	4 (8)	4 (10)	2 (8)	2 (11)
Q	4 (12)	4 (10)	5 (12)	6 (13)	3 (9)	4 (13)	3 (8)	3 (8)	3 (9)	4 (12)	5 (9)	4 (12)	3 (16)	2 (9)
E	4 (14)	5 (16)	6 (16)	4 (8)	5 (15)	4 (10)	4 (15)	5 (13)	4 (12)	5 (14)	6 (12)	4 (11)	3 (12)	3 (14)
L	4 (15)	5 (13)	6 (15)	5 (12)	5 (13)	5 (15)	4 (13)	6 (15)	4 (13)	5 (16)	6 (16)	5 (15)	3 (15)	3 (13)
B	4 (13)	5 (14)	6 (13)	8 (16)	5 (14)	5 (16)	4 (16)	5 (14)	4 (14)	5 (13)	6 (13)	5 (14)	3 (14)	3 (16)
G	4 (16)	5 (15)	6 (14)	7 (15)	5 (16)	5 (14)	4 (14)	7 (16)	5 (16)	5 (15)	5 (10)	5 (16)	3 (13)	3 (15)
Number of categories used by participant	4	5	6	8	5	5	4	7	5	5	6	5	3	3
Corrected rater–total correlation for rankings	0.94	0.96	0.92	0.74	0.90	0.93	0.92	0.78	0.86	0.87	0.72	0.91	0.89	0.94

* The profiles are narrative descriptions of residents that were created to represent the range of resident competence that clinical faculty might encounter during supervision. The categories represent levels of resident competence assigned to the profiles by each participant after reviewing the profiles. The table shows the sorting of the profiles based on highest to lowest average ranks given by the groups' 14 participants. See Table 2 and Chart 1 for descriptions of the categories.

rank across the 14 participants). As can be seen in the table, the number of categories used by participants ranged from a minimum of three (participants 2 and 3 in Group Four) to a maximum of eight (participant 4 in Group One). However, the modal response was five categories. Despite this range of categories used, the 14-rater alpha for the category assignment was 0.97 with a single-rater ICC of 0.81, suggesting that there was very high agreement on category level "scores" among the 14 participants. Similarly, the 14-rater alpha for the rankings themselves was 0.98 with a single-rater ICC of 0.86. To guard against the potential that these ICCs are inflated by inclusion of a wide range of profiles, we restricted the range systematically and recalculated each reliability coefficient. Whereas the single-rater ICC values were sensibly lower when we looked at restricted ranges of the profiles, the ICCs were generally similar (and still reasonably high) when looking at faculty rankings of the top eight profiles (ICC = 0.63), the bottom eight profiles (ICC = 0.64), and the middle eight profiles (ICC = 0.55).

Chart 1 presents the final negotiated ranking of the 16 profiles by each of the four groups, the categories generated during the group discussions, and the descriptions of each category offered by each group. As can be seen, there were some discrepancies in overall ranking and in the overall number of categories generated, with two groups generating

Chart 1

Final Negotiated Ranking of 16 Profiles Generated by 14 Faculty Participants in Four Groups, the Negotiated Categories Generated During the Group Discussions, and the Descriptions of Each Category Offered by Each Group, Three Canadian Medical Schools, 2009*

	Group 1		Group 2		Group 3		Group 4	
Overall Rank	Profile	Category descriptions	Profile	Category descriptions	Profile	Category descriptions	Profile	Category descriptions
1	H	outperformer, dream resident, meets/exceeds expectations, not yet ready for practice, performs well in supported environment,	R	very good except…; role models; great but not perfect; minor tune up	A	outstanding; see once in a lifetime; exemplary; keeper; leader in the field; would like to have on staff	H	excellent, high performer, outstanding, "chosen ones"
2	A	minimal guidance required, beyond typical resident	H	needed	H		A	
3	J	needs guidance/fine-tuning, identifiable areas can be improved with minor	A		R	excellent, good patient care, exceeds expectations, above average, meets all the CanMeds	R	
4	K	intervention/attention	F		F	criteria	F	
5	N	lone wolf, problems in group setting/with team building, maturity/personality issues, social boundary/rapport issues, acceptable but needs improvement in certain	K	average; solid; teachable; safe; one major domain needs work; need a little instruction	J		K	
6	R	areas, identifiable gaps but	J	and will improve	C	meets expectations but needs	J	
7	F	can be fixed	N		N	improvement, has some issues	N	acceptable, average, "could be fixed"/ "can get better"
8	P	borderline, motivational deficits, professionalism issues, bare minimum but still pass, minimal competence, doesn't go beyond minimum,	M	they pass; cruise by; safe underachiever; needs improvement;	P	but generally ok, minor deficiencies that needs to work on	P	
9	C	just tries to fly through the	D	remediation	K		M	
10	D	month	Q	possible; no fatal	D	can't translate book knowledge,	C	
11	M		C	flaw; could improve with	Q	really needs improvement, unsafe	D	
12	Q		P	direction	M		E	unacceptable, poor, unsafe, fail, below average
13	L	competence/professionalism	E	unsafe; critical	E		Q	
14	B	issues/unprofessional, bad	L	flaw; make your	B		L	
15	G	communicator, unacceptable, Jerry Springer, dangerous/	B	heart sink; personality gaps;	G	problematic, unsafe, fail, unsatisfactory	B	
16	E	unsafe, requires high degree of remediation, multiple deficits, fail, competence lacking	G	unbridgeable synapses; can't trust	L		G	

* The profiles are narrative descriptions of residents to show the range of resident competence that clinical faculty might encounter during supervision. The categories represent levels of resident competence assigned by each participant after reviewing the profiles and then negotiating to create final categories.

five categories, one group generating four categories, and one group generating only three categories. There were many consistencies, however, in the language participants used in their discussion and categorization of the resident profiles. For example, safety was a common theme in the lowest-ranked profiles, as were professionalism issues and presumed personality defects. The issue of "remediability" and response to feedback arose as important distinguishing features between the lowest-ranked profiles and those ranked slightly higher, as did the degree of supervision required. On the other hand, in the higher-ranked profiles, participants commented on readiness for practice and suitability as a colleague/consultant. There were interesting idiosyncrasies noted in some instances as well; for example, for one participant the issue of improvement took on the greatest importance and weight, so that any profile in which a resident showed evidence of improvement or response to feedback was rated relatively higher, and those that indicated no response were ranked much lower. For another participant, the issue of treating work as a "9-to-5 job" seemed to be a major issue, and those profiles were ranked relatively lower.

