Family Medicine Mandatory Assessment of Progress

Results of a pilot administration of a family medicine competency-based in-training examination

Fok-Han Leung MD CCFP MHSc   Jodi Herold MA PhD   Karl Iglar MD CCFP

Abstract

Objective To report the results of a pilot in-training progress test, the Family Medicine Mandatory Assessment of Progress, taken by first- and second-year postgraduate family medicine trainees.

Design Assessment of resident performance on a key-features approach multiple-choice progress test. Test questions were developed by competency content area experts.

Setting University of Toronto in Ontario.

Participants First- and second-year family medicine residents.

Main outcome measures Construct validity was assessed based on performance on the test by first- and second-year residents, Canadian and international medical graduates, and residents with more or less than 1 month of relevant clinical experience.

Results Pilot progress testing of family medicine residents (N = 255) at the University of Toronto revealed a significant 1.6% difference ($P<.01$) in mean scores between first- and second-year postgraduate family medicine trainees and achieved construct validity across many parameters studied. The agreement coefficients for residents being identified as the poorest performers ranged from 0.88 to 0.90 depending on the domain of practice assessed.

Conclusion Competency-based progress testing using the key-features model is a valid means of assessing the progress of family medicine residents.

EDITOR'S KEY POINTS

• Competency-based education is becoming the standard of training in family medicine in Canada. As part of its evaluation process, the University of Toronto in Ontario used a key-features approach to develop in-training progress testing. This study was a pilot assessment of the progress testing among first- and second-year family medicine residents.

• As expected, second-year residents performed better than first-year residents, and Canadian medical graduates performed better than international medical graduates. There was also a trend toward more clinical exposure leading to higher scores, with differences reaching statistical significance for care of children and adolescents, maternity care, and end-of-life care ($P<.05$). These results suggest that such a key-feature problem approach is a valid method of assessment of the family medicine expert role in a competency-based curriculum.

Évaluation obligatoire des progrès en médecine familiale

Résultats d'un examen pilote évaluant la compétence en médecine familiale durant la formation

Fok-Han Leung MD CCFP MHSc   **Jodi Herold** MA PhD   **Karl Iglar** MD CCFP

Objectif Présenter les résultats d'un examen pilote évaluant les progrès accomplis en cours d'étude, le *Family Medecine Mandatory Assessment of Progress*, auquel ont participé des résidents 1 et 2 en médecine familiale.

Type d'étude Évaluation des progrès obtenus par des résidents à l'aide d'un examen à choix multiples portant sur des domaines clés. Les questions de l'examen ont été formulées par des experts du domaine de la compétence.

Contexte L'Université de Toronto, en Ontario.

Participants Des résidents 1 et 2 en médecine familiale.

Principaux paramètres à l'étude On a évalué la validité conceptuelle de l'examen en comparant les résultats des résidents 1 et des résidents 2, ceux des médecins diplômés au Canada ou à l'étranger, et ceux des résidents ayant plus d'un mois ou moins d'un mois d'expérience clinique.

Résultats L'examen pilote mesurant les progrès de 255 résidents en médecine familiale de l'Université de Toronto a révélé une différence significative de 1,6 % ($P <$,01) entre les scores moyens des résidents 1 et 2 en médecine familiale; l'examen a aussi démontré une validité conceptuelle à l'égard de plusieurs des paramètres à l'étude. Les coefficients d'accord pour les résidents identifiés comme les moins performants variaient entre 0,88 et 0,90 selon le domaine de pratique évalué.

Conclusion Évaluer les progrès en compétence à l'aide du modèle basé sur des domaines clés est une façon valide d'évaluer les progrès des résidents en médecins familiale.

POINTS DE REPÈRE DU RÉDACTEUR

• En médecine familiale, une formation axée sur les compétences est en voie de devenir la norme au Canada. Dans le cadre de son processus d'évaluation, l'Université de Toronto, en Ontario, a utilisé une approche basée sur des domaines clés pour développer une évaluation des progrès durant les études. Cette étude se voulait une évaluation pilote des progrès observés chez des résidents 1 et 2 en médecine familiale.

