Use of In-Training Examination Scores as a Fellowship Candidate Evaluation Metric: Time for a Change

Abstract In the field of obstetrics and gynecology (OB/GYN), the Council on Resident Education in Obstetrics and Gynecology (CREOG) administers an annual in-training examination to all OB/GYN residents as a formative educational tool for assessing medical knowledge and promoting self-improvement. Although the CREOG examination is not designed or intended for knowledge certification, many OB/GYN subspecialty fellowship programs request and use CREOG examination scores as a metric to evaluate fellowship candidates. Among the 57 gynecology-based urogynecology fellowship programs, 30 programs (53%) request CREOG examination scores to be submitted by candidates, as of March 2023. Although the use of CREOG examination scores as an evaluation metric may constitute a minor component within the fellowship match process, this practice fundamentally contradicts the intended purpose of the examination as an educational self-assessment. In addition, it introduces the potential for bias in fellowship recruitment, lacks psychometric validity in predicting specialty board examination failure, and shifts the CREOG examination from its original intention as low-stakes self-assessment into a high-stakes examination akin to a certification examination. For these reasons, we call upon the urogynecology community to prioritize the educational mission of the CREOG examination and reconsider the practice of requesting or using CREOG examination scores in the fellowship match progress.

Urogynecology 2024;30:394–398

DOI: 10.1097/SPV.0000000000001489

Christopher X. Hong, MD, Colin B. Russell, MD, MPH, Elizabeth A. Southworth, MD, and Pamela S. Fairchild, MD

Author affiliations, Conflicts of Interest, and article information are provided at the end of this article.

COMMENTARY

Obstetrics and gynecology (OB/GYN) residents who seek subspecialty fellowship training are frequently counseled to pursue certain goals that can bolster their competitiveness. These goals typically include engaging in scholarly activities, obtaining strong letters of recommendation, and performing well on "the CREOGs" examination. The pressure to achieve high scores on this examination is evident in the proliferation of paid subscription test preparation services and the popularity of not-for-profit initiatives like "CREOGS Over Coffee," a study podcast designed specifically for Council on Resident Education in Obstetrics and Gynecology (CREOG) examination preparation.[1] Given the importance of this examination to many trainees, it is crucial to understand its intended purpose, its current role as a metric for evaluating candidates

Copyright © 2024 American Urogynecologic Society. Unauthorized reproduction of this article is prohibited.

for OB/GYN fellowships, and the potential implications of this practice to equity and inclusion in our specialty.

What Is "the CREOGs" Examination, and What Is Its Purpose?

In OB/GYN, the CREOG administers an annual in-training examination (ITE) to all residents in OB/GYN residency training programs that are accredited by the Accreditation Council for Graduate Medical Education, colloquially referred to as "the CREOGs." The CREOG examination, established in 1970, was designed as an educational tool to assess medical knowledge and facilitate self-improvement in OB/GYN training. It has continued to be an essential component of residency training programs in the United States.[2]

Although the examination has undergone changes in its structure and content, its fundamental purpose as a formative rather than summative assessment has remained. Dr Harrison Visscher, a former director of education for the American College of Obstetricians and Gynecologists, succinctly summarized this intention, stating "The objectives of the examination are to enable directors of residency programs in obstetrics and gynecology to evaluate the cognitive knowledge of their individual residents and their respective educational programs. A secondary objective is to assist in their self-evaluation of faculty members and private practitioners who serve as teachers. The emphasis is on education, not certification."[3] This educational mission distinguishes the CREOG examination from certifying examinations administered by the United States Medical Licensing Exam and the American Board of Obstetrics and Gynecology (ABOG). Simply stated, the CREOG examination is not designed or intended for knowledge certification. Performance reports are distributed to residents and residency program directors as a formative reference with teaching points for further education. They are not intended for external parties to compare knowledge acquisition among residents from different residency training programs.

How Is the CREOG ITE Used Today?

Despite the original purpose of the CREOG examination, examination scores are now commonly used by OB/GYN fellowships for candidate evaluation. Fellowship training programs often request submission of CREOG examination scores via email to program coordinators in addition to the standardized Electronic Residency Application Service application. At present, there is considerable variation in this practice among different OB/GYN subspecialties and between fellowship programs within each subspecialty, and the prevalence in the request of CREOG examination scores has not been examined.

