SHORT REPORTS

Validity and Utility of the CanMEDS "Resident as Teacher Multisource Feedback" Assessment Tool for Resident-led Structured Teaching

Janice C.L. Lee, MD, MScCH[1,2], Yu Qing Huang, MDCM[1,2], Camilla L. Wong, MD, MHSc[1,2]

[1]*Division of Geriatric Medicine, Temerty Faculty of Medicine, University of Toronto;* [2]*Unity Health Toronto, Toronto, ON*

https://doi.org/10.5770/cgj.28.795

ABSTRACT

We evaluated the validity of using the CanMEDS Resident as Teacher Multisource Feedback (RaTMSF) assessment tool to gather learner feedback from structured resident-led teaching within the University of Toronto's postgraduate geriatric medicine residency program. The RaTMSF consists of 10 rated items and narrative comments. Completed RaTMSF evaluations from resident teachers were analyzed by descriptive statistics for internal consistency and inter-rater reliability, and narrative comments were reviewed for thematic content. Resident teachers were surveyed on the acceptability of the tool to develop teaching competencies. A total of 132 evaluations were collected prospectively from 11 residents from April 2021 to April 2022, and retrospectively from seven graduates from 2016 to 2019. The overall performance rating, 4.75 (SD 0.47), was very positive for all resident teachers. The RaTMSF demonstrated high internal consistency with Cronbach's alpha of 0.97, 95% CI 0.89–0.99 between all 10 items, and good inter-rater reliability with Fleiss kappa of 0.73 (95% CI 0.13–0.80). The most common themes of narrative comments also captured in the rated items were organization to teach (n=53) and openness to questions (n=36). Written comments regarding delivery style (n=52) and audience interactivity (n=44) were not captured on the rated items. While most resident teachers surveyed found the RaTMSF acceptable to use, we suggest opportunities to improve the RaTMSF by restructuring focus onto written feedback and revising rated items to better reflect themes found in narrative comments. The RaTMSF can be a valuable feedback tool to help residents gather high-quality feedback on their teaching skills.

Key words: medical education, postgraduate, teaching, assessment, validity

Introduction

The University of Toronto (UofT)'s postgraduate geriatric medicine residency program transitioned to the Competence by Design (CBD) medical education model in 2019.(1) Residents are required to demonstrate evidence of competency to "teach students, residents, the public, and other health professionals".(2,3) In the CBD model, small frequent low-stakes assessments are crucial to the curriculum. Residents' teaching skills are now being assessed by entrustable professional activities (EPAs) completed by faculty members. The implementation of EPAs has not changed the program's ongoing use of the CanMEDS Resident as Teacher Multisource Feedback (RaTMSF) assessment tool to enhance the collection of quality feedback from learners since 2016.(4,5) However, its validity has not been studied. The RaTMSF results are not formally factored into competency decision-making process of residents. We wanted to investigate the utility of the RaTMSF as a complementary assessment tool to EPAs for resident-led teaching.

The aim of our study was to explore whether UofT's geriatric medicine residency program should continue using the RaTMSF as-is for resident teachers conducting structured small group didactic teaching. We evaluated the validity of using the RaTMSF for instrument validity, interpretation of scores, and its utility of actions resulting from the tool to justify continued use in its current form within the program.(6)

METHODS

Our study used mixed methods by combining qualitative content analysis to identify common themes and quantitative analyses to derive validity measures of the RaTMSF. We followed the validity framework in educational testing proposed by Russel *et al*.(6)

Study Setting

This study was conducted within the two-year geriatric medicine subspecialty residency program at UofT, whose curriculum is situated between four academic sites in Toronto. During each four-week clinical rotation, residents were encouraged to use the RaTMSF for scheduled small group teaching presentations.

© 2025 Author(s). Published by the Canadian Geriatrics Society. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial No-Derivative license (https://creativecommons.org/licenses/by-nc-nd/4.0/), which permits unrestricted non-commercial use and distribution, provided the original work is properly cited.

