Canadian Medical Education Journal

Brief Reports

Use the right words: evaluating the effect of word choice and word count on quality of narrative feedback in ophthalmology competency-based medical education assessments

Rachel Curtis,[1] Christine C Moon,[2] Tessa Hanmore,[1,3,4] Wilma M Hopman,[5] Stephanie Baxter[1]

[1]Department of Ophthalmology, Queen's University, Ontario, Canada; [2]School of Medicine, Queen's University, Ontario, Canada; [3]Department of Physical Medicine and Rehabilitation, Queen's University, Ontario, Canada; [4]Department of Psychiatry, Queen's University, Ontario, Canada; [5]KGH Research Institute and Department of Public Health Sciences, Queen's University, Ontario, Canada

Correspondence to: Dr. Rachel Curtis, MD, FRCSC; Department of Ophthalmology, Kingston Health Sciences Centre-Hotel Dieu Hospital Site and Queen's University, 166 Brock Street, Kingston, ON, K7L 5G2; email: 8rc16@queensu.ca

Published ahead of issue: Apr 29, 2024; CMEJ 2024 Available at https://doi.org/10.36834/cmej.76671

Abstract

Background: The purpose of this study was to investigate the effect of word choice on the quality of narrative feedback in ophthalmology resident trainee assessments following the introduction of competency-based medical education at Queen's University.

Methods: Assessment data from July 2017-December 2020 were retrieved from Elentra™ (Integrated Teaching and Learning Platform) and anonymized. Written feedback was assigned a Quality of Assessment for Learning (QuAL) score out of five based on this previously validated rubric. The correlation between QuAL score and specific coaching words was determined using a Spearman's Rho analysis. Independent samples t-tests were used to compare the QuAL score when a specific word was used, and when it was absent.

Results: A total of 1997 individual assessments were used in this analysis. The number of times the identified coaching words were used within a comment was significantly and positively associated with the total QuAL score, with the exception of "next time" (rho=0.039, p=0.082), "read" (rho = 0.036, $p = 0.112$), "read more" (rho = -0.025, $p = 0.256$) and "review" (rho = -0.017, $p = 0.440$). The strongest correlations were for "continue" (rho = 0.182, $p < 0.001$), "try(ing)" (rho = 0.113, $p < 0.001$) and "next step" (rho = 0.103, $p < 0.001$). The mean value of the QuAL score increased when coaching words were used vs. not used with the largest mean difference of 1.44 ($p < 0.001$) for "reflect". A clear positive relationship was demonstrated between word count and QuAL score (rho = .556, $p < 0.001$).

Conclusions: The use of certain coaching words in written comments may improve the quality of feedback.

Résumé

Résumé français à venir.

Introduction

Narrative comments comprise a large part of assessment in Competency-Based Medical Education (CBME) and provide a record of faculty feedback and coaching directed towards the learner.[1] While there is an extensive body of research that identifies narrative comments as an essential part of CBME,[1-5] few studies have explored what language contributes to quality in written assessments. As medical residency training programs transition from traditional time-based models to competency-based and hybrid models,[6-9] there is a growing need to understand how feedback delivery may be optimized. Our study uses a quantitative method of evaluating qualitative written narrative feedback; at present there are four validated tools to evaluate the quality of narrative comments in the context of CBME.[10-15]

When thoughtfully composed, narrative feedback is a personalized commentary on resident performance.[16] Effective feedback has been described as timely, specific, and actionable, with an emphasis on coaching behaviors versus high-stakes assessment.[17]

Understanding the ingredients that contribute to excellent quality feedback may help guide evaluators to refine their word choice and length of comment to be the most effective. Presented at the International Conference on Residency Education (ICRE) in 2015, Ross introduced five words/phrases commonly seen in high quality narrative feedback.[18] More recently, Branfield Day et al. identified similar phrases in assessment comments that conveyed coaching language to foster learning.[3] These phrases help frame strategies to assist residents in building their skills and knowledge; for example, beginning a sentence with "remember that…" was often followed by specific, actionable and detailed suggestions for improvement.[3] Feedback that uses coaching language instead of generalized descriptions of the learning interaction is more effective, and signals recommendations for resident improvement.[19]

