

Validation of a large-scale clinical examination for international medical graduates

Susan Glover Takahashi MA PhD Arthur Rothman EdD Marla Nayer MEd PhD
Murray B. Urowitz MD FRCP(C) Anne Marie Crescenzi

Abstract

Objective To evaluate a new examination process for international medical graduates (IMGs) to ensure that it is able to reliably assign candidates to 1 of 4 competency levels, and to determine if a global rating scale can accurately stratify examinees into 4 levels of learners: clerks, first-year residents, second-year residents, or practice ready.

Design Validation study evaluating a 12-station objective structured clinical examination.

Setting Ontario.

Participants A total of 846 IMGs, and an additional 63 randomly selected volunteers from 2 groups: third-year clinical clerks (n=42) and first-year family medicine residents (n=21).

Main outcome measures The accuracy of the stratification of the examinees into learner levels, the impact of the patient-encounter ratings and postencounter oral questions, and between-group differences in total score.

Results Reliability of the patient-encounter scores, postencounter oral question scores, and the total between-group difference scores was 0.93, 0.88, and 0.76, respectively. Third-year clerks scored the lowest, followed by the IMGs. First-year residents scored highest for all 3 scores. Analysis of variance demonstrated significant between-group differences for all 3 scores ($P < .05$). Postencounter oral question scores differentiated among all 3 groups.

Conclusion Clinical examination scores were capable of differentiating among the 3 groups. As a group, the IMGs seemed to be less competent than the first-year family medicine residents and more competent than the third-year clerks. The scores generated by the postencounter oral questions were the most effective in differentiating between the 2 training levels and among the 3 groups of test takers.

EDITOR'S KEY POINTS

- In 2003, the Canadian Task Force on Licensure of International Medical Graduates (IMGs) recommended that provinces increase the number of IMGs being assessed and that the assessment process be made able to reliably differentiate between various levels of competence. This study sought to evaluate a new examination process for IMGs.
- The results of this study provide evidence of the validity of the clinical examination and the rating scale used to determine whether the level of competence of the IMG candidates was at the clerkship or first-year-resident level. The study suggests that incorporating postencounter questions that address clinical problem solving related to the patient just seen provides important information related to the candidate's competence. The study also strongly supports the use of a common rating scale to assess multiple levels of competence.

This article has been peer reviewed.
Can Fam Physician 2012;58:e408-17

Validation d'un examen clinique de grande envergure pour les médecins diplômés à l'étranger

Susan Glover Takahashi MA PhD Arthur Rothman EdD Marla Nayer MEd PhD
Murray B. Urowitz MD FRCP(C) Anne Marie Crescenzi

Résumé

Objectif Évaluer un nouveau type d'examen pour les médecins diplômés à l'étranger (MDÉ) afin de s'assurer qu'il permette de répartir avec fiabilité les candidats selon les 4 niveaux de compétence et déterminer si une échelle d'évaluation globale peut répartir les candidats en 4 niveaux d'apprentissage: stagiaires de clinique, résidents 1, résidents 2 ou prêts à pratiquer.

Type d'étude Étude de validation d'un examen clinique structuré objectif comportant 12 stations.

Contexte L'Ontario.

Participants Un total de 846 MDÉ en plus de 63 volontaires choisis au hasard parmi 2 groupes: stagiaires cliniques de troisième année (n=42) et résidents 1 de médecine familiale (n=21).

Principaux paramètres à l'étude La précision de la répartition des candidats dans les différents niveaux d'apprentissage, l'impact des scores attribués aux stations de rencontre avec des patients, les scores pour les questions orales post rencontre et les différences entre groupes pour le score total.

Résultats Les scores attribués pour les rencontres avec les patients, pour les questions orales suivant les rencontres et pour les différences totales entre groupes avaient une fiabilité de 0,93, 0,88 et 0,76, respectivement. Les stagiaires de clinique de troisième année avaient les scores les plus faibles, suivis des MDÉ. Les résidents 1 avaient les scores les plus hauts pour les 3 catégories. L'analyse de variance révélait des différences significatives entre les groupes pour les 3 scores ($P < ,05$). Les scores pour les questions post rencontre pouvaient différencier les 3 groupes.

Conclusion Les scores obtenus à l'examen clinique pouvaient différencier les 3 groupes. Comme groupe, les MDÉ semblaient moins compétents que les résidents de première année de médecine familiale et plus compétents que les stagiaires cliniques de troisième année. Les scores obtenus aux questions post rencontre étaient les plus efficaces pour différencier les 2 niveaux de formation et les 3 groupes ayant subi ces tests.

