
Article:
Lockyer, J, Sargeant, J, Richards, SH orcid.org/0000-0003-1416-0569 et al. (2 more authors) (2018) Multisource Feedback and Narrative Comments: Polarity, Specificity, Actionability, and CanMEDS Roles. Journal of Continuing Education in the Health Professions (1). pp. 32-40. ISSN 0894-1912

https://doi.org/10.1097/CEH.0000000000000183

Multisource Feedback and Narrative Comments: Polarity, Specificity, Actionability and CanMEDS roles

Abstract

Background: Multisource feedback (MSF) is a questionnaire based assessment tool that provides physicians with data about workplace behaviors and may combine numeric and narrative (free-text) comments. Little attention has been paid to wording of requests for comments, potentially limiting its utility to support physician performance. This study tested the phrasing of two different sets of questions.

Method: Two sets of questions were tested with family physicians, medical and surgical specialists and their medical colleague and co-worker respondents. One set asked respondents to identify one thing the participant physician does well and one thing the physician could target for action. Set 2 questions asked what does the physician do well and what might the physician do to enhance practice. Resulting free-text comments provided by respondents were coded for polarity (positive, neutral, or negative), specificity (precision and detail), actionability (ability to use the feedback to direct future activity) and CanMEDS roles (competencies) and analyzed descriptively.

Results: Data for 222 physicians (111 physicians per set) were analyzed. 1824 comments (8.2/physician) were submitted, with more comments from co-workers than medical colleagues. Set 1 yielded more comments and were more likely to be positive, semi specific, and very actionable than Set 2. However, Set 2 generated more very specific comments. Comments covered all CanMEDS roles with more comments for collaborator and leader roles.

Discussion: The wording of questions inviting free-text responses influences the volume and nature of the comments provided. Individuals designing MSF tools should carefully consider wording of items soliciting narrative responses.

Keywords: Multisource feedback; narrative comments; free-text comments; qualitative comments; assessment; feedback; continuing professional development

Background

Multisource feedback (MSF) is an assessment process that provides physicians with questionnaire-based data about workplace behaviors, and seeks to stimulate reflection and guide improvement. MSF is usually a multistage process. Data about an individual's observable workplace behaviors are collected from those who work with the physician. Data are aggregated and these data are provided to the physician, who is expected to reflect upon them and use them to guide practice improvement. The physician may also meet with a colleague/supervisor to discuss the report and develop a plan for improvement.[1] MSF has been shown to help physicians identify and respond to feedback, and is particularly useful in addressing core medical competencies such as communicator, collaborator and professional competencies.[2]

MSF may fall short of its goals, particularly when the feedback is numeric and does not contain narrative information. Physicians have indicated that reports containing only numerical scores were not specific enough to identify needed improvements.[3-5] A systematic review found that narrative comments, credible raters, and feedback facilitation were critical factors leading to acceptance and use of MSF.[6]

Despite pleas for the inclusion of narrative comments,[3-5] there is limited information from MSF studies in which free-text comments are a component of the process.[7-10] However, these studies are consistent in finding that the majority of comments are positive endorsements of the physician's practice and provide few negative comments. Where adverse statements are offered, they may be presented within a comment providing both positive and negative content.[8] Furthermore, comments provided by respondents may focus on global behaviors rather than on specific, actionable, behaviors, and may thus offer few strategies for improvement.[7] These findings are consistent within the broader literature of MSF in industry and performance appraisal which has also shown greater numbers of positive than negative comments,[11, 12] and that comments are unlikely to be precise with specific examples offered to guide behavior,[12] nor provide specific recommendations.[13] These findings occur despite the fact that people pay attention to narrative comments more often than they do to numerical scores.[12,13]

An examination of the wording of requests for free-text comments suggests they are often neutrally worded, or encourage positive or general feedback,[7-10] potentially explaining the lack of direction offered. Examples include: "Anything especially good?";[10] "Provide comments that refer to specific events, and that are supportive and constructive";[7] and "Please feel free to add any other comments you have about this doctor".[8] One notable exception however, invites raters to "Complement your responses with narrative "positive comments" as well as "suggestions for improvement".[9] Given these observations, it appears that close attention needs to be paid to the wording of requests for comments. Without specific wording, free-text comments are unlikely to yield specific and actionable comments to guide physician improvement in the domains covered by physician core competencies, which are generally the focus of MSF assessments.[2]

The three concepts of polarity, specificity and actionability are particularly germane to examinations of MSF comments. Polarity, sometimes termed 'valence',[13] has been examined in several studies,[7-12,14] and refers to whether the comments are positive or negative.[13] Positive comments provide an endorsement of the physician's professional performance while negative comments indicate problematic performance.[8] In studies of medical student and resident assessment, comments with negative polarity have been found to provide more discriminating information[15,16] and are viewed more seriously than those with positive or neutral polarity.[16,17] Specificity refers to the level of precision and detail provided,[13-15,18] with particular attention to concreteness and elaboration on the behavior.[19] Actionability refers to the extent to which recipients can use the feedback to identify what they can do

to direct future behavior.[12,20,21] In MSF, polarity, specificity, and numbers of comments have been shown to explain a small but significant portion of individual improvement.[14] Goal setting been shown to be critical in translating feedback into performance but relies on specific directive data to enable the person to take action.[12] Content also matters in MSF. The framework provided by CanMEDS[22] and its roles is particularly relevant within the Canadian context as it guides residency training and assessment and subsequently positions reporting of continuing professional activities.