One of the commonly cited barriers to faculty participation in CBME is time; there is an ever-expanding amount of clinical, teaching, and academic duties for a teaching physician.[16,20] With competing interests and depleting resources, making feedback delivery efficient by using words/phrases with the most impact ensures that coaching quality does not suffer under these constraints. In addition, the relationship between the quantity of words applied to a comment and the quality of feedback is relatively unexplored in assessment.[21] Those that have explored this relationship have found that longer written comments are correlated with better quality feedback.[13,22,23] However, a "sweet-spot" of length that is not formulaic, but provides guidance to optimize quality, allows evaluators to be aware of an approximate length of comment before plateauing into "extra words" for the sake of length.

In July 2017, Queen's University, in Kingston, Ontario, Canada implemented CBME for all 28 postgraduate specialty training programs.[8] As such, the Queen's University DOO was the first ophthalmology program in Canada to be fully immersed in CBME and has assessment data of trainees over this time period. The purpose of this study was to investigate the relationship between coaching language, word count and the quality of written feedback in resident assessments. Inter-rater agreement for the total QuAL score was previously established as excellent.[24] Ultimately, by guiding purposeful word choice and length of written feedback we hope to optimize the effectiveness and efficiency of feedback delivery in the context of CBME.

Methods

Study design

This retrospective cohort study was conducted at Queen's University and was approved by the Queen's University and Affiliated Hospitals Health Sciences Research Ethics Board (TRAQ 6029081). Ophthalmology resident assessment data from July 2017 to December 2020 were included.

Sample size

A total of 1997 assessments contained narrative comments, and were scored and analyzed.

Study protocol

Ophthalmology resident assessment data were retrieved from Elentra™ (Integrated Teaching and Learning Platform) and anonymized. The data were coded with unique identifiers and names were removed. The anonymized data were entered into an Excel sheet by a research assistant separate from the grading process. Written feedback was assigned a QuAL score.[15] The QuAL score consists of three components. The first (Evidence) is a 4-level item that asks, "Does the rater provide sufficient evidence about resident performance?," where zero indicates no comment at all, and three a full description. The second (Suggestion) and third (Connection) are binary, where zero indicates "no" and one indicates "yes" in response to the questions, "Does the rater provide a suggestion for improvement?" and "Is the rater's

suggestion linked to the behavior described."[15] All individual assessments were scored by an ophthalmology faculty member (SB), and a randomized sample of 10% was independently rescored by a final year ophthalmology resident (RC) to ensure inter-rater reliability. Both raters were blinded to any identifying information and graded independently of one another. The Intra-class correlation coefficient (ICCs) for the two graders was excellent at 0.90 (95% CI 0.88-0.92, $p < 0.001$).[24] All QuAL scores were completed prior to the coaching word analysis; the two raters did not have specific knowledge of identified coaching words in the literature prior to scoring the narrative comments.

Outcome measures

The primary outcomes of our study were the associations between QuAL score and specific coaching language ("suggest," "try(ing)," "because," "consider," "next step," "continue," and "next time"). These words/phrases were selected based on preliminary work by Ross (2015), with overlap from research conducted by Branfield Day.[3,18] "Continue" and "next time" were included as they were pre-existing prompts in the comments section of the evaluation forms. In addition, the words "discuss," "recognize," "demonstrate," "remember," "reflect," and "practice" were chosen by our research group as language that was potentially associated with better quality feedback. Commonly used phrases generally perceived as components of poor quality feedback ("read," "read more," and "review") were examined and were intended to represent negative controls.[25] QuAL scores were assigned to each assessment prior to the identification of coaching words and negative control phrases.