• Comme on pouvait s'y attendre, les résidents 2 ont mieux réussi que les résidents 1, tandis que les médecins formés au Canada ont été meilleurs que ceux formés à l'étranger. On observait aussi une tendance selon laquelle une exposition plus importante à la clinique entraînait des meilleurs scores, cette différence devenant significative dans le cas des soins aux enfants et aux adolescents, des soins obstétricaux et des soins de fin de vie ($P <$,05). Ces résultats suggèrent qu'une évaluation basée sur des problèmes clés est une méthode valide pour évaluer les connaissances en médecine familiale dans un curriculum fondé sur la compétence.

Adequate assessment of medical trainees requires a number of tools.[1,2] Reliance on a single method of assessment might lead to incomplete evaluation.[2] When program directors in family medicine were asked to predict the performance of trainees on the national certification examination, they were only able to accurately identify 25% of the top and bottom performers.[3] In-training examinations have been shown to be predictive of success on certification examinations.[4,5]

Progress testing is a longitudinal assessment process used to test functional knowledge and competence.[6,7] It has been used in the undergraduate medical curriculum at a number of institutions.[8-10] There are several reports of the use of this type of assessment method in postgraduate education, most notably in obstetrics and gynecology,[11] radiology,[12] internal medicine,[13] and osteopathic surgical residency programs.[14]

In 2009 the Department of Family and Community Medicine (DFCM) at the University of Toronto in Ontario introduced a competency-based curriculum in family medicine based on the CanMEDS–Family Medicine framework.[15,16] As one means of evaluating the family medicine expert role, specifically knowledge acquisition and application, the DFCM implemented a series of progress tests to serve as formative feedback for residents in planning future learning experiences.

In this article we will describe our experience in the DFCM at the University of Toronto with the pilot administration of an in-training examination in family medicine residency and provide initial evidence of construct validity.

METHODS

Setting
The family medicine training program in Canada is 24 months in length. The DFCM has the largest training program in family medicine in Canada, with more than 250 residents enrolled in the 2-year program at the time of the study. It consists of 14 geographically distributed sites, as well as a rural residency program. Residents are assigned to one of these sites for the duration of the training program.

Development of the progress assessment
The Family Medicine Mandatory Assessment of Progress (FM-MAP) was introduced to assess knowledge acquisition and application in a competency-based curriculum. The tests are written twice a year. The tests, administered during academic half-days (weekly academic time protected for all trainees in the program), comprise more than 120 questions and are 4 hours in length. The questions are case-based and use the key-feature problem approach.[17,18] The key-feature items were developed

based on the learning outcomes in our competency-based curriculum. Identified family medicine experts in 12 competency domains produced documents detailing the expected competencies to be achieved for the practice-ready physician.

Similar groups to those writing the competency documents were assembled and provided with a half-day key-features writing orientation and workshop. Writing groups were instructed to write questions aimed at the practice-ready physician.

The test was administered in the fall of 2009, with an alternative date provided 1 week later. Residents were informed that the test was mandatory. Each local hospital training site was provided with the test materials a day in advance and was instructed to find appropriate testing facilities. Detailed test distribution and invigilation instructions were provided. All papers were collected afterward and returned to a central location.

The scoring key for the multiple-choice scanner was developed in advance of the test. After the results were analyzed, the individual results were distributed to each resident electronically. The scores were reported as deciles. Each resident received an overall decile score, a decile score based on level of training (first or second year), and a percent score relating to each competency domain, with the mean (SD) score for all residents as a comparator. Site program directors received site-specific scores for overall performance and for each competency domain. Both residents and site directors also received general feedback on competency domains where the performance of the entire cohort was lower relative to other domains.

As a means of gathering evidence of construct validity, total (percent correct) scores were compared between first- and second-year residents, between Canadian medical graduates (CMGs) and international medical graduates (IMGs), and between residents identified by their site directors as the "top 5%" and all other residents. Competency-area score comparisons were performed between first- and second-year residents, and between residents who had less than 1 month of clinical experience in that area and those who had 1 month or more of experience. All group comparisons were done using ANOVA (analysis of variance) in SPSS.

The study was approved by the Research Ethics Board of St Michael's Hospital in Toronto.

