

The General Internal Medicine Certification Examination: Does it Measure What it Should?

Debra Pugh MD, MHPE, FRCPC, Bert Govig MD, MPH, FRCPC, Neil Gibson MD, MSc, FRCPC, Gary Cole PhD



Pugh



Govig



Gibson

About the Authors

Debra Pugh practises general internal medicine at the University of Ottawa. Bert Govig practises community-based general internal medicine in Amos, Quebec. Neil Gibson practises internal medicine at the University of Alberta. All three authors serve on the GIM examination Board at the Royal College of Physicians and Surgeons of Canada. Gary Cole is Manager of Assessment at the RCPSC. Correspondence may be directed to dpugh@toh.on.ca

Summary

A certification exam for the new sub-specialty of General Internal Medicine (GIM) was recently developed by an examination board at the Royal College of Physicians and Surgeons (RCPSC). The board sought to create an exam that would reflect the challenges faced by practising general internists, while minimizing repetition of material already assessed by the Internal Medicine certification exam. In this paper, the authors present evidence for the validity of the content of the exam.

Résumé

Un examen de certification concernant la nouvelle subsécialité de Médecine interne générale a récemment été élaboré par un comité d'examen du Collège royal des médecins et chirurgiens du Canada. Dans cette démarche, le comité a cherché à créer un examen qui reflète les défis auxquels sont confrontés les internistes généraux praticiens, tout en réduisant au maximum la répétition de matériel déjà évalué par l'examen de certification de Médecine interne. Dans le présent article, les auteurs expliquent les fondements de la validité du contenu de l'examen.

Background

In 2010, the Royal College of Physicians and Surgeons of Canada approved the creation of the new subspecialty of General Internal Medicine (GIM). Subsequently, an examination board comprised of 10 physicians who practise as General Internists was tasked with creating a certification examination. Recognizing the significant knowledge overlap between the specialty of Internal Medicine (IM) and the sub-specialty of GIM, the board sought to create an exam that would reflect the challenges faced by practising General Internists, while minimizing repetition of material already assessed by the IM examination.

Because this is a new certification examination being created from “scratch,” there is an imperative to demonstrate evidence for the validity of the scores obtained from this examination.¹ One important source of validity evidence for

an examination is “content evidence”. Content evidence refers to ensuring that the construct being assessed is accurately and completely represented on a test.² Some arguments for content validity relate to: item writer qualifications (e.g., recruiting content experts); quality of test questions (e.g., having a review process); examination blueprint (e.g., demonstrating that it reflects the objectives of training); and the match of item content to blueprint (e.g., getting feedback from test takers).

In creating this examination, a group of content experts practising GIM in diverse settings across Canada was assembled. This group collaboratively developed an exam blueprint that represented both testable objectives of training and GIM in practice. With the guidance of educational experts, the board iteratively developed, vetted, and revised questions resulting in a question bank that meshed with the blueprint. A sample of questions was selected for the exam, and these questions were

then pilot-tested and modified based on feedback from 5 physicians practising as “General Internists”. As a final source of evidence to support content validity, input from the first cohort of test-takers was sought.

The purpose of this paper is to review the responses to this first post exam survey of GIM candidates specifically with respect to content validity (i.e., how well this exam represents the objectives of training and the unique challenges faced by General Internists).

Methods

Examination format

The examination consisted of 43 short-answer questions. Candidates had 3 hours to complete the exam. Test items were mostly case-based questions developed from the real life experience of a practising General Internist. The exam was designed to assess clinical problem solving, rather than rote memorization.

Examination content

Through a consensus-driven exercise, the board developed an examination blueprint, based on the Objectives of Training for GIM.³ However, rather than using a traditional blueprint based on sub-specialties (e.g., Cardiology, Endocrinology, etc), the board developed a content grid (deemed the Purple Print) based on: 1) the settings in which GIM is practised, and 2) the cognitive skills, tasks and attitudes that are particularly germane to the practice of GIM. Although there was an underlying assumption that all questions on the exam would

require the use of “medical expertise,” there was a concerted effort to develop questions that also tested non-medical expert roles.³

As demonstrated in Figure 1, the vertical axis of the grid includes the various settings where GIM is practised. The horizontal axis of the grid includes the foci that represent the higher order skills, tasks and attitudes that one would expect from a consultant practising GIM.⁵

Participants

All examinees who wrote the RCPSC GIM certification examination in the fall of 2014 ($n = 71$) were invited to participate in an anonymous post-examination survey.