Data analysis

Data were imported into IBM SPSS (Version 27.0, Armonk, NY, 2021) for statistical analysis. The correlation between the number of words and the QuAL score was explored using Spearman's Rho. The correlation between the number of *times* each comment contained the specific words or phrases, and the QuAL score, was also assessed with the Spearman's Rho. Independent samples t-tests were used to compare the mean QuAL scores. To supplement the initial Spearman correlation and provide more detail about the QuAL score at different levels, the word count was subdivided into groups of approximately 20% (10% categories after 55 words due to the large range up to 283) including 1-15, 16-30, 31-55, 56-80, and 81+. One-way ANOVA was used to examine the mean QuAL score for each of the five groups, with Tukey's post hoc tests utilized to compare each category to all others. Differences were considered statistically significant if $p < 0.05$, and no adjustment was made for multiple comparisons.

Results

Assessments were collected from 20 different residents spanning postgraduate training years 1-5. The average QuAL score for all 1997 assessments was 3.07.

Frequency of coaching word use

The number of times that a coaching word was used within each comment ranged from zero to three. Table 1 provides the number of times coaching words were used once or twice in each assessment.

Table 1. Frequency of use of coaching words in all available assessments with narrative feedback (N = 1997), and correlation between total QuAL score and the number of times a coaching word was used within the comments

Word or phrase	Frequency used once *n* (%)	Frequency used twice *n* (%)	Total Frequency *n* (%)	Spearman's Rho	p-value
Suggest	67 (3.4)	0	67 (3.4)	0.063	0.005*
Try(ing)	95 (4.8)	2 (0.1)	97 (4.9)	0.113	<0.001*
Because	30 (1.5)	1 (0.1)	31 (1.6)	0.055	0.015*
Consider	137 (6.9)	2 (0.1)	139 (7.0)	0.078	0.001*
Next Step	78 (3.9)	0	78 (3.9)	0.103	<0.001*
Continue	289 (14.5)	18 (0.9)	307 (15.4)	0.182	<0.001*
Next Time	15 (0.8)	0	15 (0.8)	0.039	0.082
Discuss	218 (10.9)	9 (0.5)	227 (11.4)	0.062	0.006*
Recognize	71 (3.6)	3 (0.2)	74 (3.8)	0.085	<0.001*
Demonstrate	27 (1.4)	3 (0.2)	30 (1.6)	0.057	0.011*
Remember	79 (4.0)	1 (0.1)	80 (4.1)	0.064	0.004*
Practice	107 (5.4)	7 (0.4)	116 (5.9)[@]	0.061	0.007*
Reflect	11 (0.6)	1 (0.1)	12 (0.7)	0.075	<0.001*
Read	86 (4.3)	3 (0.2)	89 (4.5)	0.036	0.112
Read More	22 (1.1)	1 (0.1)	23 (1.2)	-0.250	0.256
Review	157 (7.9)	6 (0.3)	163 (8.2)	-0.017	0.440

*[@] "Practice" was used three times in one assessment. p-values are based on the Spearman Rho. *=Statistically significant.*

Correlation between total QuAL score and coaching word use

The number of times a coaching word was used within a comment was significantly and positively associated with the total QuAL score for all coaching words. The strongest correlations were for the words/phrases "continue," "try(ing)," and "next step." The negative control words/phrases "read more" and "review" were negatively correlated with the QuAL score, see Table 1.

The effect of coaching words on mean QuAL score

As shown in Table 2, the mean value of the QuAL score increased when coaching words were present; this mean difference was significant for all words except for "next time" and "read."

Table 2. Mean value of the QuAL score when coaching words were used, versus not used

Word or phrase	QuAL Score when Present	QuAL Score when Absent	Mean Difference (Present-Absent)	p-value
Suggest	3.57	3.05	0.52	0.006*
Try(ing)	3.80	3.03	0.77	<0.001*
Because	3.71	3.06	0.65	0.017*
Consider	3.49	3.04	0.45	0.001*
Next Step	3.83	3.04	0.79	<0.001*
Continue	3.71	2.95	0.76	<0.001*
Next Time	3.73	3.06	0.67	0.087
Discuss	3.31	3.04	0.27	0.011*
Recognize	3.72	3.05	0.67	<0.001*
Demonstrate	3.77	3.06	0.71	0.011*
Remember	3.53	3.05	0.48	0.006*
Practice	3.43	3.05	0.40	0.008*
Reflect	4.50	3.06	1.42	<0.001*
Read	3.31	3.06	0.25	0.118
Read More	2.74	3.07	-0.33	0.290
Review	2.98	3.08	-0.10	0.404