POINTS DE REPÈRE DU RÉDACTEUR

- En 2003, le Groupe de travail sur la diplomation des médecins diplômés à l'étranger (MDÉ) recommandait que les provinces augmentent le nombre de MDÉ évalués et que l'évaluation soit capable de différencier de façon fiable leurs divers niveaux de compétence. La présente étude visait à évaluer un nouveau type d'examen pour les MDÉ.
- Les résultats de cette étude apportent des preuves de la validité de l'examen clinique et de l'échelle de scores utilisée pour déterminer si le niveau de compétence des MDÉ évalués correspond au niveau de stagiaire clinique ou de résident 1. Les résultats laissent entendre que le fait d'introduire des questions portant sur la solution du problème clinique présenté par le patient qui vient d'être rencontré fournit des informations importantes sur la compétence du candidat. L'étude favorise aussi fortement l'utilisation d'une échelle d'évaluation commune pour les multiples niveaux de compétence.

Cet article a fait l'objet d'une révision par des pairs.
Can Fam Physician 2012;58:e408-17

The Canadian Task Force on Licensure of International Medical Graduates (IMGs) was established in 2002 with a mandate to aid in integrating qualified IMGs into the Canadian work force.¹ The task force involved key stakeholders from the medical community—including medical students and residents; federal, provincial, and territorial government representatives; Human Resources and Skills Development Canada; Citizenship and Immigration Canada; and IMGs—in the development of the recommendations. *International medical graduates* were defined as individuals holding medical degrees from schools not accredited by the Committee on Accreditation of Canadian Medical Schools or the Liaison Committee on Medical Education. A number of challenges were identified for IMGs entering the work force, one of which was that they were often unable to demonstrate their skills owing to work force policies, limited access to assessment or training, and lack of support for understanding the licensure requirements.¹

In September 2003 the task force presented 6 recommendations to the federal Advisory Committee on Health Delivery and Human Resources. One recommendation was to increase the capacity to assess and prepare IMGs for licensure. A report from the task force on IMG assessment and licensure identified characteristics of medical licensure policies and practices that were effective, fair, and cost-effective.¹ The following characteristics were included:

- having the capacity to reliably differentiate between competent and incompetent;
- being based upon a reasonable standard that is applied evenly to all;
- having sufficient capacity to evaluate all who wish to be assessed; and
- being efficient and cost-effective, and decreasing time delays or unjustified costs so that these are not barriers.

Selection of IMGs into residency or practice has been a matter of interest for some time. While estimates vary, IMGs comprise up to one-third of the Canadian²⁻⁵ and American^{6,7} physician work force. Before moving to Canada, most IMGs had been in independent practice, 42.8% had practised for 1 to 5 years, and 45.6% had practised for 6 to 20 years.⁴

Testing of IMGs' clinical skills began in the United States in July 1998, in response to concerns that IMGs might be lacking basic clinical skills.^{3,8}

As regulation is provincially mandated in Canada, each province has its own process for assessing and integrating IMGs. British Columbia⁹ and Alberta¹⁰ assess IMGs for entry into residency programs, while Manitoba,¹¹ Nova Scotia,¹² and Quebec¹³ assess IMGs for entry to practice (Table 1). Ontario has a long history of programs designed to assess, educate, and

integrate IMGs into medical practice. These include the Ontario Pre-internship Program (PIP) (1986-1999), the Ontario International Medical Graduate Program (1999-2003), the Assessment Program for International Medical Graduates (2003-2004), and International Medical Graduates Ontario (IMGO) (2003-2007).

Starting in 1986, PIP introduced a vigorous screening process that included both a written examination and an objective structured clinical examination (OSCE). This program provided 24 positions at the clerkship level to IMGs. Candidates then applied for residency positions in the same way as Canadian medical graduates (CMGs). This was followed by the Ontario International Medical Graduate Program, which continued with the PIP program processes while increasing the number of clerkship spots from 24 to 50.

The Assessment Program for International Medical Graduates was the beginning of practice-ready assessment (PRA) positions. This evolved with IMGO into assessments for entry at the second-year-resident level. The Clinical Examination 1 (CE1) was developed by IMGO and launched in 2003. The validation study reported here occurred under IMGO in 2005. During IMGO, the CE1 was administered at multiple sites across Ontario.

The IMGO was an organization formed under the leadership of several partners: the Council of Ontario Faculties of Medicine, the College of Physicians and Surgeons of Ontario, and the Ontario Ministry of Health and Long-Term Care. The IMGO was a provincial program developed in consultation with the Royal College

Table 1. Provincial international medical graduate assessment programs

PROVINCE OR PROGRAM	ASSESSMENT PROCESS
Alberta International Medical Graduate program	<ul style="list-style-type: none"> • File review • 10-station OSCE • 9-station multiple mini-interview
Nova Scotia Clinical Assessment for Practice Program	<ul style="list-style-type: none"> • OSCE • Therapeutics examination • 13-mo period of monitored CPD
International Medical Graduates of British Columbia	<ul style="list-style-type: none"> • 16-station OSCE • 1-wk orientation • 12-wk clinical assessment
Manitoba International Medical Graduate Program has a Clinical Assessment Professional Enhancement component	<ul style="list-style-type: none"> • 3-d assessment • Multiple-choice test • Therapeutics assessment • Structured oral interview • Comprehensive clinical encounter
Quebec	<ul style="list-style-type: none"> • Evaluation while working under a restricted licence to practise

CPD—continuing professional development, OSCE—objective structured clinical examination.

of Physicians and Surgeons of Canada, the College of Family Physicians of Canada, and the Medical Council of Canada (MCC). The IMGGO provided access to professional practice in Ontario for IMGs who met the Ontario regulatory requirements.

