Objective structured clinical examinations (OSCE) performance among Quebec urology residents: A retrospective study from 2008–2019

Laurianne Rita Garabed, MD[1]; Ahmad Almarzouq, MD[2]; Jason Hu, MD[2]; Sero Andonian, MD[2]; Mohamed El-Sherbiny, MD[2]; Nader Fahmy, MD[2]

[1]Department of Urology, Université de Montréal, Montreal, QC, Canada; [2]Department of Urology, McGill University, Montreal, QC, Canada

Cite as: Garabed LR, Almarzouq A, Hu J, et al. Objective structured clinical examinations (OSCE) performance among Quebec urology residents: A retrospective study from 2008–2019. *Can Urol Assoc J* 2020;14(9):E435-44. http://dx.doi.org/10.5489/cuaj.6246

Published online March 30, 2020

Abstract

Introduction: We aimed to compare objective structured clinical examinations (OSCE) performance of residents from four Canadian urology programs, based on resident and station characteristics. We also aimed to evaluate OSCE contents by station type and subspecialty.

Methods: Scores of 109 postgraduate year (PGY)-3 to PGY-5 residents were retrospectively reviewed from 19 OSCEs from May 2008 to February 2019. Scores were grouped by station type/subspecialty, PGY level, medical graduate type (Canadian medical graduate [CMG], international medical graduate [IMG]), sex, and choice of fellowship/practice. Linear mixed modelling was performed to obtain least square means to account for repeated measures.

Results: Score increases from PGY-3 to PGY-5 were significant for all station types and subspecialties ($p\leq0.001$). Scores were similar between male and female residents, and between CMGs and IMGs, except in visual recognition examinations (VREs) (males: 44.3±1.0, females: 39.0±1.6, $p=0.005$; IMG: 47.3±1.7, CMG: 41.6±0.9, $p=0.004$). Relative to uro-oncology stations, scores were lower in andrology ($p=0.010$) and functional urology ($p<0.001$). More female residents chose pediatric (14.3% vs. 1.5%, $p=0.024$) and functional urology fellowships (17.9% vs. 2.9%, $p=0.021$). More male residents chose endourology/robotic fellowships (30.9% vs. 10.7%, $p=0.042$). No associations between subspecialty scores and choice of fellowship/practice were found. Oral stations and VREs were more frequent than telephone stations. Uro-oncology and pediatric urology were more frequent than other subspecialties.

Conclusions: Scores improved with higher PGY level. IMGs and male residents scored better in VREs. Scores were lower in functional urology. There was no correlation between subspecialty score and choice of fellowship/practice. Subspecialties and forms of evaluation were not equally represented.

Introduction

Objective structured clinical examinations (OSCEs) have been used for several years as standardized evaluation tools in medical training. While initially developed for medical students,[1,2] OSCEs have been used as a valid and reliable assessment tool in residency, including in surgical specialties[3] such as urology.[4] Urology OSCEs may include various forms of examination, including standardized oral questions, visual recognition examinations (VREs), surgical simulation stations, and telephone consultations, while covering a range of urological subspecialties. Considering the extensive use of OSCEs in urological training, more data is warranted on OSCE content and performance among urology residents and OSCE content to identify relative strengths and weaknesses in trainees' knowledge and skills, and score progress.

Studying performance differences by sex could guide efforts into urology curriculum and training personalization, considering that urology has historically been a male-represented specialty despite recent increases in female representation in the field.[5] Notably, women are particularly underrepresented when it comes to educational and administrative positions in urology, but there is an increasing number of female urology residents who choose to pursue an academic career and fellowship training;[6] however, experience training female urologists is also limited.[7] More insight is, therefore, needed on sex-specific learning outcomes and patterns. Indeed, it has been previously shown that resident sex has an impact on resident-selected learning objectives, as the needs of female residents in surgery tend to lean towards knowledge-based rather than skill-based objectives.[8] Female applicants in urology have been shown to be mostly interested in particular urological subspecialties — such as pediatric, reconstructive, and functional urology — early in their training, even as medical students.[9] For all these reasons, the impact of sex on in-training performance and choice of fellowship is important to study given the potential training implications that they might have.

Studying the performance differences between international medical graduates (IMGs) and Canadian medical graduates (CMGs) would allow evaluation of whether country of medical school graduation has an impact on performance, building the bases to create an equalizing working environment where knowledge and skills could be shared between residents with different educational backgrounds in respective areas of relative strength. As international and Canadian medical school curricula are not identical, we might expect some differences in scores in certain subspecialties or modes of evaluation, but the direction and areas in which those will occur are difficult to predict, as current literature on the topic is limited.

Previously published studies in other medical subspecialties show greater in-training objective knowledge scores[10] among residents pursuing a fellowship, as well as a subjective relationship between areas of strong personal knowledge and fellowship choice.[11] On the other hand, a previously published study in urology reported that two-thirds of urology residents choose to enter fellowship with the goal of countering perceived training deficiencies.[12] Therefore, it is unclear in which direction in-training performance is associated to urology fellowship choice in Canada. Studying the relationship between in-training performance in urological subspecialties and choice of practice is essential, as it could both guide and strengthen resident confidence towards their career or fellowship choices, and help faculty support residents in their decision through improved and individualized training.

OSCE results are particularly interesting to study given their proven validity in other medical fields. In medical publications, OSCE scores have been shown to be predictive of future performance on national high-stakes examinations,[13] to demonstrate progression of most clinical skills during residency training,[14] and to have high reliability over time, as evidenced by improvements in performance by PGY level,[15] including in urology for oral stations,[4] a trend we would expect to observe in our study. Previous studies on OSCEs have also proven the ability of OSCEs to give insights on curriculum design, as OSCEs can be used as evaluation tools when comparing different training methods; they have been used in urology for evaluating urodynamic study interpretation skills between residents who used different learning media[16] but also in other specialties, such as orthopedic surgery, to compare near-peer and other forms of teaching.[17] Finally, previous studies showing resident dissatisfaction with OSCE format in general surgery, in spite of it being an adequate measure of clinical knowledge, has led the way to reconsideration of OSCE format in a dynamic way, such as by tailoring OSCE content and format to level of training.[18] Given all the mentioned advantages of studying OSCE performance and its relationship with various variables, the present study could potentially lay the ground to improved OSCE design in Canadian urology programs, notably in the context of future training developments in competence by design (CBD) education.