*p-values are based on the independent samples t-tests. *=Statistically significant.*

Word count and QuAL score

There was a significant correlation between the number of words used and the QuAL score, with a Spearman's Rho value of 0.556 ($p < 0.001$). The number of words included in feedback comments ranged from 0 (these 481 assessments were excluded from the analysis), to 283 words. The word count was subdivided into groups of 20% to determine the relationship between increasing word count and QuAL score as seen in Figure 1 and subdivided into 10% categories after 55 words**.** The one-way ANOVA and Tukey's post-hoc tests indicated that each category represented a significant increase from the previous ($p < 0.001$ for all), with the exception of the last two categories, 55-80 and 81+ ($p = 0.444$).

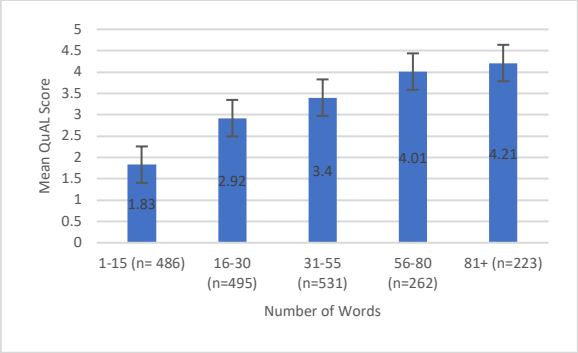

Figure 1. The number of words used in assessment narrative feedback versus the mean QuAL score. Standard error bars are represented on the graph. Tukey's post-hoc tests indicate statistically significant differences between all levels ($p<0.001$) for all but 56-80 versus 81+, where $p=0.444$.

Discussion

As the first ophthalmology program in Canada to fully integrate CBME into the core of their residency training, this study offers a unique and early perspective to help inform program development.

Our most compelling result is that when specific coaching words are used in narrative feedback, the QuAL score is consistently increased. This relationship was most notable for the words "next step," "try(ing)," and "continue." The phrases "next time," "read," "read more," and "review" were unsurprisingly poorly or negatively correlated with the QuAL score.[25] These generic phrases are non-specific and less helpful for targeted learner development.

We suggest that coaching language be encouraged to help guide and frame narrative comments. At our center we have recently modified the structure of some of our assessment forms to include a list of suggested prompts to encourage the use of coaching phrases in the free text feedback boxes.

Our analysis yielded a few surprising results. The phrases "next time" and "next steps" were infrequently used in our pool of narrative feedback; however, our forms use the phrases "next steps," "next time," and "continue" as prompts for the text-field, and we surmise that these words were underutilized in the body of the comments due to repetition. Predictably, there was a clear relationship initially between greater mean QuAL score and increasing word count. We had anticipated a plateau to this trend much earlier than demonstrated in our analysis; this was eventually seen, but not until after 80 words (Figure 1). This may seem discouraging that outstanding feedback quality can seemingly only be achieved with lengthy comments.

However, we argue that with increased use of strategic coaching language, the length of comment can be shorter while achieving the same quality of feedback. In our analysis of early assessment data, the densest concentration of high achieving QuAL scores (4/5 and 5/5 grades) is not at the far end of the number of words spectrum; there are numerous succinct written comments achieving high QuAL scores in our dataset. Roberts et al. found that written feedback could be both succinct (on average less than 20 words per comment) and categorized as coaching feedback with recommendations for next steps.[19]

As the culture of assessment shifts and CBME becomes engrained in PGME across Canada, we have the opportunity to direct focus on optimizing narrative feedback to train not just competent, but excellent physicians.

Limitations

All assessments were from a single surgical subspecialty at a single center. Some comments may have been composed by the resident receiving the feedback; one option for assessment completion on Elentra™ allows both the resident and the assessor to contribute to the form before final submission. Although we believe that the majority of comments were not generated by the resident, it is impossible to deduce who wrote what components of the narrative feedback. Despite excellent inter-rater reliability, both graders were physicians, familiar with the clinical context of the feedback and department experts in CBME.