Post-examination survey

Survey questions sought to gain information about how well the content on the exam reflects the scope of practice of a general internist versus a non-GIM Internist

Survey questions were developed through an iterative process by the three members of the GIM Board of Examiners and an assessment expert from the RCPSC (DP, BG, NG, and GC, respectively).

Immediately following the administration of the 2014 RCPSC GIM certification examination, candidates were emailed the post-examination survey. Respondents were asked to select their responses using three to five-point rating scales, and they were encouraged to provide written comments.

Figure 1. GIM Exam Content Grid

Settings	Foci								
	Palliative Care	Therapeutics	Medical Procedures	Disease Complexity	Polypharmacy	Data Interpretation	Social Determinants of Health	Health care Systems	Professionalism
Peri-operative Medicine									
Critical Care									
Obstetrical Medicine									
Acute and Urgent Care									
Chronic Disease Prevention and Management									

NB: a subsequent iteration of the Purple Print merged critical care with acute and urgent care, and moved chronic disease to the horizontal axis of the grid.

Table 1. Survey Responses: how well various practice settings were represented on the GIM certification examination (n=54)

	Under-Represented	Appropriately Represented	Over-Represented
Peri-operative Medicine	62.9%	35.2%	1.9%
Obstetrical Medicine	57.4%	38.9%	3.7%
Chronic Disease Prevention and Management	35.2%	61.1%	3.7%
Acute and Urgent Care	18.5%	61.1%	20.4%
Critical Care	13.0%	53.7%	33.3%

Table 2. Survey Responses: how well various foci were represented on the GIM versus IM certification examination (n=36)

	Under-Represented	Appropriately Represented	Over-Represented
Palliative Care	52.8%	25.0%	22.2%
Data Interpretation	8.3%	16.7%	75.0%
Therapeutics	40.0%	17.1%	42.9%
Medical Procedures	27.8%	25.0%	47.2%
Disease Complexity	58.3%	11.1%	30.6%
Polypharmacy	47.2%	33.3%	19.5%
Social Determinants of Health	94.4%	5.6%	0%
Health Care Systems	91.7%	8.3%	0%
Professionalism	91.7%	8.3%	0%

Analysis

Descriptive statistics were calculated for quantitative data. Written comments were reviewed to identify any emerging themes.

Results

Seventy one candidates wrote the fall 2014 RCPSC GIM certification examination. Fifty-six candidates (78.9%) completed the post-examination survey.

Examination Content - General

The majority of respondents reported that the examination was too long, with 80% disagreeing that they had enough time to complete the examination. About half of the respondents indicated the examination difficulty was appropriate (50.9%), while 7.3% indicated it was too easy, and 41.8% indicated it was too difficult (although it is important to note that they completed the survey before learning of their status). Only 32.7% felt the questions were clear and unambiguous.

Examination Content – Settings

Respondents reported that most of the settings identified in the blueprint were appropriately represented on the GIM examination (Table 1). Overall, 57.4% felt that all the important content areas had been assessed. Regarding content that was not represented on the examination, there were 21 written comments. In these comments, the following areas were identified as lacking representation: Peri-operative Medicine, Rheumatology, and Cardiology (each mentioned 5 times); Obstetrical Medicine and Respiriology (each mentioned 4 times); Nephrology and Chronic Disease Prevention (each mentioned 2 times); Toxicology, Hypertension, Venothromboembolic Disease, Nephrology, Neurology, Medical Oncology, and Palliative Care (each mentioned once).

The comments also indicated that respondents were surprised, and frustrated, to see the inclusion of non-medical expert roles on the examination. Of 8 written comments about the content of the exam, 4 related to the over-emphasis of these roles.

Exam Content - Foci

Respondents who had recently taken the Internal Medicine (IM) examination (i.e., within the past 5 years) were asked to compare how well the foci, targeted by the blueprint, had been represented on the GIM versus the IM examination ($n=36$, 65.5%). A total of 77.2% respondents felt that there was little to no overlap between these higher level cognitive skill sets comparing the IM exam to the GIM exam. Table 2 illustrates a clear split in terms of the perceptions of how different foci were represented on GIM versus IM exam.