Based on these observations, this study aims to describe and identify trends in OSCE performance of urology residents from four Canadian urology programs, from 2008–2019 based on resident characteristics (PGY level, sex, medical graduate type [CMG vs. IMG], and choice of urological practice), as well as OSCE station characteristics (station subspecialty and form of evaluation). Study choice of fellowship/specialty will also be stratified by sex. This study also aims to evaluate the overall representation of urological subspecialties and forms of evaluation in the content of OSCE stations from 2008–2019, which will provide a view of past OSCE content and serve as a reference point for future improvements in OSCE design.

Methods

Study design

This study is a retrospective analysis of Quebec urology resident performance in oral subspecialty stations, VREs, and telephone consultation stations in OSCE sessions from 2008–2019. Data was obtained through Excel sheets with raw scores by resident and station for each OSCE session, which were created in the weeks following each OSCE examination session by an assigned clinician urologist member of the OSCE-organizing committee and compiled over the years since 2008. Stations were linked to respective scenarios, and station subspecialty was determined based on scenario topic. Therefore, for each examination session, scores were obtained by subspecialty for each participating resident.

Study population

OSCE performance of 109 urology residents from McGill University, Université de Montréal, Université Laval, and Université de Sherbrooke was retrieved from 19 OSCE examinations sessions across 13 consecutive cohorts from May 2008 to February 2019. Each OSCE examination session takes place at the same site — McGill University or Université de Montréal — for all participating residents and involves the participation of up to all four programs, depending on invited programs for a particular session. The number of participants by PGY level and participating programs for each OSCE examination session are outlined in Supplementary Table 1. Analyzed OSCE examination sessions include data up to and including February 2019, as this examination session was the last one of the academic year for a particular postgraduate year (PGY). PGY-4 and PGY-5 residents were trained at McGill University, Université de Montréal, and Université Laval. PGY-3 were trained at any of the four urology programs, including Université de Sherbrooke, as Université de Sherbrooke trains residents up to PGY-3 level, after which they train at one of the three other Quebec programs. The proportions of CMGs vs IMGs and female vs. male residents for each OSCE examination session are included in Supplementary Table 2.

OSCE format

Each year, one to two OSCE examination sessions are organized for Quebec PGY-3, PGY-4, and PGY-5 urology residents. Invited residents all take the same OSCE examination on the same day at the same site. Each examination session includes a variable number of OSCE stations, which may be oral stations, VREs, telephone consultation stations, or surgical simulations. In this study, grading was included for oral stations, VREs, and telephone stations. Surgical simulations were excluded from the analysis, as OSCE Excel sheets focused on oral stations, VREs, and telephone stations, and did not include data for all surgical simulation sessions since 2008. Oral stations were categorized by subspecialty: uro-oncology, pediatric urology, endourology, andrology, functional urology, reconstructive urology/trauma, and transplant. Scenarios and questions were developed by academic clinician staff urologists from any of the four institutions based on their subspecialties. Grading was based on predetermined answer sheets based on a checklist system, with allocated points for each given answer. Questions and grading sheets were reviewed by staff urologists from the OSCE organizing committee prior to examination. During the examination, answer sheets were checked by academic clinician staff urologists from any of the four Quebec urology programs. Examiners were different from the scenario authors for each respective station.

Future urology fellowship/practice

Chosen urology practice/fellowship was determined for all alumni from Quebec urology residency programs (96 residents out of 109). This was done through inquiring collateral information from respective urology residency programs, if possible. Otherwise, this information was obtained through a Google search of resident names, looking for fellowship-specific or hospital-specific pages. Future urology practice/fellowship was categorized under one of the following: andrology, endourology/robotic, functional urology, uro-oncology, pediatric urology, reconstructive urology, transplant, and general urology fellowship/direct practice. In the event where future urology practice was not consistent with fellowship specialty, fellowship specialty was retained.

Outcomes

The primary aim of the study was to compare resident OSCE scores by population-specific data: PGY level, sex, medical graduate type, and choice of urological practice. Consequently, primary outcomes included PGY level-stratified scores, female vs. male scores, CMG vs. IMG scores for each station type, as well as score comparison between residents by chosen fellowship/practice in order to evaluate potential correlations between in-training scores and choice of career. Additionally, we aimed to compare resident scores by station-specific data. Related outcomes included comparison of scores between the various subspecialties, and between oral, VRE, and telephone stations.

The secondary aim of this study was to evaluate the content of OSCE examination sessions between 2008 and 2019, specifically looking at the representation of urological subspecialties in oral stations, and at the representation of oral, VRE, and telephone stations in the 19 OSCE examination sessions. A further secondary outcome was the choice of fellowship/specialty stratified by sex.

An exploratory aim was added when analyzing OSCE results by station subspecialty data, with the goal of comparing PGY-5 scores in functional urology before and after introduction of the pan-Canadian functional urology course in October 2017 in order to evaluate for potential changes in graduating resident performance following this change in curriculum.

Statistical analysis

All scores are presented as percentages. For each station, comparison of scores across PGY levels was done by evaluating least square means and standard errors obtained from linear mixed modelling while considering PGY level as a fixed effect and the individual resident as a random effect to account for repeated measures.

The comparison of station types was performed by evaluating least square means obtained from linear mixed modelling while considering station type as a fixed effect and the individual resident as a random effect. The oral uro-oncology stations were chosen as the reference group for this analysis, given they were the only oral station included in every OSCE session.

Additionally, comparisons of least square means by sex for each station were also performed through linear mixed modelling while including sex and PGY level as fixed effects and individual resident as a random effect. Similar analyses of scores by medical graduate type (CMG/IMG) were also conducted. Oral transplant stations were not included in any of the previously mentioned analyses given the relative lack of observations pertaining to them, as they were only present in two OSCE sessions.

For oral subspecialty stations, Kruskal-Wallis tests were performed to compare mean PGY-5 scores stratified by chosen fellowship; univariable logistic regression models were also fitted to explore associations between PGY-5 scores on subspecialty station and choice of fellowship.

Fisher's exact tests were used to evaluate the association between choice of fellowship/subspecialty and sex.

