Validity and reliability of an in-training evaluation report to measure the CanMEDS roles in emergency medicine residents

Aliya Kassam, PhD*; Tyrone Donnon, PhD*; Ian Rigby, MD[†]

ABSTRACT

Background: There is a question of whether a single assessment tool can assess the key competencies of residents as mandated by the Royal College of Physicians and Surgeons of Canada CanMEDS roles framework.

Objective: The objective of the present study was to investigate the reliability and validity of an emergency medicine (EM) in-training evaluation report (ITER).

Method: ITER data from 2009 to 2011 were combined for residents across the 5 years of the EM residency training program. An exploratory factor analysis with varimax rotation was used to explore the construct validity of the ITER. A total of 172 ITERs were completed on residents across their first to fifth year of training.

Results: A combined, 24-item ITER yielded a five-factor solution measuring the CanMEDs role Medical Expert/Scholar, Communicator/Collaborator, Professional, Health Advocate and Manager subscales. The factor solution accounted for 79% of the variance, and reliability coefficients (Cronbach alpha) ranged from $\alpha = 0.90$ to 0.95 for each subscale and $\alpha = 0.97$ overall. The combined, 24-item ITER used to assess residents' competencies in the EM residency program showed strong reliability and evidence of construct validity for assessment of the CanMEDS roles.

Conclusion: Further research is needed to develop and test ITER items that will differentiate each CanMEDS role exclusively.

RÉSUMÉ

Contexte: On se demande si un outil d'évaluation unique peut permettre d'apprécier les principales compétences des résidents, comme l'exige le cadre des rôles CanMEDS du Collège royal des médecins et chirurgiens du Canada.

Objectif: L'étude visait à évaluer la fidélité et la validité d'un rapport d'évaluation en cours de formation (REF) en médecine d'urgence (MU).

Méthode: Des données de REF recueillies sur des stagiaires pendant les années 2009–2011 et couvrant les 5 années du programme de formation en MU au niveau de la résidence ont été mises en commun. Une analyse factorielle exploratoire avec rotation Varimax a été utilisée pour examiner la validité conceptuelle du REF. Au total, 172 REF ont été remplis sur des résidents en MU, et ce, de la première à la cinquième année de stage.

Résultats: Un REF global en 24 points a produit une solution à cinq facteurs permettant de mesurer les rôles CanMEDS, soit les sous-groupes Expert médical/Érudit, Communicateur/Collaborateur, Professionnel, Promoteur de la santé, et Gestionnaire. La solution factorielle représentait 79% de la variance et des coefficients de fidélité (coefficients alpha [α] de Cronbach); ainsi, la valeur α variait de 0.90 à 0.95 dans chacun des sous-groupes et était égale à 0.97 dans l'ensemble. Le REF global en 24 points, utilisé pour évaluer les compétences des stagiaires acquises dans le cadre du programme de résidence en MU a révélé une bonne fidélité et fourni des données probantes sur la validité conceptuelle en ce qui concerne l'évaluation des rôles CanMEDS.

Conclusion: Il faudrait approfondir la recherche pour élaborer et vérifier d'autres points du REF, qui permettraient de distinguer chacun des rôles CanMEDS, exclusivement.

Keywords: clinical competence, emergency medicine, in-training evaluation report

In 1996, the Royal College of Physicians and Surgeons of Canada (RCPSC) presented a competencies framework for the assessment of all residents in resident training programs across Canada.[1] The competencies expected of residents are designed to address the question, "What attributes do physicians need in order to be effective in their practice?"