RESULTS

A total of 255 of the 261 residents participated in the first iteration of the FM-MAP. The 6 residents who were excluded were unable to return to their home sites to write the test. Overall, 132 (51.8%) were first-year residents; 198 (77.6%) were CMGs and 57 (22.4%) were IMGs.

The first iteration of the FM-MAP achieved an α coefficient of 0.76. The mean score on the examination was 75.6%. Second-year residents achieved a mean (SD) score of 76.4% (4.7%) compared with 74.8% (4.8%) for first-year residents ($F_{1,253} = 7.3$, $P < .01$, partial $\eta^2 = 0.03$). The mean (SD) score was 76.7% (3.8%) for CMGs compared with 71.7% (5.9%) for IMGs ($F_{1,253} = 59$, $P < .001$, partial $\eta^2 = 0.19$). Program directors were asked to predict their "top 5%" of residents who would perform best on the FM-MAP (before knowing the results). These identified residents had a mean score of 77.7% compared with 74.9% for the rest of the cohort ($F_{1,253} = 15.8$, $P < .001$).

The agreement coefficients for residents being identified as the poorest performers ranged from 0.88 to 0.90 depending on the domain of practice assessed.

Table 1 shows the performance of residents on the test as a function of year of training and previous clinical exposure to the relevant domain of care. Exposure to care of children and adolescents, maternity care, and end-of-life care corresponded to statistically significantly higher scores.

DISCUSSION

This study describes promising results from the pilot administration of a written examination for family medicine residents in the largest training program in Canada. The historical mainstay of written examinations in medicine has been multiple-choice questions.[19] With the advantage of being able to cover a large breadth of clinical material and being easy to mark, multiple-choice questions are employed at all points of the training process. However, this style of question lacks the ability to probe the depth of knowledge of trainees and inadequately addresses knowledge application. Conversely, patient management problems and modified essay questions, while assessing the clinical process, are very long, and only a low number of cases can be sampled.

Our examination follows a multiple-choice format using key-feature problems as the basis for the questions and is intended to assess the family medicine expert role in our competency-based curriculum.

Key-features style questions are gaining in popularity[20-22] because of their ability to assess clinical decision-making skills; beyond simple medical knowledge recall, knowledge application is also tested. Optically, a key-feature question can be similar to a multiple-choice question; but rather than knowledge recall, a key-feature question will involve a case-based clinical decision. In developing a key-features question, a question writer would consider a particular clinical expert competency and ask "What are the critical or essential steps in the resolution of this problem?" This would form the list of key features or critical steps around which the case would be written.

Our results provide promising evidence of test validity. As expected our second-year residents outperformed our first-year residents. Although the difference in mean score is relatively small (1.6%), it is statistically significant and in keeping with the literature; a previous study assessing in-training examinations in internal medicine

Table 1. Effect of year of training and related experience on mean scores for competency areas on the FM-MAP test

COMPETENCY AREA	MEAN (SD) SCORE BY YEAR OF TRAINING		MEAN (SD) SCORE BY EXPERIENCE IN RELATED ROTATION	
	FIRST	SECOND	<1 MO	≥1 MO
Behavioural medicine	80.1 (11.1)	78.1 (12.1)	NA*	NA*
Care of children and adolescents	59.1 (8.1)	60.1 (8.1)	58.9 (8.0)	61.1 (8.2)†
Care of the elderly	79.1 (9.1)	79.1 (9.1)	79.5 (9.3)	79.1 (7.0)
Emergency medicine	69.1 (9.1)	72.1 (9.1)†	70.0 (8.6)	71.8 (9.3)
End-of-life care	75.1 (7.1)	75.1 (8.1)	74.8 (7.5)	79.1 (7.6)†
Family medicine	89.1 (7.1)	89.1 (8.1)	NA*	NA*
In-hospital care	69.1 (15.1)	71.1 (14.1)	68.9 (15.2)	71.1 (13.8)
Maternity care	84.1 (9.1)	87.1 (8.1)§	83.5 (9.0)	81.1 (7.7)§
Mental health care	79.1 (9.1)	81.1 (8.1)	79.8 (8.8)	80.8 (5.9)
Musculoskeletal medicine	64.1 (11.1)	65.1 (12.1)	NA*	NA*
Surgical skills	79.1 (12.1)	82.1 (13.1)‡	80.1 (12.7)	83.1 (11.7)
Women's health	80.1 (10.1)	83.1 (9.1)‡	81.1 (9.8)	82.2 (6.8)