The primary purpose of our study was to examine how the phrasing of reviewer prompt questions affected the quantity, types and nature of information physicians received. Specifically, we wished to examine the ability of 2 different sets of prompts posed to MSF reviewers to generate free-text comments. As a secondary purpose of the study, we sought to compare the frequency of free-text comments by set and by question related to the comments' actionability, specificity, polarity, and the degree to which they map on to Royal College of Physicians and Surgeons of Canada (RCPSC) CanMEDS roles (competencies).[22]

Methods

Participants and tasks

Study participants were physicians in one of three specialty groups: family physicians, medical specialists (e.g. medicine, pediatrics, and psychiatry) and surgical specialists, and their respondent-reviewers (8 medical colleagues and 8 co-workers) who participated in the College of Physicians and Surgeons of Alberta, Physician Assessment Review (PAR) program between January 1, 2015 and December 31, 2015. Pivotal Research Inc, the independent survey organization managing the MSF process, handled recruitment as part of its routine activity. They were instructed to randomly select a sample of physicians, using a computerized algorithm, with a goal of identifying 288 physicians who would be invited to participate in PAR that year (110 family physicians, 88 medical and 88 surgical specialists). Once the random sample of physicians was ascertained, a second round of computerized randomization was undertaken to allocate the physician to one out of two experimental colleague questionnaires prepared for the study. The two colleague questionnaires differed only in respect of the wording of two sets of questions, each set seeking to elicit a free-text response. All of a physician's nominated medical colleagues and co-workers received either Set 1 or Set 2 questions.

Each set had two questions:

- 'Set 1' questions were: "What is one (1) thing the physician does particularly well?" (Q1) and "What is one (1) thing you would have the physician target for action?" (Q2)
- 'Set 2' questions used more general wording: "What does this physician do well?" (Q1), and "What might this physician do to enhance his/her practice?" (Q2).

The questions were developed by a committee comprising six physicians and a medical educator. The committee sought to ensure that the physician received feedback from medical colleagues and co-workers that was affirming or positive about their work as well helping them to identify things that could be done to improve practice.

Data collection

Pivotal Research Inc. captured the free-text comments from colleagues and provided an anonymized spreadsheet dataset. Each physician could receive a maximum of 32 comments with their review (maximum two comments from eight medical colleague reviewers, and two comments from 8 eight co-

worker reviewers). The research team were provided with an anonymised dataset that included specialty (i.e. surgery, medicine, or family medicine); the data source (medical colleague or a co-worker); the question set (i.e. Set 1 or Set 2) and question number (i.e., Q1 or Q2).

From the data in the spread sheet, we coded each comment by polarity, specificity, and actionability. We also coded for the 7 CanMEDS role(s) of medical expert, leader, scholar, professional, communicator, collaborator, and health advocate.[11] For polarity, coding options were positive (or affirmative) providing endorsement of the physician's professional performance, negative indicating problematic performance, or mixed (i.e., containing both positive and negative data), an approach consistent with a previous study.[8] For specificity, coding options were not specific, semi-specific, and very specific. In coding for specificity, the task was to determine whether the information in the comment was sufficiently clear, precise and detailed to guide the physician with very specific used to elaborate on behaviour.[13-15,18,] Given the wording of the questions posed, only Q2's for both sets were worded such that they could be coded for actionability. Coding options were not actionable, semi actionable (i.e., general areas provided for improvement without naming a specific behaviour), and very actionable (i.e., enough detail that the physician could act upon the information provided).[12,20,21]

For the polarity coding, two researchers independently coded half of the comments with a third researcher coding all of the comments. The ratings were compared and differences discussed. In coding for specificity, two researchers coded 10% of the comments and then discussed the comments, resolving minor differences and agreeing on the coding framework. Following that, one of the two then coded the remaining comments, consulting with the second on uncertain coding. For actionability, three researchers each coded a quarter of the comments with a fourth researcher coding all comments. The ratings were compared and differences discussed. In coding the comments for CanMEDS roles, all roles (i.e., for medical expert, communicator, collaborator, scholar, leader, professional, and health advocate) identified within a single comment were coded. This required careful reading of the comments and discussion of the roles.[22] After the initial coding and discussion by two researchers, a further discussion between the two researchers and a medical educator from the Royal College of Physicians and Surgeons of Canada was held to ensure their understanding of each of the roles was correct. This led to the development of a final coding structure and further discussion. One researcher then undertook the majority of the coding and a second researcher reviewed 10% of the comments and verified the more complex comments. The approach taken to coding through iterative rounds of discussion and negotiation to create a consistent understanding of the data and its coding framework is in alignment with thematic analysis which recommends regular meetings and discussions.[23-25] Discussions of codes occurred through several conference calls and e-mails.

Data Analysis

We used the coded data, including the derived variables, for analysis. Two versions of this data were analysed: the first used the physician as the level of observation, and the second used each comment as the level of observation. Thus in the second dataset there could be multiple observations (i.e., comments) per physician reviewed.

For the data at the physician level, we conducted descriptive analyses of the data to determine the number of physicians who received each set of questions and the mean numbers of comments each received from medical colleagues and co-workers. For the data at the comment level, we conducted descriptive analyses of data to determine the numbers of comments that each set and question generated, as well as the numbers of comments provided by medical colleagues and co-workers with regard to polarity, specificity, actionability and CanMEDs role(s). Continuous variables were described

using means and standard deviations, and categorical variables were described using frequencies and proportions.