Postgraduate educators, however, are held responsible for using the Canadian Medical Education Directives for Specialists (CanMEDS) roles framework as the basis for developing medical curricula and measurement tools to assess residents throughout their training programs. The seven CanMEDS roles are Medical Expert (the central role), Communicator, Collaborator, Health Advocate, Manager, Scholar, and Professional.[2]

Postgraduate training programs in Canada use a variety of assessment tools to evaluate competency-based medical education specific to the CanMEDS roles. A survey of program directors showed concern about how the non–Medical Expert or intrinsic CanMEDS roles are evaluated in postgraduate medical education programs.[3] Although the use of multiple tools may be recommended when assessing the CanMEDS roles, the utility and feasibility of such a task within resident training curricula needs further investigation. As Lurie suggests,[4] the use of multiple and different assessment methods need to be based on the comparability of these measures as a function of whether the performance measured represents a stable characteristic of the examinee rather than unique elements derived from prescriptive testing situations. Nevertheless, the use of more than a single assessment method to triangulate assessments of the various CanMEDS role competencies is recommended rather than a single tool or method. For example, the assessment of the Communicator role as a function of the patient-physician relationship can be measured using multiple evaluators and a combination of tools or methods such as the in-training evaluation report (ITER), daily encounter cards (DECs), objective structured clinical examination (OSCE), and multi-source feedback, to name a few.[2]

ITERs are the most commonly used tool to evaluate the clinical performance of residents because they allow the measurement of intrinsic domains of clinical competence such as collaboration and health advocacy. ITERs also allow for longitudinal assessment.[2,3,5] To date, the ITER (as a summative measure based on preceptors' observations over typical block rotation periods) has been recognized as the one tool that can best capture an assessment of all seven CanMEDS roles.[2,3] The ITER is completed by the clinical assessor based on a number of observed clinical performances over the course of a resident's training period (e.g., 8-week block rotation) and can

be an efficient way to conduct a broad evaluation of domain and expected behaviour areas. It allows the evaluator to use multiple sources of information and provides a summative evaluation of the resident's clinical performance.

The ITER also has several disadvantages. The reliability and validity of the ITERs may be compromised because of small sample sizes, inaccurate documentation of resident performance, evaluator inconsistency (lenient v. stringent raters), and the lack of variability in the sample of residents selected. The time of evaluation may also impact the reliability and validity of ITER scores because a large gap between observation and evaluation may introduce recall bias.[6] Kogan and colleagues showed that the use of in-training assessment of residents' clinical skills may be influenced by the faculty's own clinical skills.[7] Faculty use various criteria to assess clinical performance and therefore assign levels of importance to what is acceptable. Faculty assessment behaviours were not significantly correlated with demographic characteristics such as age, gender, or clinical or previous experience as preceptors, whereas faculty clinical skills were. From a resident perspective, the ITER process is seen as "inauthentic" due to the debatable credibility of the evaluator. Residents may also believe that their own self-assessment is a more accurate reflection of their performance.[8]

COMPETENCY-BASED MEDICAL EDUCATION IN EMERGENCY MEDICINE

There is controversy regarding the applicability and assessment of competency-based education during clinical rotations in emergency medicine (EM). A focus group study of EM faculty found that the Medical Expert and Professional roles were seen as most relevant to an EM rotation, whereas the Health Advocate, Manager, Scholar, and Collaborator roles were least relevant.[9] Another study that surveyed emergency physicians from two Canadian academic tertiary health sciences centres found a strong preference for Medical Expert topics, whereas Professional, Health Advocate, Collaborator, and Communicator topics were not a reported priority for professional development.[10]

Although being a Medical Expert is central to every competent physician, other competencies are just as important. EM physicians are in an ideal position to liaise between the public and the medical system. They are often the initial point of contact for patients

entering the hospital system and are at an advantage to identify and act on the social, environmental, and biologic factors that influence the health of individuals and society.[11]

It is important to measure clinical competence in EM residents to ensure that resident training program curricula are providing material and encounters for all aspects of clinical care. Although there has been research on the strengths and weaknesses of the different types of assessment tools,[12] at present, there is no literature on how Canadian residency programs effectively teach and assess their residents' clinical competencies in EM across all seven domains. As a result, training programs must be creative in meeting teaching goals with the resources at their disposal, and the assessments need to be comprehensive in addition to being reliable and valid.[13]

The objective of the present study was to investigate the reliability (internal consistency and intraclass correlation [ICC]) and validity (construct) of using a common ITER that can be used to measure the clinical competencies of EM residents across all 5 years of residency training and the seven CanMEDS roles or competency domains.