The IMGGO screening process was straightforward. Candidates who met basic eligibility requirements were invited to participate in a comprehensive written examination and a comprehensive OSCE. Basic eligibility requirements included the following: being a graduate of a medical school that was, at the time of the applicant's graduation, listed in the World Directory of Medical Schools published by the World Health Organization or the Foundation for Advancement of International Medical Education and Research; submission of transcripts and certificates; language proficiency in English or French; successful completion of the MCC Evaluating Examination (MCCEE) and the MCC Qualifying Examination 1 (MCCQE1); submission of a curriculum vitae; and payment of registration fees. Based on an evaluation of the candidates' academic and professional histories and the results of these assessments (specifically the results of the written examination and the OSCE), candidates were classified as being in 1 of 4 competency levels: ready for entry to training at the clerkship level (pre-residency); ready for first-year residency; practice ready (ready for practice pending 6 months in practice assessment); or as being not competent or unacceptable to the IMGGO program.

The OSCE provided important input to the IMGGO selection process. There is an extensive body of literature attesting to the validity of OSCE-type examinations¹⁴⁻¹⁶ used at different levels of clinical training in medicine¹⁷ and in the assessment of IMGs.¹⁸ The use of different ratings in OSCE-type examinations has also been described.¹⁹⁻²² The unique feature of the IMGGO comprehensive CE1 was the requirement that its results assign candidates to 4 different levels of competence (unsatisfactory, ready for clerkship, ready for residency, and practice ready). The focus of the examination is on the CanMEDS medical expert and communicator roles. Candidates who are more competent would be assessed as being ready for first-year residency or as practice ready; those less competent would be classified as unsatisfactory or as being at the clerkship level.

The next iteration of IMG assessments was the launch of the Centre for the Evaluation of Health Professionals Educated Abroad (CEHPEA) in 2007, which continues to the present day. The CEHPEA has a dedicated examination facility for administering the examinations, and the CE1 has been offered at this single site with up to 8 administrations in an application cycle. The CEHPEA administered the CE1 through 2010, as well as continued with advanced-level assessments

for second-year resident or PRA positions in up to 11 specialties. Most IMG candidates taking the CE1 aspired to enter first-year family medicine residency programs. Following completion of the CE1, candidates would apply through the Canadian Residency Matching Service for residency positions. While the process is the same as that for CMGs, IMGs candidates are not competing against CMGs for positions but are applying for specific IMG positions funded by the Ministry of Health. By 2010 there were 75 family medicine positions allocated for IMGs.

Objective and hypotheses

One source of evidence of the validity of the results generated by this examination would be the demonstration that Canadian trainees are appropriately differentiated and classified by the examination. The study reported here examined the ability of the CE1 results to validly differentiate between 2 levels of competence—clerkship and first-year residency—using a common rating form. To this end, samples of third-year clerks and first-year family medicine residents from the University of Toronto in Ontario were included in the October 2005 test administration, and their results were compared with the results of the IMG candidates.

There were 3 hypotheses for this study:

- that the residents would demonstrate significantly higher results (at least 0.5 SD) than the third-year clerks, with only modest overlap between the 2 score distributions;
- that in the results of those components of the examination that related specifically to resident-level skills and knowledge, the differences between the results of the 2 groups would be considerable (ie, at least 0.5 SD); and
- that a global rating form would function effectively to discriminate among different levels of performance.

Our use of a 0.5 SD seemed reasonable given the distributions involved. It appeared that a difference larger than this would be "clinically" significant (ie, would have clinical meaning).

METHODS

Participants

The study participants were randomly selected volunteers from 2 groups of students at the University of Toronto. Group 1 consisted of third-year clinical clerks and group 2 consisted of first-year family medicine residents. The recruitment notice specified that only CMGs were eligible to volunteer.

Students at these levels were selected for the study because the primary entry point for IMGs was first-year

residency; hence the intent of the CE1 is to differentiate between those who are ready to enter residency (first-year residents) with those who are not (clerks who have not yet finished medical school).

The examination was administered on October 1, 2005, to 846 IMG candidates in 5 Ontario cities simultaneously: Ottawa, Kingston, Toronto, Hamilton, and London. The largest administration site was Toronto.