Table 2 presents the final negotiated rankings and categorizations for the four groups, again ordered by overall rank of the profile. Again, despite some differences in the rankings and categorizations across groups, the intergroup reliability was very high, with a single-group ICC of 0.87 (four-group alpha = 0.96) for category membership scores and a single-group ICC of 0.91 (four-group alpha = 0.98) for the overall rankings, suggesting that there is more consistency than inconsistency in the decisions made. Table 2 also presents the results of the latent partition analysis, which generated four categories overall with two profiles in the top category, five in each of the middle categories, and four in the lowest-rated category. See Chart 1 for a presentation of the profiles, categories, and descriptions.

Finally, when asked about their experience in reading and sorting the profiles, participants in each group felt that the profiles were realistic and authentically captured most, if not all, of the characteristics of the residents whom they typically supervise. There was a sense from two groups that the profiles were skewed a bit to the negative (i.e., there was a disproportionate number of profiles reflecting problematic performance). Others noted that these residents were more difficult to evaluate, so having more options in those categories was helpful. Similarly, in one group there was concern that there was no "perfect" profile—that is, a resident with no flaws or weaknesses. However, in a subsequent group this was explicitly probed, and those participants felt that there is no such thing as a resident without any deficiencies. Another issue raised was that the profiles focus on the actual behavior observed, without any hint as to the cause or context of that behavior. For example, some participants recalled residents who cause "95% of the grief" because "they have personal problems, their dog died," or "their wife was sick for two weeks, but what you're seeing is the end result of that 'background noise' that you may not be aware of." In the context of this discussion, one participant questioned whether they should mark someone differently based on the reasons for that person's deficiencies.

Despite these minor issues, faculty felt they could readily "see" or "find" their residents in the profiles provided. Several commented that the profiles "nicely captured things that are hard to evaluate." In one group, faculty discussed the idea of rank-ordering the profiles and concluded that figuring out the category a resident belonged to was more important than the rank-order, and that it was the categories that were probably the most meaningful in terms of assessment.

Discussion

Much of the effort in improving the evaluation of clinical competence over the last few decades has focused on deconstructing competence into a list of "speciated" competencies that are believed to be separately evaluable on a corresponding evaluation instrument. Embedded in this activity is the assumption that by deconstructing competence into separate, behaviorally anchored competencies, we will be able to achieve greater precision in the evaluation of each, and that the aggregation of these separate, precise evaluations will more accurately and objectively represent the overall competence of the individual being evaluated. These assumptions have been questioned on both a theoretical basis[8,14,17,18,21] and an empirical basis,[9,22] and some researchers have begun to search for approaches that more effectively capture the clinical supervisor's integrated, subjective clinical impression of a trainee in a way that offers standardization and meaningful comparison across trainees.

The work we have described here is another effort in this direction. We created a set of standardized narratives, or "profiles," of residents representing various levels of competence, using the language and descriptive style of experienced faculty telling stories about actual residents they had supervised. We then "scaled" these integrated representations of resident performance based on a consensus of clinical education experts. We would note that the scaling process in which we have engaged is strongly reminiscent of a multiple-cut-point, standard-setting process often used in more classic testing formats. In particular, the Angoff method, which asks experts to define the characteristics of the borderline performer is, in essence, asking those experts to create a "profile" at one point on the performance scale. This method has been used previously in the context of performance-based assessments.[23] Further, the "contrasting groups method" asks expert judges to make a (usually dichotomous) pass/fail decision about a number of candidates on the basis of an overall understanding of each candidate's actual performance on the test. It, too, has been applied to performance-based assessments.[24] Thus, there is some precedent for our procedure, which, in a sense, combines these two approaches by asking experts to reflect on hypothetical performances and make (in our case multiple) cut-point categorizations. Unlike typical standard-setting situations, which eventually abstract the expert categorizations into a cut point for numeric scores produced by the test, our profiles would themselves be the summative representation of the resident being assessed. The goal of our research was to establish the feasibility of such an approach to this modified "standard-setting" procedure applied to this unique form of "scale."

Table 2

Negotiated Category Assignments and Overall Rankings by 14 Faculty Participants for Each of 16 Resident Profiles by Each of Four Groups, Three Canadian Medical Schools, 2009*

	Negotiated category assignments by number (and overall rank from 1 to 16) for each profile, by group				
	Group One	Group Two	Group Three	Group Four	Descriptions of categories
Profile identifier					
H	1 (1)	1 (2)	1 (2)	1 (1)	Outstanding, excellent, exemplary
A	1 (2)	1 (3)	1 (1)	1 (2)	
R	3 (6)	1 (1)	2 (3)	1 (3)	Solid, safe
F	3 (7)	1 (4)	2 (4)	1 (4)	Needs fine tuning
J	2 (3)	2 (6)	2 (5)	1 (6)	
K	2 (4)	2 (5)	3 (9)	1 (5)	
N	3 (5)	2 (7)	3 (7)	2 (7)	
C	4 (9)	3 (10)	3 (6)	2 (10)	Borderline, bare minimum, acceptable
D	4 (8)	3 (11)	3 (8)	2 (8)	Safe underachiever
P	4 (10)	3 (9)	4 (10)	2 (11)	Cruise by, fly through, remediable
M	4 (11)	3 (8)	4 (12)	2 (9)	
Q	4 (12)	3 (12)	4 (11)	3 (13)	
E	5 (16)	4 (13)	4 (13)	3 (12)	Unacceptable, unsatisfactory, fail
L	5 (14)	4 (15)	4 (14)	3 (15)	Multiple deficits, critical flaw
B	5 (13)	4 (14)	5 (16)	3 (14)	Unsafe
G	5 (15)	4 (16)	5 (15)	3 (16)	
Number of categories used	5	4	5	3	—
Corrected group–total correlation for rankings	0.91	0.94	0.92	0.97	—

* The profiles are narrative descriptions of residents that were created to represent the range of resident competence that clinical faculty might encounter during supervision. The "number of categories used," which are of levels of resident competence that participants assigned to each profile, were negotiated within each group and were based on categories created earlier by the 14 faculty participants in the four groups after reviewing the profiles. The authors used latent partition analysis (described in the text) to generate the groups' negotiated categories of competence into four overall categories; the right-hand column presents descriptions of those four categories offered by the group's participants.