There were 13 comments associated with this question. Again, several of the comments suggested that respondents objected to the inclusion of questions testing non-medical expert CanMEDS roles (21.6% of comments).

Discussion and Lessons Learned

The survey responses indicate that the GIM certification examination succeeded in assessing the higher order knowledge, skills, and attitudes that are required by general internists practising in Canada, and avoided the redundancy of retesting what was assessed by the core IM examination.

It was the intention of the examination board to create questions that would represent relatively realistic complex cases in order to better assess the clinical reasoning skills required of a consultant.⁶ Interestingly, many candidates appeared to struggle with the ambiguous nature of the questions, which were often developed to deliberately evoke uncertainty.

A recurrent theme related to the respondents' surprise and frustration with the emphasis on non-medical expert CanMEDS roles. Although these are emphasized in the avowed curriculum (i.e., representing more than half of the content in the RCPSC objectives of training for GIM), these roles are somehow devalued by trainees, perhaps due to a "hidden curriculum" which minimizes their importance.⁷ However, studies have linked poor performance on assessments of non-medical expert roles with future complaints to regulatory bodies suggesting that these are important constructs to assess in high stakes examinations.⁸

It was clear from the surveys that many candidates felt that there was not enough time allocated to complete the examination. Although the number of questions was comparable to other written certification examinations administered by the RCPSC, the amount of information presented in the stems was often significantly greater due to the complexity of the cases. In future iterations, the examination board will need to be mindful of this when setting the examination while still maintaining diverse enough

content to avoid case specificity.⁹ The goal, after all, is to assess trainees' clinical acumen, rather than the speed in which they can complete an examination.

The results of the survey were enlightening, and will serve to guide the board in improving the quality of the GIM certification examination and ensuring that it measures what it should (i.e., the skills, knowledge and attitudes required of general internists). In addition, since assessment drives learning, training, and often curriculum reform, we expect that future candidates will be better prepared for the content assessed by this examination.

Ethics

Given that the research was done as part of usual quality control practices for the exam, ethical approval was not sought (as per article 2.5 from the Government Panel on Research Ethics). <http://www.pre.ethics.gc.ca/eng/policy-politique/initiatives/tcps2-epc2/chapter2-chapitre2/#toc02-1a>

References

1. Downing SM. Validity: On meaningful interpretation of assessment data. *Medical Education* 2003; 37: 830-837.
2. Cook D. & Beckman T. Current concepts in validity and reliability for psychometric instruments: theory and application. *The American Journal of Medicine* 2006; 119: 166.e7-166.16.
3. Royal College of Physicians and Surgeons of Canada. Objectives of training in the subspecialty of General Internal Medicine (PDF). 2012. Accessed on February 2, 2015 at: <http://www.royalcollege.ca/cs/groups/public/documents/document/y2vk/mdaw/~edisp/tztest3rcpsced000901.pdf>
4. Frank JR, Snell LS, Sherbino J. The draft CanMEDS 2015 physician competency framework (PDF). Royal College of Physicians and Surgeons of Canada. 2014. Accessed on February 2, 2015 at: http://www.royalcollege.ca/portal/page/portal/rc/common/documents/canmeds/framework/canmeds2015_framework_series_III_e.pdf
5. Miller GE. The assessment of clinical skills/competence/performance. *Academic Medicine* 1990; 65: S63-7.
6. Eva K. What every teacher needs to know about clinical reasoning. *Medical Education* 2004; 39: 98-106.
7. Holmes CL, Harris IB, Schwartz AJ, et al. Harnessing the hidden curriculum: a four-step approach to developing and reinforcing reflective competencies in medical clinical clerkship. *Advances in Health Sciences Education: Theory and Practice* 2014; Oct 16 [Epub ahead of print].
8. Tamblyn R, Abrahamowicz M, Dauphinee D, et al. Physician scores on a national clinical skills examination as predictors of complaints to medical regulatory authorities. *JAMA* 2007; 298: 993-1001.
9. Norman G, Bordage G, Keane D. How specific is case specificity? *Medical Education* 2006; 40: 618-23.