



Assessing postgraduate trainees in Canada: Are we achieving diversity in methods?

Sophia Chou, Jocelyn Lockyer, Gary Cole & Kevin McLaughlin

To cite this article: Sophia Chou, Jocelyn Lockyer, Gary Cole & Kevin McLaughlin (2009) Assessing postgraduate trainees in Canada: Are we achieving diversity in methods?, Medical Teacher, 31:2, e58-e63, DOI: [10.1080/01421590802512938](https://doi.org/10.1080/01421590802512938)

To link to this article: <https://doi.org/10.1080/01421590802512938>



Published online: 03 Jul 2009.



Submit your article to this journal [↗](#)



Article views: 869



View related articles [↗](#)



Citing articles: 2 View citing articles [↗](#)

WEB PAPER

Assessing postgraduate trainees in Canada: Are we achieving diversity in methods?

SOPHIA CHOU, JOCELYN LOCKYER, GARY COLE & KEVIN McLAUGHLIN

University of Calgary, Canada

Abstract

Background: Resident evaluation is a complex and challenging task, and little is known about what assessment methods, predominate within or across specialties.

Aims: To determine the methods program directors in Canada use to assess residents and their perceptions of how evaluation could be improved.

Methods: We conducted a web-based survey of program directors from The Royal College of Physicians and Surgeons of Canada (RCPSC)-accredited training programs, to examine the use of the In-Training Evaluation Report (ITER), the use of non-ITER tools and program directors' perceived needs for improvement in evaluation methods.

Results: One hundred forty-nine of the eligible 280 program directors participated in the survey. ITERs were used by all but one program. Of the non-ITER tools, multiple choice questions (71.8%) and oral examinations (85.9%) were most utilized, whereas essays (11.4%) and simulations (28.2%) were least used across all specialties. Surgical specialties had significantly higher multiple choice questions and logbook utilization, whereas medical specialties were significantly more likely to include Objective Structured Clinical Examinations (OSCEs). Program directors expressed a strong need for national collaboration between programs within a specialty to improve the resident evaluation processes.

Conclusions: Program directors use a variety of methods to assess trainees. They continue to rely heavily on the ITER, but are using other tools.

Introduction

Postgraduate medical education and assessment have undergone many changes. While it is recognized that the ultimate goal of a postgraduate training program is to produce competent physicians ready for independent medical practice (Epstein & Hundert 2002), the perspective of what constitutes a 'competent' physician and what evidence is required to attest that someone is 'competent' have changed. Professional organizations, such as the Royal College of Physicians and Surgeons of Canada (RCPSC) (Frank 2005), the Accreditation Council for Graduate Medical Education (ACGME) of the United States (ACGME 2007), the Modernizing Medical Careers Programme of the United Kingdom (UK Foundation Programme Office 2007) and the Confederation of Postgraduate Medical Education Councils of Australia (Commonwealth Department of Health and Ageing 2003), have each identified the competencies which physicians will attain through their postgraduate residency training. These competencies, while varying from one jurisdiction to another, include communication skills, collegiality and professionalism in addition to the medical expert role.

Traditionally in Canada, the assessment of the trainee has relied upon the In-training Evaluation Report (ITER). The ITER provides systematic, longitudinal observation and feedback on resident performance in real clinical settings. It can be completed part way through or at the end of a rotation to

Practice points

- ITERs are the popular assessment tools used in Canadian postgraduate training programs, but they may not be valid due to limited rater training and lack of non-physician assessors.
- Of the non-ITER tools, multiple choice questions and oral examinations were most utilized, whereas essays and simulations were least used across all specialties.
- Only half of the programs use OSCEs in resident evaluation.
- Surgical specialties had significantly higher multiple choice questions and logbook utilization, whereas medical specialties were significantly more likely to incorporate OSCEs in resident evaluation.
- According to program directors, the best way to improve resident evaluation would be by national collaboration under the guidance of the RCPSC.

document how well the trainee is doing with physician preceptors, other residents, allied health professionals, and patients. ITER formats include checklists, global rating scales and behavioural assessment forms (Gray 1996b). At the end of training, a Final In-Training Evaluation Report (FITER) is completed enabling the candidate to proceed to the RCPSC