Study Participants

All 11 geriatric medicine subspecialty residents participated in the study between April 2021 and April 2022. They were assessed as resident teachers by learners including medical students, residents, fellows, allied health members, and geriatricians. We obtained informed consent from all resident teachers and learners. This study received research ethics board approval from the University of Toronto (protocol #40636).

Assessment Tool

The RaTMSF was created by expert Canadian medical educators in 2015 and distributed in print and on their website.[5] This two-page assessment consists of 10 rated items on a five-point Likert scale, and three narrative comments for areas of strength and areas for improvement (see Appendix A). In addition to the print format, the RaTMSF was made available to the study participants as an online Qualtrics survey.

Data Collection

Resident teachers received monthly email reminders to use the RaTMSF during their scheduled teaching sessions. Completed RaTMSFs were collected prospectively between April 2021 and April 2022. Amalgamated feedback from this study was returned to each resident teacher at the end of the study period to help them build a teaching portfolio. Completed RaTMSFs of past geriatric medicine resident graduates between 2016 and 2019 were collected retrospectively. At the end of the study, resident teachers and past graduates were surveyed via email to comment on the RaTMSF's strengths, weakness, suggestions for improvement, and acceptability of the tool (see Appendix B for survey contents).

Statistical Analysis

Data collected were deidentified by JCL so that the outcome assessor YQH was blinded for the quantitative data analysis. We used descriptive statistics to summarize baseline characteristics of participants, resident teachers' mean performance, and its associated standard deviation (SD). We used Cronbach's alpha to determine internal consistency for the rated items. We used Fleiss kappa statistic to determine inter-rater reliability as it allows for assessing of reliability within more than two raters.[7] We made an a priori decision to use methods of imputation to complete non- and partially rated items to address the potential bias and loss of precision that could result from complete assessment analysis. For items that a learner did not rate, we replaced the missing data with the assessment's mode. We then conducted a sensitivity analysis comparing kappa of two models: the first model contained only completed assessments, and the second model contained all assessments with missing values replaced. All analyses were performed using R statistical software, version 4.3.3 (R Foundation for Statistical Computing; https://www.r-project.org/foundation/). Narrative comments of the RaTMSF and the end-of-study survey responses were reviewed by JCL for thematic content.

RESULTS

A total of 32 teaching sessions consisting of 132 RaTMSF assessments were collected prospectively from 11 residents from April 2021 to April 2022, and retrospectively from seven graduates from 2016 to 2019. The median number of assessments per teacher was six prospectively (IQR 6.25) and six retrospectively (IQR 3). Twenty-seven assessments were completed by health-care professionals, 60 by medical residents, 19 by medical students, and 14 by others. Twelve assessments did not identify the learner's role. The most common themes of narrative comments also captured in the rated items were organization to teach (n=53), openness to questions (n=36), and agreeing on expectations (n=19). Written comments regarding delivery style (n=52), audience interactivity (n=44), and content knowledge (n=43) were not captured on the rated items (Table 1).

RaTMSF Instrument Validity

Of the 132 RaTMSF assessments, there were 67 incomplete assessments consisting of 11 assessments with ≥2 missing ratings and 56 assessments with ≥2 items rated "not able to comment" that we labeled "partially rated assessments" (PRAs). Assessments that were fully complete or missing <2 rated items were labeled as "completed assessments" (CAs).

We constructed two models for sensitivity analysis. Model 1 excluded the 67 incomplete PRAs (n=65). In the imputation Model 2, we replaced any assessment's unrated item with the rating of the assessment's mode (n=132). Sensitivity analysis on both models obtained similar results (Table 2). The RaTMSF demonstrated high internal consistency with Cronbach's alpha of 0.97, 95% CI 0.89–0.99 between all ten items. The RaTMSF also demonstrated good inter-rater reliability in Model 1 with Fleiss kappa 0.73 (95% CI 0.13–0.80), and very good inter-rater reliability in imputation Model 2 with Fleiss kappa 0.81 (95% CI 0.52–0.87). Due to a large number (45.5%) of items not being completed by at least two learners for the same resident teacher, per-item inter-rater reliability kappa statistic was not possible.

