

Validity, Reliability and Acceptability of the Team Standardized Assessment of Clinical Encounter Report*



Camilla L. Wong, MD, MHSc^{1,2}, Mireille Norris, MD, MHSc³, Samir S. Sinha, MD, PHd⁴, Maria L. Zorzitto, MD, MSc², Sushma Madala, MBBS⁵ Jemila S. Hamid, PhD^{1,6}

¹*Li Ka Shing Knowledge Institute, St. Michael's Hospital, Toronto, ON, Canada;* ²*Division of Geriatrics, St. Michael's Hospital, Toronto, ON, Canada;* ³*Division of Geriatrics, Sunnybrook Health Sciences Centre, Toronto, ON, Canada;* ⁴*Division of Geriatrics, Mount Sinai Hospital, Toronto, ON, Canada;* ⁵*The Wright Center for Graduate Medical Education, Pennsylvania, PA, USA;* ⁶*Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, ON, Canada*

DOI:<http://dx.doi.org/10.5770/cgj.19.234>

ABSTRACT

Background

The Team Standardized Assessment of a Clinical Encounter Report (StACER) was designed for use in Geriatric Medicine residency programs to evaluate Communicator and Collaborator competencies.

Methods

The Team StACER was completed by two geriatricians and interdisciplinary team members based on observations during a geriatric medicine team meeting. Postgraduate trainees were recruited from July 2010–November 2013. Inter-rater reliability between two geriatricians and between all team members was determined. Internal consistency of items for the constructs Communicator and Collaborator competencies was calculated. Raters completed a survey previously administered to Canadian geriatricians to assess face validity. Trainees completed a survey to determine the usefulness of this instrument as a feedback tool.

Results

Thirty postgraduate trainees participated. The prevalence-adjusted bias-adjusted kappa range inter-rater reliability for Communicator and Collaborator items were 0.87–1.00 and 0.86–1.00, respectively. The Cronbach's alpha coefficient for Communicator and Collaborator items was 0.997 (95% CI: 0.993–1.00) and 0.997 (95% CI: 0.997–1.00), respectively. The instrument lacked discriminatory power, as all trainees scored "meets requirements" in the overall assessment. Ninety-three per cent and 86% of trainees found feedback useful for developing Communicator and Collaborator competencies, respectively.

Conclusions

The Team StACER has adequate inter-rater reliability and internal consistency. Poor discriminatory power and face validity challenge the merit of using this evaluation tool. Trainees felt the tool provided useful feedback on Collaborator and Communicator competencies.

Key words: communication, collaboration, postgraduate, assessment, feedback

INTRODUCTION

The CanMEDS framework outlines physician competencies adopted by the Royal College of Physicians and Surgeons of Canada (RCPSC), which is the specialty physician certification organization in Canada.⁽¹⁾ The role of the Communicator includes patient-centred therapeutic communication through shared decision-making and effective interactions with other professionals.⁽¹⁾ As a Collaborator, the physician must be able to collaborate effectively with an interprofessional team for the provision of optimal care, education, and scholarship.⁽¹⁾ Poor interprofessional collaboration can have a negative impact on the quality of patient care.⁽²⁾

The Task Force on Resident Training in Geriatrics Interdisciplinary Team Care of the American Geriatrics Society Education Committee developed guidelines in 1999 which recommended evaluation strategies.⁽³⁾ These recommendations included observations of the resident by core interdisciplinary team members. Canadian standards of accreditation for residency programs indicates: "There must be mechanisms in place to ensure the systematic collection and interpretation of evaluation data on each resident enrolled in the program."

*This study was previously published as an abstract (*Can Geriatr J.* 2014;17(4)).

One mechanism to collect and evaluate data on a resident's performance as a Collaborator and Communicator is to use Standardized Assessment of a Clinical Encounter Reports (StACERs). The Team StACER was developed by consensus of a sub-committee of the Geriatric Medicine Specialty Committee of the RCPSC. It is a checklist designed for Geriatric Medicine residency programs to evaluate Communicator and Collaborator competencies.

