

STUDENT TEACHING AND EVALUATION

Daily encounter cards facilitate competency-based feedback while leniency bias persists

Glen Bandiera, MD, MEd;^{*†} David Lendrum, MD[‡]

ABSTRACT

Objective: We sought to determine if a novel competency-based daily encounter card (DEC) that was designed to minimize leniency bias and maximize independent competency assessments could address the limitations of existing feedback mechanisms when applied to an emergency medicine rotation.

Methods: Learners in 2 tertiary academic emergency departments (EDs) presented a DEC to their teachers after each shift. DECs included dichotomous categorical rating scales (i.e., “needs attention” or “area of strength”) for each of the 7 CanMEDS roles or competencies and an overall global rating scale. Teachers were instructed to choose which of the 7 competencies they wished to evaluate on each shift. Results were analyzed using both staff and resident as the units of analysis.

Results: Fifty-four learners submitted a total of 801 DECs that were then completed by 43 different teachers over 28 months. Teachers’ patterns of selecting CanMEDS competencies to assess did not differ between the 2 sites. Teachers selected an average of 3 roles per DEC (range 0–7). Only 1.3% were rated as “needs further attention.” The frequency with which each competency was selected ranged from 25% (Health Advocate) to 85% (Medical Expert).

Conclusion: Teachers chose to direct feedback toward a breadth of competencies. They provided feedback on all 7 CanMEDS roles in the ED, yet demonstrated a marked leniency bias.

Key words: medical education, daily encounter cards, feedback, evaluation, practice-based assessment

RÉSUMÉ

Objectif : Nous avons cherché à déterminer si le nouveau concept de fiche de rencontre quotidienne (FRQ), axée sur les compétences et conçue pour réduire au minimum le biais d’évaluation et maximiser l’évaluation indépendante des compétences, pourrait remédier aux limites des mécanismes de rétroaction existants appliqués à une rotation en médecine d’urgence.

Méthodes : Les apprenants dans deux services d’urgence de soins tertiaires à vocation universitaire ont présenté une FRQ à leurs enseignants après chaque quart de travail. Les fiches comportaient, d’une part, des échelles d’évaluation à variables catégorielles dichotomiques (c.-à-d. « attention à apporter » ou « point fort ») pour chacune des sept compétences CanMEDS et, d’autre part, une échelle d’évaluation générale. On a demandé aux enseignants de choisir une de ces sept

From the *Department of Medicine and the Wilson Centre for Research in Education, University of Toronto, Toronto, Ont., the †Department of Emergency Medicine, St. Michael’s Hospital, Toronto, Ont., and the ‡University of Toronto FRCP(EM) Residency Program, Toronto, Ont.

Received: Mar. 17, 2007; revisions received: Aug. 15, 2007; accepted: Aug. 15, 2007

This article has been peer reviewed.

CJEM 2008;10(1):44-50

compétences et de l'évaluer pendant chaque période de travail. Puis, les résultats ont été analysés, les enseignants et les résidents constituant les unités d'analyse.

Résultats : Sur une période de 28 mois, 54 apprenants ont soumis 801 FRQ, qui ont été remplies par 43 enseignants. Le schéma selon lequel les enseignants ont choisi les compétences CanMEDS à évaluer était similaire dans les deux services d'urgence. Les enseignants ont choisi en moyenne trois rôles par FRQ (gamme de 0 à 7). La note « attention à apporter » n'a été accordée que sur seulement 1,3 % des fiches. La fréquence selon laquelle chaque compétence a été sélectionnée variait de 25 % (promoteur de la santé) à 85 % (expert médical).

Conclusion : Les enseignants ont choisi de donner de la rétroaction à l'égard d'une variété de compétences. Ils ont fourni des commentaires sur chacune des sept compétences CanMEDS dans un service d'urgence et ont par ailleurs fait montre d'indulgence marquée dans leur évaluation.

Introduction

There are numerous challenges to providing regular, quality feedback to learners in the emergency department (ED). Multiple simultaneous demands on a teacher's time, the unpredictability of emergency medicine (EM) practice, the mismatching of learners' and teachers' schedules, shift work and the heterogeneity of learner characteristics impede meaningful feedback sessions.¹⁻⁴ The integrity of residency programs and steady learner progress toward competency rely on a feedback system that minimizes the effects of the aforementioned barriers to feedback for learners in the ED.