METHOD

The EM residency program at the University of Calgary is a 5-year integrated program in which residents are evaluated on an ongoing basis by means of direct observation, as well as in-training written and oral examinations. The Department of Emergency Medicine has developed ITERs for each of the postgraduate (PGY) first to fifth year levels of the program that share 24 common items used to assess each of the seven CanMEDS roles. Items were developed by the EM program director in addition to a team of EM preceptors with the help of the objectives of training for emergency medicine outlined by the RCPSC.[14,15] Although each ITER was reviewed and revised until a consensus was reached, the items for the intrinsic roles (e.g., Communicator, Collaborator, Heath Advocate) are similar for each PGY level. More specifically, items will vary across levels on the clinical aspect for the Medical Expert role as the expectations for a PGY-1 resident are not as extensive as they would be for a PGY-5 resident. For example, the item for "Resuscitation Skills" for PGY 1 reads, "Is able to perform CPR, defibrillation, and basic airway

management including intubation with close staff supervision," and the item for PGY 5 reads, "Supervises all resuscitation cases."

All items are rated on a 5-point scale from 1 = fails to meet to 5 = exceeds expectations, with the option of not able to assess/applicable ("NA") checkbox. Completed at the end of rotation, ITER forms are entered by the EM clinical evaluators electronically through a health care administration online program. The ITER data for the EM residents were collected from July 2009 to June 2011 across all 5 years of the residency training program at the University of Calgary. The ITER data were imported into *SPSS* version 19.0 for Windows (SPSS Inc, Chicago) and combined for analysis across the 5 years of the EM residency program.

Exploratory factor analysis using varimax rotation and a Cronbach alpha was also computed. Items loading on a common factor were then analyzed and grouped together. Items with a factor loading of 0.320 (10% of variance) or greater were retained.[16] The exploratory factor analysis on the 24 items was conducted using the principal components method of extraction with an orthogonal varimax rotation. As selection of the number of factors is a critical phase of the analysis, the following criteria were used: eigenvalues > 1, a scree test of eigenvalues plotted against factors, and review of the residual correlation matrix.[16] As the number of factors was theoretically defined by the seven CanMEDS roles and the scree plot is not an exact measure, we performed several factor analyses to derive at a parsimonious factor solution. Each time a different number of factors was specified, the scree test was repeated, and the residual correlation matrix was examined for any large residuals (i.e., > 0.10) that would suggest the presence of another factor. The resulting factors identified were analyzed for internal consistency using Cronbach alpha to derive reliability coefficients (internal consistency) for each subscale. In addition, ICC analysis was conducted to investigate the consistency in rating across the various EM evaluators.

RESULTS

From 2009 to 2011, 172 ITER forms were completed across the 5 years of the EM residency program. The number of items that were used to measure each of the seven CanMEDS roles and were common across all postgraduate years was 24. Specific items that were not common across the ITERs pertaining to skills such as

intravenous insertion, venipuncture, suturing, catheter insertion, urinalysis, and lumbar puncture that were relevant to the PGY-1 ITER did not appear on the PGY-4 and -5 ITERs, where more focus was placed on decision making and expert use of consultation skills.

Factor analysis resulted in a five-factor solution accounting for 79% of the variance, and the salient loadings for the items assigned to one of the five factors were found to range from 0.323 to 0.865 (Table 1). Only item loadings ≥ 0.320 (accounting for 10% or more of the variance on any one factor) are reported. A combined, 24-item ITER yielded a five-factor solution measuring the Medical Expert, Communicator and Collaborator, Professional, Health Advocate, and Manager subscales. The internal reliability coefficients (Cronbach alpha) ranged from α = 0.73 for the Manager subscale to α = 0.95 for the Professional subscale and an overall reliability of α = 0.97 for all 24 items. The average (or sum) of the scores of the 35 EM evaluators is highly reliable, with an ICC of 0.97 with 95% confidence intervals from 0.96 to 0.98, suggesting that the ratings obtained from the evaluator were consistent despite the apparent differences in scoring the ITERs across residents' levels of performance.