Our first important finding was that faculty participants were quite content with the descriptions of residents provided in the 16 profiles. In no case was a profile identified as unrealistic or unrepresentative of a "real" resident. Further, several participants spontaneously noted a sense that they were reading about actual residents they had worked with, some with groans because they felt they had identified "their" resident in one of the lower-rated profiles. The only noted absence in the set of profiles was "the perfect resident" who has no foibles at all. Thus, clearly, the language, style, and range of descriptions resonated well with this group of experienced clinical faculty, regardless of the medical school at which they were supervising residents.

Further, we found that when clinical faculty were asked to review this set of 16 profiles representing residents across a range of competence and to rank these narrative representations relative to each other, the faculty were quite comfortable with the task and highly consistent in the resulting rankings. Remarkably, this was true not only among faculty within a given institution but also across institutions that might be said to have quite different institutional cultures. As an interesting additional note, the interrater reliability was fairly consistent throughout the various levels of performance, as indicated by the ICCs calculated when the range of profiles was restricted.

This is not to say that there was unfailing uniformity of opinion in our participants. In particular, three participants showed slightly more idiosyncratic sort patterns, with each using a greater number of categories to distinguish levels of competence, and each demonstrating slightly lower (though still impressive) corrected item–total correlations for their rankings ($r = 0.74$, 0.78, and 0.72, respectively). Notably, however, this group did not represent a coherent alternate pattern of sorting because their sorts also correlated lower with each other than with those of participants demonstrating a more standard pattern. So, although some idiosyncratic patterns of sort were observed (e.g., one participant placed the highest value on evidence of improvement, whereas another downgraded any profile where the resident seemed to be treating the rotation like a 9-to-5 job), the pattern of responses we observed suggests that, similar to the findings of Bogo et al,[16,17] experienced clinical faculty have fairly consistent constructions of what the continuum of performance looks like even when their focus on what constitutes exemplary/poor performance might be somewhat variable. Interestingly, these opinions and constructions were present without requiring specific training or instruction. That is, attending physicians' experience in the field was sufficient to allow them to judge residents' level of performance, at least when the full range of performance levels was presented to them at once.

This finding is promising for future efforts to document meaningful evaluations of clinical competence because it implies that the problem of inconsistent evaluation may have less to do with individual faculty members' presumed idiosyncratic (or uninformed) understanding of competence and more to do with the manner in which our faculty are expected to represent this overall clinical impression using the evaluation tools with which they are provided: evaluation forms based on distinct, individual competencies, each of which must be separately rated. Faculty in our study, from three institutions, had remarkably similar conceptualizations of different levels of performance, and what those levels mean, despite having had no

specific training for this task. This suggests that the solution to improving evaluations may not lie in training faculty to observe and document better[22] or to make minor modifications to existing tools and scales. Rather, consistent with Schuwirth and van der Vleuten's[8] suggestion, our findings suggest that efforts at improving clinical performance measures might more profitably focus on fundamentally rethinking the structure of the tools we are using, to ensure that the instruments authentically represent the way in which faculty functionally conceptualize their residents' clinical competence on a day-to-day basis. What is needed now is the development of methods that will allow faculty members' subjective representations of their residents' performance to be smoothly translated into some form of documentation.

We should note that this finding of high consistency among faculty in the rankings they assigned in our current data set is an interesting contrast to previous findings from our own work that found strong idiosyncrasies among faculty in interpreting individual behaviors, particularly in the context of professionalism.[25–27] One explanation for this discrepancy is that our earlier studies explored responses to single challenging scenarios without giving the raters a larger perspective on the student with which to interpret the performance they were seeing. In the current study, the profiles that faculty were reviewing represented a summary of an entire rotation's worth of behaviors and encompassed a more comprehensive range of clinical performance. We take this to be further evidence that purely behavior-based descriptions of performance are unlikely to be the solution to the pitfalls of "objective" evaluation. Single observations of behavior are always interpreted in light of a larger set of contextual factors, and competence is most consistently understood by faculty through patterns of behavior rather than on the basis of any single observation. Thus, finding ways to represent this more integrated, synthetic, "pattern-based" interpretation of competence will be an important consideration in the development of future evaluation instruments.

It would appear, therefore, that this approach of creating authentic-sounding standardized narrative profiles of residents at various levels of performance is possible. These profiles resonated strongly with faculty attendings, who felt that the profiles captured areas of competence that are otherwise difficult to evaluate. Further they seemed to "scale" effectively and with high reliability using techniques reminiscent of other standard-setting procedures. How (and whether) the resulting "narrative-based scale" could be used in actual practice remains to be seen. Our participants certainly seemed to indicate (at least anecdotally) that they could see clear correspondences between particular profiles and actual residents they had supervised. This suggests some possibility for using such a tool as a form of summative assessment whereby faculty match their residents to one (or more) scaled profile. However, additional work would clearly be needed to assess the reliability, validity, generalizability, and feasibility issues that would have to be addressed if such an assessment were to be performed on every resident supervised. Alternatively, this set of profiles might be a mechanism to enable a supervisor who is struggling with a difficult resident to more effectively articulate some of the nature of that difficulty by finding in the matching profile some language to express the manner in which the resident is struggling. Or, perhaps this set of profiles might simply be another tool in the faculty development armament that would better prepare supervisors for interacting with (and perhaps evaluating) residents who are performing problematically. Thus, we are not promoting the use of profiles as the "solution" to the evaluation problem. There are clearly several issues remaining and hurdles to address before any such system might maximally benefit evaluation procedures in the clinical setting. However, although perhaps not a solution, we do feel that the results of the current study do offer a richer understanding of the problem of codifying clinical performance.

Funding/Support: Funding for this study was provided through a peer-reviewed grant from the Medical Council of Canada, Ottawa, Ontario, Canada.

Other disclosures: None.