Interpretation of RaTMSF Scores

The overall performance rating was very positive, 4.75 (SD 0.47), for all resident teachers. Almost all learners rated

TABLE 1.
Of 132 learner assessments, rated items were reflected in some, but not all, of the common themes in written narrative comments

Written Narrative Comments Captured in Rated Items	*Written Narrative Comments Not Captured in Rated Items*
Was organized to teach (clinical or structured) (n=53)	Delivery: pace, clarity, learning aids (n=52)
Asked for, and welcomed my questions (n=36)	Engagement and interactivity (n=44)
Ensured we agreed on expectations early (n=19)	Content, knowledge, and evidence-based (n=43)

resident teachers as “4-Skillful” to “5-Exemplary”, with high concordance of ratings to the 10th item of overall performance rating, 4.60 (SD 0.52). Therefore, the RaTMSF rating system had poor discriminatory power to identify differences in teaching performance.

Utility of Actions from the RaTMSF

Eleven of 13 resident teachers who responded to the email survey found the RaTMSF acceptable as a tool to gather high-quality actionable feedback on their teaching skills. Resident teachers valued written narrative comments more than rated items, appreciated its brevity and ease of use, and commented positively on how it can contribute to building a teaching portfolio. Weaknesses of the RaTMSF were a lack of rated items specific to structured teaching, confusing instructions for the learner, and its length at two pages. Lastly, there was no formal integration of the RaTMSF with the residency program’s entrustable professional activity platform and, therefore, some resident teachers felt the RaTMSF was inconsequential (see Table 3 for thematic content of survey results).

DISCUSSION

In this study, geriatric medicine residents at UofT were uniformly skilled teachers when assessed by the RaTMSF. The RaTMSF demonstrated high internal consistency and inter-rater reliability of its rated items. Resident teachers valued feedback provided by the standard five-point Likert scale and written comments. Most resident teachers felt the RaTMSF is acceptable to use for gathering feedback. We believe there are opportunities for modifying the RaTMSF to better suit the needs of our residents conducting structured teaching. It is uncertain whether these findings are generalizable to other programs or teaching environments.

TABLE 2.
Mean performance of geriatric medicine residents in each rated item is similar in both Partially Rated Assessments (PRAs) Model 1 and Completed Assessments (CAs) Model 2

Rated Items	*PRAs Model 1: Mean (SD) of the 65 Evaluations that had ≤1 Unrated Items*	*CAs Model 2: Mean (SD) of all 132 Evaluations Applying Imputation Technique*
Item 1	4.91 (0.29)	4.75 (0.47)
Item 2	4.91 (0.29)	4.75 (0.45)
Item 3	4.83 (0.37)	4.70 (0.51)
Item 4	4.78 (0.42)	4.69 (0.48)
Item 5	4.84 (0.41)	4.70 (0.50)
Item 6	4.95 (0.21)	4.82 (0.38)
Item 7	4.93 (0.24)	4.85 (0.36)
Item 8	4.90 (0.29)	4.82 (0.38)
Item 9	4.86 (0.35)	4.76 (0.44)
Item 10 (Overall)	4.75 (0.47)	4.60 (0.52)

SD = standard deviation.

Proposed Changes to Rated Items

Item ratings allow raters to provide feedback that is organized and constructive.[8] Based on the narrative comments analyzed, domains that should receive focus on the rated items are: organization to teach, openness to questions, establishing expectations, delivery style, audience interactivity, and content knowledge.