There have not been any studies of the validity, reliability or acceptability of this instrument to date, in spite of its mandatory use since 2007. The RCPSC requires the Team StACER to be completed by a geriatrician. It is unknown whether the geriatrician's assessment reflects the observations and assessments made by the interdisciplinary team. Assessments by other members of the clinical team may provide insight into a trainee's capacity for team work and interpersonal sensitivity.

Interprofessional team rounds present a pragmatic opportunity for interdisciplinary team members to observe and assess a resident's Communicator and Collaborator competencies. In this study, we determined the inter-rater reliability, internal consistency, face validity, and acceptability of the Team StACER in assessing and providing feedback on Collaborator and Communicator competencies among postgraduate trainees during interprofessional rounds.

METHODS

Study Population

Postgraduate trainees in Family Medicine, Internal Medicine and Geriatric Medicine at the University of Toronto rotating through the Geriatric Medicine service at St. Michael's Hospital, Mount Sinai Hospital, or Sunnybrook Health Sciences Centre were eligible for the study. Recruitment occurred from July 2010–November 2013 based on convenience sampling. Participation was voluntary. A rotation is four weeks in duration but up to twelve weeks for Geriatric Medicine residents. Consent status was not disclosed to the study investigators until the data analysis phase was completed and after the resident's formal residency rotation evaluation was performed.

Intervention

Weekly interprofessional meetings were attended by an interprofessional team including a geriatrician, postgraduate medical trainees, and depending on the composition at each site, a nurse, occupational therapist, physiotherapist, pharmacist, registered dietician, chaplain, and social worker. A second geriatrician was also present. Two geriatricians and members of the interprofessional team observed the trainee during an interprofessional team meeting and independently completed the Team StACER based on observations. A page of standard instructions explaining the procedure for the conduct of the

interdisciplinary team meeting and administration of the Team StACER was provided. Immediately after the meeting, one geriatrician reviewed the completed StACER with the trainee to provide feedback on strengths and weaknesses on Communicator and Collaborator competencies.

Data Collection

The Team StACER

The Team StACER was completed by the two geriatricians and interprofessional team present at the meeting. The Team StACER assesses Communicator and Collaborator competencies using six and seven elements, respectively, where options for performance include "below expectations", "meet expectations", and "not applicable". There is a global assessment measurement dichotomized into "fails to meet requirements" and "meets requirements". There is room for free-text comments (see Figure 1).

Face Validity Survey

After study recruitment ended, the geriatricians and interprofessional team members completed an anonymous survey to rate the extent to which the Team StACER addressed the overall concepts of Collaborator and Communicator. This eight-item 5-point Likert scale survey was previously developed and administered to geriatricians across Canada to assess the face validity of the Team StACER. There is a free-text area for comments.

Acceptability Survey

Immediately, after receiving feedback, the trainee completed an anonymous survey asking if the feedback process was useful in developing competencies as a Communicator and Collaborator. This survey consisted of two questions: 1) Did you find receiving feedback using the Team StACER useful in developing your competency as a Communicator?, and 2) Did you find receiving feedback using the Team StACER useful in developing your competency as a Collaborator?

Data Analysis

Descriptive statistics were expressed in frequencies and proportions for categorical variables. The majority of the evaluations scored "meet expectations" in most items and the prevalence of "below expectations" was very low and hence the data was sparse; thus, inter-rater reliability was determined between two geriatricians and between all team members, including the geriatricians, using the prevalence-adjusted bias-adjusted kappa (PABAK), which is a function of percentage agreement and is an alternative approach when the traditional kappa statistics fails to capture agreement due to prevalence and bias.⁽⁴⁾ When the number of raters was more than two, the average PABAK was calculated and the 95% CIs adjusted. If an evaluator had less than 5 evaluations, the evaluator was excluded from the inter-rater reliability