Correspondence: S Chou, Faculty of Medicine, University of Calgary, 1403-29th Street N.W., Calgary, AB T2N 2T9, Canada. Tel: 403 944 2667; fax: 403 270 0055; email: Sophia.chou@calgaryhealthregion.ca

certifying examination. However, along with the development of CanMEDS roles has come the recognition that a combination of assessment tools must be used to evaluate trainees on a broader range of competencies. Accordingly, the RCPSC has published *The CanMEDS Assessment Tools Handbook: An Introductory Guide to Assessment Methods for the CanMEDS Competencies* (Bandiera et al. 2006). The handbook is designed to help residency education program directors and committees select the optimal methods for assessment. It provides general information and suggests how methods such as the ITERs, written tests, structured oral examinations, direct observation, Objective Structured Clinical Examinations (OSCEs), standardized patients, multi-source feedback, portfolios and logbooks, simulation-based assessment and encounter cards can be informative. It also suggests which tools are most effective for each of the competencies.

Despite the availability of a handbook and training offered by the RCPSC and within medical schools, it is not known what assessment methods predominate within or across specialties. The general standards of accreditation of the RCPSC (RCPSC 2006) states that the assessment system must be based on the goals and objectives of the program and must clearly identify the methods by which the residents are to be evaluated. It further specifies that residency programs must assess medical knowledge with appropriate written and performance-based assessment, clinical skills with direct observation and other CanMEDS competencies with observation and interviews with peers, allied health professionals and patients. However, the actual ways that residents are assessed have been left to the discretion of program directors. This approach is very different than that adopted in the United Kingdom, where the Foundation Programme has mandated the use and administration frequency of four specific tools, namely, mini clinical evaluation exercise, direct observation of procedural skills, case-based discussion and multi-source feedback (UK Foundation Programme Office 2007). In Canada, it is particularly important to identify the approaches being used in evaluation as this will provide a baseline on which the adoption of different tools can be monitored over time. Sharing 'best practices' across and within specialties can inform and improve assessment.

The purpose of this study was to examine evaluation practices in Canadian postgraduate training programs. Our first objective was to determine the assessment methods used currently, with a particular focus on the ITER due to its traditional role in evaluation. Our second objective was to identify the ways that program directors believe residency evaluation could be improved. We also examined whether the medical, surgical and investigative specialties differed in their approaches to resident evaluation.

Methods

Study sample

We conducted a web-based survey of program directors for the RCPSC accredited specialty postgraduate training programs from the 13 English-speaking medical schools for the academic year of 2006–2007. Due to the feasibility and practicality

considerations, subspecialties and French-speaking programs were excluded. There were a total of 284 eligible training programs with 280 program directors surveyed; 4 program directors had responsibility for 2 programs each.

Data collection. The survey asked program directors to describe how often they administered ITERs. Firstly, they were asked whether the ITERs were program or rotation specific and who completed the ITERs (i.e. physician preceptors only versus nurses, other health care professionals, resident peers, medical students, patients and self). They were also asked about ITER training for physician raters. Secondly, program directors were queried about their use of non-ITER assessment tools, including Multiple Choice Questions (MCQs), Short Answer Questions (SAQs), essays, oral examinations, OSCEs, simulations and logbooks. They were also queried about the mean number of non-ITER assessment tools used. Lastly, program directors were asked to use a 5-point scale to indicate their perceived need for the development of new assessment tools, systematic integration of currently available tools, national collaboration between programs within a specialty, local collaboration between residency programs within a university and for specific guidelines for tools by the RCPSC.

The survey was administered on-line using Survey Console (*SurveyConsole* 2007). The on-line survey, along with a cover letter that explained the nature of the study and its potential impact, were first distributed in early March 2007 by e-mail to the specialty program directors included in the study. When the response rate fell at approximately 2.5 weeks after the first e-mail, a reminder e-mail was sent to all program directors. A third e-mail invitation was sent 2 days after the study deadline to those who had not responded. Data collection was completed by April 2007.

Statistical analyses. Programs were grouped into three categories: medical, surgical and investigative specialties (Table 1). The mean number of trainees in each program along with the range (minimum and maximum) were determined. Descriptive analyses were calculated for the items on the survey. Differences between specialty groups were analysed using one-way analysis of variance (ANOVA) and $p < 0.001$ was considered statistically significant.