Conclusions

Using this QuAL score, we have shown that strategically used coaching words can enhance the quality of narrative feedback in assessments. Although increased word count is associated with a higher QuAL score, there is a demonstrated plateau to this relationship.

Conflicts of Interest: No authors have conflicts of interest to disclose. The authors do not have any proprietary interests in the materials described in the article. This project was completed at Queen's University.
Presentations: The contents of this manuscript have not been presented at any meetings prior to submission, nor have they been published elsewhere. The Queen's ophthalmology CBME research group has presented findings from the same dataset at COS June 2022 and ICRE 2022, focusing on distinct analyses, in addition to an abstract presented at ICAM in 2023. Our manuscript "Evaluating the effect of assessment form design on the quality of feedback in one Canadian ophthalmology residency program as an early adopter of CBME," has been published as a research letter in the Canadian Journal of Ophthalmology and contains the Intra Class Correlation value for the total QuAL score also mentioned in this manuscript; this has been stated and cited in the paper.
Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.
Edited by: Cindy Schmidt (senior section editor); Marcel D'Eon (editor-in-chief)

References

1. Holmboe ES, Sherbino J, Long DM, Swing SR, Frank JR, International CBME Collaborators. The role of assessment in competency-based medical education. *Med Teach*. 2010 Aug 1;32(8):676-82. https://doi.org/10.3109/0142159X.2010.500704
2. Marcotte L, Egan R, Soleas E, Dalgarno N, Norris M, Smith C. Assessing the quality of feedback to general internal medicine residents in a competency-based environment. *Can Med Ed J.* 2019 Nov;10(4):e32. https://doi.org/10.36834/cmej.57323
3. Branfield Day L, Rassos J, Billick M, Ginsburg S. 'Next steps are…': An exploration of coaching and feedback language in EPA assessment comments. *Med Teach*. 2022 Aug 8:1-8. https://doi.org/10.1080/0142159X.2022.2098098
4. Tekian A, Borhani M, Tilton S, Abasolo E, Park YS. What do quantitative ratings and qualitative comments tell us about general surgery residents' progress toward independent practice? Evidence from a 5-year longitudinal cohort. *Amer J Surg*. 2019 Feb 1;217(2):288-95. https://doi.org/10.1016/j.amjsurg.2018.09.031
5. Ginsburg S, van der Vleuten CP, Eva KW. The hidden value of narrative comments for assessment: a quantitative reliability analysis of qualitative data. *Acad Med*. 2017 Nov 1;92(11):1617-21. https://doi.org/10.1097/ACM.0000000000001669
6. Stodel EJ, Wyand A, Crooks S, Moffett S, Chiu M, Hudson CC. Designing and implementing a competency-based training program for anesthesiology residents at the University of Ottawa. *Anesthesiol res pract*. 2015 Dec 21;2015. https://doi.org/10.1155/2015/713038.
7. Caccia N, Nakajima A, Scheele F, Kent N. Competency-based medical education: developing a framework for obstetrics and gynaecology. *J obstet gyn Can*. 2015 Dec 1;37(12):1104-12. https://doi.org/10.1016/S1701-2163(16)30076-7.
8. Stockley D, Egan R, Van Wylick R, et al. A systems approach for institutional CBME adoption at Queen's University. *Med Teach.* 2020 Aug 2;42(8):916-21. https://doi.org/10.1080/0142159X.2020.1767768.
9. Jurd S, de Beer W, Aimer M, Fletcher S, Halley E, Schapper C, Orkin M. Introducing a competency based fellowship programme for psychiatry in Australia and New Zealand. *Austral Psych*. 2015 Dec;23(6):699-705. https://doi.org/10.1177/1039856215600898.
10. Dudek NL, Marks MB, Wood TJ, Lee AC. Assessing the quality of supervisors' completed clinical evaluation reports. *Med Ed.* 2008 Aug;42(8):816-22. https://doi.org/10.1111/j.1365-2923.2008.03105.x.
11. Cheung WJ, Dudek N, Wood TJ, Frank JR. Daily encounter cards—evaluating the quality of documented assessments.