New competency-based medical education models require learners to demonstrate progression toward achieving multiple goals. One such model is the Royal College of Physicians and Surgeons of Canada (RCPSC) CanMEDS framework, which includes 7 different physician roles.⁵ This model requires residents to receive feedback on their performance with respect to all 7 roles during the course of their residency. Direct observation methods remain the keystone for the assessment of postgraduate medical learners and are commonly regarded as the best methods to assess multiple competencies in a clinical practice environment.⁶ Previous literature, however, has suggested that most clinical teachers can assess at best 2 or 3 separate aspects of a learner's performance following a single encounter.^{7,8} The simultaneous assessment of all 7 CanMEDS roles would be problematic. Developing a summative assessment of learners following an ED rotation is particularly challenging as they often work with multiple teachers during their rotation, each of whom may provide a fragmented assessment of the 7 CanMEDS roles.

In most cases, a teacher's assessment of a learner's performance of multiple roles will be determined by 2 or 3 main overriding perceptions of the learner, a source of error known as the halo effect (if the impression is swayed by a positive experience) or millstone effect (if the

impression is swayed by a negative experience).⁷⁻¹⁶ Another common source of error in direct observation assessment is the leniency or range restriction effect, wherein teachers provide "inflated" or overly favourable assessments.⁹⁻¹⁶ Teachers may be reluctant to provide poor feedback to learners or identify areas requiring attention because of the fear of retribution, concern over the need to justify the assessment and the emotional difficulty associated with providing what is often considered to be "bad news."⁹ Therefore, teachers, in effect, tend to use only the upper or positive portions of an evaluation scale.

Daily encounter cards (DECs) are 1 method used to regularly assess learner performance in a dynamic clinical setting. This method has been successfully used in multiple ambulatory and inpatient environments.¹⁷⁻²¹ It is not clear how well teachers assess multiple competencies using these cards, nor is it clear how their use guides the feedback given directly to learners at the completion of an ED shift. Furthermore, using the CanMEDS framework as a basis for providing feedback has also not been explored. Accordingly, it is not known if learners receive regular meaningful feedback on all of the CanMEDS roles on EM rotations. A specific challenge in the use of DECs is the provision of feedback to learners on all 7 CanMEDS roles while minimizing the sources of error mentioned above. One method to address this challenge is to frame the assessments provided to the learners as constructive feedback rather than evaluative or punitive, and to allow teachers to identify a focused selection (2 or 3) of the 7 roles on which they wish to provide feedback for a given shift encounter.

Study question

We sought to test the hypotheses that 1) EM teachers will collectively provide learners with comprehensive feedback covering all of the CanMEDS roles over the course of a rotation despite having the individual ability to choose which

roles to comment on after a shift; and 2) EM teachers will identify areas for improvement when provided with DEC's that are designed specifically to reduce the negativity associated with such assessments.

Methods

This study was carried out at 2 tertiary care academic adult EDs with 36 000 and 58 000 patient visits per year, respectively. The study took place between July 2004 and October 2006. We implemented a system wherein EM teachers were required to provide postgraduate year 1 (PGY-1) learners with individual feedback at the completion of each ED shift, with documentation of the discussion on DEC's. Learners comprised Canadian PGY-1 residents enrolled in family medicine or RCPSC specialty programs other than emergency medicine, and were completing rotations in the 2 study EDs. Faculty consisted of staff emergency physicians with academic appointments at the University of Toronto. At each site, learners were matched with a single staff physician for the entire shift, but worked with many different physicians during their entire rotation. Rotations were anywhere from 4 to 8 weeks long, depending on the curriculum needs of the resident's home program. We designed a DEC with several features to maximize the frequency and quality of feedback and to minimize the halo effect (Fig. 1). The cards first included a linear analog scale on which teachers assessed overall learner performance on an anchored normative scale. This was followed by a list of 7 roles based on the RCPSC CanMEDS training paradigm. For each role, there were specific descriptors outlining what types of behaviour would constitute performance in that area. Teachers could choose either "needs attention" or "area of strength" for each role. A space for the evaluator to include specific examples upon which the assessment was based was provided for each role. To allow faculty to focus only on those areas of learner performance for which they had formed an opinion and to avoid the aforementioned halo and millstone effects, which could be imposed by requiring completion of all areas of the DEC after every encounter, faculty were asked to assess only 2 or 3 roles per shift. This design element allowed us to determine what roles the faculty felt were demonstrated (either positively or negatively) during typical shifts. The choices were dichotomous, reflecting the ability of teachers to make relatively crude assessments on multiple domains. "Needs attention" was chosen because it was felt that this terminology was more prescriptive and constructive than the more commonly used "below expectations" and was in keeping with the distinction between feedback