To address the primary study objective of exploring the relationship between the phrasing of questions posed to reviewers (i.e. Set 1 and Set 2) and the number of comments reviewers generated from each Set, we used a t-test to compare the mean number of comments provided by co-workers and medical colleagues for both datasets. Analyses were stratified by Question, where "Q1" and "Q2" data were analysed separately. This analysis took place using data at the physician-level. To address the secondary study objectives comparing characteristics of the comments, we used data at the comment-level. We used Chi-Square tests to compare the proportion of comments generated between reviewer type (i.e., medical colleague or co-worker source) and Set 1 and 2 stratified by Q1 or Q2. We used a Chi-square to assess differences in proportions between Sets 1 and 2 stratified by question number (Q1 or Q 2) and in the polarity, specificity, and actionability of the comments; by reviewer type and characteristics of the comments; between Sets 1 and 2 stratified by question number (Q1 or Q2) and CanMEDS roles; and by reviewer type and CanMEDs role.

Results

Comments were available for analysis for 222 physicians (Table 1); 90 family physicians (45 each for set 1 and set 2), 63 medical specialists (34 for Set 1, 29 for Set 2) and 69 surgeons (32 for set 1, 37 for set 2). There were equal numbers of physicians who received each set. More comments were received for Set 1 than for Set 2 item-wording.

All comments received in the original dataset (n=2,133) were reviewed. We excluded comments that contained punctuation only (n=7), or stated "not applicable" (n=24), provided no information other than "I have nothing to comment" (n=219), or "this physician is a good doctor" (n=59), leaving 1824 comments available for analysis. For the CanMEDS analysis, a further 24 comments were excluded as they could not be mapped onto a CanMEDS competency, leaving 1800 comments available for analysis. See figure 1.

Of 1824 comments, 1013 were provided by co-workers and 811 by medical colleagues. The mean number of comments received per physician was 8.2 (standard deviation [SD] 4.0) with a larger number of comments generated by co-workers compared with medical colleagues ($p<0.001$). No statistical differences were found in the number of comments provided by medical colleagues or co-workers in response to the first or second question in either presentation set. A total of 1800 comments were analysed for their CanMEDS roles. The numbers of comments yielded in response to both sets and to both questions was similar for medical colleagues and co-workers. See Table 1.

Association between question phrasing ("Set") and number of comments generated

We compared the mean frequency of comments generated by Sets 1 and 2, finding differences between sets for Q2 but not for Q1 (Table 2). For Q1, Set 1, the mean number of comments produced was 5.98 (standard deviation [SD] 2.72), where Q1, Set 2 produced a mean of 5.44 (SD 2.55). The difference between means for Q1 data was not significantly different (95% confidence interval [CI] -0.15, 1.24, $p=0.129$). For Q2, Set 1, the mean number of comments generated was 3.13 (SD 1.80) whereas for Q2, Set 2 the mean was 2.67 (SD 1.35). The difference between Sets for Q2 was statistically significant (95% CI 0.01, 0.91, $p=0.047$).

Characterization of comments for polarity, specificity, and actionability

Physicians received more positive than negative or mixed comments (1394, 76%; 329, 18%; 101, 6% respectively). No difference in polarity of responses was observed between set for Q1. However, respondents provided a greater proportion of negative responses for Q2 Set 2. ($p < 0.001$). No differences were found for polarity between medical colleagues and co-workers. See tables 3 and 4.

Selected examples of comments for polarity include:

> *Positive: If he could somehow clone himself so he is able to work locally and internationally at the same time, many more would benefit from their relationship with him! (#280854)*
> *Negative: Timely clear communication with nursing staff and colleagues about patient concerns and availability. I often overhear nurses stating they have had trouble contacting Dr. Y about results or changes in a patient's condition, or that they don't know when he'll be coming in to do rounds. (#280781)*
> *Mixed: Not run behind on time so much but that would be difficult because she is so amazing and listens to your concerns. (#280875)*

Respondents provided similar numbers of very specific and semi specific comments (47% and 50% respectively), with 3% being not specific. Set 2 yielded more specific comments than did set 1 in response to Q1 ($p < 0.001$), but no differences were observed between sets for Q2. Overall, more specific comments were provided by co-workers than by medical colleagues ($p <0.001$). See tables 3 and 4.

Selected examples of how specificity varied include:

> *Not specific: Always humble and quiet, thus difficult to evaluate on some of the questions (#280399)*
> *Semi specific: Maintain the desire to continually improve and grow as a physician. (#280455)*
> *Very specific: Dr. X is an incredibly thorough doctor, she takes time to make sure the patient has all of the information they require and is caring and compassionate in all she does. It is truly a pleasure to work with her (#280560)*

The examination of comments for actionability included only the comments for Q2. The majority (300, 54%) of comments provided for Q2 were 'very' actionable. Set 1 yielded substantially more very actionable comments compared to set 2 (61% vs 45%, $p < 0.001$); set 2 yielded more comments classified as semi actionable or not actionable (30% vs 25% and 25% vs 13%), $p < 0.001$). Overall, co-workers provided fewer semi-actionable comments when compared with medical colleagues (21% vs 37%, $p<0.001$). See tables 3 and 4.