	<i>Below Expectations</i>	<i>Meets Expectations</i>	<i>Not A pplicable</i>	<i>Comments</i>
<i>Communicator</i>				
1a.				Listens actively to speakers
1b.				Non-verbal communication conveys interest/commitment
1c.				Uses language understood by all team members
1d.				Uses appropriate communication techniques (e.g. summarizing, clarifying questions, bridging to move meeting forward)
1e.				Keeps self and team focused on meeting purpose (e.g. does not engage in storytelling, or side conversations)
1f.				Avoids disruptive behavior (e.g. interrupting, hostile, humor, negative comments)
<i>Collaborator</i>				
2a.				Ensures decisions are made and clearly articulated
2b.				Ensures decisions reflect the input of the team
2c.				Ensures patient goals/wishes are considered
2d.				Clarify/negotiate roles/responsibility for implementation of team decisions
2e.				Demonstrates appropriate techniques for resolving conflict (e.g. clarifying issues, negotiating resolution, etc.)
2f.				Encourages full and equitable participation of all team members
2g.				Manages meeting time
3.				Overall Assessment of Candidate
		Fails to Meet Requirements	Meets Requirements	

Summary of resident’s strengths and weaknesses: _____

Comments on the process: _____

Date: _____

Evaluator’s Name: _____ Evaluator’s Signature: _____

Resident’s Name: _____ Resident’s Signature: _____

Copyright ©2009 The Royal College of Physicians and Surgeons of Canada. Referenced and produced with permission.

FIGURE 1. Assessment grid for a team conduct observation based on the team Standardized Assessment of a Clinical Encounter Report (StACER)

analysis. Internal consistency of the six and seven items for the constructs Communicator and Collaborator competencies, respectively, were determined using Cronbach’s alpha.⁽⁵⁾ Due to the sparseness of the data and zero variability in most StACER items, the correlation coefficient on which the Cronbach’s alpha coefficient is based could not be calculated. Rather, we used similarity matrices, which are a measure of association for categorical data. Cronbach’s alpha for both Communicator and Collaborator items was calculated based on the simple

matching similarity matrix.⁽⁶⁾ The alpha coefficients were averaged across multiple raters. Comments in the face validity questionnaires were analyzed by two reviewers independently for themes. Each Team StACER was analyzed by a study author to determine whether individual comments were concordant or discordant with the corresponding item on the Team StACER; for a subset of Team StACERs, this was examined in duplicate independently by a second study author. Statistical analyses were performed using the R statistical software.⁽⁷⁾

Ethics, Consent and Permissions

Ethics approval for the study was granted from the Research Ethics Board (REB) (St. Michael's Hospital REB 10-013, Mount Sinai Hospital REB 11-0123-E, Sunnybrook Health Sciences Centre REB 173-2011). Informed consent was obtained from all study participants.

RESULTS

A convenience sample of 30 postgraduate trainees was approached for consent and all 30 (100%) participated in the study. There were 7 family medicine, 14 internal medicine, and 9 geriatric medicine trainees distributed among three study sites (23, 4, and 3 trainees). 170 Team StACERs were completed by 19 unique evaluators (9 geriatricians, 2 nurses, 2 physiotherapists, 2 pharmacists, 1 occupational therapist, 1 dietician, 1 social work, and 1 chaplain). Ninety-three per cent of study participants had a Team StACER completed by two geriatricians. The average number of Team StACERs completed by an evaluator was 8.9 (SD = 9.2) and an average of 5.7 (SD = 1.37, range = 2 to 7) Team StACERs were completed for each trainee (see Table 1).

Inter-Rater Reliability

Almost perfect agreement was observed among evaluators, where the PABAK estimates among all evaluators for Communicator and Collaborator items were in the range of 0.87–1.00 and 0.86–1.00, respectively.⁽⁸⁾ A slightly higher inter-rater reliability was observed between geriatricians compared to the overall agreement among all evaluators; the PABAK estimates for Communicator and Collaborator items ranged from 0.90–1.00 and 0.83–1.00 (Table 2).