and evaluation. "Area of strength" was chosen to encourage faculty to select it only when performance clearly stood out for that learner on that role. The anchors were pilot tested on a small number of faculty, who felt they would accomplish the intended goal. Finally, a narrative section was provided for the inclusion of overall comments. The system required residents to present the DEC to their teachers at the end of each shift. The card was completed, discussed with the learner and then returned to them for signature and deposition into a secure mail slot. The rotation coordinator retrieved the completed cards, which were used as the basis for the summative evaluation for each learner, at the end of the rotation. Submission of the cards was mandatory and considered a formal record of attendance. Learners with more than 2 cards missing were interviewed by the rotation coordinator and the cards either retrieved from the learner at that time or solicited after the fact from the teacher. Shifts for which no cards were available would be considered incomplete.

The study was initially performed at a single site. To assess for site specificity and to enhance the generalizability of our study findings, a second site was recruited for a 4-month period, from July to October 2006, at the end of the study period. Ethics approval for the study was provided by the research ethics boards of both study institutions.

All numerical and categorical data were collected and entered into an Excel spreadsheet (Microsoft Corp., Redmond, Washington). The descriptive analysis was carried out on the entire data set and also by site. Based on previous experience and expert consultation, we considered each CanMEDS role to be adequately addressed by faculty, in general, if at least 20% of DEC's contained comments pertaining to the role. Similarly, we deemed that faculty members were able to use the DEC's to identify areas requiring attention on any of the CanMEDS roles if at least 10% of assessments were "needs attention." Finally, the data were analyzed by the learner and the faculty member to determine how pervasive the patterns across learners and teachers were. We sought to determine the proportion of residents who received feedback on all 7 of the CanMEDS roles during the rotation as an indicator of the protocol's utility in an operational sense for individual residents.

Results

Over a period of 24 months at one site and 4 months at another site, 54 learners submitted a total of 801 DEC's, completed by 43 different teachers. Cards were received for all learners during the study period, and no learners had more

than 2 missing DECs. As demonstrated in Table 1, the patterns of selection of CanMEDS roles did not differ between sites. Teachers selected an average of 3 CanMEDS roles per DEC (range 0–7) (Table 2). The overall frequency with which each CanMEDS role was selected ranged from 25% of DECs for Health Advocate to 85% of DECs for Medical Expert (Table 1). Very few of the ratings (36 of 2420, 1.3%) were “needs attention,” and the majority of these were in the Medical Expert role (Table 3). Of

the 43 faculty members, 33 (77%) did not select “needs attention” on any of the cards during the study period.

Discussion

Our primary hypothesis was that EM teachers would collectively provide feedback to individual learners covering the breadth of competencies over the course of a rotation, even though the teachers were given the opportunity to

Junior Resident Shift Feedback Card — Emergency Medicine (FRONT)

JUNIOR RESIDENT: _____ DATE _____ SUPERVISING PHYSICIAN: _____

1. Rate the Overall performance ON THIS SHIFT with a circle or 'x' on the line below.

AS EXPECTED

2. Using 2 or 3 of the following CanMEDS Roles, explain where the resident needs to improve to move up a level on the above scale. Provide specific examples for the assessments.
N = NEEDS ATTENTION; S = PARTICULAR STRENGTH

	Assessment		Example(s)
MEDICAL EXPERT:			
- Can APPROACH most presenting problems	N	S	_____
- Safe basic procedural skills	N	S	_____

Junior Resident Shift Feedback Card — Emergency Medicine (BACK)