This study received ethics approval from the Conjoint Health Research Ethics Board at the University of Calgary.

Results

Of the 280 eligible program directors, 11 could not be contacted due to inaccurate e-mail addresses. A total of 167 (62.1%) questionnaires were returned and 149 (55.4%) were included in the final analysis. Sixteen duplicate questionnaires were excluded and additional two questionnaires were excluded due to the lack of information on specialty and university. Because none of the program directors from the general pathology and medical biochemistry returned completed questionnaires, these two specialties were not included in the analyses. Of the 131 non-respondent program directors, 52 were from the 126 eligible programs in medical specialties, 41 were from the 99 programs in surgical specialties and

Table 1. Response rate and mean number of non-ITER assessment tools used by programs.

Specialty (No. eligible programs)	No. respondents (% returned)	No. assessment tools used, mean (SD)
Medical specialties	74 (59.2)	3.64 (1.22)
Anesthesiology (13)	10 (76.9)	3.80 (0.92)
Community medicine (8)	5 (62.5)	2.80 (0.84)
Dermatology (6)	3 (50.0)	4.00 (1.00)
Emergency medicine (11)	6 (54.5)	3.83 (1.17)
Internal medicine (13)	9 (69.2)	4.44 (1.01)
Medical genetics (6)	6 (100)	3.50 (1.05)
Neurology, adult (13)	6 (46.2)	3.50 (0.55)
Neurology, paediatric (7)	7 (100)	3.00 (0.82)
Paediatrics (13)	6 (46.2)	4.50 (1.05)
Physical Medicine & Rehabilitation (11)	2 (18.2)	4.50 (3.54)
Psychiatry (13)	7 (53.8)	3.57 (1.51)
Radiation Oncology (11)	7 (63.6)	2.57 (1.40)
Surgical specialties	58 (58.0)	3.72 (1.40)
Cardiac surgery (10)	3 (30.0)	3.33 (2.31)
General surgery (13)	11 (84.6)	3.00 (0.78)
Neurosurgery (11)	8 (72.7)	3.50 (1.60)
Obstetrics & Gynaecology (13)	8 (61.5)	4.75 (1.58)
Ophthalmology (11)	8 (72.7)	3.88 (0.84)
Orthopaedics (13)	7 (53.8)	2.57 (0.98)
Otolaryngology (10)	3 (30.0)	4.67 (1.53)
Plastic Surgery (9)	3 (33.3)	3.33 (0.58)
Urology (10)	7 (70.0)	4.86 (1.07)
Investigative specialties	17 (28.8)	3.35 (1.37)
Anatomical pathology (12)	5 (41.7)	3.20 (1.92)
Diagnostic radiology (13)	6 (4.0)	3.83 (1.33)
General pathology (10)	0 (0)	N/A
Haematological pathology (6)	1 (16.7)	2.00 (N/A)
Medical biochemistry (2)	0 (0)	N/A
Medical microbiology (9)	1 (11.1)	2.00 (N/A)
Nuclear medicine (7)	4 (57.1)	3.50 (0.58)
Total	149 (55.4)	3.64 (1.31)

38 were from the 55 programs in investigative specialties. Investigative specialties were significantly less likely to participate in the study; only 17 of the 55 (31%) eligible program directors responded, compared to the 58.7% and 58.6% response rates in the medical and surgical specialties, respectively. The mean numbers of residents were similar between respondent and non-respondent programs (21.8 ± 23.5 versus 17.9 ± 16.7 , respectively, $p = 0.406$). Information about the response rates by programs are provided in Table 1.

In-training evaluation reports

In-training evaluation reports were used by all but one respondent program. Of the respondent programs, 47 programs (32.4%) administered ITERs at least monthly, 39 programs (26.9%) administered ITERs bimonthly and 58 programs (40.0%) administered ITERs quarterly. Significant differences in ITER administration frequency were observed between the three specialty categories. The medical specialties administered ITERs most frequently (almost monthly), while surgical specialties were more likely to administer them quarterly. In keeping with the RCPSC evaluation requirement, 128 (85.9%) of the respondent programs had ITERs developed specifically for their programs. Furthermore, 85 programs

(58.2%) had specific ITERs tailored for different clinical rotations. These results are presented in Table 2. A total of 37 programs (24.8%) provided ITER training for physician raters. Compared to physician raters, very few programs had non-physician assessors involved in ITER completion. These results are presented in Table 3.