Proposed Change to Narrative Feedback

We propose modifying the RaTMSF to focus on narrative feedback. Not only did rated items demonstrate poor discriminatory power, but resident teachers also found narrative comments to be more valuable than ratings. While narrative comments can enhance learning, the value of ratings can be negligible due to respondent fatigue.[9] To mitigate respondent fatigue that can negatively affect the completion of latter components, we suggest placing the narrative feedback before the rated items.[10]

RaTMSF as an Assessment Tool in Our Residency Program

Participants were uniformly excellent resident teachers and the consequences of poor feedback on the RaTMSF for a resident’s educational progression is not clear. Based on our results, we would not recommend the RaTMSF be used

TABLE 3.
Thematic content of the strengths and weaknesses of the RaTMSF tool from a survey of the resident teachers

Strengths	*Weaknesses*
Acceptable to use to gather feedback on teaching skills. (n=11)	Rated items lacked questions specific for formal teaching presentations. (n=4)
Able to collect high quality actionable feedback. (n=7)	Instructions are lengthy and unclear. (n=4)
Written narrative feedback for strengths and areas of improvement were most valuable than ratings. (n=6)	Ten rated items and 2-page assessment form is too long. (n=3)
Short and brief. (n=5)	No integration with entrustable professional activities. (n=2)
Easy to use and understand. (n=4)	Narrative comments at the end were not consistently completed. (n=2)
Helpful to build a teaching portfolio. (n=4)	
Standardized 5-point Likert scale. (n=3)	
Has both ratings and narrative comments. (n=2)	
Focused on learner’s experience. (n=1)	

as an assessment method to inform competency decisions because of the poor discriminatory power of the rated items. However, the RaTMSF is a valuable feedback tool to help residents gather high-quality feedback, and potentially can have a role in developing a resident's teaching portfolio for career advancement.

ACKNOWLEDGEMENTS

Not applicable.

CONFLICT OF INTEREST DISCLOSURES

We have read and understood the *Canadian Geriatrics Journal*'s policy on conflicts of interest disclosure and declare that we have none.

FUNDING

JCL and YQH were supported by funding from the Margaret Watson Award from Department of Geriatric Medicine, University of Toronto during the conduct of this study. YQH is supported by the Eliot Phillipson Clinician-Scientist Training Program at University of Toronto. The publication fee of this article is funded by the Julius and Marie Molinaro Elder Care Education Fund. Funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

REFERENCES

1. The Royal College of Physicians and Surgeons of Canada. What is competence by design? [Internet]. 2024 [cited 2024 Feb 7]. Available from: https://www.royalcollege.ca/en/standards-and-accreditation/competence-by-design.html
2. The Royal College of Physicians and Surgeons of Canada. Geriatric medicine competencies 2019 [Internet]. 2018 [cited 2024 Feb 7]. Available from: https://www.royalcollege.ca/content/dam/documents/ibd/geriatric-medicine/geriatric-medicine-competencies-e.pdf
3. Faculty of Medicine, University of Toronto. Competence by design: geriatric medicine COD 11B—Teaching other learners. Part B: Formal scheduled teaching [Internet]. n.d. [cited 2024 Feb 7]. Available from: https://deptmedicine.utoronto.ca/sites/default/files/inline-files/COD%2011B_Teaching%20other%20learners_Part%20B%20Formal%20Scheduled%20teaching_0.pdf
4. Berk RA. Using the 360° multisource feedback model to evaluate teaching and professionalism. *Med Teach*. 2009 Jan 1; 31(12):1073–80.
5. Glover Takahashi S, editor. CanMEDS teaching and assessment tools guide. Ottawa, ON: Royal College of Physicians and Surgeons of Canada; 2015.
6. Russell M. Clarifying the terminology of validity and the investigative stages of validation. *Educ Meas Issues Pract*. 2022 Jun; 41(2):25–35.
7. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Medica*. 2012 Oct 15;22(3):276–82.
8. Dahlquist HC, Hafler JP, Hersh DC. A pilot study of a direct teaching observation tool for residents. *Yale J Biol Med*. 2020 Aug 31;93(3):419–22.
9. Shute VJ. Focus on formative feedback. *Rev Educ Res*. 2008 Mar;78(1):153–89.
10. Lavrakas PJ. Respondent fatigue. In: E*ncyclopedia of survey research methods*, 1st ed. Thousand Oaks, CA: SAGE Publications; 2008. p. 742–43.