> *Not actionable: I am not sure, I think she works within her comfort zone and skill level. (#283724)*
> *Semi actionable: Understanding different patients have different needs (280266)*
> *Very actionable: In some of his responses to patients and co-workers he can come off very blunt, whether he means to or not, but recognizing that could really make a difference in the way patients see him. (#283977)*

Comments' content in relation to CanMEDS competencies

Many comments reflected more than one CanMEDS role classification for each doctor, and 3282 codes were provided for the 1800 comments. Set 1 yielded more codes than Set 2 for both Q1 and Q2. Of

these, the most frequent CanMEDS roles identified were collaborator and leader followed by medical expert (16%), professional and communicator (15% each), and scholar (13%). Comments relating to health advocacy were fewer. Except for the communicator and health advocate roles, Set 2 provided a greater number of comments for Q1. The scholar role showed the only differences between sets and only for Q2. See table 5. Co-workers provided more comments for CanMEDS roles, particularly for the roles of communicator, collaborator, leader, and professional, while medical colleagues provided more comments about medical expert. See table 6. Examples of comments provided for CanMEDS roles are shown in table 7 (on-line) and demonstrate the variability in comments but also the richness of some of the comments.

Discussion

This study contributes to the research on the utility of capturing free-text comments with MSF. The wording of questions inviting a free-text response appeared to influence the quantity and nature of the colleague comments provided. It is thus of importance to test the wording of questions intended to generate free-text responses to ensure that the questions posed are fit for purpose. MSF recipients have noted the limitations of MSF that contain only numeric data.[3-5] Free text comments offer the potential to provide the elaboration that is needed to understand scores, and to either affirm the good work that the physician is doing or to identify behaviours that might be targeted to improve practice. Both sets of questions generated more comments than in other studies, potentially related to asking two questions, rather than one question. As found in another study,[8] co-workers were more likely to provide comments than were medical colleagues.

Comments were more likely to be positive than negative, although our wording generated more negative comments per physician than has been observed in previous MSF studies.[7-10] Eliciting a greater volume of negative comments was probably attributable to the design, i.e. the wording used in the two questions. However, it is not clear why Set 1, Q2 would generate more comments than the more general wording option.

Given the importance of providing physicians with feedback specific enough to guide improvement, obtaining higher proportions of comments that were very specific and very actionable than other studies affirms the importance of asking the right questions to enhance the utility of MSF data. Indeed, our line of questioning with two questions in each set did provide twice as many comments which were very specific in nature, when compared to other studies.[7,9] It is possible that the open-endedness of Set 2 Q1 may have enabled respondents to be more expansive in their responses and for those responses to be coded as more specific than responses in which respondents commented on only one thing the physician did well. In contrast, for actionability, it appeared easier for respondents to consider and focus on just '1' thing the physician could target for action, in comparison to a broader question focused on enhancing practice. There are cautions in this. Examinations of comments must also take into consideration the cognitive processes involved in generating comments which came at the end of answering very specific questions. Providing narrative comments is a task requiring more cognitive effort than providing numeric ratings. Respondents had to draw on their long term memory, do problem solving different task than providing numeric ratings. Respondents had to draw on their long term memory, do problem solving to determine what the person could do to improve, and write down their thoughts.[13] Their ideas about improvement were likely influenced by their motivation, experience providing comments, writing skill, the perceived consequences of providing feedback, and social desirability pressures.[11,13] The cognitive processes involved in the task of generating comments may also

explain why physicians received about 8 comments when their 16 respondents might have generated 32 comments, had they answered both questions.

Respondents provided data about all of the CanMEDS roles. The largest numbers of comments were for collaborator and leader, perhaps because the MSF questionnaires for both medical colleague and co-workers included several items focusing on collaboration and teamwork. Conversely, there are relatively few items on the questionnaires addressing the health advocate role, an aspect of practice that is difficult to observe by other colleagues. There were differences by set with Set 1 producing more codes for all of the CanMEDS roles, but that may be associated with the larger numbers of comments associated with Set 1 and the ease of providing 'one thing' the physician did well or could target for action. There were differences in CanMEDS coding by source. Medical colleagues provided more comments than co-workers regarding the CanMEDS medical expert role, but this is likely an artefact of the PAR program as co-worker questionnaires do not include items pertaining to the medical expert role. While questionnaires for both sources contained items related to communicator, collaborator, leader, and professional; co-workers provided more comments about these roles and may have felt more comfortable commenting on these aspects. The comments provided very rich and complementary insights into the work the physician was doing as well as areas that could benefit from attention.

The study has limitations. It was conducted in one Canadian province. While we developed the coding framework collaboratively, the act of coding is subject to and influenced by the coder's interpretation of the data. To minimize bias, we adopted several approaches (e.g., double coding and discussion) to ensure coding was accurate and defensible.[23-25] Unlike other studies,[8,9] data were not available to analyse the association between scores physicians received on the PAR questionnaires and characteristics of the comments which may have yielded greater insights into the ways that respondents used comments. While the PAR program requires patient participation, the questionnaires used to elicit patient feedback were paper based and it wasn't feasible to solicit and electronically transcribe free-text comments from patients. We did not have an opportunity to follow up with the physicians who received the comments to determine their perceptions of the comments and whether the comments helped them use their PAR feedback data more effectively.

Conclusions and Implications

The study suggests that careful thought should be given to the wording of questions seeking to elicit narrative comments. Narrative comments are valued and they act in a support capacity for numerical ratings, providing context and helping the individual interpret their scores.[12,13] Carefully considering the number and nature of questions posed appears to offer potential to increase the volume and utility of the resulting comments. Being asked to identify one thing to target for action appeared to influence the numbers of responses received. Requesting comments from co-workers also increased the number, specificity and actionability of the comments. Our study garnered relatively more negative or corrective comments than earlier studies, potentially providing physicians with more guidance on how to improve. Comments elicited could be mapped across all of the CanMEDS roles. The numbers of comments related to collaborator, leader and communicator roles are particularly notable, given the critical role these competencies play in ensuring patient safety. From a performance monitoring perspective, comments are valuable as they provide insight into the different aspects of practice which medical colleagues and co-workers see as areas for the physician to address. For the physician, they demonstrate that their medical colleagues and co-workers are observing all aspects of their work and are able to offer constructive and useable guidance.