	Assessment		Example(s)
SCHOLAR: (Inquisitive, Evidence – based, uses IT well)	N	S	_____
MANAGER: (Considers resource issues, efficient)	N	S	_____
HEALTH ADVOCATE: (Knows determinants of health, advocates for patient)	N	S	_____
COMMUNICATOR: (Patient interactions, case presentations, charting)	N	S	_____
COLLABORATOR: (Collegial, understands roles)	N	S	_____
PROFESSIONAL: (Punctual, responsible, has integrity, ethical)	N	S	_____

3. GENERAL COMMENTS:

STAFF SIGNATURE: _____ **RESIDENT SIGNATURE:** _____

Fig. 1. Daily encounter card used in this study.

limit their shift-based feedback to 2 or 3 key competencies per shift, and this was supported by our data. Physician performance is complex, yet can be broken up into definable observable behaviours representing core competencies, such as those defined in the RCPSC CanMEDS training model.^{5,6} Ideally, learners will be given regular feedback on how well they perform with respect to all of these fundamental competencies. It has been demonstrated, however, that some of the most reliable assessments of learner performance are global impressions.^{7,8,11,12} Relying on these general assessments does not allow directed focused feedback to learners about how they can improve performance. Alternatively, requesting teaching staff to provide accurate feedback on a number of different areas of performance based on a single encounter, often in the form of an itemized list on an evaluation form, burdens them with the unfair expectation that they can independently judge more than 2 or 3 aspects of a complex interaction.^{7,10,13–15} Allowing teachers to select which competencies to comment on from among the CanMEDS roles for a given shift can help avoid the forced attempt to assess too many separate areas of performance but could result in learners getting infrequent feedback on key areas of

competence. This concern was not borne out in our study because we observed that each of the 7 CanMEDS roles were mentioned on at least 25% of DECAs. Direct observation of residents in practice is the foundation of the North American postgraduate medical education system.^{10,12,22,23} Moreover, competency-based education and evaluation are becoming more popular in an age of increased accountability. These 2 realities of medical education mean that formal ongoing assessment of key competencies is required of today's teachers. Our study demonstrates that DECAs can underpin a system that allows teachers to provide regular, directed, specific feedback to learners that is tailored to their performance on a shift, yet still results in an overall assessment of the learner that incorporates all CanMEDS roles. In our study, every learner had at least 6 of the 7 roles commented on during feedback, and only 6 of 54 learners (11%) received no comments for 1 of the CanMEDS roles (1 on the Collaborator role, 1 on the Manager role, 1 on the Scholar role and 3 on the Health Advocate role). Given that all roles were frequently identified by teachers, the lack of comment on these roles for these learners may reflect their performance rather than a systematic avoidance of feedback by teachers.

Two implications stem from these findings. Faculty demonstrated a willingness and ability to choose from a standardized list of potential competencies to construct their feedback, selecting an average of 3 out of 7 roles in our study. Our system may encourage teachers to put more thought into what they decide to document, resisting the temptation that is sometimes created on other forms to just provide the same score for all items, which would be in keeping with an overall global impression and would forego an attempt to provide item-specific assessments (the vertical line approach). The second implication is that the ED is a practice environment wherein all 7 CanMEDS roles are demonstrated on a regular basis. This concept may be important for program directors and curriculum planners to consider when deciding on goals and objectives for various clinical rotations. Furthermore, the high

Table 1. Percentage of cards with ratings for each CanMEDS role, by teaching site

CanMEDS role	Site 1, % of ratings; <i>n</i> = 634		Site 2, % of ratings; <i>n</i> = 167	
	N	S	N	S
Medical expert	2.1	83.1	1.8	86.8
Scholar	0.3	25.4	1.8	32.9
Manager	0.5	26.2	1.2	27.5
Health advocate	0.3	23.0	0.0	34.7
Communicator	0.2	35.3	0.0	37.7
Professional	0.0	40.7	0.0	53.9

N = needs attention; S = particular strength.