Ethical approval: For all aspects of this study, IRB approval was obtained from all institutions involved.

Dr. Regehr is professor, Department of Surgery, and associate director, Centre for Health Education Scholarship, Faculty of Medicine, University of British Columbia, Vancouver, British Columbia, Canada.

Dr. Ginsburg is associate professor of respirology and internal medicine and scientist, Wilson Centre for Research in Education, University Health Network, Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada.

Dr. Herold is assistant professor, Department of Medicine, University of Toronto, and scientist, Wilson Centre for Research in Education, University Health Network, Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada.

Dr. Hatala is clinical associate professor, Department of Medicine, University of British Columbia, Vancouver, British Columbia, Canada.

Dr. Eva is associate professor, Department of Medicine, and senior scientist, Centre for Health Education Scholarship, Faculty of Medicine, University of British Columbia, Vancouver, British Columbia, Canada.

Ms. Oulanova is a PhD candidate, Department of Adult Education and Counseling Psychology, Ontario Institute for Studies in Education, University of Toronto, Toronto, Ontario, Canada.

References

1 Chaudhry SI, Holmboe ES, Beasley BW. The state of evaluation in internal medicine residency. J Gen Intern Med. 2008;23:1010–1015.
2 Turnbull J, van Barneveld C. Assessment of clinical performance: In-training evaluation. In: Norman GR, van der Vleuten CPM, Newble DI, eds. International Handbook of Research in Medical Education. London, UK: Kluwer Academic Publishing; 2002:793–810.
3 Gray JD. Global rating scales in residency education. Acad Med. 1996;71(10 suppl): S55–S63.
4 Dudek NL, Marks MB, Regehr G. Failure to fail: The perspectives of clinical supervisors. Acad Med. 2005;80(10 suppl):S84–S87.
5 Cohen G, Blumberg P, Ryan N, Sullivan P. Do final grades reflect written qualitative evaluations of student performance? Teach Learn Med. 1993;5:10–15.
6 Speer AJ, Solomon DJ, Ainsworth MA. An innovative evaluation method in an internal medicine clerkship. Acad Med. 1996; 71(10 suppl):S76–S78.
7 Holmboe ES, Hawkins RE, Huot SJ. Effects of training in direct observation of medical residents' clinical competence: A randomized trial. Ann Intern Med. 2004;140:874–881.
8 Schuwirth L, van der Vleuten CPM. Merging views on assessment. Med Educ. 2004;38: 1208–1210.
9 Regehr G, Bogo M, Regehr C, Power R. Can we build a better mousetrap? Improving the measures of practice performance in the field practicum. J Soc Work Educ. 2007;43:327–343.
10 Hodges B. Medical education and the maintenance of incompetence. Med Teach. 2006;28:690–696.
11 Frank JR. The CanMEDS 2005 Physician Competency Framework. Ottawa, Ontario, Canada: Royal College of Physicians and Surgeons of Canada; 2005.

12 The ACGME's Outcomes Project. http://www.acgme.org/acwebsite/RRC_280/280_corecomp.asp. Accessed February 2, 2012.

13 Jones MD Jr, Rosenberg AA, Gilhooly JT, Carraccio CL. Perspective: Competencies, outcomes, and controversy—Linking professional activities to competencies to improve resident education and practice. Acad Med. 2011;86:161–165.

14 Leach DC. ACGME e-Bulletin. August 2006. http://www.acgme.org/acWebsite/bulletin/bu_index.asp. Accessed December 14, 2011.

15 Hurwitz B. Narrative and the practice of medicine. Lancet. 2000;356:2086–2089.

16 Bogo M, Regehr C, Power R, Hughes J, Woodford M, Regehr G. Toward new approaches for evaluating student field performance: Tapping the implicit criteria used by experienced field instructors. J Soc Work Educ. 2004;40:417–426.

17 Bogo M, Regehr C, Woodford M, Hughes J, Power R, Regehr G. Beyond competencies: Field instructors' descriptions of student performance. J Soc Work Educ. 2006;42:579–593.

18 Ginsburg S, McIlroy J, Oulanova O, Eva KW, Regehr G. Toward authentic clinical evaluation: Pitfalls in the pursuit of competency. Acad Med. 2010;85:780–786.

19 Wiley DE. Latent partition analysis. Psychometrika. 1967;32:183–193.

20 Miller D, Wiley DE, Wolfe R. Categorization methodology: An approach to the collection and analysis of certain classes of qualitative information. Multivariate Behav Res. 1986; 21:135–167.

21 van der Vleuten CP, Norman GR, Graaff E. Pitfalls in the pursuit of objectivity: Issues of reliability. Med Educ. 1991;25:110–118.

22 Lurie SJ, Mooney CJ, Lyness JM. Measurement of the general competencies of the Accreditation Council for Graduate Medical Education: A systematic review. Acad Med. 2009;84:301–309.

23 Norcini J, Stillman P, Sutnick A, et al. Scoring and standard setting with standardized patients. Eval Health Prof. 1993;16:322–332.

24 Clauser BE, Clyman SG. A contrasting-groups approach to standard setting for performance assessments of clinical skills. Acad Med. 1994;69(10 suppl):S42–S44.

25 Ginsburg S, Regehr G, Lingard L. Basing the evaluation of professionalism on observable behaviors: A cautionary tale. Acad Med. 2004; 79(10 suppl):S1–S4.

26 Ginsburg S, Lingard L, Regehr G, Underwood K. Know when to rock the boat: How faculty rationalize students' behaviors. J Gen Intern Med. 2008;23:942–947.

27 Ginsburg S, Regehr G, Mylopoulos M. From behaviours to attributions: Further concerns regarding the evaluation of professionalism. Med Educ. 2009;43:414–425.

Standardized Narratives of GIM Residents

Glenn Regehr, Shiphra Ginsburg, Jodi Herold, Rose Hatala, Kevin Eva, Olga Oulanova

The following 16 narratives were developed as part of a program of research funded by the Medical Council of Canada.

Construction and scaling of these narratives is described in detail in Regehr et al (2012).