Internal Consistency

A high level of inter-correlation between items was observed in both Communicator and Collaborator questionnaires, leading to nearly perfect similarity. The minimum observed similarity (averaged across the seven evaluators) observed between any two items was very high (0.92) and was

between two Collaborator items (Q2c: Ensures patient goals/wishes are considered, and Q2d: Clarify/negotiate roles/responsibility for implementation of team decisions). The alpha coefficient estimated using data from different evaluators ranged from 0.99–1.00, indicating a very strong internal consistency.⁽⁹⁾ This same result was found for both Communicator and Collaborator items. The internal consistency of the six items for the Communicator competency was 0.997 (95% CI: 0.993–1.00). The internal consistency of the seven items for the Communicator competency was 0.997 (95% CI: 0.997–1.00).

Concordance Between Ratings and Comments

One study author analyzed 106 StACERs and a second study author analyzed 83 StACERs to determine whether individual comments were concordant or discordant with the corresponding item rating on the Team StACER. Nineteen StACERs were examined in duplicate independently by two study authors; the StACERs were from three trainees with 14 items, yielding a total of 266 concordance evaluations. Among the 27 concordance evaluations for which evaluations were provided by both concordance evaluators, a 96.3% agreement was observed. The comments were concordant to an “overall meets requirement” rating in 92.7% of evaluations examined. The Communicator item with the greatest discordance between the rating and comments was “uses appropriate communication techniques” (37.8%). The Collaborator item with the greatest discordance between the rating and comments was “ensures decisions are made and clearly articulated” (31.0%) (see Table 3 and Figure 2).

Discriminatory Power

The Team StACER lacked discriminatory power as all trainees, regardless of level of training, scored “meets requirements” in the overall assessment.

Face Validity

A total of 17 face validity questionnaires (89.5% response rate) were completed (by 8 geriatricians, 2 nurses, 1 occupational therapist, 2 physiotherapists, 1 dietician, 2 pharmacists and 1 social worker). Common themes identified in the comments included an assumption of the instrument is that the physician is leading the team meeting, the tool cannot differentiate between levels of trainees, and a scale rather than only “yes” and “no” options may have provided better detail on competencies (see Table 4).

Acceptability

Twenty-eight acceptability questionnaires (93.3% response rate) were completed (by 7 family medicine, 12 internal medicine, and 9 geriatric medicine trainees), of which 92.9%

TABLE 1.
Study demographics

Trainee level, N (%)	30
Family medicine	7 (23.3%)
Internal medicine	14 (46.7%)
Geriatric medicine	9 (30%)
Distribution of trainees, N	
Site 1	23
Site 2	4
Site 2	3

TABLE 2.
Inter-rater reliability (PABAK) of individual StACER items among all raters and between geriatricians

<i>StACER Item</i>	<i>Inter-rater Reliability Among All Raters^a</i>		<i>Inter-rater Reliability Between the MDs</i>	
	<i>N^b (min,max)</i>	<i>PABAK (95% CI)</i>	<i>N</i>	<i>PABAK (95% CI)</i>
Communicator				
Q1a	18.05 (14,21)	1.00 (0.56,1.00)	21	1.00 (0.81,1.00)
Q1b	18.05 (14,21)	1.00 (0.56,1.00)	21	1.00 (0.81,1.00)
Q1c	18.05 (14, 21)	1.00 (0.56,1.00)	21	1.00 (0.81,1.00)
Q1d	17.48 (13,21)	0.86 (0.38,0.97)	20	0.90 (0.67,0.98)
Q1e	17.24 (13,21)	0.97 (0.51,1.00)	21	0.95 (0.78,1.00)
Q1f ^c	3 (1,10)	1.00 (-0.48,1.00)	10	1.00 (0.66,1.00)
Collaborator				
Q2a	17.76 (14,25)	0.97 (0.51,1.00)	20	0.95 (0.73,1.00)
Q2b	17.86 (14,25)	0.97 (0.52,1.00)	20	0.95 (0.79,1.00)
Q2c	17.24 (14,21)	0.86 (0.36,0.97)	18	0.83 (0.58,0.96)
Q2d	14.24 (7,20)	0.91 (0.33,1.00)	10	0.90 (0.54,1.00)
Q2e ^c	2 (1,6)	1.00 (-0.67,1)	1	1.00 (0.06,1.00)
Q2f	18.05 (14,21)	0.96 (0.51,1.00)	21	1.00 (0.80,1.00)
Q2g	17 (14,20)	0.93(0.45,1.00)	19	1.00 (0.79,1.00)
Meets Requirement				
Q3	17 (14,21)	1.00 (0.56,1)	21	1.00 (0.81,1.00)

