Discrepancies Between Preceptor and Resident Performance Assessment: Using an Electronic Formative Assessment Tool to Improve Residents’ Self-Assessment Skills

Karen Schultz, MD, Tara McGregor, MD, MSc, Rob Pincock, MD, Kathleen Nichols, MD, PhD, Seema Jain, MD, and Joel Pariag, MD,

K. Schultz is professor, Department of Family Medicine, and associate postgraduate dean, Queen’s University, Kingston, Ontario, Canada; ORCID: https://orcid.org/0000-0001-7041-1700.

T. McGregor is assistant professor and assessment director, Department of Family Medicine, Queen’s University, Kingston, Ontario, Canada.

R. Pincock is assistant professor and curriculum director, Department of Family Medicine, Belleville Family Medicine Site, Queen’s University, Kingston, Ontario, Canada.

K. Nichols is assistant professor and assessment lead, Peterborough-Kawartha Program, Department of Family Medicine, Queen’s University, Kingston, Ontario, Canada.

S. Jain is assistant professor and assessment lead, Bowmanville-Oshawa-Lakeridge Program, Department of Family Medicine, Queen’s University, Kingston, Ontario, Canada.

J. Pariag is assistant professor and assessment lead, Bowmanville-Oshawa-Lakeridge Program, Department of Family Medicine, Queen’s University, Kingston, Ontario, Canada.

Correspondence should be addressed to Karen Schultz, Department of Family Medicine, Queen’s University, 220 Bagot Street, Kingston, Ontario, Canada K7L 5E9; telephone: (613) 533-9300 ext. 73973; email: karen.schultz@queensu.ca.
Acknowledgments: The authors would like to thank and acknowledge Rachelle Porter, software developer, who designed the software for the assessment system, including adding reporting functions as the program evolved and generated the data for the exploration described in this article; as well as Dr. Jane Griffiths, who was instrumental in conceiving of and establishing our assessment program.

Funding and Support: Support for Ms. Porter’s time was provided by the Queen’s University Department of Family Medicine and is also gratefully acknowledged.

Disclosures: None reported.

Ethical approval: Approval was received December 2, 2019, following a delegated review from the authors’ institutional Health Sciences and Affiliated Hospitals Research Ethics Board (File # 6028253).

Previous presentations: These data were presented as a poster at the College of Family Physician’s of Canada’s national Family Medicine Forum in Vancouver, BC, October 30, 2019.
Abstract

Problem

Accurate self-assessment is a critical skill for residents to develop to become safe, adaptive clinicians upon graduation. Physicians need to be able to identify and fill in knowledge and skill gaps to deal with rapid expansion of medical knowledge and unpredicted novel emerging medical issues. Residency training to date has not consistently focused on building these overarching skills, nor have the burgeoning assessment data that competency-based medical education (CBME) affords been used beyond their initial intent to inform summative assessment decisions. Both are important missed opportunities.

Approach

The Queen’s University family medicine program adopted CBME in 2010. In 2011 it added the capacity for residents to electronically self-assess their daily performance, with preceptors reviewing and modifying as needed before submitting. In 2018 it designed software to report discordance between residents’ self-assessment and preceptors’ assessment of performance.

Outcomes

From 2011–2019, 56,585 field notes were submitted, 11,429 by residents, with 28% of those (3,200/11,429) showing discordance between residents’ and preceptors’ performance assessments. When discordant, residents assessed their performance as less competent (undercalled) than their preceptor did 73% of the time (2,336/3,200 field notes). For the 864 field notes (27% of 3,200 discordant notes) where residents rated their performance higher than their preceptor did (overcalled, for 162/1,120 (14%) residents), 6 residents overcalled performance to a dangerous extent (2 or 3 levels of supervision higher than what their supervisors assessed them at) and 26 repeatedly (greater than 5 times) overcalled their level of performance by 1 supervisory level.
Next Steps

Inaccurate self-assessment (both overcalling and undercalling performance) has negative consequences. Awareness is a first step in addressing this. Discrepancy reports will now be used during regular academic reviews with residents to discuss the nature, degree, and frequency of discrepancies, with the intent of fostering improved self-assessment of performance.
Problem