The resulting ranks and categorizations of the narratives can be found in the original publication, but in brief the results are as follows:

Narrative Label	Rank (1=best)	Category (1=highest)	Category Labels
H	1	1	Outstanding, Exemplary
A	2	1	
R	3	2	Solid, Safe, But Needs Fine-tuning
F	4	2	
J	5	2	
K	6	2	
N	7	2	
C	8	3	Safe Underachiever, Remediable
D	9	3	
P	10	3	
M	11	3	
Q	12	3	
E	13	4	Unsafe, Unacceptable, Unsatisfactory
L	14	4	
B	15	4	
G	16	4	

For further details or information, please contact:

Dr. Shiphra Ginsburg
Mount Sinai Hospital
433-600 University Ave.
Toronto, Ontario, Canada, M5G 1X5
e-mail: shiphra.ginsburg@utoronto.ca.

Regehr G, Ginsburg S, Herold J, Hatala R, Eva K, Oulanova O. Using "standardized narratives" to explore new ways to represent faculty opinions of resident performance. Acad Med. 2012 Apr;87(4):419-27.

Vignette A

From the start, A demonstrates keen interest by taking every opportunity to read relevant literature, learn from complicated cases, and ask many questions at seminars and teaching sessions. Although at the beginning A's knowledge base relevant to this rotation had some gaps, A undergoes significant growth in this domain. When a question comes up that A does not know the answer to, A takes the initiative to read up in the area and seems to want to understand more deeply what's going on. A appears excited, curious, and enthusiastic. A welcomes challenges and regards difficult clinical problems as learning opportunities, rather than threats. As A's knowledge base grows, A also effectively applies theoretical knowledge and recent literature to individual cases.

Throughout the rotation, A is always on time, and does not leave until everything is under control. Even when things are very busy on the service, A seems to know what's going on down to the last detail and always appears in control of the patient data. A has good organizational skills and assigns work load effectively, taking into account limitations of more junior residents. When working with a weaker team member, A provides appropriate supervision and follows up to make sure that nothing is missed.

During rounds you observe A interact with patients and seldom feel that you need to interject. A explains things carefully to patients, relates information in a clear and comprehensive manner and establishes a warm and empathic connection with patients and families. In communicating with patients, A comes across as patient, caring, and knowledgeable.

A communicates effectively with other house staff. A treats nurses and other team members with respect and is generally well liked by the allied health staff.

A exudes warmth and caring for everyone around, including patients, patient families, staff, and other residents. A also demonstrates effective non verbal communication such as making eye contact with the patient when it is appropriate and using therapeutic touch. In your opinion, this resident will need to continue to improve in terms of the knowledge base, but you have no doubt that this will happen.

Regehr G, Ginsburg S, Herold J, Hatala R, Eva K, Oulanova O. Using "standardized narratives" to explore new ways to represent faculty opinions of resident performance. Acad Med. 2012 Apr;87(4):419-27.

Vignette B

B's knowledge base is fine and B spends a lot of time in the library reading the literature. B can usually list differentials for common medical problems. However, as you observe B's clinical skills, you worry about B's performance and the translation of B's 'book knowledge' to the bedside. For example, reviewing B's approach to collecting patient information, you become concerned. In taking history and carrying out a physical assessment B misses a significant amount of relevant information; it appears that B either does not know what questions to ask or does not make an effort to find the answers (e.g., B does not go back in medical records, or call the family doctor to obtain missing information). When you point out these concrete deficiencies to B, B always has a reason for why things did not get done (usually the reason involves other people, or the nature of the situation). B does not acknowledge these deficiencies and does not seem to have insight into his/her problematic performance, or to take responsibility for it. B becomes defensive when you offer feedback on his/her performance and does not appear to integrate your suggestions for improvement. After several weeks of going through similar cases B has made little progress. Overall, you do not have confidence in B's assessment skills.

B does not recognize other people's expertise and does not seem to acknowledge the opinions of allied health. Juniors come to you as the attending for help with problems that you would expect B should be able to handle.

B is very punctual and often stays late to ensure that the patients are stable and things are under control. B puts in the time and seems to be trying but just doesn't seem to be able to put it all together and does not appear to understand his/her deficiencies.

Regehr G, Ginsburg S, Herold J, Hatala R, Eva K, Oulanova O. Using "standardized narratives" to explore new ways to represent faculty opinions of resident performance. Acad Med. 2012 Apr;87(4):419-27.

Vignette C

Throughout the rotation C works hard and demonstrates responsibility for the patients. Although sometimes C misses a few things on history or physical, on the whole you are not concerned, and you consider this resident's performance to be at a safe level. Importantly, C recognizes gaps in his/her knowledge base and reads the necessary literature. However, C seldom goes beyond what is required for that particular case. Your impression is that C seeks to fill these knowledge gaps because it is expected, and not because C feels passionate about acquiring knowledge in an area, or enthusiastic about a challenging patient. C sometimes seems to treat work like a job that needs to get done but overall does do a good job.

Although C always treats patients with respect, C does not radiate warmth for patients and as a result is sometimes not perceived as being very involved or interested. While C may not place great value on the role of the patient's family and relationship-building, C certainly does not dismiss these elements. C maintains a high level of professionalism in interactions (i.e. C's grooming, language, and overall presentation are always appropriate and C is punctual) but does not establish a deep connection with his/her patients.

C's written communication skills are very good and C's clinical notes are concise yet contain all the essential information. While at the start C tends to spend a lot of time writing notes, C eventually improves his/her time management.

C gets along with other members of the team and has a general sense of camaraderie particularly with other medical colleagues. Overall you believe that C can improve in some areas but at the same time you do not have serious concerns about C's performance. You realize, however, that you will likely not remember much about this resident once he/she finishes the rotation.

Regehr G, Ginsburg S, Herold J, Hatala R, Eva K, Oulanova O. Using "standardized narratives" to explore new ways to represent faculty opinions of resident performance. Acad Med. 2012 Apr;87(4):419-27.

Vignette D

D's knowledge in traditional domains is solid and this resident works hard in the clinic and gets most tasks done in an efficient manner. However, you notice that D occasionally misses things on history or physical exam. Although you do not consider this to be a serious problem, D's work is sometimes not quite at the level that you'd expect of a resident at this stage. D does not seem to have any academic interest that is over and above what is required for exams. As a result, D is sometimes too simplistic in his/her decision making, and often fails to consider all the possibilities. When you tell this resident, "Go and read about this" D does it but in a brief way, not attending to detail and without demonstrating curiosity or enthusiasm. It appears that D's other responsibilities make it difficult for this resident to read extra and put in more time into his/her learning than what is minimally required.