The study has implications for MSF practice and further research. MSF has been criticized when data is limited to numeric scores as they are insufficient to provide direction.[3-5] Increasing the questions and altering them can increase the number of comments provided. However, such questions do need to be targeted, and fit for the purpose for which they were intended. Work will need to be done to ensure that the comments are high yield. Perhaps, having the comments first might reduce the fatigue effects, decrease the priming that comes from having the ratings first, and increase the data available to affirm behaviour and guide improvement.[13] Also, further research with physicians receiving such narrative comments to determine how they respond to them and what they do with them, will further inform the work of making MSF data more useful to the physician.

Lessons for Practice

- Narrative comments from medical colleagues and co-workers can supplement numeric data provided with multisource feedback assessments.
- Care must be taken through careful design and wording of requests for narrative comments to ensure the data that medical colleagues and co-workers provide: (1) contains both positive (identifies what physicians do well) and negative content (identifies what they could improve), (2) are specific, and (3) are actionable to guide physician professional development.
- Multisource feedback data can cover the full spectrum of CanMEDS roles.
- Testing of questions is needed to ensure that physicians find narrative data helpful and guides practice improvement.

References

1. Lockyer J, Multisource Feedback, In ES Holmboe, SJ Durning, RE Hawkins (Eds), Practical Guide to the Evaluation of Clinical Competence, 2nd edition, Philadelphia: Elsevier, 2017, pp 204-214
2. Lockyer J, Sargeant J, Implementing Multisource Feedback, in D Boud and E Molloy (Eds), Feedback in Higher and Professional Education, London, Routledge, 2013.
3. Sargeant J, Mann K, Ferrier S. Exploring family physicians reactions to MSF performance assessment: perceptions of credibility and usefulness. Med Educ, 2005. 39:497-504
4. Sargeant J, Mann K, Sinclair D, van der Vleuten C, Metsemakers J. Challenges in multi source feedback: Intended and unintended consequences, Medical Education 2007, 41(6):581-596
5. Overeem K, Lombarts MJ, Arah OA, Klazinga NS, Grol RP, Wollersheim HC. Three methods of multi-source feedback compared: a plea for narrative comments and co-workers' perspectives. Med Teach. 2010;32(2):141-7
6. Ferguson J, Wakeling J, Bowie P. Factors influencing the effectiveness of multisource feedback in improving the professional practice of medical doctors: a systematic review. BMC Med Educ. 2014;14:76.
7. Canavan C, Holtman MC, Richmond M, Katsufrakis PJ, The Quality of Written Comments on Professional Behaviors in a Developmental Multisource Feedback Program, Acad Med, 2010, 85(10S): S106-S109.
8. Richards S H, Campbell JI, Walshaw E, Dickens A, Greco M, A multi-method analysis of free-text comments from the UK General Medical Council Colleague Questionnaires, Medical Education 2009; 43: 757-766.
9. van der Meulen MW, Boerebach BC, Smirnova A, Heeneman S, Oude Egbrink MG, van der Vleuten CP, Arah OA, Lombarts KM. Validation of the INCEPT: A Multisource Feedback Tool for Capturing Different Perspectives on Physicians' Professional Performance. J Contin Educ Health Prof. 2017;37(1):9-18.
10. Whitehouse A, Hassell A, Wood L, Wall D, Walzman M, Campbell I, Development and reliability testing of TAB, a form for 360 degrees assessment of senior house officers' professional behavior, as specified by the General Medical Council. Medical teacher 2005; 27(3): 252-258
11. Wilson KY, An analysis of bias in supervisor narrative comments in performance appraisal, Human Rel'ns. 2010; 63(12): 1903-1923.
12. David EM, Examining the role of narrative performance appraisal comments on performance, Human Performance 2013; 26:430-450.
13. Brutus S, Words versus numbers: A theoretical exploration of giving and receiving narrative comments in performance appraisal, Human Resource Management Review 2010; 20: 144-157
14. Smither JW, Walker AG, Are the characteristics of narrative comments related to improvement in multirater feedback over time? Journal of Applied Psychology 2004; 89: 575-581
15. Plymale MA, Donnelly MB, Lawton J, Pulito AR, Mentzer RM. Faculty evaluation of surgery clerkship students: important components of written comments. Acad Med. 2002;77(10 Suppl):S45-7.
16. Gulbas L, Guerin W, Ryder HF. Does what we write matter? Determining the features of high- and low-quality summative written comments of students on the internal medicine clerkship using pile-sort and consensus analysis: a mixed-methods study. BMC Med Educ. 2016 May 13;16:145.