Fifty third-year medical students and 50 first-year family medicine residents from the 2005 to 2006 cohorts at the University of Toronto were selected randomly and invited to participate in the study. Sixty-three students accepted the invitation and participated in the clinical examination: 42 were third-year clinical clerks and 21 were first-year family medicine residents. Students were paid an honorarium for their participation. The study participants, all from the University of Toronto, attended the examination at the Toronto site. The physician examiners and IMGs were blind to whether participants were IMGs, clerks, or residents.

The third-year clinical clerks were students at the beginning of their formal hospital-based clinical clerkship. Their performance in the CE1 would define the minimum level of acceptable performance for the IMG candidates.

Ethics approval for this project was received from the University of Toronto Ethics Review Office. Approval for this project was also provided by the appropriate program directors and associate deans of medical education. Participants were informed that individual results would not be disclosed to the Toronto Program Director or Associate Dean of Postgraduate Medical Education. Participants were assigned once-only identification numbers and these numbers were attached to examination materials and results. Participants received confidential individual performance reports.

Examination

The IMG CE1 was a 12-station OSCE with standardized patients portraying the cases and experienced physician educators rating the performance of the candidates. These examiners were all faculty at 1 of the 5 Ontario medical schools. An overview of the examination design is found in **Box 1**.

Each station lasted 10 minutes, divided into 8 minutes for the encounter with a standardized patient and 2 minutes for postencounter oral questions posed by the physician examiner. As a rule, most of the postencounter questions addressed resident- rather than clerkship-level competencies (eg, investigations and patient management).

The objectivity of the clinical examination was achieved by using standardized guidelines for the administration of the examination, training of physician

Box 1. Overview of the examination design

General features

- Examination evaluates primary care medical knowledge, is "comprehensive," and is generalist in nature (ie, *not* specialist examination)
- Examination evaluates postgraduate and practitioner medical knowledge (ie, beyond medical clerk)

Clinical domains evaluated

- Medicine: 2 stations
- Pediatrics: 2 stations
- Obstetrics and gynecology: 2 stations
- Surgery: 2 stations
- Psychiatry: 2 stations
- Prevention: 1 station
- Other: 1 station

Competencies evaluated

- Communication: All cases
- History: 10 cases
- Physical: 4 cases
- Investigation: 9 cases
- Judgment and decision making: All cases
- Management: 9 cases
- Health promotion: 5 cases

examiners and standardized patients, and common rating forms rather than station-specific checklists. Each site received all documents from a central location. This included specific instructions and material for orientation of candidates and examiners and all administrative materials (eg, candidate labels and booklets, station material, test sheets, and training materials for the standardized patients).

All examiners at the study site (Toronto) were faculty in the University of Toronto Faculty of Medicine. Orientation for examiners was conducted the morning of the examination and included a discussion of the type of ratings being used. Examiners rated the candidates' performances in up to 11 domains and also provided an overall rating (**Table 2**). Possible ratings included 4 different levels of competence (unsatisfactory, clerkship, first-year resident or higher, and practice ready). The form also included the opportunity to indicate if the candidate demonstrated any unprofessional behaviour and to identify specific weaknesses.

The orientation included a discussion of the expected performance for each level of competence. Examiner notes were provided for each station, which identified the specific expectations for performance (eg, history questions that should be asked, physical maneuvers that should be conducted, or specific management or counseling that was appropriate for that station).

Table 2. Common rating form

ELEMENT	NOTES
Patient encounter	
<ul style="list-style-type: none"> • History taking and data collection • Physical examination • Verbal communication skills • Nonverbal communication skills • Response to patient's feelings, needs, and values • Organization • Management of standardized patient and case (ie, during 8-min patient encounter) 	<p>Facilitates patient's telling of story; uses questions or directions to obtain accurate, complete information</p> <p>Conducts appropriate examination(s), demonstrates appropriate techniques</p> <p>Demonstrates fluency in verbal communications (eg, grammar, vocabulary, tone, volume)</p> <p>Demonstrates responsiveness; demonstrates appropriate nonverbal communications (eg, eye contact, gesture, posture, use of silence)</p> <p>Shows respect, establishes trust; attends to patient's needs of comfort, modesty, confidentiality, information</p> <p>Approach is coherent and succinct</p> <p>Explains rationale for test, treatment, or approach; obtains patient consent; educates or counsels regarding management; performs appropriate management; considers risks and benefits</p>
Postencounter probe	
<ul style="list-style-type: none"> • Working diagnosis • Differential diagnoses • Investigations • Treatment or management plan 	<p>Obtains correct diagnosis</p> <p>Provides most likely alternative diagnoses</p> <p>Selects appropriate diagnostic studies or treatments; considers risks, benefits, costs.</p> <p>Selects appropriate treatments (monitoring, counseling, medications); considers risks, benefits, costs</p>
Overall rating	
<ul style="list-style-type: none"> • Rating based on the OVERALL performance; the candidate's current competence 	<p>CLERKSHIP: Could function as a clerk; could complete accurate history taking and physical examination; able to provide diagnosis and differential diagnoses; and able to select appropriate investigations for common and life-threatening conditions</p> <p>FIRST-YEAR RESIDENT or higher: Could function as a resident; demonstrates skill in judgment, synthesis, caring, effectiveness; is efficient and organized in history taking and physical examination; skilled and selective in treatment planning; understands limits and abilities; demonstrates appropriate presence, confidence, and assuredness; shows balanced demonstration of technical and humanistic skills</p> <p>PRACTICE READY: Could practice without supervision; could function as an independent practitioner in primary care setting</p>
Unprofessional behaviour	Indicate if there was a critical incident of unprofessional behaviour; please describe in detail
Key weaknesses	
<ul style="list-style-type: none"> • Inadequate knowledge • Provided misinformation to patient • Could not focus on patient's problem • Verbal language skills • Poor organization • Other 	Describe the candidate's key weaknesses; provide detailed comments on all unprofessional behaviour and all unacceptable ratings