On a personal level, D is a very pleasant individual, and has natural charm which makes it easy for D to connect with others. You often hear positive feedback about this resident from patients and the team and people seem to really enjoy interacting with D. D has a very strong bedside manner and appears to genuinely enjoy talking to patients.

When presented with opportunities to teach more junior residents, D does not take up these chances to demonstrate leadership skills and does not seem interested in doing or learning extra. This resident is reluctant to take on extra responsibilities and is often eager to leave work by 5pm.

Although you have some concerns about D's clinical skills, this resident has a warm and agreeable personality, and you find that you sometimes do not notice or dismiss the minor errors that this resident makes.

Regehr G, Ginsburg S, Herold J, Hatala R, Eva K, Oulanova O. Using "standardized narratives" to explore new ways to represent faculty opinions of resident performance. Acad Med. 2012 Apr;87(4):419-27.

Vignette E

E comes across as a very bright person from an intelligence point of view. From the start of the rotation, E demonstrates extensive knowledge of relevant literature and in group discussions quotes the most recent publications. But you are somewhat concerned that E tends to spend a lot of time reading in the library instead of trying to learn by spending more time in the clinic and seeing complex cases.

E's communication with patients is clear and he/she says all the right words. But E does not seem to connect with patients on a personal level and does not demonstrate a lot of empathy or compassion. E treats interactions with patients and clinical work overall as a job and does not appear to go above what is required to provide minimally adequate level of care. E presents as professional in his/her conduct and appearance.

As you observe E's interactions with the rest of the team, you become very concerned about E's interpersonal skills. E does not seem to be able to relate to the nurses, or to the other residents. You also receive feedback that E has at times been rude with the staff. You are told that this resident does not listen to other team members and only does what he/she believes is right instead of following directions of and taking into account the opinions of others. When the more senior residents suggest that something needs to be done, E fails to follow their directions, and simply carries on in the way that he/she wants to. E also appears to be overconfident at times and has a poor sense of his/her own limitations. E never calls for help because this resident appears to think that he/she knows more than everybody else.

E's interpersonal skills and E's interactions with others do not improve. This creates an overall tense atmosphere on the rotation. Indeed, there is a sense of relief among the team members when this resident's rotation is over.

Regehr G, Ginsburg S, Herold J, Hatala R, Eva K, Oulanova O. Using "standardized narratives" to explore new ways to represent faculty opinions of resident performance. Acad Med. 2012 Apr;87(4):419-27.

Vignette F

F is a resident who is highly organized, efficient, and energetic. F manages time well and is able to prioritize tasks effectively. This ability to prioritize and F's efficiency improve even further during the rotation. F handles demanding situations well and does not appear stressed or fatigued on busy days. Even though F is efficient, this resident does not appear rushed. Clinically, F has a very strong knowledge base and a sound understanding of how different drugs and devices work, what their indications and contraindications are, the way in which different patients might respond to these, the side effect profile, and the expected benefits. Technically F is very competent. F is comprehensive in the approach to obtaining a medical history and summarizing what is wrong with the patient. F is also inclusive in generating a management plan. When reporting on patients, this resident expresses ideas clearly and succinctly both verbally and in writing, and is not over inclusive but communicates the core and essential data. As the rotation progresses, F gains more confidence, makes quicker decisions, and achieves greater finesse in his/her judgment.

F comes across as scholarly because this resident routinely looks up relevant information and is very good at critically appraising the literature and generating discussion about it. F often has a good sense of the applicability of the literature.

F has good relationships with other team members, but can sometimes be demanding when he/she perceives that someone is not responding quickly enough to requests. Similarly, during the rotation you have a chance to observe F's teaching skills and notice that F describes things well, but can become frustrated with juniors who don't catch on as quickly. However, F has a very strong interpersonal approach with patients, is compassionate, and practices patient-centered care as well as family-centered care.

Regehr G, Ginsburg S, Herold J, Hatala R, Eva K, Oulanova O. Using "standardized narratives" to explore new ways to represent faculty opinions of resident performance. Acad Med. 2012 Apr;87(4):419-27.

Vignette G

G presents as reserved and quiet, to the point of seeming disengaged in rounds. G demonstrates a weak knowledge base, which manifests in part in G's superficial answers and lack of participation during teaching sessions. G struggles with prioritizing and cannot adequately formulate what is wrong with the patient, and which problems need attention first. When it comes to problem management, G is often unsure about which test to order, or which meds to give. G doesn't know how to perform some procedures that you would expect a resident at this level to be very familiar with.

You are especially worried about this resident because G does not ask questions or seek out guidance when G does not know or understand something. For example, G does not communicate with the team when patients are really sick. This makes you concerned about G's judgment and ability to assess a complex situation. As a result of G's lack of insight into his/her own performance and G's tendency to not call for help, this resident occasionally creates dangerous situations that put patients at risk.

The feedback from nurses and other allied health staff is that they do not have confidence in G's knowledge and abilities. In communicating with other staff G is often brief and does not maintain appropriate eye contact.

With patients and families G is polite and respectful but also vague and often unhelpful in terms of answering their questions and communicating information. G communicates the minimal necessary information to families and they appear dissatisfied after talking with G. You often feel that this resident is a bad reflection on you as the attending.

Overall, G does not seem to enjoy the work and you impression is that G does not really want to be there or learn. While G is not dismissive of criticism, there appears to be little improvement in this resident's performance as a result of feedback.

Regehr G, Ginsburg S, Herold J, Hatala R, Eva K, Oulanova O. Using "standardized narratives" to explore new ways to represent faculty opinions of resident performance. Acad Med. 2012 Apr;87(4):419-27.

Vignette H

H seems to get along with everyone. This resident listens to the opinions of juniors in a respectful way and provides them with helpful and objective feedback. H has a great sense of humour which makes people feel relaxed around this resident. H often takes initiative with juniors, embracing any opportunity to teach or provide guidance to more junior members of the team. H is always available and approachable so juniors feel comfortable coming to this resident for help.