The exploratory comparison of least square means of PGY-5 functional urology scores before and after 2017 were also performed through linear mixed modelling while considering the year of the OSCE session (pre/post-October 2017) as a fixed effect and individual resident as a random effect.

Continuous data are presented as least square means ± standard errors and categorical data as counts and percentages. Significance testing was two-sided, with a threshold set at p<0.05. All statistical analyses were performed with SAS version 9.4 (SAS Institute, Cary, NC, U.S.).

Results

In total, OSCE scores of 109 residents from 19 OSCE examination sessions were included. The cohort was comprised mostly of male residents (75 male [69%] and 34 female residents [31%]) and mostly of CMGs (91 CMGs [83%] and 18 IMGs [17%]). The number of CMGs and IMGs, as well as male and female residents, sitting at each respective OSCE examination session is included in Supplementary Table 2.

Comparison of least square mean scores for all stations by PGY level are shown in Table 1. For every form of examination and station subspecialty, scores increased from PGY-3 to PGY-5 (all p≤0.001). All stations, except telephone stations (p=0.058), experienced significant increases from PGY-3 to PGY-4 (p<0.050). All stations, except telephone stations (p=0.285) and endourology stations (p=0.06), also experienced significant increases from PGY-4 to PGY-5 (p<0.050). Male and female resident scores across most forms of examination and subspecialties were similar, except for VREs (males: 44.3±0.9 vs. females: 39.0±1.6, p=0.005) (Table 2). When stratifying by medical graduate type, resident scores were comparable across all urological subspecialties in oral stations and in telephone stations, but IMGs performed better in VREs than CMGs (IMGs: 47.3±1.7 vs. CMGs: 41.6±0.9, p=0.004) (Table 3).

Relative to uro-oncology subspecialty stations (70.7±0.9), scores on andrology (67.1±1.1, p=0.010), functional urology (62.0±0.9, p<0.001), and VRE stations (43.0±0.9, p<0.001) were significantly lower (Supplementary Table 3). When comparing functional urology scores before and after the start of the pan-Canadian functional urology course for PGY-5s (before October 2017 vs. starting from October 2017), an improvement was found for PGY-5 scores, although it was not statistically significant (65.4±1.7 vs. 91.7±3.0, p=0.082).

Fellowship/practice choice was determined among 96 urology residents/alumni, the other ones having not graduated yet. Most residents chose general urology fellowships/direct practice (33.3%) or endourology/robotic urology fellowships (25.0%). The least pursued urological practice/fellowship was transplant (2.1%). A greater proportion of female residents undertook pediatric (14.3% vs. 1.5%, p=0.024) and functional fellowships (17.9% vs. 2.9%, p=0.021) with statistical significance, whereas male residents chose endourology/robotic fellowships more frequently than female residents (30.9% vs. 10.7%, p=0.042) (Fig. 1). Mean scores in all oral subspecialty stations were similar when stratified by chosen fellowship (Fig. 2). In univariable logistic regression analyses, PGY-5 score in a specific subspecialty station was not associated with choice of fellowship (Supplementary Fig. 1).

Among the 19 OSCE examination sessions, oral and VRE stations were continuously represented, whereas telephone

Table 1. Scores by PGY level and station type

Station type	PGY-3 score	PGY-4 score	PGY-5 score	p		
				PGY-3 vs. PGY-5	PGY-3 vs. PGY-4	PGY-4 vs. PGY-5
Uro-oncology	63.2 (1.1) (n=102, N=76)	69.7 (1.1) (n=96, N=72)	78.7 (1.1) (n=108, N=70)	<0.001	<0.001	<0.001
Pediatric urology	63.2 (1.2) (n=100, N=74)	73.0 (1.3) (n=94, N=70)	79.7 (1.2) (n=105, N=67)	<0.001	<0.001	<0.001
Endourology	63.7 (2.1) (n=44, N=36)	71.0 (2.1) (n=41, N=34)	76.7 (1.9) (n=53, N=14)	<0.001	0.021	0.06
Andrology	59.0 (1.9) (n=65, N=56)	66.0 (2.0) (n=62, N=55)	75.7 (1.9) (n=69, N=58)	<0.001	0.013	0.001
Functional urology	53.2 (1.9) (n=95, N=69)	59.9 (1.9) (n=90, N=66)	72.1 (1.8) (n=103, N=65)	<0.001	0.015	<0.001
Reconstructive/trauma	62.8 (1.6) (n=46, N=41)	70.0 (1.6) (n=46, N=41)	78.5 (1.5) (n=54, N=42)	<0.001	0.004	0.001
VRE	35.0 (1.4) (n=102, N=76)	43.2 (1.4) (n=96, N=72)	50.4 (1.4) (n=108, N=70)	<0.001	<0.001	<0.001
Telephone	69.7 (1.9) (n=58, N=48)	74.9 (1.9) (n=56, N=45)	77.7 (1.8) (n=60, N=44)	0.001	0.058	0.285

Data are presented as least square means (standard error) for every station type obtained from linear mixed modeling with PGY level as a fixed effect and individual resident as a random effect. A lowercase n represents the number of observations, while an uppercase N represents the number of residents from which the observations are drawn. PGY: postgraduate year; VRE: visual recognition examination.

Table 2. Comparison of scores by sex

Station type	Male		Female		p
Oral uro-oncology	70.4	(0.7) (n=226, N=75)	71.1	(1.2) (n=80, N=34)	0.586
Oral pediatric urology	72.7	(0.8) (n=220, N=72)	70.0	(1.4) (n=79, N=34)	0.091
Oral endourology	71.6	(1.4) (n=104, N=60)	67.2	(2.4) (n=34, N=24)	0.112
Oral andrology	67.4	(1.3) (n=143, N=62)	65.5	(2.1) (n=53, N=25)	0.464
Oral functional urology	61.8	(1.3) (n=211, N=69)	61.4	(2.1) (n=77, N=34)	0.891
Oral reconstructive/ trauma	71.0	(1.1) (n=106, N=68)	68.9	(1.7) (n=40, N=30)	0.308
VRE	44.3	(0.9) (n=226, N=75)	39.0	(1.6) (n=80, N=34)	0.005
Telephone	73.7	(1.2) (n=129, N=56)	75.3	(2.1) (n=45, N=25)	0.495

Data are presented as least square means (standard error) obtained from linear mixed modeling with sex and PGY level as fixed effects and individual resident as a random effect. A lowercase n represents the number of observations, while an uppercase N represents the number of residents from which the observations are drawn. PGY: postgraduate year; VRE: visual recognition examination.

consultations were only present in 11 sessions out of 19. For oral questions, subspecialties were variedly represented across the 19 OSCE examination sessions, with transplant being the least represented (2/19), followed by endourology (8/19) and reconstructive urology/trauma (8/19). Uro-oncology (19/19) and pediatric urology (18/19) questions were the most represented oral stations in the OSCE examination sessions (Table 4).