FM-MAP—Family Medicine Mandatory Assessment of Progress, NA—not applicable.
*There were 3 competency areas that could not be linked to a single rotation for experiential analysis.
†Effect is significant ($P < .05$).
‡Effect is significant ($P < .01$).
§Effect is significant ($P < .001$).

demonstrated a mean difference of 5.1% (95% CI 4.1% to 6.2%) between adjacent training years.[23]

Our residents in the CMG group outperformed our residents in the IMG group by 7.0%. This is again an expected outcome and supported by the literature, in which internationally trained residents usually achieve less favourable certification examination scores compared with locally trained residents.[24-26] In fact, one study found that 56.0% of IMG physicians in Quebec were successful on the Certification Examination in Family Medicine compared with 93.5% of Canadian and American medical graduates.[27]

A previous study noted that residency program directors in family medicine were not able to accurately predict top and bottom performers.[3] Our results suggest otherwise; further study will be required to understand this difference.

Clinical exposure should lead to better outcomes on the test. There was indeed a trend in this direction across all the domains examined, although differences only reached statistical significance in care of children and adolescents, maternity care, and end-of-life care. The small differences in performance can be partly explained by the fact that even residents who indicated they had a lack of clinical exposure to the various domains of practice examined indeed were exposed to these areas in the course of their training in family medicine environments.

Limitations

This study reports the findings of our pilot implementation. Future studies are needed and will include continuing to examine the progress of trainees with repeated iterations of the FM-MAP in the course of the 2-year residency program. While our normative standards demonstrated construct validity, next steps should also include standard setting using external standards such as comparison with the Certification examination.[28]

Conclusion

The FM-MAP using a key-feature problem approach is a valid method of assessment of the family medicine expert role in a competency-based curriculum. We demonstrated evidence of construct validity across several parameters. Designed for use with a multiple-choice scanner, the FM-MAP is a good formative assessment tool that can be efficiently and economically implemented.

Dr Leung is a staff physician at St Michael's Hospital in Toronto, Ont, and Assistant Professor in the Department of Family and Community Medicine and Associate Program Director of the Family Medicine Postgraduate Program at the University of Toronto. Dr Herold is Education Assessment and Psychometrics Consultant in the Postgraduate Medical Education Office at the University of Toronto. Dr Iglar is a staff physician at St Michael's Hospital, and Associate Professor in the Department of Family and Community Medicine and Program Director of the Family Medicine Postgraduate Program at the University of Toronto.