To assess the prevalence of this practice in urogynecology and reconstructive pelvic surgery, we performed a cross-sectional assessment on the 57 OB/GYN-based urogynecology fellowship programs accredited by the Accreditation Council for Graduate Medical Education. We searched program websites for in-training score submission policies, and for programs without a clear policy, we contacted program coordinators to obtain further clarity. Through this process, we acquired CREOG score submission policies from all 57 fellowship programs.

We found that, as of March 2023, 30 programs (53%) request CREOG examination scores to be submitted by applicants, whereas 27 programs (47%) do not have a requirement. Among the 30 programs that request CREOG examination scores to be submitted, 22 (73%) considered their request to be mandatory, meaning that candidates who do not submit scores are not considered for an interview invitation.

There are likely several factors that contribute to this practice. First, given the competitive nature of subspecialty fellowship matching, programs may seek additional objective metrics to aid in candidate selection for interviews and in creating their rank-match lists. Second, programs may use CREOG examination scores as a means to identify candidates at risk of failing the ABOG written examination based on prior research that proposed a score threshold of 200—below which candidates are thought to be at higher risk of failure.[4] The failure of a fellow to pass the ABOG written examination could have heightened significance for fellowship programs with institutional policies for board eligibility and certification, particularly when fellows are also anticipated to function as independent providers in a faculty capacity.

Why Should Fellowship Programs That Use CREOG ITE Scores as a Candidate Evaluation Metric Reconsider Their Use?

In the absence of national guidelines on the appropriate use of CREOG ITE scores, we acknowledge the autonomy of individual fellowship programs to create their own evaluation criteria for fellowship candidates. We recognize that the authority to provide guidance

Copyright © 2024 American Urogynecologic Society. Unauthorized reproduction of this article is prohibited.

does not rest with us but with governing organizations. However, from a trainee perspective, there are concerns associated with the use of CREOG examination scores in fellowship candidate evaluation.

Biases and Inequity

It is important for us to be aware of known biases in medical recruitment practices, which may place underrepresented minorities at an disadvantage.[5] The use of CREOG examination scores as an evaluation metric may introduce and perpetuate biases in trainee recruitment, as has been observed with standardized tests in medicine such as the United States Medical Licensing Examination.[5] The resources available to residents for CREOG examination preparation may not be distributed equitably among individuals of different socioeconomic and training backgrounds. As trainees, we have observed this firsthand. Placing significant emphasis on score results may create undue pressure for residents seeking fellowship training to invest more financial resources in test preparation services. Consequently, this practice may preferentially benefit residents who possess the personal resources to engage in extra preparation and those in residency training programs that can support additional study time and resources. While the socioeconomic circumstances faced by disadvantaged groups have been recognized in other types of standardized testing and can be considered as part of holistic recruitment practices, it is currently not feasible to account for differences in resources among various residency programs. Further research is needed to investigate potential biases in fellowship recruitment that may result from use of CREOG exam scores without accounting for disparities in resource access. Nevertheless, it would be naive to regard our specialty as immune from problems of bias and inequality associated with use of exam scores in many other recruitment practices.

Psychometric Validity

Furthermore, the use of a CREOG examination score threshold to identify residents at risk of failing the ABOG qualifying examination, also known as the "written" portion of the certifying process, lacks psychometric validity. A study by Withiam-Leitch and Olawaiye[2] revealed a weak correlation ($r = 0.38$) between CREOG examination scores and ABOG qualifying examination scores. This study concluded that, like ITEs in other specialties, the psychometric qualities of the CREOG examination do not fulfill the criteria necessary for it to be considered a valid predictor of performance on the ABOG qualifying examination.[2] Although Withaim-Leitch and Olawaiye[2] and other studies have found that achieving a high score on the CREOG examination is predictive of success on the ABOG written examination, the reverse is not true.[6] The most extensive and recent study investigating the association between CREOG examination scores and ABOG written examination outcomes found that a score exceeding an optimal cutoff value of 204 in postgraduate year 3 was associated with a strong positive predictive value for passing the ABOG written examination (100%; 95% confidence interval, 90–100%).[4] However, the negative predictive value was weak (38%; 95% confidence interval, 22–56%). It is crucial to interpret this study in the context of its population, which consisted of only 80 residents from 2 institutions, and the study period of 2002 to 2012, during which first-time pass rates for the ABOG written examination were in the low 80%. Given the subsequent increase in first-time pass rates to more than 90%, it is anticipated that the negative predictive value of comparable score thresholds would be even lower than previously reported. Those who place a high value on the future ABOG examination success of their fellows can find reassurance in the fact that more than 90% of residents taking the ABOG written examination will pass it on their first attempt, including the large majority of residents who did not meet previously suggested CREOG score thresholds.