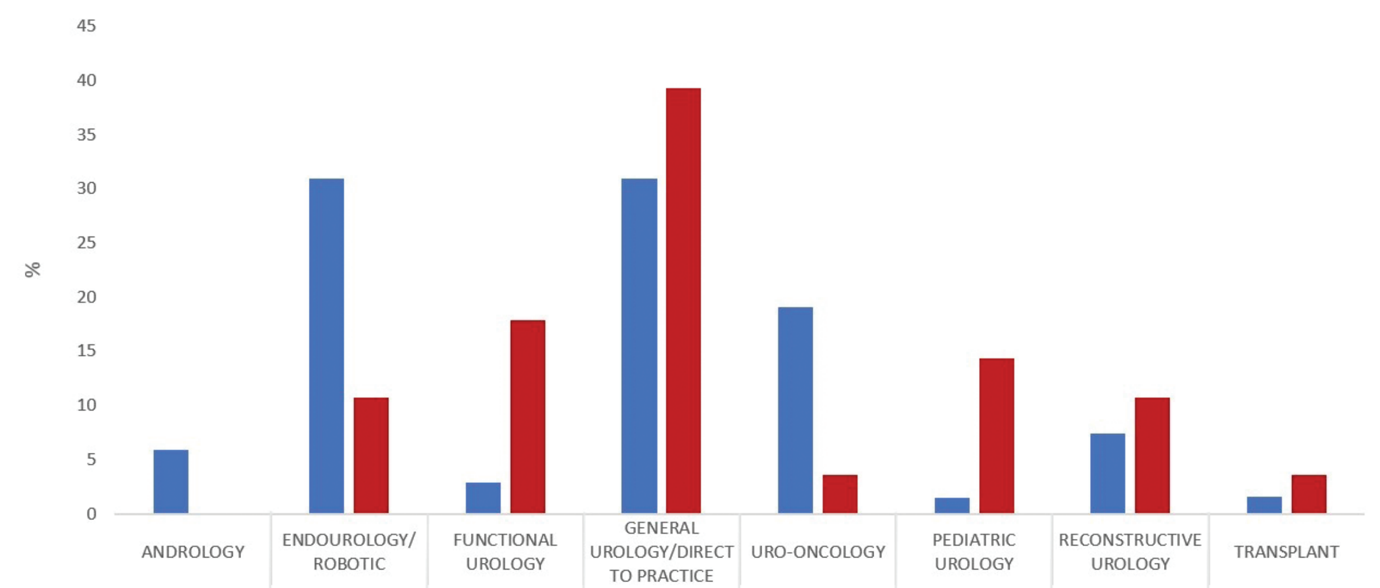
Table 3. Comparison of scores by medical graduate type

Station type	CMG		IMG		p
Oral uro-oncology	70.5	(0.7) (n=239, N=91)	70.7	(1.3) (n=67, N=18)	0.909
Oral pediatric urology	72.0	(1.9) (n=235, N=89)	76.1	(1.7) (n=64, N=17)	0.096
Oral endourology	70.7	(1.4) (n=109, N=69)	74.6	(1.9) (n=29, N=15)	0.104
Oral andrology	66.8	(1.3) (n=153, N=69)	67.2	(2.4) (n=43, N=18)	0.887
Oral functional urology	62.6	(1.2) (n=228, N=88)	58.4	(2.4) (n=60, N=15)	0.115
Oral reconstructive/ trauma	70.6	(1.0) (n=121, N=84)	69.7	(2.2) (n=25, N=14)	0.705
VRE	41.6	(0.9) (n=239, N=91)	47.3	(1.7) (n=67, N=18)	0.004
TELEPHONE	74.3	(1.2) (n=131, N=66)	73.3	(2.2) (n=43, N=15)	0.678

Data are presented as least square means (standard error) obtained from linear mixed modeling with medical graduate type and PGY level as fixed effects and individual resident as a random effect. A lowercase n represents the number of observations, while an uppercase N represents the number of residents from which the observations are drawn. CMG: Canadian medical graduate; IMG: international medical graduate; OSCE: objective structured clinical examination; PGY: postgraduate year; VRE: visual recognition examination.

Discussion

From 2008–2019, OSCE performance of Quebec urology residents reflects higher scores with more advanced postgraduate training for all urological subspecialties and



	ANDROLOGY	ENDOUROLOGY/ ROBOTIC	FUNCTIONAL UROLOGY	GENERAL UROLOGY/DIRECT TO PRACTICE	URO-ONCOLOGY	PEDIATRIC UROLOGY	RECONSTRUCTIVE UROLOGY	TRANSPLANT
Total, n (%)	4 (4.2)	24 (25.0)	7 (7.3)	32 (33.3)	14 (14.6)	5 (5.2)	8 (8.3)	2 (2.1)
■ Male, n (%)	4 (5.9)	21 (30.9)	2 (2.9)	21 (30.9)	13 (19.1)	1 (1.5)	5 (7.4)	1 (1.5)
■ Female, n (%)	0 (0.0)	3 (10.7)	5 (17.9)	11 (39.3)	1 (3.6)	4 (14.3)	3 (10.7)	1 (3.6)
p-value	0.318	0.042	0.021	0.479	0.060	0.024	1.000	0.500

Fig. 1. Chosen urological fellowship/specialty by sex.