When it comes to patient care, H is proactive in anticipating problems and is prepared to deal with complicated situations in a calm and efficient manner. H effectively assesses difficult cases and prioritizes well what needs to get done and in what order. While H's knowledge is not as extensive as some other residents, this knowledge is pragmatic and well applied.

H's sense of humour and warm personality make it easy for this resident to establish strong rapport with patients. H is able to connect with patients on a non-medical level because he/she is well-rounded and has interests outside professional ones. H appears to be interested in patients as people which facilitates alliance building. Patients seem comfortable around H.

H demonstrates a profound sense of responsibility and your initial impression about this resident is that H can be trusted. As the rotation progresses, you are convinced even further that H can be relied upon, and that if something needs to be done, it will be done. You trust H's evaluations because when this resident does not know something, he/she says so. H can assess his/her own abilities well and is not afraid to ask for help when the situation demands a more senior and experienced individual.

While H comes to the rotation reasonably confident, with experience, H develops in being more sure of him/herself and becomes more confident in his/her own judgment. You believe that H is the sort of doctor that you would want your family to go to.

Regehr G, Ginsburg S, Herold J, Hatala R, Eva K, Oulanova O. Using "standardized narratives" to explore new ways to represent faculty opinions of resident performance. Acad Med. 2012 Apr;87(4):419-27.

Vignette J

Overall, J is a resident who can be trusted and relied on as a safe pair of hands. You are not concerned about this resident's knowledge base or clinical skills, which are right at the level you'd expect for this stage of training. You trust J's assessments and do not feel the need to check up on his/her work.

Throughout the rotation J works hard and is reasonably efficient. J never expresses frustration about staying late if there is work that needs to be done, and always follows up on assigned tasks. J doesn't always spontaneously read up on cases, but will do so if asked or reminded.

J is a fine teacher and has sound leadership skills. J is patient and spends a lot of time teaching more junior residents. They seem to feel comfortable in coming to J with questions since J is approachable and willing to help. J establishes friendly and professional relationships with other team members and you never get negative feedback about this resident. J is polite and respectful but tends to be quiet in group discussions. Sometimes J has to be encouraged to share opinions with the group.

J really makes the effort to connect with patients on a personal level, and comes across as very empathic, caring and respectful. This resident's calm and supportive manner is much appreciated by the patients on the team.

Regehr G, Ginsburg S, Herold J, Hatala R, Eva K, Oulanova O. Using "standardized narratives" to explore new ways to represent faculty opinions of resident performance. Acad Med. 2012 Apr;87(4):419-27.

Vignette K

K seems to be able to take care of any patient that he/she is presented with. You notice that this resident really shines when things become chaotic on the service because K remains calm, composed, and efficient. K demonstrates an extensive knowledge base, and is great at synthesizing the history and developing concise differential diagnoses. K appears to have an easy time with prioritizing patients' care. K is very adept at using the hospital system and is able to get things done for the patients. This resident has a good understanding of the hospital system and utilizes resources in an efficient manner.

As you observe K's performance, you notice that K seems to do a lot of direct hands on patient care. Although this is great for the patients, the junior members seem to become frustrated with K's tendency to micromanage. At times K's enthusiasm and way of handling patient care can lead to tensions among the team members. The juniors sometimes feel that they are not given enough chances to make decisions on their own. When you provide feedback to K about this issue, K acknowledges the concern and adjusts his/her approach accordingly, but at times K will still get carried away and fall back into his/her routine of micromanaging.

In his/her interactions with allied health, K seems to be more directive and less collaborative, almost treating them as employees. However, K presents as respectful, caring, and compassionate in his/her communication with patients and families. K provides information to patients in a clear manner and takes time to explain things thoroughly. When interacting with patients, K consistently integrates the impact of the patient's social and ethnic background and adjusts his/her communication style accordingly. Patients tend to love K and often volunteer positive feedback about K.

Regehr G, Ginsburg S, Herold J, Hatala R, Eva K, Oulanova O. Using "standardized narratives" to explore new ways to represent faculty opinions of resident performance. Acad Med. 2012 Apr;87(4):419-27.

Vignette L

L is a bright enough individual and generally you are not particularly concerned about L's clinical skills. However, you worry about L's interactions with others. Nursing staff and other allied health members complain because L treats them with arrogance and does not seem open to their suggestions. While other team members often know much more than L about particular areas, you sense that they feel devalued by L because of his/her attitude. At times you feel like L is dismissing you when you offer suggestions for treatment plans or during procedures and chooses to carry on in his/her own way. Sometimes L takes short cuts and it feels like L presumes too much so you are not always sure you can trust this resident's assessments. You find you check up on L's patients more than you expected to for his/her level of training. Although most of the time you believe that L's clinical skills are fine, sometimes you do wonder whether L is 'safe.'

L's behaviour in the clinic is disconcerting to you as you really worry about L's lack of insight into his/her own behavior. When you provide feedback, L seems to listen but nothing really changes in this resident's performance. You worry that because this resident does not appear to be receptive to feedback, L will not actually fix his/her problem areas.

L is punctual but appears to treat work as a '9-5 job' and is often eager to leave at the end of the day. L appears impatient when forced to stay behind because things are not under control. L does not take time to teach junior residents and does not seek out opportunities to demonstrate leadership skills. When L does conduct teaching sessions, L appears to lack patience and you see that the junior members avoid going to L for help.

Regehr G, Ginsburg S, Herold J, Hatala R, Eva K, Oulanova O. Using "standardized narratives" to explore new ways to represent faculty opinions of resident performance. Acad Med. 2012 Apr;87(4):419-27.

Vignette M

M has good assessment and diagnostic skills but has a hard time coming up with realistic management plans and struggles with prioritizing what is most important in a particular situation. In addition, M has a poor understanding of how the hospital system works and often seems unaware of the resources available and how to take advantage of them.

When it comes to describing a clinical situation to other team members, M has difficulties with communication. Especially at the start of the rotation, M's statements do not follow a logical sequence and allied health members sometimes express frustration about their interactions with this resident. Although M comes across as a well-meaning and caring individual in his/her interactions with patients, patients sometimes have a confused look after communicating with M. When you observe M explain something to patients, you often feel the need to interject as you are concerned that M does not deliver information in a coherent way.