Mean patient-encounter, postencounter oral question, and total scores were calculated for each station and for the total test. In each case, scores could range from a minimum of 0 to a maximum of 3.

Between-group (clerks, first-year residents, and IMGs) differences in scores (total, patient-encounter, and postencounter oral questions) were investigated using ANOVA and Scheffé post hoc comparisons ($\alpha = .05$).

RESULTS

A total of 909 candidates took the CE1, with more than 90% being IMGs (63 CMGs, 846 IMGs) (Table 3). Reliability of the total, patient-encounter, and postencounter oral question scores was 0.76, 0.93, and 0.88, respectively.

Table 3. Groups of candidates

GROUP	N (%)
IMGs	846 (93.1)
Clinical clerks	42 (4.6)
First-year residents	21 (2.3)
Total	909 (100.0)

IMG—international medical graduate.

Table 4. Comparison of patient-encounter, postencounter oral question, and total scores: IMGs scored between clerks and first-year residents.

GROUP	N	MEAN	SD
Patient interaction score			
• IMGs	846	1.797	0.375
• Clerks	42	1.771	0.293
• First-year residents	21	2.141	0.219
• Total	909	1.804	0.372
Postencounter question score			
• IMGs	846	1.557	0.354
• Clerks	42	1.330	0.326
• First-year residents	21	1.872	0.242
• Total	909	1.554	0.357
Total score			
• IMGs	846	1.709	0.352
• Clerks	42	1.604	0.284
• First-year residents	21	2.042	0.216
• Total	909	1.712	0.351

IMG—international medical graduate.

The total and component scores (ie, 8-minute patient encounter and 2-minute postencounter oral questions) and SDs are presented in **Table 4**. The third-year clerks scored the lowest, followed by the IMGs; the first-year residents had the highest scores for both component and total scores. Standard deviations were highest for the IMGs and lowest for the first-year residents.

The ANOVA demonstrated significant between-group differences, with the total and the 2 component scores all at the $P < .05$ level (**Table 5**).

Scheffé post hoc comparisons with each score indicated that the total and patient-encounter results do not differentiate between the clerks and the IMG candidates, while the postencounter question results differentiate among all 3 groups (**Table 6**).

DISCUSSION

There was interest in the ability of the global rating scale to accurately and reliably stratify the examinees into the

benchmarked levels of postgraduate learners (ie, clerks, first-year residents, second-year residents, or practice ready). Additionally, there was interest in the relative contribution of the 2 parts of the assessment (ie, the observed and rated 8-minute patient encounter and the 2-minute standardized oral questions posed by examiners specifically exploring clinical decision making and intervention planning) in the accurate stratification of examinees into residency learner levels. The accurate assignment of candidates is what makes appropriate placement into residency programs possible.

Previous work has demonstrated that IMGs were deficient in clinical skills,⁸ and that they pass certification examinations at a lower rate than CMGs,²³⁻²⁵ so a clearer estimation of the level of the candidate's ability, as compared with the Canadian levels of training, was desirable.

The notion of stratification into multiple levels was novel and had not been used in high-stakes examinations before; generally, checklists were used²⁶ and there was a single bar of competent or "good enough" (ie, pass or fail). Checklists might miss important aspects of clinical care, which cannot be described in discrete checklist items,²⁷ while global ratings have been shown to provide reliable evaluation of candidates.^{20,21}

Additionally, checklists do not necessarily coincide with evidence-based items identified in the literature,²⁸ and some candidates might complete many checklist items while still missing key tasks,²⁹ decreasing the validity of checklists for assessment of clinical skills. In particular, global ratings have been shown to be more appropriate for experienced clinicians compared with students or residents³⁰ because the experienced clinicians might not require completion of all the detailed tasks of the checklist to arrive at the correct diagnosis and treatment plan. This is an important point given that most of the IMGs who are taking the CE1 have had many years of practice in their home countries. This is an important validation study that demonstrated that it is possible to classify candidates into multiple levels of competence using a common examination instrument.