Accurate self-assessment is a critical skill in the self-regulated profession of medicine. The Dunning-Kruger effect, described in 2003, is the phenomenon of poor performers overcalling their abilities (when self-assessment indicates higher performance than external assessment) and good performers undercalling their abilities (when self-assessment rates performance as lower than does external assessment).\(^1\) In health care, overconfidence in abilities can lead to patient safety concerns, while underconfidence can result in physicians restricting their practice, overusing resources, and/or experiencing anxiety with potential burnout. In addition to accurately knowing one’s limits, given the exponential expansion of medical knowledge and the inevitable emergence of unforeseen novel medical issues, physicians need to be adaptive, learning and applying new information and skills throughout their careers.\(^2\) A critical first step for this is accurately self-assessing current abilities to identify the need for self-guided education. Competency-based medical education (CBME), where frequent formative assessment data about residents’ competency development is gathered, lends itself to building these self-assessment skills. Reassuringly, there is some evidence that deliberate a curriculum to foster self-assessment can improve this skill.\(^3\) A strategy we have adopted at the Queen’s University Department of Family Medicine residency program to develop self-assessment skills is to empower residents to start documenting verbal feedback about their performance provided by their preceptor, including a self-assessment of that performance. This is then reviewed by the preceptor, changed as needed, signed off and saved in the resident’s portfolio. Repeated or concerning discordance between resident and preceptor assessments presents a learning opportunity for residents to hone self-assessment skills. In this article, we report on the tool developed within our program that allows us to do that.
Approach

Our Canadian family medicine residency program implemented competency-based medical education in 2010. Our main method of generating frequent formative assessment of performance for our residents is through electronic field notes. These are brief notes that document a resident’s performance in a clinical encounter and summarize the preceptor’s verbal feedback given about his or her performance during that encounter. In addition, a competence judgment about the resident’s performance is also captured by the preceptor assessing them as being at one of 4 levels: flagged (concerning behavior), needing close supervision, minimal/reactive supervision, or supervision for refinement only. Narrative descriptions of the expected performance for a particular patient encounter are readily available to our preceptors by hovering over the “details” link beside each level of supervision. Doing so results in a pop-up description of performance expected for that type of encounter. These descriptions are not meant to reduce complex tasks to a checklist of actions, but to paint a picture of competence at the different levels of supervision. Once completed, the field note is submitted and populates that resident’s electronic portfolio. Interpreting this gathered formative competency data to make summative decisions is done through our portfolio assessment support system, which has 2 equally important components: the resident’s electronic portfolio and an academic advisor for each resident, with whom the resident meets with 3 times a year. While the portfolio collects several types of assessments (e.g., rotation assessments, objective structured clinical exam results, simulation course results, resident-as-teacher assessments, multisource feedback), the most important tool for evaluating most competencies are the field notes. Because these are electronic, multiple preceptors in multiple settings over time can enter information, and that information can be collated and sorted according to the desired competencies. During the 3 times a year meetings, the academic advisor, in discussion with the resident, interprets the collated
data, makes competency decisions, and decides next steps to adjust the resident’s learning plan to further their competency development. At key junctures, the academic advisor passes on competency decisions to the program director, who determines program completion and who forwards these declarations to the regulatory bodies. (A more robust description of our field notes and how they are used both formatively and to inform summative decisions about competency development is available6).

Initially, all our field notes were preceptor-initiated. One year after implementing CBME, in 2011, we realized that allowing residents to initiate a field note wherein they documented the verbal feedback given to them by their preceptor and then choose the level of supervision they felt the preceptor gave them for that encounter, would increase the number of field notes, the number of preceptors contributing to the field notes, and the contexts field notes were created in—all things that add to the trustworthiness of this qualitatively grounded workplace-based assessment. In response to the existing literature about the poor accuracy of self-assessment,7,8 we built in mandatory verification functionality such that any resident-initiated field note is sent to the preceptor as an email with a one-time link to that field note. The preceptor reviews the resident-initiated field note, and, having been the person to witness the performance, modifies the written feedback recorded by the resident on that same field note if desired, and changes the level of supervision if they feel it did not reflect their assessment of the resident (Figure 1). Preceptors can access the link in the field note that pops up a description of expected performance for each level of supervision if they are uncertain which level to choose. They then submit the field note and the field note then becomes part of the resident’s electronic field note collection.
In 2018 we also tapped into residents’ self-perception of performance to deliberately assess their self-assessment skills. We did this by creating software that would detect when a resident’s self-initiated field note was changed by the preceptor and in what way (e.g., did the preceptor move the level of supervision chosen by the resident higher or lower, and if so by how many categories). A report can now be generated for each resident identifying how often their self-assessment on their resident-initiated field notes varies from their preceptors and to what degree. Tapping into our assessment data this way has allowed us to use learning analytics to further inform our understanding of resident’s development. Although this is low-level analytics, that of descriptive analytics to look at “what is happening,” it is the beginning of a deeper dive into our growing assessment data and is using these data in a way beyond what they were initially intended for (using formative assessment to inform summative assessment of our residents).