Non-ITER assessment tools used

When queried about assessment tools other than ITERs, all respondent programs reported using one or more tools for resident evaluation. Multiple choice questions and oral examinations were most frequently used, and were also more likely to have pre-determined pass/fail. Essays and simulations were least utilized. About half of the respondent programs reported the use of OSCEs. These results are summarized in Table 4. Among the non-ITER assessment tools, oral examinations were administered most frequently, with a mean of 2.20 (SD = 1.48) times per year as shown in Table 5. When considering the annual administration frequency of all non-ITER assessment tools as a sum, residents were evaluated from twice per year in two respondent programs (1.5%) to 29 times per year in one program (0.7%), with a mean of 7.95 (SD = 4.18) times per year. Of the respondent programs, the mean number of non-ITER assessment tools utilized was 3.64 (SD = 1.31) (Table 1).

Ways of improving resident evaluation

When asked about potential areas for improvement in evaluation processes, respondent program directors indicated that national collaboration between programs within their own specialty, and leadership from the RCPSC regarding assessment tools were their greatest needs.

Discussion

This is one of the first studies of Canadian residency program directors designed to examine current national practices of resident evaluation. The response rate includes program directors from the 28 RCPSC accredited specialties across 13 English-speaking medical schools with the RCPSC accredited residency training programs at the time of the study. Investigative specialties were less likely to participate in the study.

In-training evaluation reports

ITERs completed by physician raters continue to be the most commonly used assessment method with most programs administering them monthly or every 2 months. The combined advantages of feasibility and face validity likely lead to this high usage (Daelmans et al. 2005). Few programs provided ITER training to physician raters, a finding consistent with the results of previous work by Epstein and Hundert (2002) and Ruedy (2006). Very few programs relied on non-physician raters to assess trainees. This last finding was again identified by Epstein and Hundert (2002) and Ruedy (2006), and is a concern as proponents of ITERs such as Gray (1996a) and

Table 2. ITER characteristics and administration frequency with physician raters by specialty category.

	Medical specialties (N = 74)	Surgical specialties (N = 58)	Investigative specialties (N = 17)	p
ITER characteristics (%)				
Program specific	86.5	82.8	94.1	NS
Rotation specific	63.9	56.1	41.2	NS
Administration frequency (%)				
≤Monthly	46.6	9.1	47.1	<0.0001
Bimonthly	23.3	30.9	29.4	
Quarterly	30.1	60.0	17.6	

Note: No programs responded to the choice of annual administration.

Table 3. ITER administration frequency by source.