There is a considerable benefit to this type of rating form, in that the developmental effort put into creating and approving checklists is substantially decreased. This makes it easier to develop stations and increase the size of item banks. A common rating form also makes it possible to use the same stations with different levels of learners. Each level will demonstrate increasing levels of competence through including more of the key components within their patient interactions and responses to the examiner questions.

To use this rating scale, the examiners must have a strong mental model of the performance expectations for each of the 4 rating levels. For this reason, it is necessary for the examiners to be faculty members who are involved in teaching the learners at the various levels of education. The standard being set

Table 5. Analysis of variance by type of ratings

SCORE	SUM OF SQUARES	DF	MEAN SQUARE	F	P VALUE
Patient encounter					
• Between groups	2.462	2	1.231	9.062	<.001
• Within groups	123.089	906	0.136		
• Total	125.551	908			
Postencounter questions					
• Between groups	4.245	2	2.122	17.252	<.001
• Within groups	111.453	906	0.123		
• Total	115.697	908			
Total test score					
• Between groups	2.784	2	1.392	11.553	<.001
• Within groups	109.143	906	0.120		
• Total	111.927	908			

Table 6. Scheffé post hoc comparisons: Means for groups in homogeneous subsets are displayed; means appearing in the same column are not significantly different.

SCORE	LEVEL OF COMPETENCE	N	SUBSET FOR $\alpha = .05$		
			1	2	3
Total test score	Clerks	42	1.6043		
	IMGs	846	1.7089		
	First-year residents	21		2.0421	
Patient encounter	Clerks	42	1.7712		
	IMGs	846	1.7971		
	First-year residents	21		2.1405	
Postencounter questions	Clerks	42	1.3300		
	IMGs	846		1.5573	
	First-year residents	21			1.8724

IMG—international medical graduate.

in the examination is thus equivalent to the standard set throughout the medical education process, again increasing the validity of the examination results.

For high-stakes examinations, a reliability of 0.8 or higher is desirable. The reliability of the 2 component scores (ie, patient encounter and postencounter oral questions) were well above this level and the total score reliability was just below. The assessment of skills with the patient and the use of knowledge related to the patient (as assessed by the examiner questions) were both high. While these component scores are not used in any decision making, the high reliability lends confidence to the validity and interpretation of the OSCE results.

The results of the ANOVA and Scheffé analyses suggest that the clinical examination scores were

capable of differentiating among third-year clerks, first-year family medicine residents, and the IMG candidates. Before this examination, candidates were assessed as passing or failing at a single level. What is new here is that candidates were classified across multiple levels, and that the decisions made based on these classifications were appropriate ones. In addition, the results suggest that the group of IMGs are less competent than first-year Canadian family medicine residents and more competent than the clerks. These differences legitimize the continued use of the evaluation, and reinforce that the postencounter probe is more discriminating, particularly at higher levels of competence, as per our original hypothesis. There is no information currently available that would indicate a specific difference that is clinically or educationally large enough to confirm that the groups are different.

The added value of the postencounter questions in differentiating among the appropriate rating levels is important. These questions add to the validity of the stations and the examination through exploring knowledge and clinical problem solving that cannot be assessed during the patient interaction. In fact, the results demonstrate that the scores generated by the postencounter oral questions were the most effective in differentiating between the 2 training levels and among the 3 groups of test takers. Candidates at the higher levels are able to provide a more appropriate list of differential diagnoses and are more likely to identify the correct diagnosis. These candidates also provided more focused and appropriate investigations and management plans. This ability to synthesize the information obtained from the standardized patient encounter and develop an appropriate plan is observed by the examiner. As the postencounter questions inform the examiners' decisions and are effective in differentiating among candidates, they also act to improve the

reliability of the examination by assisting the examiner in selecting the more appropriate rating.

As the CMGs were trained at a single university and have a consistent and standard training program, the variation among these candidates would be expected to be low. Clerks are still early in their training and there is a broader range to their abilities at this point, resulting in a larger SD than for the first-year residents. Residents were approximately 4 months into their residency at the point of taking the CE1. There is less variability among this group than among the clerks, resulting in a lower SD. On the other hand, the IMGs have been trained in programs around the world with different curricula, different methods of teaching, different methods of assessment, and different criteria for evaluation. The IMGs are a very heterogeneous group, resulting in larger SDs.

The CE1 used a standardized global rating form for all stations. One benefit of this approach is that it simplifies item development. The standard domains are used for all stations, and item writers develop guidelines regarding expectations for that station. This eliminates the need to create a detailed checklist and the ensuing discussions regarding weighting of specific items. Checklists evaluate a limited range of competencies and tend to benefit candidates who are thorough, rather than those who approach patients from a broader perspective.^{21,27} Checklists might be appropriate for lower-level students, where the focus is on learning the details of procedures; however, for more advanced practitioners global ratings allow for an improved assessment of current clinical skills.