As an aside, in addition to resident-initiated field notes increasing their fieldnote collections, we have also realized another helpful use of this resident self-recording function. All residents can and do self-record their performance on field notes (it is preferentially used by residents on off-service rotations where, to date, preceptor-initiated field notes have been scarce). Additionally now, for the struggling resident, where insight into performance and/or their ability to process feedback as intended can be a part of their performance issues, we now preferentially ask those residents to reflect on the feedback given to generate a self-initiated field note. Preceptors can then quickly see if the resident is processing verbal feedback accurately. If not, they can then correct the resident-written feedback, allowing residents to process the feedback through another modality (written instead of verbal). It also prompts the preceptor to check the resident’s comprehension with any further important verbal feedback.
Outcomes

Between 2011 and 2019, 56,585 field notes were submitted on the 1,120 residents going through our 2-year program. Of these, 80% (45,156) were preceptor-initiated and 20% (11,429) were resident-initiated. Over their 2-year residency, residents receive anywhere from 0–381 field notes, with the average being 76 field notes per resident over 2 years (Table 1). This number varies between our 4 sites, mainly due to the different curriculum structures at those sites. Residents at the 2 sites with longitudinal curricula (without rotations and therefore no end-of-rotation assessments) receive more field notes, these more evenly distributed throughout their 2 years with a more even distribution of resident-initiated and preceptor-initiated field notes. For the 2 rotationally based sites, residents receive most of their field notes in the first year, with most resident-initiated field notes tending to occur on off-service rotations.

Of the 11,429 resident-initiated field notes, most (8,229/11,429 (72%)) showed concordance between resident and preceptor assessments of level of supervision needed (e.g., the preceptor did not change the level of supervision chosen by the resident). Of the 28% (3,200) that were discordant (e.g., the preceptor changed the level of supervision chosen by the resident), preceptors assessed the resident higher than the resident assessed themselves 73% (2,336 field notes) of the time. For the 864 field notes (27% of discordant field notes) where residents overcalled their performance, which happened for 162/1,120 residents (14%), 6 residents overcalled this to a dangerous extent (2 or 3 levels of supervision higher than what their supervisors assessed them at) and 26 repeatedly overcalled their level of performance by 1 level of supervision (greater than 5 times).
Next Steps

To date, our numbers are too small to associate residents who overcall their performance with adverse patient safety outcomes or residents in identified difficulty (using the blunt tool of residents requiring remediation). Exploring this association as data accumulate will be an important next step. Given our finding of a small but important subset of residents who overcall or undercall performance, we have now built our system to flag when a resident significantly and/or consistently over- or undercalls their performance where significantly equals varying by 2 or 3 levels and consistently means varying by 1 level 5 or more times. Their academic advisor, as well as site director and program director, are automatically notified when this happens. This allows for discussing these discrepancies, with the hope that doing so will allow for better calibration of self-assessment skills, decreasing the risk to patient safety for those overconfident in their abilities, and reducing anxiety with its attendant consequences for those undercalling their abilities. Monitoring the psychological impact of this intervention on residents will be an important element of evaluating this function. Next steps will also involve tracking those residents demonstrating difficulties with accurate self-assessment to see if this intervention does indeed result in improvement in this skill. We also realize that development of other strategies to better support residents in developing this important overarching skill of being able to accurately self-assess performance are also required.
Analysis of the data we gathered is limited by being carried out at a single institution in a single specialty where an increased emphasis on feedback, both how to give and receive it, has been an active part of faculty and resident development for the past 10 years. It is possible in other settings that there may be differing discordance between residents’ and preceptors’ assessment of clinical performance. It does, however, demonstrate the ability to leverage CBME data beyond an initial goal of generating robust low stakes formative assessment, expanding its use to now potentially improve self-assessment. Given that this is a critical skill for the adaptive physician, to do so is a helpful use of these data.
References


Figure Legend

Figure 1

A hypothetical testing resident’s summary of their understanding of feedback provided by their preceptor for their care of a patient with chest pain. The resident feels they are “ready for independence, supervision for refinement only.” The preceptor, receiving the resident’s assessment via one-time email, has changed the level of supervision to “close supervision” and explained this in the comments box at the bottom. On submission, this field note now populates the resident’s portfolio. The discrepancy can be discussed with the resident by both the preceptor and the resident’s academic advisor.
Table 1