Indeed, a systematic review of ITEs from multiple specialties concluded that weak performance on ITEs does not necessarily indicate risk of failure on subsequent board examinations.[7] The reviewers advised caution and restraint in using ITE scores for making important decisions. Given the recent change in CREOG examination scoring from a mean-based score centered on 200 to an equated percentage correct score out of 100, this is an ideal time to cease using unsubstantiated score thresholds.

Testing Conditions and Cheating

It is also important to recognize that the current high-stakes nature of the CREOG examination may induce examinees to consider cheating because of the intense pressure to achieve a high score, undermining the examination's original intent as a self-assessment tool. Although proctors are assigned to supervise each examination session, the examination can be taken on personal computers, conducted virtually, or involve breaks with less stringent oversight compared with certifying examination testing environments. The

Copyright © 2024 American Urogynecologic Society. Unauthorized reproduction of this article is prohibited.

testing conditions for the CREOG examination are not intended for certification and lack robust measures to discourage or identify cheating. The potential for cheating and its relative ease is another compelling reason why CREOG examination scores should not be considered a reliable metric for candidate evaluation or used by outside parties for direct comparisons between examinees.

Examination Intent and Purpose

The current use of CREOG examination scores in evaluating fellowship candidates has reshaped the examination from a low-stakes, formative self-assessment tool to a high-stakes examination resembling a certification examination. Although some may note that CREOG examination scores are just 1 factor among many in candidates' evaluations, any use of CREOG examination scores as an evaluative or comparative metric contradicts the intended purpose of the examination.

We acknowledge concerns regarding potential disadvantages faced by residents from programs with limited opportunities for performing research and obtaining strong letters of recommendation. In view of some, a strong performance on the CREOG examination might be the primary means to enhance a fellowship application when other opportunities are lacking. We recognize the disparities in resource access that traditionally serve as indicators of a strong fellowship candidate. However, a solution should revolve around ensuring more equitable access to opportunities and considering an individual's training environment during the evaluation process, rather than continuing with the use of CREOG examination scores, which, in our view, likely perpetuates other biases.

How Have Requests for CREOG ITE Scores Affected the Urogynecology Fellowship Application Process?

Irrespective of one's opinion on the use of CREOG examination scores, the variation in program-specific policies on the use of CREOG examination scores and the requirements for score submission has introduced complexity and heterogeneity into the fellowship application process. Although this may not be evident to individual training programs, it is readily apparent to any recent urogynecology fellowship applicant seeking to collect data across numerous training programs.

Although fellowship programs often provide information about application requirements on their websites, we found that only 40% of fellowship program websites included details on their CREOG examination score submission policy. Obtaining further clarification on policies required multiple email exchanges with program coordinators for more than half of programs. Notably, the absence of a score submission requirement on a fellowship program's website does not necessarily indicate that scores are not required; we found that 10 programs with a policy of requesting scores did not include CREOG examination scores as part of the listed fellowship application requirements on their respective websites.

Conversely, the inclusion of CREOG examination scores as an application requirement does not necessarily indicate that submission is mandatory, although applicants would be unlikely to decline requests. Because there is no centralized repository of information on CREOG examination score submission policies, candidates must individually search for policies on program websites and reach out to programs for clarification, which places a burden on candidates and fellowship coordinators alike. The current variation and incomplete disclosure of CREOG examination score policies among urogynecology fellowships highlight a need and opportunity for improvement.