17. Ginsburg S, Regehr G, Lingard L, Eva KW. Reading between the lines: faculty interpretations of narrative evaluation comments. Med Educ. 2015;49(3):296-306.
18. van der Leeuw RM, Schipper MP, Heineman MJ, Lombarts KM. Residents' narrative feedback on teaching performance of clinical teachers: analysis of the content and phrasing of suggestions for improvement. Postgrad Med J. 2016;92(1085):145-151.
19. Brouwers M, Kho ME, Browman GP, Cluzeau F, feder G, Fervers B, Hanna S, Makarski J on behalf of the AGREE Next Steps Consortium. AGREE II: Advancing guideline development, reporting and evaluation in healthcare. Can Med Assoc J. 2010: 182:E839-842.
20. Ratner RK, Riis J. Communicating science-based recommendations with memorable and actionable guidelines. Proc Natl Acad Sci U S A. 2014 Sep 16;111 Suppl 4:13634-13641.
21. Shoemaker SJ, Wolf MS, Brach C. Assessment Tool (PEMAT): a new measure of understandability and actionability for print and audiovisual patient information. Patient Educ Couns 2014; 96(3):395-403
22. Frank J, Snell L, Sherbino J, CanMEDS 2015: Physician Competency Framework, Ottawa: Royal College of Physicians and Surgeons of Canada, 2015.
23. Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. BMC Med Res Methodol. 2013; 18;13:117.
24. Braun V, Clarke V, Terry G, Thematic Analysis, In P Rohleder, AC Lyons, Qualitative Research in Clinical and Health Psychology, New York, Palgrave MacMillan, 2015
25. Braun V, Clarke V, Successful Qualitative Research: A Practical Guide for Beginners, London: Sage Publications, 2013

413

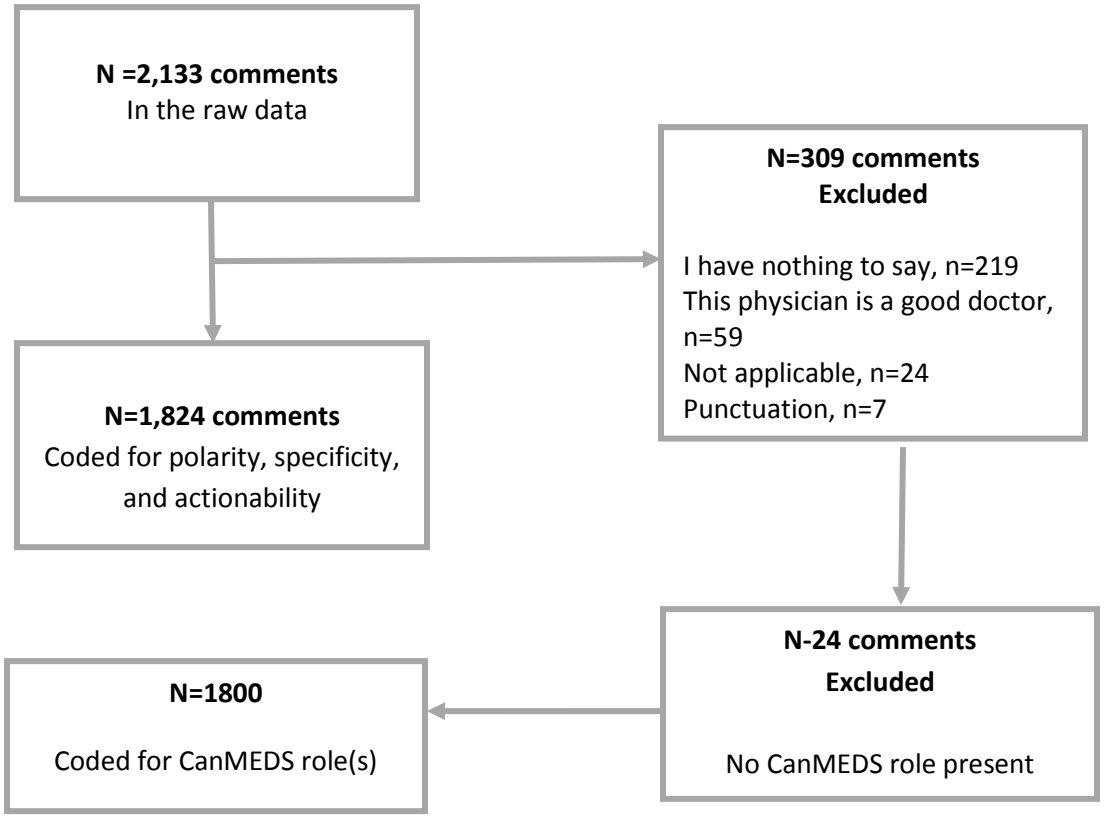

Figure 1: Flow diagram depicting comments available for coding by characteristic (polarity, specificity, and actionability) and CanMEDS role(s)

Table 1. Characterization of the "cohort" of comments.

	All comments	Comments received from co-worker reviewers	Comments received from medical colleague reviewers	Test statistic
Set 1 questions received, # physicians	111	106	98	
Set 2 questions received, # physicians	111	103	105	p=0.47 (NS)
Polarity, Specificity, Actionability				
Mean # comments (SD)	8.2 (4.0)	4.8 (2.7)	4.0 (2.3)	
Median (range) # comments	8 (1,24)	5 (1,17)	4 (1,11)	p<0.001
CanMEDS				
Mean # comments (SD)	8.1 (4.0)	4.8 (2.6)	4.0 (2.3)	p<0.001
Median (range) # comments	8 (1,23)	4 (1, 16)	4 (1,11)	
Comments generated (n, row %)	1824	1013	811	
Set 1, Q1	664	360 (54%)	304 (45%)	p=0.98 (NS)
Set 2, Q1	604	328 (54%)	276 (45%)	
Set 1, Q2	300	183 (61%)	117 (39%)	p=0.19 (NS)
Set 2, Q2	256	142 (55%)	114 (44%)	
All comments	1800	996	804	
Set 1, Q1 (n=660)	660	358	302	p=0.98 (NS)
Set 2, Q1 (n=604)	604	328	276	
Set 1, Q2 (n=290)	290	175	115	p=0.20 (NS)
Set 2, Q2 (n=246)	246	135	111	

*Statistical significance at p<0.05

Table 2. T-test results comparing mean number of comments generated per physician between Sets 1 and 2.