Limitations

As the study participants were all from one university and were assessed at that same site, it is possible that some examiners could have recognized the third-year students or first-year residents. This could have biased the examiners and affected the scoring for those individuals. Given the number of examiners (n=126) and the number of students, the likelihood of this happening was low and therefore the likelihood of it having an effect on the overall results is also very low.

The examination had high stakes for the IMGs, because doing well on the examination was important to their application for a first-year residency position. The study participants were volunteers and were paid for their participation. They would not likely have put the same effort into preparing for the examination as the IMGs did. This could have affected the scores for the clerks and first-year residents, likely resulting in lower scores than if they had had a stronger interest in performing as well as possible on the examination. If this had been the case, the difference between the first-year residents and IMGs would


have been accentuated and the difference between the clerks and IMGs would have been decreased.

Physician examiners were not told which candidates were IMGs, clerks, or residents. Despite this, older candidates and those for whom English was clearly a second (or third) language could have been identified by examiners as IMGs.

Finally, while the third-year clerks and first-year residents who participated were randomly invited to participate from their cohorts, it is possible that those who did volunteer were not representative of the cohort. Some might have been drawn to the study because they perceived an educational benefit from their participation (eg, having experience at an OSCE or being provided with feedback on their performance) and others might have volunteered simply to earn money. Weaker candidates might not have volunteered. These factors could have affected the score differences between the different groups. No analysis was conducted to compare these volunteers with the balance of their classes.

Conclusion

These results provide evidence of the validity of the clinical examination and the rating scale used to determine whether the level of competence of the IMG candidates was at the clerkship or first-year-resident level. It also suggests that incorporating postencounter questions that address clinical problem solving related to the patient just seen provides important information related to the candidate's competence. The study strongly supports the use of a common rating scale for assessing across multiple levels of competence.

Owing to the large numbers and the logistical challenges of conducting a large-scale examination on a single day, plans are under way for the examination to be administered multiple times a year. Future work will investigate the equivalence of the examinations administered over a period of months. Additionally, analysis of various factors that could be predictors of scores on the examination (such as age, location of undergraduate degree, or years since graduation) could provide insight into the assessment of future IMGs. 

Dr Glover Takahashi is Director of Education and Research for Postgraduate Medical Education at the University of Toronto in Ontario. **Dr Rothman** is Professor Emeritus in the Faculty of Medicine at the University of Toronto. At the time this work was conducted, **Dr Nayer** was Director of Assessment Operations at the Centre for the Evaluation of Health Professionals Educated Abroad (CEHPEA), and Assistant Professor in the Department of Physical Therapy at the University of Toronto. **Dr Urowitz** is Director of Health Professional Affairs at the CEHPEA and Professor of Medicine at the University of Toronto. At the time this work was conducted, **Ms Crescenzi** was Executive Director and CEO at the CEHPEA.

Contributors

All authors contributed to concept and design of the study; data gathering, analysis, and interpretation; and preparing the manuscript for submission.

Competing interests

None declared

Correspondence

Dr Marla Nayer, University of Toronto, Post Graduate Medical Education, 500 University St, Toronto, ON M5G 1V7; telephone 416 931-6861; e-mail marla.nayer@utoronto.ca