Characteristics of Resident and Preceptor-Initiated Field Notes, by Program Site, Queen’s University Family Medicine Program, Ontario, Canada 2011–2019

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Program totals</th>
<th>Site A</th>
<th>Site B</th>
<th>Site C</th>
<th>Site D</th>
</tr>
</thead>
<tbody>
<tr>
<td>PGY1 and PGY2 residents each year, no.</td>
<td>140</td>
<td>12</td>
<td>16</td>
<td>12</td>
<td>100</td>
</tr>
<tr>
<td>Total field notes 2011–2019, no.</td>
<td>56,585</td>
<td>6,618</td>
<td>12,598</td>
<td>3,554</td>
<td>33,815</td>
</tr>
<tr>
<td>Number of field notes over each resident’s 2-year program, mean</td>
<td>76</td>
<td>122</td>
<td>246</td>
<td>68</td>
<td>59</td>
</tr>
<tr>
<td>Total preceptor-initiated field notes 2011–2019, no.</td>
<td>45,156</td>
<td>6,290</td>
<td>7,779</td>
<td>2,204</td>
<td>28,883</td>
</tr>
<tr>
<td>Number of preceptor-initiated field notes per resident over their 2-year program, mean (range)</td>
<td>61 (0–377)</td>
<td>116 (0–245)</td>
<td>152 (0–377)</td>
<td>42 (0–161)</td>
<td>51 (0–346)</td>
</tr>
<tr>
<td>Total resident-initiated field notes 2011–2019, no.</td>
<td>11,429</td>
<td>328</td>
<td>4,819</td>
<td>1,350</td>
<td>4,932</td>
</tr>
<tr>
<td>Number of resident-initiated field notes per resident over their 2-year program, mean (range)</td>
<td>15 (0–381)</td>
<td>6 (0–40)</td>
<td>94 (0–381)</td>
<td>26 (0–115)</td>
<td>8 (0–143)</td>
</tr>
</tbody>
</table>

Abbreviations: FN, field notes; PGY1, postgraduate year 1 resident; PGY2, postgraduate year 2 resident;

*Sites A and B have a longitudinal structure to their curriculum and hence no end-of-rotation assessments, site D has a rotational structure, and site C has a hybrid structure with both a longitudinal component and rotations. Both sites C and D utilize end-of-rotation assessments as part of their program of assessment.*
Figure 1

![Image of Family Medicine Field Notes](image-url)

---

**FAMILY MEDICINE FIELD NOTES**

Approve a Field Note submitted on your behalf

Resident: Testing Resident

Date of encounter: 2020-07-13

Setting: Ambulatory/office practice

Resident: Testing Resident

email: testing.dfmresident@dfm.queenau.ca

On Behalf of: Karen Schultz

email: karen.schultz@cfm.queenau.ca

Type of encounter/situation: 45 yo M, chest pain

Phase Observed: History

Direct Observation - Yes

level of performance chosen by evaluator

level of performance chosen by Resident

Ready for independence; supervision for refinement

Skills Dimensions / CanMEDs Roles

Clinical Reasoning Skills

Family Medicine Expert

Domains of Clinical Care

Care of Adults

Linked EDA Care of the Adult with an Acute Serious Illness

Feedback

Continue (strengths): great history re chest pain

Change (or do less):

Consider (next steps):

This field note is: Not Flagged

Submitted via: Intranet

Comments

<table>
<thead>
<tr>
<th>Date</th>
<th>Submitted by</th>
<th>Status/Flag</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>2020-07-13</td>
<td>Testing Resident</td>
<td>Not Flagged</td>
<td></td>
</tr>
</tbody>
</table>

Approve this Field Note:

Please select the Level of performance:

- [ ] Flagged/ concerning performance (details)
- [ ] Requires close supervision (details)
- [ ] Requires minimal supervision (details)
- [ ] Ready for independence; supervision for refinement (details)
- [ ] Would you like to mark this as an outstanding field note?

Add comments to this field note:

I do agree that you did a thorough history around CAD causes for CP (location, radiation, other symptoms, pitting events, risk factors), however we also discussed the importance of exploring other dangerous causes of CP (e.g. PE) and what you would look for on hx and pe. I have changed the level of supervision as this is an important thing to do before you are ready for independence.