DISCUSSION

The combined, 24-item ITER used to assess clinical competency in EM residents showed strong reliability and evidence of construct validity for the assessment of specific CanMEDS roles. The overlapping or cross-loading of items across a factor solution is not uncommon or unexpected as a measure of a resident's clinical competence in one role or domain (e.g., the ability to communicate effectively) may also reflect the individual's abilities in other domains (e.g., abilities to collaborate with or manage others). The common ITER items used across the 5 years of EM residency training were shown to measure some of the CanMEDS roles (i.e., Medical Expert and Professional) better than others (i.e., Communicator and Collaborator, Health Advocate, and Manager).

Although some items are clearly associated with specific roles, 14 of the items were also shown to overlap with one or more factors. It must be emphasized that the CanMEDS roles have distinct competencies that can be assessed independently through observations by appropriate evaluators. In the case of the ITERs in our study, there were items that clearly focused on the assessment of a specific aspect of one particular role and other items that overlapped between the roles or factors (i.e., item loadings > 0.32 or 10% of the variance across roles). Although discrete measures of residents' particular competences within any one role are important, as illustrated in the diagram that depicts the CanMEDS roles framework, these domains or constructs also overlap (e.g., the item "explores patient's beliefs, concerns and expectations" incorporates aspects of being an effective communicator/collaborator with being a professional and a health advocate). As shown in the factor analysis results, the item "[a]pplies new knowledge to daily practice" reflects the residents' competency as a Medical Expert (item loading = 0.609, accounting for 37% of the variance) while acknowledging the importance of using this new knowledge effectively in daily practice as part of the Communicator and Collaborator roles (0.586, 34% of the variance). Nevertheless, although an argument may be made that the Communicator and Collaborator roles are synonymous, they are also aspects of each role that make these constructs distinct. The combination of the Communicator and Collaborator roles warrants further investigation with the addition of modified or new items that are more content specific to define and measure each role distinctly.

Regarding the Health Advocate factor, the item "explicitly integrating opinions of the patient and caregivers into management plans" loads across three other factors and warrants rewording to better reflect an assessment of this role that does not evoke other domains, such as Communicator and Collaborator and Professional. Nevertheless, this item also illustrates the balance that must be taken into consideration when the objective is to measure competencies or abilities across a number of the roles or a mandate to provide a clear assessment of these domains as discrete constructs or stand-alone roles. Items believed to pertain to the Scholar role that were not identified as a separate factor also warrant further investigation, and the addition of new items should be developed and tested in future iterations.

There are several limitations to this study. First, combining the items across each year may have had an influence on the results. For example, items that were combined were measuring the same concept but were worded slightly differently and may have been evaluated differently based on the resident's year of

Table 1. Factor analysis of common items in emergency medicine in-training evaluation reports (R1–R5)

Item description	Factor				
	1	2	3	4	5
1. Medical Expert					
• Demonstrates a good general clinical knowledge in common problems	**0.788**				
• Develops a systematic and cognitive approach to clinical reasoning in common cases	**0.785**				
• Medical history: performs independently in common cases—appropriate, thorough, and directed	**0.749**				
• Physical examination: performs independently in common cases—appropriate, thorough, and directed	**0.836**				
• Diagnostic tests: understands and orders appropriate, cost-effective, basic laboratory tests and diagnostic images	**0.537**				0.438
• Consultancy skills: shows solid evidence of self-assessment by recognizing when personal limits are exceeded and demonstrates timely consultation with the staff emergency physician	**0.597**				
• Applies new knowledge to daily practice	**0.609**	0.586			
2. Communicator and Collaborator					
• Establishes and maintains rapport with patient and family		**0.828**			
• Explores patient's beliefs, concerns, and expectations		**0.648**	0.351	0.453	
• Informs and counsels patients about care decisions		**0.750**	0.323	0.353	
• Communicates effectively with other health workers in both verbal and written form		**0.628**	0.422		
• Maintains collegial and respectful relationships with medical, nursing, paramedical, and support staff		**0.601**	0.456		
• Understands the roles/expertise of the other individuals involved in patient care	0.436	**0.508**			0.340
• Identifies learning needs and makes use of available learning resources	0.559	**0.596**			
3. Professional					
• Is punctual for shifts and educational events and shows enthusiasm for learning			**0.706**		
• Follows through on assigned tasks, transfer of patient care; informs supervisor of errors or concerns			**0.806**		
• Demonstrates good self-assessment ability by being aware of own limitations and seeking feedback; reflects on practice			**0.733**		
• Delivers highest quality care with integrity, honesty, and compassion with sensitivity to racial, cultural, and societal issues		0.347	**0.763**		
• Upholds the legal and ethical standards of the profession		0.328	**0.779**		
4. Health Advocate					
• Explicitly integrates the opinions of the patient and caregivers into management plans	0.393	0.422	0.406	**0.527**	
• Identifies the patient's status with respect to one or more of the determinants of health	0.342		0.340	**0.764**	
• Assesses the patient's ability to access various services in the health and social system	0.345			**0.776**	
5. Manager					
• Effectively cares for several patients during a shift and follows them through their ED course					**0.865**
• Develop self-assessment skills to formulate realistic expectations and a balanced lifestyle		0.466		0.322	**0.549**
% of variance accounted for	21.5	19.5	19.5	10.8	8.2
Reliability coefficient (Cronbach alpha)	0.92	0.93	0.95	0.90	0.73