References

- Canadian Task Force on Licensure of International Medical Graduates. *Report of the Canadian Task Force on Licensure of International Medical Graduates*. Ottawa, ON: Canadian Task Force on Licensure of International Medical Graduates; 2004.
- Dauphinee WD. The circle game: understanding physician migration patterns within Canada. *Acad Med* 2006;81(12 Suppl):S49-54.
- Whelan GP, Gary NE, Kostis JB, Boulet JR, Hallock JA. The changing pool of international medical graduates seeking certification training in US graduate medical education programs. *JAMA* 2002;288(9):1079-84.
- Crutcher RA, Banner SR, Szafran O, Watanabe M. Characteristics of international medical graduates who applied to the CaRMS 2002 match. *CMAJ* 2003;168(9):1119-23.
- Pollett A. Laboratory medicine residency training 2007-2008. *Can J Pathol* 2009;1(2):28-9.
- Gastel B. Impact of international medical graduates on U.S. and global health care: summary of the ECFMG 50th Anniversary Invitational Conference. *Acad Med* 2006;81(12 Suppl):S3-S6.
- Hallock JA, Kostis JB. Celebrating 50 years of experience: an ECFMG perspective. *Acad Med* 2006;81(12 Suppl):S7-16.
- Peitzman SJ, McKinley D, Curtis M, Burdick W, Whelan G. International medical graduates' performance of techniques of physical examination, with a comparison of U.S. citizens and non-U.S. citizens. *Acad Med* 2000;75(10 Suppl):S115-7.
- University of British Columbia [website]. *Welcome to the IMG-BC program*. Vancouver, BC: University of British Columbia; 2011. Available from: www.imgbc.med.ubc.ca. Accessed 2011 Mar 8.
- Alberta International Medical Graduate Program [website]. *Welcome to the Alberta International Medical Graduate Program*. Calgary, AB: Government of Alberta; 2011. Available from: www.aimg.ca. Accessed 2011 Mar 8.
- University of Manitoba [website]. *International Medical Graduate Program*. Winnipeg, MB: University of Manitoba; 2011. Available from: <http://umanitoba.ca/faculties/medicine/education/imgp/index.html>. Accessed 2011 Mar 8.
- College of Physicians and Surgeons of Nova Scotia. *Clinician Assessment for Practice Program*. Halifax, NS: College of Physicians and Surgeons of Nova Scotia; 2011. Available from: www.capprogram.ca/index.html. Accessed 2011 Mar 8.
- Santé et services sociaux Québec [website]. *Recrutement santé Québec. For physicians who earned their degrees outside Canada and the United States*. Québec, QC: Santé et services sociaux Québec; 2011. Available from: www.msss.gouv.qc.ca/sujets/organisation/medecine/rsq/index.php?home. Accessed 2011 Mar 8.
- Newble D, Dawson B, Dauphinee D, Macdonald M, Swanson D, Mulholland H, et al. Guidelines for assessing clinical competence. *Teach Learn Med* 1994;6(3):213-20.
- Harden RM, Gleeson FA. Assessment of clinical competence using an objective structured clinical examination (OSCE). *Med Educ* 1979;13(1):41-54.
- Dupras DM, Li JT. Use of an objective structured clinical examination to determine clinical competence. *Acad Med* 1995;70(11):1029-34.
- Petrusa ER, Blackwell TA, Ainsworth MA. Reliability and validity of objective structured clinical examination for assessing the clinical performance of residents. *Arch Intern Med* 1990;150(3):573-7.
- Cohen R, Rothman AI, Ross J, Poldre P. Validating an objective structured clinical examination (OSCE) as a method for selecting foreign medical graduates for a pre-internship program. *Acad Med* 1991;66(9 Suppl):S67-9.
- Cohen R, Rothman AI, Ross J, Poldre P. Validating and objective structured clinical examination (OSCE) as a method for selecting foreign medical graduates for a pre-internship program. *Acad Med* 1991;66(9 Suppl):S67-9.
- Cohen D, Colliver J, Robbs RS, Swartz MH. A large-scale study of the reliabilities of checklist scores and ratings of interpersonal and communication skills evaluated on a standardized-patient examination. *Adv Health Sci Educ Theory Pract* 1997;1:209-13.
- Cunnington J, Neville A, Norman G. The risks of thoroughness: reliability and validity of global ratings and checklists in an OSCE. *Adv Health Sci Educ Theory Pract* 1997;1(3):227-33.
- Rothman A, Blackmore D, Dauphinee D, Reznick R. The use of global ratings in OSCE station scores. *Adv Health Sci Educ Theory Pract* 1997;1:213-9.
- Andrew RF. How do IMGs compare with Canadian medical school graduates in a family practice residency program? *Can Fam Physician* 2010;56:e318-22. Available from: www.cfp.ca/content/56/9/e318.full.pdf+html. Accessed 2012 Jun 7.
- MacLellan AM, Brailovsky CA, Rainsberry P, Bowmer I, Desroches M. Examination outcomes for international medical graduates pursuing or completing family medicine residency training in Quebec. *Can Fam Physician* 2010;56:912-8.
- Bowmer MI, Roy M, Wood TJ. How do Canadians studying abroad (CSAs) compare to other examinees on a high-stakes licensure examination? Paper presented at: 14th Ottawa Conference. Miami, FL, 2010 May 15-19.
- Whelan GP, Boulet JR, McKinley D, Norcini JJ, van Zanten M, Hambleton RK, et al. Scoring standardized patient examinations: lessons learned from the development and administration of the ECFMG Clinical Skills Assessment (CSA). *Med Teach* 2005;27(3):200-6.
- McKinley RK, Strand J, Ward L, Gray T, Alun-Jones T, Miller H. Checklists for assessment and certification of clinical procedural skills omit essential competencies: a systematic review. *Med Educ* 2008;42(4):338-49.
- Hettinga AM, Denessen E, Postma CT. Checking the checklist: a content analysis of expert- and evidence-based case-specific checklist items. *Med Educ* 2010;44(9):874-83.
- Payne NJ, Bradley EB, Heald EB, Maughan KL, Michaelsen VE, Wang XQ, et al. Sharpening the eye of the OSCE with critical action analysis. *Acad Med* 2008;83(10):900-5.
- Hodges B, Regehr G, McNaughton N, Tiberius R, Hanson M. OSCE checklists do not capture increasing levels of expertise. *Acad Med* 1999;74(10):1129-34.

————— * * * —————