References
1. Sherbino J, Bandiera G, Frank JR. Assessing competence in emergency medicine trainees: an overview of effective methodologies. *CJEM* 2008;10(4):365-71.
2. Epstein RM, Hundert E. Defining and assessing professional competence. *JAMA* 2002;287(2):226-35.
3. Brailovsky CA, Grand'Maison P, Lescop J. Residency directors' predictions of candidates' performances on a licensing examination. *Acad Med* 1995;70(5):410-4.
4. Kearney RA, Sullivan P, Skakun E. Performance on ABA-ASA in-training examination predicts success for RCPSC certification. *Can J Anaesth* 2000;47(9):914-8.
5. Holmboe ES, Hawkins RE. Methods for evaluating the clinical competence of residents in internal medicine: a review. *Ann Intern Med* 1998;129(1):42-8.
6. McHarg J, Bradley P, Chamberlain S, Ricketts C, Searle J, McLachlan JC. Assessment of progress tests. *Med Educ* 2005;39(2):221-7.
7. Schuwirth LW, van der Vluten CP. The use of progress testing. *Perspect Med Educ* 2012;1(1):24-30. Epub 2012 Mar 10.
8. Blake JM, Norman GR, Keane DR, Mueller CB, Cunnington J, Didyk N. Introducing progress testing in McMaster University's problem-based medical curriculum: psychometric properties and effect on learning. *Acad Med* 1996;71(9):1002-7.
9. Arnold L, Willoughby TL. The quarterly profile examination. *Acad Med* 1990;65(8):515-6.
10. Van der Vleuten CP, Verwijnen GM, Wijnen WH. Fifteen years experience with progress testing in a problem-based learning curriculum. *Med Teach* 1996;18(2):103-9.
11. Dijksterhuis MG, Scheele F, Schuwirth LW, Essed GG, Nijhuis JG, Braat DD. Progress testing in postgraduate medical education. *Med Teach* 2009;31(10):e464-8.
12. Ravesloot C, van der Schaaf M, Haaring C, Kruitwagen C, Beek E, Ten Cate O, et al. Construct validation of progress testing to measure knowledge and visual skills in radiology. *Med Teach* 2012;34(12):1047-55. Epub 2012 Aug 30.
13. Pugh D, Touchie C, Wood TJ, Humphrey-Murto S. Progress testing: is there a role for the OSCE? *Med Educ* 2014;48(6):623-31.
14. Shen L. Progress testing for postgraduate medical education: a four year experiment of American College of Osteopathic Surgeons resident examinations. *Adv Health Sci Educ Theory Pract* 2000;5(2):117-29.
15. Iglar K, Whitehead C, Glover Takahashi S. Competency-based education in family medicine. *Med Teach* 2013;35(2):115-9. Epub 2012 Oct 26.
16. Working Group on Curriculum Review. *CanMEDS–Family Medicine.* Mississauga, ON: College of Family Physicians of Canada; 2009. Available from: **www.cfpc.ca/uploadedFiles/Education/CanMeds%20FM%20Eng. pdf**. Accessed 2016 Apr 11.
17. Page G, Bordage G, Allen T. Developing key-feature problems and examinations to assess clinical decision-making skills. *Acad Med* 1995;70(3):194-201.
18. Farmer EA, Page G. A practical guide to assessing clinical decision-making skills using the key features approach. *Med Educ* 2005;39(12):1188-94.
19. Gierl MJ, Lai H, Turner SR. Using automatic item generation to create multiple-choice test items. *Med Educ* 2012;46(8):757-65.
20. Nikendei C, Mennin S, Weyrich P, Kraus B, Zipfel S, Schrauth M, et al. Effects of a supplementary final year curriculum on students' clinical reasoning skills as assessed by key-feature examination. *Med Teach* 2009;31(9):e438-42.
21. Hatala R, Norman GR. Adapting the Key Features Examination for a clinical clerkship. *Med Educ* 2002;36(2):160-5.
22. Page G, Bordage G. The Medical Council of Canada's Key Features Project: a more valid written examination of clinical decision-making skills. *Acad Med* 1995;70(2):104-10.
23. McDonald FS, Zeger SL, Kolars JC. Factors associated with medical knowledge acquisition during internal medicine residency. *J Gen Intern Med* 2007;22(7):962-8. Epub 2007 Apr 28.
24. Boulet JR, Swanson DB, Cooper RA, Norcini JJ, McKinley DW. A comparison of the characteristics and examination performances of U.S. and non-U.S. citizen international medical graduates who sought Educational Commission for Foreign Medical Graduates Certification: 1995-2004. *Acad Med* 2006;81(10 Suppl):S116-9.
25. Benson JA Jr, Meskauskas JA, Grosso LJ. Performance of U.S. citizen-foreign medical graduates on certifying examinations in internal medicine. *Am J Med* 1981;71(2):270-3.
26. Hechel H, Thompson Bowles L. Specialty certification in North America: a comparative analysis of examination results. *J Med Ed* 1979;54:69-74.
27. MacLellan AM, Brailovsky C, Rainsberry P, Bowmer I, Desrochers M. Examination outcomes for international medical graduates pursuing or completing family medicine training in Quebec. *Can Fam Physician* 2010;56:912-8.
28. Ricketts C, Freeman AC, Coombes LR. Standard setting for progress tests: combining external and internal standards. *Med Educ* 2009;43(6):589-93.