ED = emergency department.
Five factors account for 79% of the variance. Only item loadings ≥ 0.320 are reported. Some of the numbers are boldface because they are greater than 0.5. This means they show good evidence of construct validity due to a high factor loading. Where there are two numbers greater than 0.5, the highest number is chosen as the appropriate factor loading.

program. Second, the resulting factor solution would have benefited from a larger sample size to meet the expectations for a more comprehensive exploratory factor analysis. Third, given that data were collected over 2 years, there may have been some overlap with residents' assessments, although multiple ITERs from the same resident should not have theoretically influenced the reliability or validity of this analysis. Fourth, it is unknown how many encounters the evaluator had with the resident as the ITER was completed after an 8-week block rotation and evaluators with more encounters may have rated residents differently from evaluators with fewer encounters. Last, no data were collected that would have allowed us to take into consideration faculty clinical skills that have been shown to correlate with assessment behaviours.[7]

This study provides a starting point for the measurement of clinical competence in EM residents in Canada using a single evaluation method to assess all seven CanMEDS roles. Additionally, this study provides evidence for the reliability and construct validity of a common ITER that may be used to assess key competencies across all years of an EM residency program.

The next logical step for future research is identifying ITER items that will differentiate each CanMEDS role exclusively. This can be done within each year of residency in EM or across all years. The importance of sample size must be addressed to ascertain construct validity using factor analysis, although there are conflicting recommendations in terms of minimum sample size and the subject to variable ratio. Although some statisticians contend that at least 10 observations are needed for each item in the instrument being used, this can be based on a sample size as small as 200.[16] Although we were able to obtain an average of only 7.2 observations per item on the common 24-item EM ITER, this number would meet the subject to variable ratio of 5 to 1 purported by other researchers.[12,17] To overcome this, longitudinal data can be collected, although this could be time consuming and would not provide unique observations across specific PGYs.

It should be acknowledged that the manner in which the forms are completed and the sociocultural context in which they are deployed may have an impact on the reliability of the process. For example, no matter how well designed an ITER form may be, if the practice is to give every resident a "5" or "excellent" rating, then the form will generate very reliable results (high internal consistency as the ratings are always the same), but they will not be valid as not every resident is always a "5." Other areas of research should investigate how educators are trained to conduct assessments in EM and what can be done to reduce bias and error when conducting assessments. In addition, all residency programs should adopt a model for ITER evaluation that includes continuous performance assessment, assessing student performance in several student domains, and ensuring that measures are consistent, current, feasible, reliable, and valid.[14] This model should integrate other effective methods and clearly outline the standards that need to be met by residents.