Source of ITERs	All programs (N = 149) (%)	Medical specialties (N = 74) (%)	Surgical specialties (N = 58) (%)	Investigative specialties (N = 17) (%)	p
Resident peers					
Never	45 (30.2)	24 (32.4)	12 (20.7)	9 (52.9)	<0.0001
Rarely	47 (31.5)	24 (32.4)	16 (27.6)	7 (41.2)	
Sometimes	36 (24.2)	21 (28.4)	14 (24.1)	1 (5.9)	
Often	15 (10.1)	4 (5.4)	11 (19.0)	0 (0.0)	
Always	6 (4.0)	1 (1.4)	5 (8.6)	0 (0.0)	
Medical students					
Never	59 (39.9)	25 (33.8)	21 (36.8)	13 (76.5)	0.004
Rarely	48 (32.4)	25 (33.8)	19 (33.3)	4 (23.5)	
Sometimes	33 (22.3)	21 (28.4)	12 (21.1)	0 (0.0)	
Often	6 (4.1)	3 (4.1)	3 (5.3)	0 (0.0)	
Always	2 (1.4)	0 (0.0)	2 (3.5)	0 (0.0)	
Nurses					
Never	36 (24.3)	16 (21.6)	9 (15.5)	11 (68.8)	<0.0001
Rarely	31 (20.9)	17 (23.0)	12 (20.7)	2 (12.5)	
Sometimes	53 (35.8)	29 (39.2)	22 (37.9)	2 (12.5)	
Often	24 (16.2)	12 (16.2)	11 (19.0)	1 (6.3)	
Always	4 (2.7)	0 (0.0)	4 (6.9)	0 (0.0)	
Other allied health					
Never	26 (17.4)	13 (17.6)	10 (17.2)	3 (17.6)	NS
Rarely	47 (31.5)	24 (32.4)	20 (34.5)	3 (17.6)	
Sometimes	54 (36.2)	27 (36.5)	20 (34.5)	7 (41.2)	
Often	20 (13.4)	10 (13.5)	7 (12.1)	3 (17.6)	
Always	2 (1.3)	0 (0.0)	1 (1.7)	1 (5.9)	
Patients					
Never	70 (47.9)	35 (47.3)	24 (42.1)	11 (73.3)	0.05
Rarely	45 (30.8)	24 (32.4)	17 (29.8)	4 (26.7)	
Sometimes	28 (19.2)	12 (16.2)	16 (28.1)	0 (0.0)	
Often	3 (2.1)	3 (4.1)	0 (0.0)	0 (0.0)	
Always	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	
Resident self					
Never	53 (36.3)	30 (40.5)	19 (33.3)	4 (26.7)	NS
Rarely	26 (17.8)	11 (14.9)	14 (24.6)	1 (6.7)	
Sometimes	28 (19.2)	14 (18.9)	11 (19.3)	3 (20.0)	
Often	20 (13.7)	11 (14.9)	5 (8.8)	4 (26.7)	
Always	19 (13.0)	8 (10.8)	8 (14.0)	3 (20.0)	

Table 4. Non-ITER assessment tool utilization by specialty and with pre-determined pass/fail systems.

Assessment tools (N = 149)	Total No. programs using (%)	Medical specialty (N = 74)	Surgical specialty (N = 58)	Investigative specialty (N = 17)	No. programs with pre-determined pass/fail (%)
MCQ	107 (71.8)	47 (63.5)	51 (87.9)	9 (52.9)	74 (61.2)
SAQ	96 (64.4)	49 (66.2)	33 (56.9)	14 (82.4)	69 (55.6)
Essays	17 (11.4)	9 (12.2)	6 (10.3)	2 (11.8)	14 (17.1)
Oral examinations	128 (85.9)	60 (81.1)	52 (89.7)	16 (94.1)	95 (70.4)
OSCE	74 (49.7)	46 (62.2)	20 (34.5)	8 (47.1)	50 (46.3)
Simulations	42 (28.2)	25 (33.8)	13 (23.1)	3 (17.6)	9 (10.1)
Logbooks	78 (52.3)	33 (44.6)	40 (69.0)	5 (29.4)	12 (12.1)

Table 5. Non-ITER assessment tools annual administration frequency.

Assessment tools	Mean (SD)
MCQ	1.14 (1.12)
SAQ	1.41 (1.39)
Essays	0.22 (0.73)
Oral examinations	2.20 (1.48)
OSCE	0.90 (1.04)
Simulations	0.67 (1.31)
Logbooks	1.53 (1.86)

Turnbull et al. (2000) have advocated for multi-disciplinary input to improve the validity and reliability of ITERs. The general standards of accreditation of the RCPSC (RCPSC 2006) strongly recommends evaluation from non-physician assessors. There were, however, differences in the use of ITERs by specialty grouping. Medical and investigative specialties administered ITERs more frequently than surgical specialties. Non-physician raters were more likely to be used by medical and surgical specialties than investigative specialties. This finding likely relates to the nature of training for investigative specialties and the opportunity these professionals get to work with nurses, patients and medical students. It does not account for a lower involvement by resident peers in ITER evaluation.