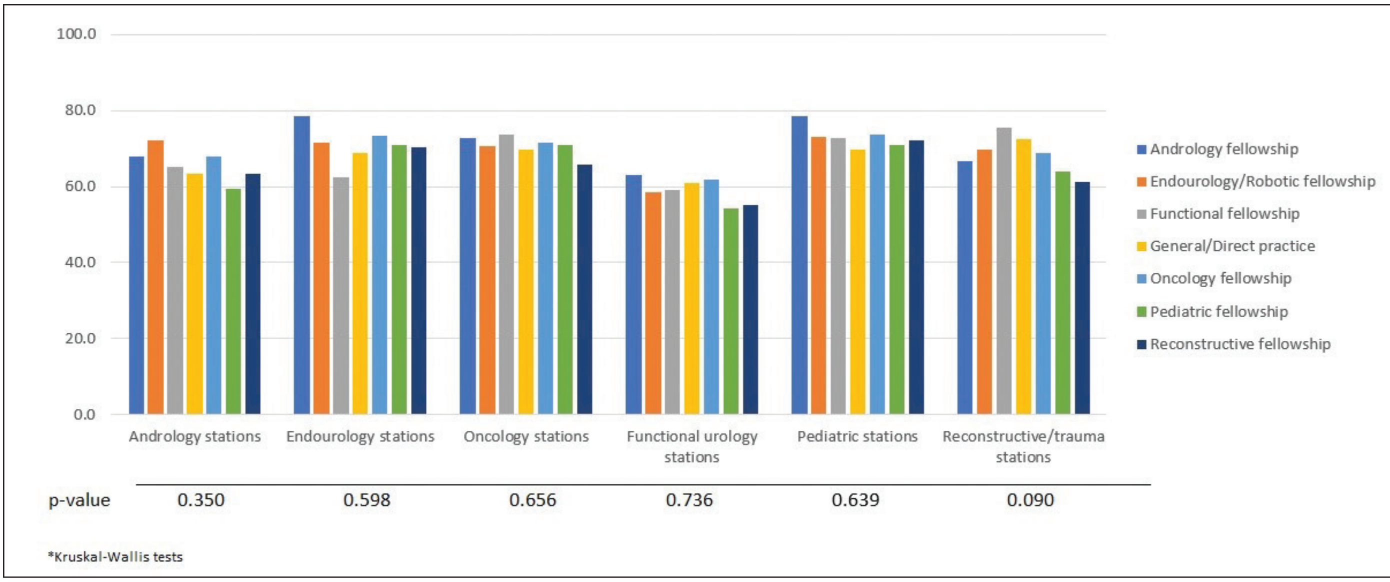


Fig. 2. Mean scores on oral subspecialty stations grouped by chosen fellowship/practice.

forms of evaluation from PGY-3 to PGY-5 level. Previous studies on surgical training,[19] including Canadian urology programs,[20,21] have shown the correlation between in-training evaluation and PGY to be particularly strong in technical skills stations. In a French study, the relationship between OSCE performance and postgraduate level of residency training has also been shown to be present in clinical oral stations in urology.[4] However, no data is available on the evolution of OSCE performance across specific urological subspecialties and in other forms of evaluation, such as VRE and telephone stations. This study highlights that performance in clinical stations improved from PGY-3 to PGY-5 across individual urological subspecialties in oral stations, in VREs, and in telephone consultations. These results reflect the validity of the use of OSCEs as a proxy for progress during the senior years of residency, not only across all urological subspecialties in oral stations, but also in other modalities of evaluations, such as VREs and telephone stations, which, to our knowledge, has not been previously shown in the literature.

Table 4. OSCE station representation across 19 OSCE examination sessions by form of examination and subspecialty

	Number of times questions were included
Oral uro-oncology	19
Oral pediatric urology	18
Oral endourology	8
Oral andrology	13
Oral functional urology	16
Oral reconstructive/trauma	8
Oral transplant	2
VRE	19
TELEPHONE	11

OSCE: objective structured clinical examinations PGY: postgraduate year; VRE: visual recognition examination.

Overall resident scores were found to be lower in VREs than in oral stations and telephone stations. This could be related to a higher level of difficulty present in VRE stations, as these stations may cover rarer pathological entities. It may also be more difficult for residents to prepare for these sessions, as they often target urology-related disciplines, such as radiology or pathology, which may be less prioritized in current urology training. VRE scores were also found to be lower among CMGs and female urology residents. While it is unclear why such a difference was found, this could be related to differences in training before residency, favoring performance of IMGs in urology-related disciplines represented in VREs. It is important to note that all IMG residents in this study are male, which could explain the difference in VRE scores between female and male urology residents. These lower VRE scores might reflect a need to redirect more training time towards urology-related disciplines, which often form the basis of VRE. However, the literature on the efficiency and relevance of VREs is virtually non-existent, such that it is difficult to assess whether lower scores are translatable into a need for more extensive teaching in these disciplines or for changes in station design. This study, therefore, underscores the need for more research on VREs in urology training and formal evaluation.

In oral stations, scores were found to be lower in functional urology than in other urological subspecialties. While this could be related to station level of difficulty, it is possible that this difference is arising from the history of paucity of standardized approaches to teach functional urology, including urodynamic skills, which had been previously noted among Canadian urology programs.[16] However, an annual functional urology course has been implemented since 2017 for all Canadian PGY-5 urology residents. We found that functional urology scores in OSCEs have been trending towards improvement after the development of this pan-Canadian annual functional urology course for PGY-5 residents, although it was not statistically significant, probably due to the limited sample size available for results after implementation of the course. More research on current functional urology training and potential standardized stategies is, therefore, needed in order to potentially enhance residents' performance in this field, with the goal of ultimately optimizing patient careThis present study found no correlation between OSCE scores and chosen fellowship/subspecialty. One may expect that residents interested in a given subspecialty, especially PGY-5 residents, would score higher than their colleagues in OSCE stations, or that residents' career choice is guided by in-training evaluation performance. A study published on internal medicine residency showed that interest in pursuing any fellowship was associated with greater general medical knowledge,[10] although the association between medical knowledge and choice of specific fellowship was not investigated. In-training examination performance has also been shown to be related to success in board examinations across various medical specialties,[22-26] but this was shown not to be the case among residents in surgery.[27] Based on the available literature and the present study, it seems possible that the association between in-training performance in formal examinations and career choice is stronger in medical than surgical specialties, the latter of which includes more technical skills that are not always evaluated in OSCEs. In addition, as mentioned in the Introduction, a previous study among urology residents stated that perceived training gaps were also a reason for residents to pursue a specific subspecialty fellowship in order to improve knowledge and skills,[12] although the relationship between perceived subjective training gaps and objective in-training performance was not studied in this article. Choice of fellowship/practice could, therefore, be influenced both by interest and perceived training gaps. In our study, no correlation between OSCE scores and choice of fellowship/practice in either direction was found. This could be related to other variables that were unmeasured in our study but potentially important factors in guiding residents from various medical and surgical specialties towards their choice of career. Notably, geographical considerations, personal factors (such as debts and the presence of children),[28] the perception of employability in a specific field,[28] the type of research publications during and prior to residency,[29] and importantly the presence of mentors in a particular subspecialty,[30,31] have all been proven to be factors influencing residents in their choice of career after completion of residency. However, these variables are less likely to be confounders in this study, as they are probably not associated with OSCE scores despite being related to career choice and, therefore, do not affect the validity of our results. Future studies measuring these factors along with OSCE performance should be undertaken to properly assess their impact on career choice.