JGME. 2016 Oct;8(4):601-4. https://doi.org/10.4300/JGME-D-15-00505.1.

12. Roshan A, Wagner N, Acai A, et al. Comparing the quality of narrative comments by rotation setting. *J Surg Educ*. 2021 Nov 1;78(6):2070-7. https://doi.org/10.1016/j.jsurg.2021.06.012
13. Roshan A, Farooq A, Acai A, et al. The effect of gender dyads on the quality of narrative assessments of general surgery trainees. *Amer J Surg*. 2022 Jul 1;224(1), 179-184. https://doi.org/10.1016/j.amjsurg.2021.12.001
14. Ross S, Hamza D, Zulla R, Stasiuk S, Nichols D. Development of and preliminary validity evidence for the EFeCT feedback scoring tool. *JGME*. 2022 Feb;14(1):71-9. https://doi.org/10.4300/JGME-D-21-00602.1
15. Chan TM, Sebok-Syer SS, Sampson C, Monteiro S. The Quality of Assessment of Learning (Qual) score: validity evidence for a scoring system aimed at rating short, workplace-based comments on trainee performance. *Teach Learn Med*. 2020 May 26;32(3):319-29. https://doi.org/10.1080/10401334.2019.1708365.
16. Braund H, Dalgarno N, McEwen L, Egan R, Reid MA, Baxter S. Involving ophthalmology departmental stakeholders in developing workplace-based assessment tools. *Can J Opthamol.* 2019 Oct 1;54(5):590-600. https://doi.org/10.1016/j.jcjo.2019.01.013
17. Tomiak A, Braund H, Egan R, et al. Exploring how the new entrustable professional activity assessment tools affect the quality of feedback given to medical oncology residents. *J Cancer Educ*. 2020 Feb;35(1):165-77. https://doi.org/10.1007/s13187-018-1456-z
18. Ross S, Huie M, Schipper S. *Use words that count: a content analysis to identify words and phrases that commonly appear in effective formative feedback*. International Conference on Residency Education 2015 Oct 22.
19. Roberts A, Jellicoe M, Fox K. How does a move towards a coaching approach impact the delivery of written feedback in undergraduate clinical education?. *Advances Health Sci Educ.* 2022 March; 1:1-5. https://doi.org/10.1007/s10459-021-10066-7
20. Steinmann AF, Dy NM, Kane GC, et al. The modern teaching physician—responsibilities and challenges: an APDIM white paper. *Am J Med*. 2009 Jul 1;122(7):692-7. https://doi.org/10.1016/j.amjmed.2009.03.020
21. Hauer KE, Nishimura H, Dubon D, Teherani A, Boscardin C. Competency assessment form to improve feedback. *Clin Teach* 2018 Dec;15(6):472-7. https://doi.org/10.1111/tct.12726
22. Bismil R, Dudek NL, Wood TJ. In-training evaluations: developing an automated screening tool to measure report quality. *Med Ed*. 2014 Jul;48(7):724-32. https://doi.org/10.1111/medu.12490
23. Chan TM, Sebok-Syer SS, Yilmaz Y, Monteiro S, Syer SS. The impact of electronic data to capture qualitative comments in a competency-based assessment system. *Cureus*. 2022. *14*(3). https://doi.org/10.7759/cureus.23480
24. Curtis R, Moon CC, Hanmore T, Hopman W, Baxter S. Evaluating the effect of assessment form design on the quality of feedback in one Canadian ophthalmology residency program as an early adopter of CBME. *Can J Ophthalmol.* 2023 Jan; 58(4): E149-150. https://doi.org/10.1016/j.jcjo.2023.01.003 .
25. Zelenski AB, Tischendorf JS, Kessler M, et al. Beyond "read more": an intervention to improve faculty written feedback to learners. *JGME*. 2019 Aug 1;*11*(4), 468-471. https://doi.org/10.4300/JGME-D-19-00058.1