Non-ITER assessment tools used

Consistent with the recognition that many methods are required to ensure a quality assessment across competencies, it was reassuring to find that all respondent programs used one or more assessment tools in resident evaluation, in addition to the ITER. Oral examinations, MCQs, and SAQs are the three most popular non-ITER assessment methods used by most programs. About half of the programs used OSCEs and logbooks. Very few programs used simulations and even fewer relied on essays. There were differences by specialty group. Medical specialties were most likely to use OSCE assessments, possibly related to the nature of the work of medical specialists and the significant investment in human resources required for OSCEs (Barman 2005), which become cost-effective when there are sufficient number of trainees. In our study, medical specialties reported higher mean number of residents than the surgical and investigative specialties. Surgical specialties had higher MCQ and logbook utilization than medical and investigative specialties. All surgical trainees (except those in ophthalmology and obstetrics and gynaecology) have to pass the Principles of Surgery Examination which consists of a series of MCQs. Most surgical programs recommend their residents maintain an updated procedural log throughout their residency in order to obtain future hospital privileges in Canada and the United States.

Ways of improving resident evaluation

Medical education researchers suggest that using a combination of assessment tools results in more comprehensive evaluation of the learner and hence a higher quality evaluation

of trainees (Miller 1990; Holmboe & Hawkins 1998; Turnbull et al. 1998; Farrell 2005; van der Vleuten & Schuwirth 2005; Cole 2006). To this end, the RCPSC requires residency training programs to provide ongoing evaluation of their residents using a variety of assessment methods. This study provides reassuring evidence that all programs were in compliance with this recommendation. Nonetheless, respondent program directors reported a strong need for improvement in resident in-training evaluation processes with support for national collaboration between programs within a specialty. One can propose that specialty societies should take a leadership role in establishing lateral collaboration between programs in resident evaluation. Such organizations exist in the United States, e.g. the Association of Program Directors in Radiology Education Committee, which assists radiology program directors to meet ACGME evaluation requirements (Collins et al. 2004). A similar approach in Canada to the formation of committees of program directors would facilitate the sharing of best practices across programs.

Study limitations

The study has a number of limitations. It was carried out in one country involving only English-speaking programs, and therefore findings may not be applicable to the French-speaking programs in Canada. Although there were data from all the 13 English-speaking institutions, program directors for the investigative specialties were significantly less likely to participate in the study compared to the medical and surgical specialties, reducing the applicability of the study results to that group. Furthermore, only program directors were surveyed and only their perspectives were presented in the study. Since program directors are only one of many stakeholders involved in resident evaluation, and as reported by Murphy et al. (2008) awareness of the perceptions of all stakeholders is important in resident evaluation. However, a study surveying all stakeholders was beyond the scope of this study.

Nonetheless, these findings suggest a number of new opportunities to improve trainee assessment. ITER assessment could be enhanced by training raters to reduce bias and errors (Holmboe & Hawkins 1998). Involving peers, patients and other health care professionals in ITER assessment would likely increase the validity of the ratings. Canada is a small country with 17 medical schools. Collaboration with the RCPSC and within specialties could help with the testing of assessment instruments and approaches to training. The importance of assessment is recognized and is too complicated to be left to the limited resources that many residency programs and specialties have.

Conclusion

In this study we found that while most of the Canadian residency programs use ITERs to evaluate resident performance, few programs provide rater training for ITERs. This increases the idiosyncrasy of observer ratings, which may reduce the validity of ITERs. Similarly, few programs use non-physician raters, which may further compromise the rating quality, particularly for competencies such as collaboration.

In addition to ITERs, multiple choice questions and oral examinations are the most popular tools in assessing resident competencies whereas essays and simulations are least used across all specialties. However, important differences were observed between specialties. Surgical specialties had higher MCQ and logbook use whereas medical specialties were more likely to incorporate the OSCE in resident evaluation. Finally, helping program directors to develop and share tools across institutions may be the optimal way of improving the diversity and quality of evaluation methods currently being used.

In this study, we found that the use of ITERs continued to predominate resident assessment. Nonetheless, other methods are being adopted and used across and within disciplines to assess trainees. This work will remain important as the task of ensuring competence across a broad set of physician roles continues. This work will require ongoing collaboration across programs, universities and within disciplines.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the article.

Notes on contributors

SOPHIA CHOU, MD, MSc, is the principal investigator responsible for study design, data analysis and drafting of this article.

JOCELYN LOCKYER, PhD, has made significant contribution to study design, data analysis and article write-up.

GARY COLE, PhD, has made significant contribution to the data acquisition and article revision.

KEVIN MCLAUGHLIN, MD, PhD, has made substantial contribution to the study design and article revision.