The various urological subspecialties and forms of examination were not equally represented in the 19 OSCE sessions. Notably, transplant, endourology, and reconstructive urology/trauma were less represented in oral stations than other subspecialties. Telephone stations were also less represented than oral stations and VREs. However, most of these differences arise from earlier OSCE examination sessions, in contrast to more recent OSCE sessions, which have included a greater variety of urological subspecialties and forms of examination. This variety is important to keep in mind when designing future examinations in urology residency training, especially in the context of CBD education, where specific objectives need to be met in all areas of the field.

Limitations

While highlighting new ideas in urology education development, this study is not without limitations.

The variability of station number and station subspecialty topics for each OSCE session was not consistent from year to year, such that some score variations may have been introduced when comparing resident scores across the 19 OSCE sessions. A smaller number of stations was present for some subspecialty categories, including transplant, endourology, and reconstructive urology, such that some residents did not have an available score for some of these subspecialties for a given OSCE examination session, which limits the sample size available for some of the analyses.

While these differences across the years may be the reflection of changes in goals of the Royal College Comprehensive Objective Examination,[32] it is difficult to assess, as official versions of previous goals are not available on the Royal College website. It is also of note that examiners were not the same during the 10 years, which is something we cannot control for, and that elements included in grading sheets are scenario-dependent, although general structure is maintained. As station format and grading method by predetermined sheets were preserved, score comparisons between available station scores was still possible.

This study is retrospective and is based on datasets created over the years. Given the aforementioned variables (presence of certain stations across OSCE sessions and the lower number of female and IMG residents), sample size limitations in certain analyses have to be considered when examining the results of those particular subgroups. Indeed, while scores from the four urology programs were present for many of the 19 examination sessions, some of the available datasets only included scores from two or three of the four urology programs. However, the number of observations is still substantial.

Finally, this study did not include technical skill stations because of restricted data, as our available datasets did not include surgical simulations raw data for most OSCE sessions. In addition, it may be difficult to compare scores of technically different skill stations over the years, as grading scheme and evaluated content may have been significantly variable over the years. As more technical skill stations are included in OSCEs and as similar skills are evaluated longitudinally in the same senior residents, it would be interesting to study the progression and characteristics of scores in surgical simulations in the future, as these skills are an integral part of surgical practice.

Conclusions

Despite its limitations, this study spans a considerable interval of time and brings important insight on trends in OSCE performance among Canadian urology residents. It supports previously published literature and explores new correlations between OSCE scores and urological subspecialties, forms of evaluation, level of residency training, resident sex, choice of urological practice, and medical graduate type in a way that will allow direct integration of results into design of current urology residency curricula. This study results also indicate areas of research deficiency in urological education, such as VRE assessment, and can be used to help guide research efforts in urological training. Notably, with recent and ongoing changes in urology curricula based on the CBD approach, parallel modifications in resident assessment and consequently OSCE design are also expected, making studies on resident performance an integral part of educational efforts across Canadian urology programs.

Competing interests: The authors report no competing personal or financial interests related to this work.