Throughout the rotation M demonstrates a strong commitment to improving and seems to genuinely want to do a better job. M appears to be aware of his/her deficiencies and is open to feedback and actually actively seeks out feedback. When M is provided with constructive criticism, there is some improvement in this resident's performance. At the end of the rotation your impression is that this resident still has a way to go but M's oral presentations and discussions of patients become more coherent and there is a better flow to them. M's ability to prioritize improves as a result of extensive feedback that this resident receives during the rotation.

Regehr G, Ginsburg S, Herold J, Hatala R, Eva K, Oulanova O. Using "standardized narratives" to explore new ways to represent faculty opinions of resident performance. Acad Med. 2012 Apr;87(4):419-27.

Vignette N

N has a very strong sense of responsibility for patients. For example, when N hands a patient over to a different service, N stays as long as required to ensure that the patient is stable and that everything is really under control. N appears to genuinely care about the well-being of patients and often stays late and comes in on week-ends to follow up on patients. Although this is great for patient care, you are sometimes concerned about this resident burning out. You also worry that N can at times become too involved in care for a particular patient. Patients tend to really connect with N and form a close bond with this resident.

N is not always up to date on every latest study, but does demonstrate a solid knowledge base, and when a challenging case comes up, N always ensures that he/she acquires the appropriate knowledge to do the job well and will consult with more senior team members, or read up on his/her own.

At the start of the rotation N seems to lack confidence in his/her abilities and can be too self-deprecating. N is often quiet in group discussions and appears reserved in interactions with house staff and other residents. While N gets along with everyone, N maintains a distance and as a result does not appear to really connect with other team members at a deeper level. When N believes that he/she has made a mistake, N readily comes forward with the error and volunteers to take corrective action. N is very receptive to criticism and seems to take feedback seriously and try to integrate it. For example, at the start of the rotation N tends to dress overly casually but is quick to improve when provided feedback on this issue. N's self-confidence improves during the rotation as N comes to trust his/her own judgment and clinical skills more readily.

Regehr G, Ginsburg S, Herold J, Hatala R, Eva K, Oulanova O. Using "standardized narratives" to explore new ways to represent faculty opinions of resident performance. Acad Med. 2012 Apr;87(4):419-27.

Vignette P

While P has solid clinical skills and is generally very safe and capable, P does not demonstrate a strong commitment to the team and does not act as a leader. For example, P looks after his/her own patients but seems to think that patients not specifically assigned to him/her are somebody else's problem. Although P is quite knowledgeable, other team members avoid going to P for help unless they do not have any other choice. Juniors do not seem comfortable interacting with P as P can be brief and impatient with them. P gets along fine with allied health and you do not hear negative feedback about this resident from nurses and other house staff.

When it comes to patient care, P is conscientious and does the tasks assigned. But you have a sense that there is an upper limit to what P could be asked to do without stretching the system and that there is a clear limit to this resident's ability. When you push P beyond this limit, you realize that things are not getting done and that this resident kind of 'maxes out' and gets frazzled. P seems to need longer to work out an issue and think things through. P does not seem to be able to quickly assess the situation, make decisions, and move on to the next issue. If there is a problem P lets you know and you feel that you can generally rely on P's assessment of his/her own ability. You trust that P will call for help when faced with a case that proves beyond his/her skill level.

When you ask this resident what he/she thinks is going on with a patient, P sticks with the just one or two tried and true things and is not comfortable thinking outside the box, coming up with good alternatives, or thinking broader.

When provided with feedback, P appears open to suggestions and constructive criticism but is slow to respond. Sometimes you need to continue providing additional feedback for P to change his/her approach. With time, P does integrate this feedback and slowly improves by the end of the rotation.

Regehr G, Ginsburg S, Herold J, Hatala R, Eva K, Oulanova O. Using "standardized narratives" to explore new ways to represent faculty opinions of resident performance. Acad Med. 2012 Apr;87(4):419-27.

Vignette Q

You get the sense that Q is happier as an observer than a participant. Q seldom takes initiative and often hesitates and waits for others to take the lead. Q's level of independence and confidence appears to be lower than what you would expect of a resident at this stage of training. When Q is asked a question, Q will answer it and when a team member asks Q for help, this resident will be thorough in offering suggestions and ideas. However, Q does not spontaneously offer to help, or engage anybody on the team in a conversation. Allied health and other staff seem to think that Q is 'nice' but do not offer much feedback about this resident. Your sense is that this resident will not be remembered after the rotation is over.

Q prefers if you run family meetings, so that he/she can observe and get the experience. When there is a chance to talk with family members who are very upset and anxious, Q falls short. Q is polite in his/her interactions with patients but seems unable to communicate deep compassion and really empathize with the patent and the family.

In terms of patient care Q is safe, with good clinical skills and solid knowledge. Q is efficient, able to prioritize, and manage his/her time well on busy days. However, by 5 o'clock Q has disappeared and generally does not seem to take much ownership of the patients.

Regehr G, Ginsburg S, Herold J, Hatala R, Eva K, Oulanova O. Using "standardized narratives" to explore new ways to represent faculty opinions of resident performance. Acad Med. 2012 Apr;87(4):419-27.

Vignette R

R is a very good resident clinically, with an excellent knowledge base and clinical skills for this level of training. R is very diligent and reliable with regards to patient care and is always on top of the details of his/her patients. In fact, R often knows a lot of details about other team members' patients as well, and will offer up opinions about their management while on rounds. Some team members have appeared a bit annoyed at these efforts, as R doesn't always wait before jumping in during group discussions. On occasion you have had to rein R in so that others have a chance to participate.

R is extremely keen and enthusiastic, and always offers to help others by taking on more work. The nurses and other allied health care professionals really enjoy working with R as their suggestions are usually followed without delay.

Even when things are busy, R always takes the opportunity to teach, and this is appreciated by the students on the team.

This resident communicates very well with patients, who spontaneously provide excellent feedback about the attention they get from R.

Regehr G, Ginsburg S, Herold J, Hatala R, Eva K, Oulanova O. Using "standardized narratives" to explore new ways to represent faculty opinions of resident performance. Acad Med. 2012 Apr;87(4):419-27.