Reference

- Accreditation Council for Graduate Medical Education. 2007. *Accreditation Council for Graduate Medical Education, ACGME Outcome Project*. Available from: <http://www.acgme.org/Outcome/> (Accessed 12 February 2008).
- Bandiera G, Sherbino J, Frank J. 2006. The CanMEDS assessment tools handbook: An introductory guide to assessment methods for the CanMEDS competencies. 1st ed. Ottawa: The Royal College of Physicians and Surgeons of Canada.
- Barman A. 2005. Critiques on the objective structured clinical examination. *Ann Acad Med Singapore* 34:478–482.
- Cole G. 2006. Effective in-training evaluation of residents. Paper presented at the The Royal College of Physicians and Surgeons of Canada Annual Conference, Ottawa, Canada.
- Collins J, Herring W, Kwakwa F, Tarver RD, Blinder RA, Gray-Leithe L, Wood B. 2004. Current practices in evaluating radiology residents, faculty, and programs: Results of a survey of radiology residency program directors. *Acad Radiol* 11(7):787–794.
- Commonwealth Department of Health and Ageing. 2003. *National training and assessment guidelines for junior medical doctors PGY1 and 2*. Available from: [http://www.health.gov.au/internet/wcms/Publishing.nsf/Content/health-workforce-new-jmonatgui.htm/\\$FILE/natassgui.pdf](http://www.health.gov.au/internet/wcms/Publishing.nsf/Content/health-workforce-new-jmonatgui.htm/$FILE/natassgui.pdf) (Accessed 12 February 2008).
- Daelmans HE, van der Hem-Stokroos HH, Hoogenboom RJ, Scherpbier AJ, Stehouwer CD, van der Vleuten CP. 2005. Global clinical performance rating, reliability and validity in an undergraduate clerkship. *Neth J Med* 63:279–284.
- Epstein RM, Hundert EM. 2002. Defining and assessing professional competence. *JAMA* 287:226–235.
- Farrell SE. 2005. Evaluation of student performance: Clinical and professional performance. *Acad Emerg Med* 12:302e306–310.
- Frank J. 2005. *The CanMEDS 2005 Physician Competence Framework, Ottawa: Royal College of Physicians and Surgeons of Canada*. Available from: http://rcpsc.medical.org/canmeds/CanMEDS2005/CanMEDS2005_e.pdf (Accessed 12 February 2008).
- Gray J. 1996a. Global rating scales in residency education. *Acad Med* 71:S55–63.
- Gray J. 1996b. Primer on resident evaluation. *Annals RCPSC* 29:91–94.
- Holmboe ES, Hawkins RE. 1998. Methods for evaluating the clinical competence of residents in internal medicine: A review. *Ann Intern Med* 129:42–48.
- Miller GE. 1990. The assessment of clinical skills/competence/performance. *Acad Med* 65:S63–67.
- Murphy DJ, Bruce D, Eva KW. 2008. Workplace-based assessment for general practitioners: Using stakeholder perception to aid blueprinting of an assessment battery. *Med Educ* 42:96–103.
- Ruedy J. 2006. Report on the processes used to evaluate residents in the postgraduate programs of a Canadian medical school with particular reference to CanMEDS roles. Ottawa: Royal College of Physicians and Surgeons of Canada.
- SurveyConsole 2007. Available from: <http://www.surveyconsole.com/> (Accessed 12 February 2008).
- The Royal College of Physicians and Surgeons of Canada (RCPSC) (2006). General standards of accreditation (June 2006). Available from: <http://rcpsc.medical.org> (Accessed 12 February 2008).
- Turnbull J, Gray J, MacFadyen J. 1998. Improving in-training evaluation programs. *J Gen Intern Med* 13:317–323.
- Turnbull J, MacFadyen J, Van Barneveld C, Norman G. 2000. Clinical work sampling A new approach to the problem of in-training evaluation. *J Gen Intern Med* 15:556–561.
- UK Foundation Programme Office. 2007. The foundation programme curriculum. Available from: <http://www.foundationprogramme.nhs.uk/pages/home/key-documents> (Accessed 12 February 2008).
- van der Vleuten CP, Schuwirth LW. 2005. Assessing professional competence: from methods to programmes. *Med Educ* 39:309–317.