This paper has been peer-reviewed

References

1. Harden RM. Revisiting "Assessment of clinical competence using an objective structured clinical examination (OSCE)." *Med Educ* 2016;50:376-9. https://doi.org/10.1111/medu.12801
2. Harden RM, Gleeson F. Assessment of clinical competence using an objective structured clinical examination (OSCE). *Med Educ* 1979;13:39-54. https://doi.org/10.1111/j.1365-2923.1979.tb00918.x
3. Cohen R, Reznick RK, Taylor BR, et al. Reliability and validity of the objective structured clinical examination in assessing surgical residents. *Am J Surg* 1990;160:302-5. https://doi.org/10.1016/S0002-9610(06)80029-2
4. Sibert L, Grand'Maison P, Doucet J, et al. Initial experience of an objective structured clinical examination in evaluating urology residents. *Eur Urol* 2000;37:621-7. https://doi.org/10.1159/000020203
5. Halpern JA, Lee UJ, Wolff EM, et al. Women in urology residency, 1978–2013: A critical look at gender representation in our specialty. *Urology* 2016;92:20-5. https://doi.org/10.1016/j.urology.2015.12.092
6. Yang G, Villalta JD, Weiss DA, et al. Gender differences in academic productivity and academic career choice among urology residents. *J Urol* 2012;188:1286-90. https://doi.org/10.1016/j.juro.2012.06.022
7. Han J, Stillings S, Hamann H, et al. Gender and subspecialty of urology faculty in department-based leadership roles. *Urology* 2017;110:36-9. https://doi.org/10.1016/j.urology.2017.07.044
8. Nebeker CA, Basson MD, Haan PS, et al. Do female surgeons learn or teach differently? *Am J Surg* 2017;213:282-7. https://doi.org/10.1016/j.amjsurg.2016.10.010
9. Lebastchi AH, Khouri Jr RK, McLaren ID, et al. The urology applicant: An analysis of contemporary urology residency candidates. *Urology* 2018;115:51-8. https://doi.org/10.1016/j.urology.2017.10.065
10. Ofoma UR, Lehman EE, Haidet P, et al. Associations between subspecialty fellowship interest and knowledge of internal medicine: A hypothesis-generating study of internal medicine residents. *BMC Med Educ* 2011;11:5. https://doi.org/10.1186/1472-6920-11-5
11. Farmakis SG, Hardy AK, Thomas KB, et al. Changes in factors influencing fellowship choices among radiology residents from 2008 to 2018 and methods that may increase interest in the pediatric radiology subspecialty. *Pediatr Radiol* 2019:1-10. https://doi.org/10.1007/s00247-019-04430-4
12. Okhunov Z, Safiullah S, Patel R, et al. Evaluation of urology residency training and perceived resident abilities in the United States. *J Surg Educ* 2019;76:936-48. https://doi.org/10.1016/j.jsurg.2019.02.002
13. Pugh D, Bhanji F, Cole G, et al. Do OSCE progress test scores predict performance in a national high-stakes examination? *Med Educ* 2016;50:351-8. https://doi.org/10.1111/medu.12942
14. Pugh D, Touchie C, Humphrey-Murto S, et al. The OSCE progress test-measuring clinical skill development over residency training. *Med Teach* 2016;38:168-73. https://doi.org/10.3109/0142159X.2015.1029895
15. Pugh D, Touchie C, Wood TJ, Humphrey-Murto SJMe. Progress testing: Is there a role for the OSCE? *Med Educ* 2014;48:623-31. https://doi.org/10.1111/medu.12423
16. Shamout S, Andonian S, Kabbara H, et al. Teaching and evaluation of basic urodynamic skills in urology residency programs: Randomized controlled study. *Neurourol Urodyn* 2018;37:2724-31. https://doi.org/10.1002/nau.23728
17. Phillips D, Pean CA, Allen K, et al. Using objective structured clinical examinations to assess intern orthopaedic physical examination skills: A multimodal didactic comparison. *J Surg Educ* 2017;74:513-8. https://doi.org/10.1016/j.jsurg.2016.10.011
18. Zyromski NJ, Staren ED, Merrick HWJCS. Surgery residents' perception of the objective structured clinical examination (OSCE). *Curr Surg* 2003;60:533-7. https://doi.org/10.1016/S0149-7944(03)00005-9
19. Cerilli GJ, Merrick HW, Staren ED, et al. Objective structured clinical examination technical skill stations correlate more closely with postgraduate year level then do clinical skill stations. *Am Surg* 2001;67:323.
20. Noureldin YA, Elkoushy MA, Fahmy N, et al. Assessment of photoselective vaporization of prostate skills during Urology Objective Structured Clinical Examinations (OSCE). *Can Urol Assoc J* 2015;9:E61-6. https://doi.org/10.5489/cuaj.2273
21. Nguyen LN, Tardioli K, Roberts M, et al. Development and incorporation of hybrid simulation OSCE into in-training examinations to assess multiple CanMEDS competencies in urologic trainees. *Can Urol Assoc J* 2015;9:32-6. https://doi.org/10.5489/cuaj.2366
22. Brateanu A, Yu C, Kattan MW, et al. A nomogram to predict the probability of passing the American Board of Internal Medicine examination. *Med Educ Online* 2012;17:18810. https://doi.org/10.3402/meo.v17i0.1881
23. Indik JH, Duhigg LM, McDonald FS, et al. Performance on the cardiovascular in-training examination in relation to the ABIM cardiovascular disease certification examination. *J Am Coll Cardiol* 2017;69:2862-8. https://doi.org/10.1016/j.jacc.2017.04.020
24. Grabovsky I, Hess BJ, Haist SA, et al. The relationship between performance on the infectious diseases in-training and certification examinations. *Clin Infect Dis* 2014;60:677-83. https://doi.org/10.1093/cid/ciu906
25. Lohr KM, Clauser A, Hess BJ, et al. Performance on the adult rheumatology in-training examination and relationship to outcomes on the rheumatology certification examination. *Arthritis Rheumatol* 2015;67:3082-90. https://doi.org/10.1002/art.39281
26. Kempainen RR, Hess BJ, Addrizzo-Harris DJ, et al. Pulmonary and critical care in-service training examination score as a predictor of board certification examination performance. *Ann Am Thorac Soc* 2016;13:481-8. https://doi.org/10.1513/AnnalsATS.201601-0150C
27. Ray JJ, Sznol JA, Teisch LF, et al. Association between American Board of Surgery in-training examination scores and resident performance. *JAMA Surg* 2016;151:26-31. https://doi.org/10.1001/jamasurg.2015.3088
28. Mok PS, Probyn L, Finlay K. Factors influencing radiology residents' fellowship training and practice preferences in Canada. *Can Assoc Radiol J* 2016;67:99-104. https://doi.org/10.1016/j.carj.2015.08.005
29. Al Omran Y. Factors influencing fellowship selection, career trajectory, and academic productivity among plastic surgeons. *Plast Reconst Surg* 2014;133:730-6. https://doi.org/10.1097/01.prs.0000438043.98762.51
30. Wang L, Mittal A, Puttmann K, et al. The changing gender landscape of pediatric urology fellowship: results from a survey of fellows and recent graduates. *J Pediatr Urol* 2019;15:51-7. https://doi.org/10.1016/j.jpurol.2018.09.001