Given the generic nature of the 24-item EM ITER investigated in this study, it could be applied to other related resident training programs. Concurrent validity testing with other tools, such as OSCEs and DECs, assessing clinical competence outlined by the CanMEDs roles is also warranted in EM and other training programs.[3,9] Although the use of an ITER may allow for the evaluation of the intrinsic domains such as the Health Advocate and Collaborator roles, which have been regarded as extraneous in EM compared to the Medical Expert, Communicator, and Professional roles, further exploration into the situational factors that affect clinical performance is also needed. From a measurement point of view, it is more feasible to measure scenarios and participants in isolation than to decipher their interactions according to Lurie, who states, "Descriptions of 'competencies' should thus specify the range of examinees and situations to which they apply. If they are specified in overly dispositional terms, or over too wide a range of unrelated situations, they will contribute to only a small fraction of the variance of any particular performance."[4]

Competing interests: None declared.

REFERENCES

1. Frank JR. *The CanMEDS 2005 physician competency framework. Better standards. Better physicians. Better care*. Ottawa: Royal College of Physicians and Surgeons of Canada; 2005.

2. Bandiera G, editor, Sherbino J, Frank JR. *The CanMEDS assessment tools handbook. An introductory guide to assessment methods for the CanMEDS competencies*. Ottawa: Royal College of Physicians and Surgeons of Canada; 2006.

3. Chou S, Cole G, McLaughlin K, et al. CanMEDS evaluation in Canadian postgraduate training programs: tools used and program director satisfaction. *Med Educ* 2009; 42:879-86, doi:10.1111/j.1365-2923.2008.03111.x.

4. Lurie S. History and practice of competency-based assessment. *Med Educ* 2011;46:49-57, doi:10.1111/j.1365-2923.2011.04142.x.

5. Chou S, Lockyer J, Cole G, et al. Assessing postgraduate trainees in Canada: are we achieving diversity in methods? *Med Teach* 2009;31:e58-63, doi:10.1080/01421590802512938.

6. Turnbull J, Van Barneveld C. Assessment of clinical performance: in-training evaluation In: Norman G, van der Vleuten C, editors. *International handbook of research in medical education*. Dordrecht (Holland): Kluwer Academic Publishers; 2002. p. 793-810.

7. Kogan JR, Hess BJ, Conforti LN, et al. What drives faculty ratings of residents' clinical skills? The impact of faculty's own clinical skills. *Acad Med* 2010;85 Suppl 10:S25-8, doi:10.1097/ACM.0b013e3181ed1aa3.

8. Watling CJ, Kenyon CF, Schulz V, et al. An exploration of faculty perspectives on the in-training evaluation of residents. *Acad Med* 2010;85:1157-62, doi:10.1097/ACM.0b013e3181e19722.

9. Watling CJ, Kenyon CF, Zibrowski EM, et al. Rules of engagement: residents' perceptions of the in-training evaluation process. *Acad Med* 2008;83(10 Suppl):S97-100, doi:10.1097/ACM.0b013e318183e78c.

10. Bandiera G, Lendrum D. Dispatches from the front: emergency medicine teachers' perceptions of competency-based education. *CJEM* 2011;13:155-61.

11. Sherbino J, Upadhye S, Worster A. Self-reported priorities and resources of academic emergency physicians for the maintenance of clinical competence: a pilot study. *CJEM* 2009;11:230-4.

12. MacCallum RC, Widaman KF, Zhang S, et al. Sample size in factor analysis. *Psychol Methods* 1999;4:84-99, doi:10.1037/1082-989X.4.1.84.

13. Lai J. Health advocacy in emergency medicine: a resident's perspective. *CJEM* 2009;11:99-100.

14. Sherbino J, Bandiera G, Frank JR. Assessing competence in emergency medicine trainees: an overview of effective methodologies. *CJEM* 2008;10:365-71.

15. Royal College of Physicans and Surgeons of Canada. *Objectives of training in emergency medicine 2011*. Available at: http://rcpsc.medical.org/residency/certification/objectives/emergmed_e.pdf (accessed April, 2012).

16. Tabachnick BG, Fidell LS. *Using multivariate statistics*. 3rd ed. Northridge (CA): Pearson Education; 2007.

17. Bryant FB, Yarnold PR. Principal components analysis and exploratory and confirmatory factor analysis. In: Grimm LG, Yarnold RR, editors. *Reading and understanding multivariale statistics*. Washington (DC): American Psychological Association; 1995. p. 99-136.