Table 2. Distribution of cards by number of roles identified

No. of roles identified per card	No. (and %) of cards; <i>n</i> = 801
0	10 (1.2%)
1	40 (5.0%)
2	255 (31.8%)
3	276 (34.5%)
4	133 (16.6%)
5	42 (5.2%)
6	15 (1.9%)
7	30 (3.8%)

Table 3. Distribution of "needs attention" ratings by CanMEDS role

CanMEDS role	% of cards identifying role as "needs attention"; <i>n</i> = 801
Medical expert	2.13
Scholar	0.63
Manager	0.63
Health advocate	0.25
Communicator	0.8
Collaborator	0.1
Professional	0.0

faculty-to-learner ratio in the ED (often 1:1) provides an opportunity for more individual observation of resident performance in each role.

The second hypothesis we sought to test was that faculty would be more likely to identify areas requiring attention for a learner if the DEC was designed to minimize the negativity associated with the rating. We provided only 2 options for each role: “needs attention” and “area of strength.” We purposely avoided any ratings that implied a judgment against a norm (such as “below expectations”) or a deficiency in performance compared with peers. In addition, we phrased the instructions to frame the assessment as constructive feedback to help the learner move up a level on the global scale at the top of the form. In doing this we hoped to emphasize the positive role of feedback. Despite these DEC design elements, only 1.3% of all ratings were “needs attention.” Furthermore, 77% of teachers never chose “needs attention.” While it is possible that some learners may have sequestered some cards with perceived negative ratings, our recovery rate of 82% suggests that our results do reflect a paucity of negative assessments, though in reality it might not be as extreme as our actual data estimate. Several previous studies have examined possible reasons for this, including faculty avoidance of negative interactions, fear of negative teaching evaluations and the perception that providing a negative rating would result in the faculty having to spend large amounts of time either justifying the comments or remediating the problem.²² We hoped to remove much of the negative stigma associated with such a rating through the design of our DEC and, in doing so, some of the legitimacy of these concerns. Unfortunately, we were unable to demonstrate widespread willingness of teachers to document areas for improvement for their learners. In fact, our results verify that the previously described widespread reluctance among medical teachers to document negative feedback exists in our practice environment as well. While many of the reasons for the observed leniency effect shown in our study are not directly attributed to our feedback system, the DEC design described here was unable to overcome these challenges. Further work exploring this interesting and important problem is required.

This study has set the stage for further lines of investigation. We are currently carrying out a series of interviews and focus groups that are intended to determine EM teachers’ attitudes and recommendations on the implementation of effective competency-based feedback to learners. In addition, further site-specific research into perceptions leading to leniency bias in feedback should be undertaken. Finally, the impact of focused faculty development to

encourage candid feedback and rigorous use of tools such as the DEC should be explored.

Limitations

This study has several limitations that should be considered when interpreting the results. First, while we did incorporate data from 2 separate teaching sites, both are affiliated with a single university and results may have been influenced by the local academic culture related to some of the disincentives for providing the perceived negative feedback to which we alluded above.^{9,22} While we have no evidence that this is a specific problem at our sites, the degree to which these impediments are perceived to be at play at the local academic institution may affect the rates of “needs attention” ratings at both sites.

Second, we relied on teachers’ documentation of their discussion to make inferences about what feedback was given to learners. We may, therefore, have underestimated the amount of feedback in the “needs attention” category that was verbally provided to learners. It can be argued that because undocumented feedback is open to challenge by learners later identified as “in trouble,” our focus on documented feedback is appropriate as an ideal. Furthermore, any differences between written and verbal feedback would represent an increase in the breadth of roles covered, further strengthening support for our conclusions about the range of feedback.

Third, some learners may not have submitted the entire collection of DECs. The average number of DECs per learner in the study was 14.9, which compares favourably with the average of 18 shifts per month worked by learners (who did not take vacation or other leaves). Our experience suggests that a number of the omissions were owing to learner or faculty forgetting to complete the forms, missing forms, and faculty working multiple shifts in a row with a given learner and deferring DEC completion until after multiple shifts (a practice that was actively discouraged yet persistent). This limitation may have resulted in an underestimation of the “real” rate of “needs attention” assessments, but should have had little effect on our conclusions about the breadth of feedback provided.

Conclusion

DECs provide evidence that learners get ongoing targeted feedback during ED rotations. EM teachers will select specific roles on which to provide competency-based feedback after individual shifts, but over the course of a rotation, teachers collectively provide feedback on a range of

competencies to their learners with very few getting no feedback on a role. The vast majority of documented assessments are areas of strength, despite the use of a DEC designed to reduce disincentives to more negatively perceived comments.