^a 7 evaluators (2 of whom MDs) are considered in the analysis as there was not enough sample size for the remaining of the 5 evaluators. All 7 evaluators are from the same site – SMH.

^b The average number of study participants used in calculating PABAK (averaged over the number of study participants used to calculate pairwise PABAK).

^c Limited sample size.

TABLE 3.
Concordance between ratings and comments for StACER items

<i>StACER Item</i>	<i>Comments Provided, N (%)</i>	<i>Concordance, N (%)</i>
Communicator		
Listens actively to speakers	26 (15.3%)	25 (96.1%)
Non-verbal communication conveys interest/commitment	23 (13.5%)	19 (82.6%)
Uses language understood by all team members	20 (11.8%)	10 (50%)
Uses appropriate communications techniques	45 (26.5%)	17 (37.8%)
Keeps self and team focused on meeting purpose	15 (8.8%)	13 (86.7%)
Avoids disruptive behaviour	5 (2.7%)	3 (60%)
Collaborator		
Ensures decisions are made and clearly articulated	29 (17.1%)	9 (31.0%)
Ensures decisions reflect the input of the team	37 (21.8%)	28 (75.7%)
Ensures patient goals/wishes are considered	11 (6.5%)	9 (81.8%)
Clarify/negotiate roles/responsibility for implementation of team decisions	18 (10.6%)	8 (44.4%)
Demonstrates appropriate techniques for resolving conflict	6 (3.5%)	6 (100%)
Encourages full and equitable participation of all team members	43 (25.3%)	28 (65.1%)
Manages meeting time	21 (12.4%)	16 (76.2%)
Overall Assessment of Candidate	82 (48.5%)	76 (92.7%)

^a The one evaluation for which disagreement was observed between the two concordance evaluators was removed both from the numerator and denominator.