	Comments generated Mean Per Physician (Standard Deviation)	P-value
Q1 Set 1 Set 2 Difference between means (95% CI)	 5.98 (2.72) 5.44 (2.55) 0.54 (-0.15, 1.24)	 0.129 (NS)
Q2 Set 1 Set 2 Difference between means (95% CI)	 3.13 (1.80) 2.67 (1.35) 0.46 (0.01, 0.91)	 0.047

*Statistical significance at $p<0.05$

Table 3. Comparison of comments by set and by question

	All comments	Q1			Q2		
		Set 1	Set 2	p-value	Set 1	Set 2	p-value
Polarity	**1824**	**664**	**604**		**300**	**256**	
Positive	1394 (76%)	659 (99%)	600 (99%)	0.255 (NS)	53 (18%)	82 (32%)	<0.001*
Negative	329 (18%)	1 (0%)	3 (0%)		193 (64%)	132 (52%)	
Mixed	101 (6%)	4 (1%)	1 (0%)		54 (18%)	42 (16%)	
Specificity	**1824**	**664**	**604**		**300**	**256**	
Very specific	856 (47%)	324 (49%)	362 (60%)	<0.001*	93 (31%)	77 (30%)	0.077 (NS)
Semi specific	905 (50%)	334 (50%)	232 (38%)		189 (63%)	150 (59%)	
Not specific	63 (3%)	6 (1%)	10 (2%)		18 (6%)	29 (11%)	
							<0.001*
Actionability**	**556**	-	-	-	**300**	**256**	
Very actionable	300 (54%)	-	-		184 (61%)	116 (45%)	
Semi actionable	153 (27.5%)	-	-		76 (25%)	77 (30%)	
Not actionable	103 (18.5%)	-	-		40 (13%)	63 (25%)	
	-						

*Statistical significance at p<=0.05 **Only Q2 assessed

Table 4. Comparison of comments by source for polarity, specificity and actionability.

	All comments	Co-worker comments	Medical colleague comments	p-value
Polarity	**1824**	**1013**	**811**	
Positive	1394 (76%)	774 (76%)	620 (76%)	0.91(NS)
Negative	329 (18%)	181 (18%)	148 (18%)	
Mixed	101 (6%)	58 (6%)	43 (5%)	
Specificity	**1824**	**1013**	**811**	
Very specific	856 (47%)	571 (56%)	285 (35%)	<0.001*
Semi specific	905 (50%)	412 (41%)	493 (61%)	
Not specific	63 (3%)	30 (3%)	33 (4%)	
Actionability**	**556**	**325**	**231**	
Very actionable	300 (54%)	186 (57%)	114 (49%)	<0.001*
Semi actionable	153 (28%)	68 (21%)	85 (37%)	
Not actionable	103 (19%)	71 (22%)	32 (14%)	

*Statistical significance at p< =0.05 **Only Q2 assessed

Table 5. Comments examined by CanMEDS roles by type of question.

		Q1 (n, column %, row %)			Q2 (n, column %, row %)		
	All comments	Set 1	Set 2	Test statistic	Set 1	Set 2	Test statistic
All codings	3282	1169	1140		364	309	
Medical expert	510 (16%)	170 (15%; 40%)	253 (22%; 60%)	P<0.001*	41 (11%; 47%)	46 (15%; 52%)	P=0.154 (NS)
Communicator	508 (15%)	218 (19%; 52%)	203 (18%; 48%)	P=0.827 NS	48 (13%; 55%)	39 (13%; 45%)	P=0.827 (NS)
Collaborator	634 (19%)	254 (22%; 47%)	288 (25%; 53%)	P=0.001*	50 (14%; 54%)	42 (14%; 46%)	P=0.959 (NS)
Leader	623 (19%)	149 (13%; 40%)	220 (19%; 60%)	P<0.001*	148 (41%; 58%)	106 (34%; 42%)	P=0.066 (NS)
Health advocate	91 (3%)	38 (3%; 47%)	43 (4%; 53%)	P=0.323 NS	5 (1%; 50%)	5 (2%; 50%)	P=0.793 (NS)
Scholar	426 (13%)	130 (11%; 39%)	206 (18%; 61%)	P<0.001*	38 (10%; 42%)	52 (17%; 58%)	P=0.013*
Professional	490 (15%)	210 (18%; 48%)	227 (20%; 52%)	P=0.031*	34 (9%; 64%)	19 (6%; 36%)	P=0.122 (NS)

*Statistical significance at p< =0.05

Table 6. Comments examined by CanMEDS roles by source

	All comments (n, column %)	Co-worker comments (n, column %, row %)	Medical colleague comments (n, column %, row %)	Test statistic
All codings	3282	1915	1077	
Medical expert	510 (16%)	220 (11%; 43%)	290 (27%; 57%)	P<0.001*
Communicator	508 (15%)	330 (17%; 65%)	178 (17%; 35%)	P<0.001*
Collaborator	634 (19%)	415 (22%; 65%)	219 (20%; 35%)	P<0.001*
Leader	623 (19%)	372 (19%; 60%)	251 (23%; 40%)	P=0.007*
Health advocate	91 (3%)	50 (3%; 55%)	41 (4%; 45%)	P=0.939 (NS)
Scholar	426 (13%)	230 (12%; 54%)	196 (18%; 45%)	P=0.523 (NS)
Professional	490 (15%)	298 (16%; 61%)	192 (18%; 45%)	P=0.004*