Acknowledgements: This research was funded by a peer-reviewed grant from the Royal College of Physicians and Surgeons of Canada.

Competing interests: None declared.

References

1. Bandiera G, Lee S, Tiberius R. Creating effective learning in today's emergency departments: how accomplished teachers get it done. *Ann Emerg Med* 2005;45:253-61.
2. Atzema C, Bandiera G, Schull MJ. Emergency department crowding: the effect on resident education. *Ann Emerg Med* 2005;45:276-81.
3. Carter AJ, McCauley WA. Off-service residents in the emergency department: the need for learner-centeredness. *CJEM* 2003;5:400-5.
4. Chisholm CD, Collison EK, Nelson DR, et al. Emergency department workplace interruptions: are emergency physicians "interrupt-driven" and "multitasking"? *Acad Emerg Med* 2000;7:1239-43.
5. Frank JR, ed. *The CanMEDS 2005 physician competency framework. Better standards. Better physicians. Better care.* Ottawa (ON): The Royal College of Physicians and Surgeons of Canada; 2005.
6. Bandiera GW, Sherbino J, Frank JR. *The CanMEDS assessment tools handbook.* 1st ed. Ottawa (ON): Royal College of Physicians and Surgeons of Canada; 2006.
7. Bandiera GW, Morrison L, Regehr G. Predictive validity of the global assessment form used in a final year undergraduate rotation in emergency medicine. *Acad Emerg Med* 2002;9:889-95.
8. Davis J, Inamdar S, Stone RK. Interrater agreement and predictive validity of faculty ratings of pediatric residents. *J Med Educ* 1986;61:901-5.
9. Dudek NL, Marks MB, Regehr G. Failure to fail: the perspectives of clinical supervisors. *Acad Med* 2005;80(10 Suppl):S84-7.
10. Resnick R, Taylor B, Maudsley R, et al. In-training evaluation — it's more than just a form. *Ann R Coll Phys Surg Can* 1991;24:415-20.
11. Streiner D. Global rating scales. In: *Assessing clinical competence.* New York (NY): Springer Verlag;1985. p. 119-41.
12. Gray JD. Global rating scales in residency education. *Acad Med* 1996;71(1 Suppl):S55-63.
13. Borman W. Effects of instruments to avoid halo error on reliability and validity of performance evaluation ratings. *J Appl Psychol* 1975;60:556-60.
14. Davis JK, Inamdar S. The reliability of performance assessment during residency. *Acad Med* 1990;65:716.
15. Scheuneman AL, Carley JP, Baker WH. Residency evaluations. Are they worth the effort? *Arch Surg* 1994;129:1067-73.
16. Cadwell J, Jenkins J. Teachers' judgments about their students: the effects of cognitive simplification strategies on the rating process. *Am Educ Res J* 1986. 23:460-75.
17. Paukert JL, Richards ML, Olney C. An encounter card system for increasing feedback to students. *Am J Surg* 2002;183:300-4.
18. Luker K, Beaver K, Austin L, et al. An evaluation of information cards as a means of improving communication between hospital and primary care for women with breast cancer. *J Adv Nurs* 2000;31:1174-82.
19. Kim S, Kogan JR, Bellini LM, et al. A randomized-controlled study of encounter cards to improve oral case presentation skills of medical students. *J Gen Intern Med* 2005;20:743-7.
20. Brennan BG, Norman GR. Use of encounter cards for evaluation of residents in obstetrics. *Acad Med* 1997;72(Suppl 1):S43-4.
21. Al-Jarallah KF, Moussa MA, Shehab D, et al. Use of interaction cards to evaluate clinical performance. *Med Teach* 2005;27:369-74.
22. Ross L, Nesbitt RE. *The person and the situation: perspectives of social psychology.* Philadelphia(PA): Temple University Press;1991.
23. Farrell SE. Evaluation of student performance: clinical and professional performance. *Acad Emerg Med* 2005;12:302e6-10.

Correspondence to: Dr. Glen W. Bandiera, St. Michael's Hospital, 1-056 Bond Wing, 30 Bond St., Toronto ON M5B 1W8; bandierag@smh.toronto.on.ca