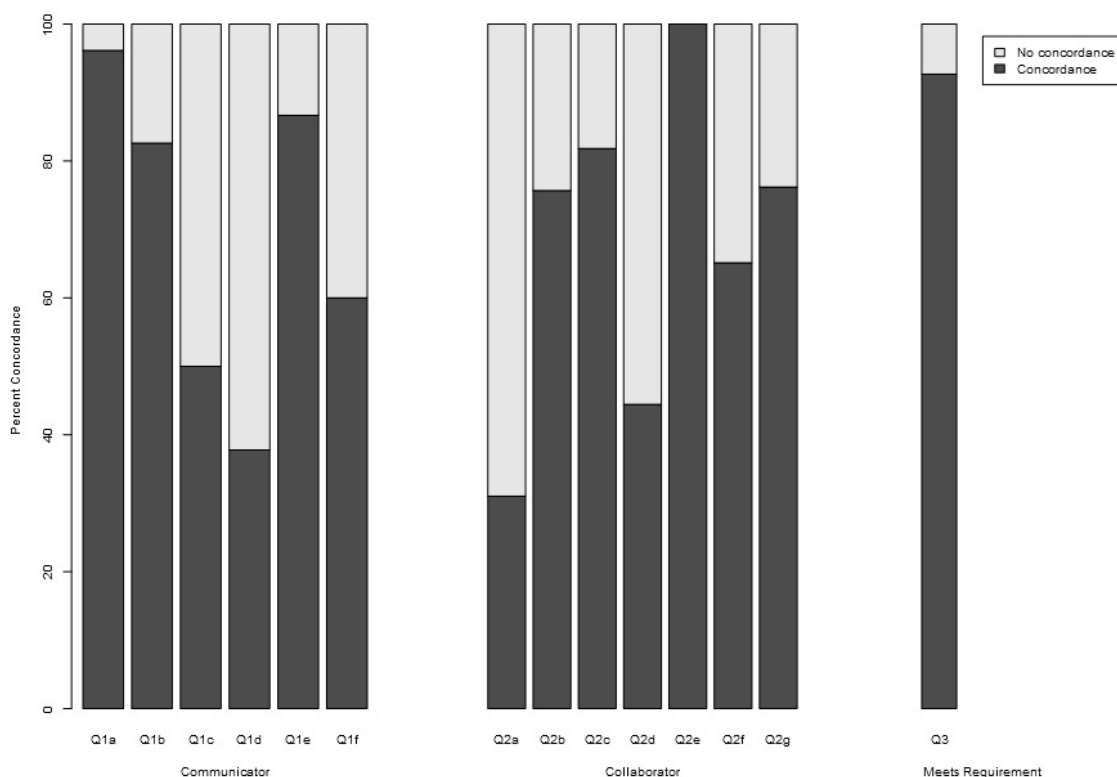


FIGURE 2. Concordance between ratings and comments provided by StACER evaluators

and 85.7% of trainees found feedback based on the instrument useful for developing Communicator and Collaborator competencies, respectively.

DISCUSSION

This is the first study to establish that there is very strong inter-rater reliability among all interprofessional team members and between geriatricians when evaluating Communicator and Collaborator competencies using the Team StACER. We thus recommend it would be reasonable for only one evaluator to complete the StACER. The main limitations of the Team StACER are the lack of discriminatory power and discordance between ratings and comments on certain items.

While this study was conducted at three sites, most of the participants were from one site. This was due to ease of recruitment as team meeting scheduling was consistent at one site. The disproportionate representation from one site might have caused the good agreement in the data, and hence did not allow direct measurement covariance and correlations among StACER items. The sample size was small and derived from convenience sampling. The small sample likely contributed to the lack of heterogeneity. This study examines face validity, but does not establish criterion validity as there is no reference standard for Communicator and Collaborator competencies. The acceptability survey may be subject to biases inherent to dichotomous responses and reporting.

Interprofessional settings are opportunities to enhance communication and collaboration between professionals. The Task Force on Resident Training in Geriatrics Interdisciplinary Team Care of the American Geriatrics Society Education Committee recommended evaluation strategies of these competencies should include trainee attendance and participation, use of a checklist to document participation in various interdisciplinary team care activities, and observations of the trainee by interdisciplinary team members and faculty.⁽³⁾ In spite of this recommendation, to our knowledge, our study is the first to determine the validity, reliability, and acceptability of such an evaluation strategy. Canada is moving towards a competency-based design to medical education where competency milestones should mark the progression of competence from transition into residency and then into practice. The milestones should enable evaluators and programs to know when a learner is ready to move onto the next stage of training. While trainees found feedback based on this instrument acceptable, the results of this study, in particular the lack of discriminatory power of the Team StACER, question the appropriateness of mandating the use of this tool in evaluating Communicator and Collaborator competencies in an era of competency-based medical education.

The lack of discriminatory power is not surprising given the dichotomous choice between “below expectations” and “meets expectations”; a trainee would need to perform very poorly to be rated “below expectations”. One proposed solution is to replace