Correspondence to: Janice C. Lee, MD, MScCH, Division of Geriatric Medicine, Department of Medicine, St. Michael's Hospital, 30 Bond Street, Toronto, ON M5B 1W8
E-mail: Janicecl.lee@utoronto.ca

APPENDIX A (part 1 of 2). Resident as teacher multisource feedback

A1. RESIDENT AS TEACHER MULTISOURCE FEEDBACK

Created for the *CanMEDS Teaching and Assessment Tools Guide* by S Glover Takahashi. Reproduced with permission of the Royal College.

See Scholar Role teacher tips appendix for this assessment tool

Instructions for Assessor:

- As Scholars, physicians demonstrate a lifelong commitment to excellence in practice through continuous learning and by teaching others, evaluating evidence, and contributing to scholarship.
- The competencies of the Scholar Role can be developed with practice and feedback. Using the form below, please help this resident physician gain insight into his/her teaching skills by providing valuable confidential feedback.
- Rest assured this information will be shared with the physician in aggregate form and for the purposes of helping the physician improve his/her leadership competencies.
- Please return this form in a confidential sealed envelope to the attention of:

  ____

RESIDENT Name: ______________________________

Postgraduate year (PGY): ______________________________

Indicate ☑ all that apply. I am a:

- ☐ Health professional team member
- ☐ Resident
- ☐ Medical student (including clerk)
- ☐ Other

Degree of Interaction

- ☐ Considerable teaching from this resident
- ☐ ***Occasional or one time teaching*** from this resident

#	This teacher...	1 Very poor	2 Needs improvement	3 Competent	4 Skilful	5 Exemplary	Not able to comment
1.	Was organized to teach (ie teaching in the clinical setting and or structured teaching)						
2.	Was available to learners so I had the support needed.						
3.	Ensured we agreed on expectations early and did his/her best to meet the expectations						
4.	Encouraged me to explore my limits safely						
5.	Provided regular, meaningful, prompt feedback to me						
6.	Demonstrated respect for me as a learner and as a person						
7.	Asked for and welcomed my questions						
8.	Asked for and welcomed my feedback						
9.	Had the educational experience to balance the work assignments and the formal learning opportunities						

Scholar

© 2015 Royal College of Physicians and Surgeons of Canada

This unmodified assessment tool is reproduced from the CanMEDS Teaching and Assessment Tools Guide by S. Glover Takahashi and is owned by the Royal College of Physicians and Surgeons of Canada. It is reproduced in accordance with the Royal College's terms for non-commercial use. Available at: https://royalcollege.ca/content/dam/document/standards-and-accreditation/scholar-a1-multisource-feedback-resident-as-teacher-e.pdf

APPENDIX A (part 2 of 2). Resident as teacher multisource feedback

CANMEDS TEACHING AND ASSESSMENT TOOLS GUIDE

A1. RESIDENT AS TEACHER MULTISOURCE FEEDBACK (continued)

Overall Rating	1 Very poor	2 Needs improvement	3 Competent	4 Skilful	5 Exemplary
	One of the worst learning experiences I have had	I learned very little of significance **or** had an unpleasant experience	Good experience **and** learned something important	Excellent experience **and** learned a great deal	One of the best teachers I have had

Areas of strength	Areas for improvement
1.	1.
2.	2.
3.	3.

Other comments:

Scholar

© 2015 Royal College of Physicians and Surgeons of Canada

This unmodified assessment tool is reproduced from the CanMEDS Teaching and Assessment Tools Guide by S. Glover Takahashi and is owned by the Royal College of Physicians and Surgeons of Canada. It is reproduced in accordance with the Royal College's terms for non-commercial use. Available at: https://royalcollege.ca/content/dam/document/standards-and-accreditation/scholar-a1-multisource-feedback-resident-as-teacher-e.pdf

APPENDIX B. Contents of the end-of-study email survey to resident teachers

(1) What are your general thoughts on the Resident as Teacher multisource feedback tool?

(2) What things did you liked about it?

(3) What things do you wish can be changed?

(4) Did you feel it is a helpful tool to inform your teaching practice?