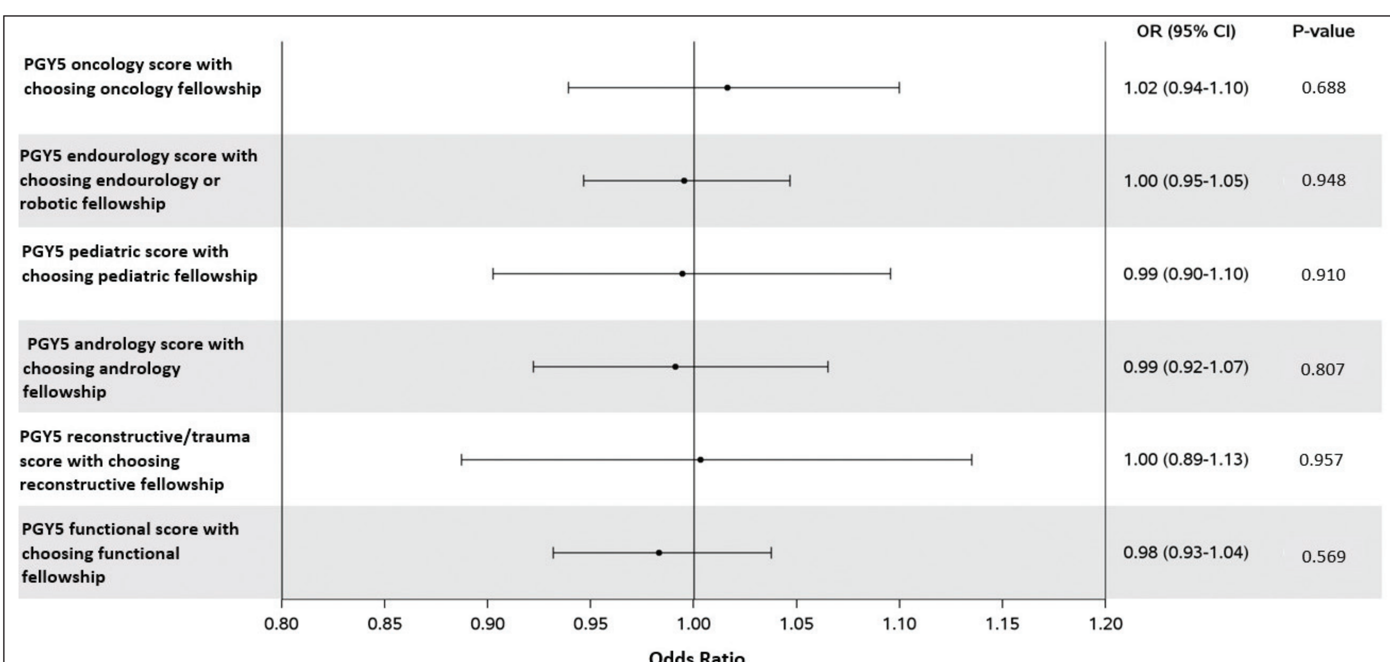
31. Kavolus JJ, Matson AP, Byrd WA, et al. Factors influencing orthopedic surgery residents' choice of subspecialty fellowship. *Orthopedics* 2017;40:e820-4. https://doi.org/10.3928/01477447-20170619-01
32. Royal College of Physicians and Surgeons of Canada. Format of the comprehensive objective examination in urology 2019 [Updated 2019 June]. Available at http://www.royalcollege.ca/rcsite/documents/ibd/urology_examformat_e. Accessed Oct. 20, 2019.

Correspondence: Dr. Laurianne Rita Garabed, Department of Urology, Université de Montréal, Montreal, QC, Canada; laurianne.rita.garabed@umontreal.ca

Supplementary Table 1. Number of residents participating at each OSCE examination session by PGY level and participating urology programs

Dates	PGY level			Participating urology programs			
	PGY-5	PGY-4	PGY-3	McGill	UdeM	Laval	Sherbrooke
May 2008	3	2	2	x			
May 2009	3	4	5	x			x
May 2010	3	5	5	x			
December 2010	8	12	10	x	x	X	x
May 2011	5	5	5	x	x		x
May 2012	3	3	4	x			
December 2012	5	5	2	x			
April 2013	3	5	4	x			
December 2013	9	7	9	x	x		x
April 2014	6	2	5	x			
November 2014	9	8	9	x	x		x
March 2015	3	3	3	x			x
December 2015	9	9	11	x	x		x
March 2016	5	4	4	x	x		x
October 2016	9	11	10	x	x		x
February 2017	4	3	5	x			x
October 2017	14	7	10	x	x	x	x
February 2018	9	3	4	x	x		
February 2019	3	3	3	x			

OSCE: objective structured clinical examinations; PGY: postgraduate year; UdeM: Université de Montréal.



Supplementary Fig. 1. Forest plot of univariable logistic regression models for association between post-graduate year 5 subspecialty station scores and choosing corresponding fellowship. CI: confidence interval; OR: odds ratio; PGY: postgraduate year.

Supplementary Table 2. Number and proportion of CMGs/IMGs and female/male residents for each OSCE examination session

OSCE examination session	Number of CMGs (%)		Number of IMGs (%)		Number of male (%)		Number of female (%)	
May 2008	5	71%	2	29%	6	86%	1	14%
May 2009	8	67%	4	33%	10	83%	2	17%
May 2010	10	77%	3	23%	10	77%	3	23%
December 2010	24	80%	6	20%	21	70%	9	30%
May 2011	11	73%	4	27%	13	87%	2	13%
May 2012	7	70%	3	30%	7	70%	3	30%
December 2012	7	58%	5	42%	10	83%	2	17%
April 2013	7	58%	5	42%	10	83%	2	17%
December 2013	20	80%	5	20%	17	68%	8	32%
April 2014	8	62%	5	38%	11	85%	2	15%
November 2014	22	85%	4	15%	18	69%	8	31%
March 2015	6	67%	3	33%	8	89%	1	11%
December 2015	25	86%	4	14%	24	83%	5	17%
March 2016	9	69%	4	31%	11	85%	2	15%
October 2016	27	90%	3	10%	18	60%	12	40%
February 2017	10	83%	2	17%	8	67%	4	33%
October 2017	29	94%	2	6%	22	67%	11	33%
February 2018	14	88%	2	13%	13	81%	3	19%
February 2019	7	78%	2	22%	6	67%	3	33%

CMG: Canadian medical graduate; IMG: international medical graduate; OSCE: objective structured clinical examination.

Supplementary Table 3. Comparison of scores by station type

Station type	Mean score		p (vs. uro-oncology)
Uro-oncology (n=306, N=109)	70.7	(0.9)	–
Pediatric (n=299, N=106)	72.1	(0.9)	0.277
Endourology (n=138, N=84)	70.9	(1.3)	0.939
Andrology (n=196, N=87)	67.1	(1.1)	0.010
Functional (n=288, N=103)	62.0	(0.9)	<0.001
Reconstructive (n=146, N=98)	70.9	(1.3)	0.931
VRE (n=306, N=109)	43.0	(0.9)	<0.001
Telephone (n=174, N=81)	74.1	(1.2)	0.021

Data are presented as least square means (standard error) obtained from linear mixed modeling with station type as fixed effect and individual resident as a random effect. P-values are for comparisons of scores to uro-oncology stations as reference. A lowercase n represents the number of observations, while an uppercase N represents the number of residents from which the observations are drawn. VRE: visual recognition examination.