TABLE 4.
Results of face validity questionnaire

Face Validity Questionnaire Item	Response, N (%)				
	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
I found the instructions for the StACERs easy to understand.	0 (0%)	0 (0%)	1 (5.9%)	15 (88.2%)	1 (5.9%)
The StACERs for the residents were performed according to the recommended procedures.	0 (0%)	0 (0%)	1 (5.9%)	14 (82.4%)	2 (11.8%)
The items on the StACERs reflected the relevant clinical skills.	0 (0%)	0 (0%)	4 (23.5%)	12 (70.6%)	1 (5.9%)
It was easy to rate the residents' skills on the StACER forms.	0 (0%)	1 (5.9%)	3 (17.7%)	13 (76.5%)	0 (0%)
I gave detailed feedback to the residents on their performances.	0 (0%)	6 (35.3%)	4 (23.5%)	6 (35.3%)	1 (5.9%)
I feel the StACERs are an appropriate evaluation of my residents' clinical skills.	1 (5.9%)	6 (35.3%)	5 (29.4%)	4 (23.3%)	1 (5.9%)
I feel the StACERs are an appropriate performance-based assessment tool.	1 (5.9%)	5 (29.4%)	3 (17.7%)	6 (35.3%)	2 (11.8%)
Overall it was NOT onerous to supervise the StACERs.	0 (0%)	0 (0%)	6 (35.3%)	10 (58.8%)	1 (5.9%)

the StACER item responses with a Likert-type scale. The reliability of Likert-type scales decreases as the number of response options is reduced, with the exception of scales with two categories is slightly higher than that obtained with three categories.⁽¹⁰⁾ As for validity, this decreases as the number of response options is reduced.⁽¹⁰⁾ Considering criteria of reliability and validity, the minimum number of response categories for items with Likert-type format should be at least four and psychometric benefits are scarce beyond seven categories.⁽¹⁰⁾ Furthermore, evaluators prefer formats with a larger number of response alternatives, as this allows more clear expression of views.⁽¹¹⁾

CONCLUSIONS

The Team StACER has adequate inter-rater reliability and internal consistency. Poor discriminatory power and face validity challenge the merit using this evaluation tool. Trainees felt the tool provided useful feedback on Collaborator and Communicator competencies.

ACKNOWLEDGEMENTS

We thank the geriatricians, interprofessional team members and medical trainees for their participation. We also thank Gary Cole, PhD, from the Royal College of Physician and Surgeons of Canada for reviewing the manuscript. This study was funded by the Postgraduate Innovation Fund from the Department of Medicine at the University of Toronto. The funding body had no role in the collection, analysis, and interpretation of data, the writing of the report, nor the decision to submit the report for publication.

CONFLICT OF INTEREST DISCLOSURES

The authors declare that no conflicts of interest exist.

REFERENCES

1. Frank JR, Snell L, Sherbino J, editors. *CanMEDS 2015 Physician Competency Framework*. Ottawa: Royal College of Physicians and Surgeons of Canada; 2015.
2. Zwarenstein M, Reeves S, Perrier L. Effectiveness of pre-licensure interprofessional education and post-licensure collaborative interventions. *J Interprofessional Care*. 2005;19(Suppl 1):148–65.
3. Counsell SR, Kennedy RD, Szwabo P, *et al*. Curriculum recommendations for Resident Training in Geriatrics Interdisciplinary Team Care. *J Am Geriatr Soc*. 1999;47(9):1145–48.
4. Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. *J Clin Epidemiol*. 1993;46(5):423–29.
5. Cronbach LJ. Internal consistency of tests: analyses old and new. *Psychometrika*. 1988;53(1):63–70.
6. Hamid JS, Meaney C, Crowcroft NS, *et al*. Cluster analysis for identifying sub-groups and selecting potential discriminatory variables in human encephalitis. *BMC Infectious Diseases*. 2010;10:364.
7. The R Foundation. What is R? . Vienna, Austria. 2012. Available from: www.R-project.org
8. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159–74.
9. Cortina JM. What is coefficient alpha? An examination of theory and applications. *J Appl Psychol*. 1993;78(1):98–104.
10. Lozano LM, García-Cueto E, Muñiz J. Effect of the number of response categories on the reliability and validity of rating scales. *Eur J Res Meth Behavioral Social Sci*. 2008;4(2):73–79.
11. Muñiz J, García-Cueto E, Lozano LM. Item format and the psychometric properties of the Eysenck Personality Questionnaire. *Personal Individual Diff*. 2005;38(1):61–69.

Correspondence to: Camilla L. Wong, MD, MHS, St. Michael's Hospital, 30 Bond Street, Toronto, ON M5B 1W8, Canada
Email: wongcam@smh.ca