*Statistical significance at p< =0.05

Table 7. Examples of comments, by CanMEDS coding and by type of reviewer. (On-line table)

	Co-worker comment example	Medical Colleague comment example
Medical Expert	It has been a pleasure to have this MD do recurring locums at our facility, so much so that he has become familiar with many of our patients, and calls to follow up even after he has left. Continuing to build on his emergency skills, like chest tubes, central lines, and intubations would enhance his practice. This is, however, not saying that Dr [removed first name] is incompetent. I am only reinforcing that these are areas most locums could use more practice at. [280581]	He understands the clinical presentation of acute surgical abdomen such as appendicitis that can save costs. Most of the time, clinicians should be able to make a clinical diagnosis. I once had a patient that I discussed with him and I made the diagnosis of appendicitis and he agreed based on clinical examination. The patient had an ultrasound that was not conclusive, but the CT finding reinforced my thinking and his thinking, patient was managed appropriately, but had to be transferred to another site for CT before being transferred back to his site. [283493] Palliative medicine care, can be complex with significant psychosocial and spiritual distress contributing greatly to a patient's (and families) distress. Dr. D. is excellent at understanding and addressing the psycho-social/spiritual issues which may (and often do so) impact the overall patient and family palliative experience and response to treatment. [283906]
Communicator	Dr. F. takes the time to make his patients feel comfortable and relaxed and informs them about their condition. He does not rush them. He answers all questions appropriately. He portrays confidence and knowledge to his patients. [283493] Explains the diagnosis, the possible causes, the treatment plan and involves the patient in the decision making of the direction they want to take. Sometimes the language barrier (accent) and speed in which he speaks can be difficult for the elderly patients. [280378]	I have heard on more than 1 occasion from multiple patients of mine how Dr T takes the time to listen, work through & explain in words they can understand their medical issues. [283577]
Collaborator	Dr G listens well. In difficult cases, when I have input on a situation, he listens to my suggestions. [284053] He is very courteous to all staff involved, being clear in what he would like and how, yet is open to discussion when applicable. [284102]	Dr. H. uses his gentle style to effectively and non-judgmentally discuss patient management improvement with referring physicians. [283605]
Leader	Dr. K. makes the working environment happy and fun, makes me enjoy my job that much more. [283563] I appreciate the leadership Dr. L. has shown, supporting clinical operations by actively participating in decisions and following up on outstanding issues. [283605]	Needs how to more effectively and efficiently run a multidisciplinary team allowing all to contribute. [284039]
Health Advocate	Advocates for X program with patients always in the forefront [284144]	Advocates for the best treatment for patients, even if it is not readily available. [284039] Explore the socioeconomic aspects of disease more and where applicable to his patients. [280518]

Scholar	He teaches, supports and encourages critical thinking skills from the nursing staff in the department. I have learned a great deal from him. [284151]	Dr J. stays current with literature and when the information is not available from her recent reviews, she is quick to inform herself (and then share with others) findings of up to date reviews and literature. [28363] Very current with medical literature, and efficiently uses [it] in patient management. [280273]
Professional	It is great pleasure to work with Dr. O.. He is one of the most professional, personable and kind physicians I have worked with. [284004] Dr. C. demonstrates exceptional compassion and concern for his patients. [284039]	Dr A. is particularly kind and caring, and demonstrates excellence in his manner and medical care in every way. I am delighted to have him as a colleague. [280245] For a relatively young inexperienced physician, she demonstrates a high level of skill and professionalism; it's very encouraging to see a young colleague with such exemplary traits. [280273]
Multiple Roles	Dr P. is an extremely collaborative, dedicated professional and very competent physician who effectively uses all the interdisciplinary team members in his clinic and others to produce the best health outcomes for his patients. Wonderful to work with. [280357 collaborator, leader] Dr. R. is an excellent communicator. Without exception, she communicates every day in a courteous, effective and respectful manner to everyone she is in contact with. This includes the housekeeping staff to nurses, other physicians, family of patients, patients, pharmacy staff, physiotherapists, absolutely no one is ever treated in an untoward manner. This behavior is unheard of in the world in general. It is a true pleasure to work with Dr. [last name removed]. She is considerate, thoughtful and intelligent, keeping herself up to date with current best practices. [280490, communicator, collaborator, scholar]	The patient's overall well-being is the focus of the care provided by Dr. M. He is very respectful of his patents and their specific needs. He is very courteous, kind and considerate about his staff and fellow workers. He handles stressful situations very calmly and professionally. [280595 , leader, professional] Dr N. is a deeply caring physician. She is a mentor to me. When we are having case discussions at rounds, she is quick to raise psychosocial concerns and to place the primary problem in the larger context of the patient. She actively encourages clear communication amongst members, and always clearly designates coverage of her patients when she is away. She also actively supports learners of all levels and encourages us to advance our own learning. [295398 , health advocate, scholar, collaborator] Extraordinary care of the patient. Goes above and beyond to reach patients and provide additional resources to them if necessary - not hesitant to consult and advocate on patient's behalf. [280770, collaborator, health advocate]