Why Should We, the Urogynecology Community, Reevaluate Our Use of CREOG ITE Scores in Fellowship Candidate Evaluation?

In urogynecology, we are well positioned to have more in-depth discussions about using CREOG examination scores for fellowship candidate evaluations and reevaluating our practices. Given the current division among urogynecology fellowship programs on the use of scores, initiating conversations among peers and programs with differing approaches will aid our collective understanding of the diverse perspectives that influence this practice within the urogynecology community.

As we engage in these conversations, we should also consider the practices of our colleagues in urology-based urogynecology fellowship training programs. Like the CREOG, the American Urological Association sponsors a urology in-service examination with the purpose to "assist residency program directors, faculty, and residents in evaluating the strengths and weaknesses of the education of their training program, as well as the performance and progress of individual residents. There is no pass or fail level for this examination in contradistinction to the Qualifying Examination (Part I) of the American Board of Urology."[8] In line with this ethos, the urology in-service examination is widely recognized as an educational tool, and

Copyright © 2024 American Urogynecologic Society. Unauthorized reproduction of this article is prohibited.

examination scores are not requested from urology residents seeking urogynecology fellowships.

Although the prevalence in request of CREOG examination scores has not been thoroughly evaluated in other OB/GYN subspecialties, recent fellowship candidates in other subspecialities anecdotally report that it is significantly higher than the 53% observed among OB/GYN-based urogynecology fellowship programs. We believe that the urogynecology community has the capacity to uphold its leadership and set a precedent for all OB/GYN subspecialties in respecting the intent and purpose of residency ITEs as an educational resource, rather than using them as a metric for evaluating fellowship candidates.

Call to Action

From our perspective, which is informed by our experiences as applicants and evidence in literature, the use of CREOG examination scores for assessing fellowship candidates in urogynecology and other OB/GYN subspecialties raises significant concerns. The potential bias in fellowship recruitment, lack of psychometric validity in predicting future ABOG examination failure, and the transformation of this educational tool into a high-stakes evaluation metric all underscore the need to reassess this practice.

Although our urogynecology fellowships continue to grow, it is crucial that we also continue to evaluate our approach to supporting residents who aspire to pursue urogynecology subspecialty training. Although the use of CREOG examination scores for evaluating fellowship candidates may constitute a minor component within the fellowship match process, this practice fundamentally contradicts the intended purpose of the examination as an educational self-assessment. Without change, OB/GYN residents who seek subspecialty training will continue to face undue pressure to perform well on the CREOG examination, on top of the already immense demands and responsibilities of their residency training.

As members of the urogynecology community, we have the opportunity to lead by example and prioritize the educational mission of the CREOG examination over its use as an unsubstantiated evaluation metric for future success in urogynecology. Let us commit to reevaluating our practices and, where necessary, reforming them to align with the true spirit of medical education.

ARTICLE INFORMATION

From the Department of Obstetrics and Gynecology, University of Michigan, Ann Arbor, MI.

Correspondence: Christopher X. Hong, MD. E-mail: cxhong@med.umich.edu.

The authors have declared they have no conflicts of interest.

© 2024 American Urogynecologic Society. All rights reserved.

REFERENCES

1. Cai F, Burns RN, Kelly B, et al. CREOGs over coffee: feasibility of an ob-gyn medical education podcast by residents. *J Grad Med Educ*. 2020;12(3):340–343. doi:10.4300/JGME-D-19-00644.1.
2. Withiam-Leitch M, Olawaiye A. Resident performance on the in-training and board examinations in obstetrics and gynecology: implications for the ACGME outcome project. *Teach Learn Med*. 2008;20(2):136–142. doi:10.1080/10401330801991642.
3. Visscher HC. Validity and purpose of in-training examination in obstetrics and gynecology, 1979–1982. *Obstet Gynecol*. Feb 1984;63(2):253–259.
4. Lingenfelter BM, Jiang X, Schnatz PF, et al. CREOG in-training examination results: contemporary use to predict ABOG written examination outcomes. *J Grad Med Educ*. 2016;8(3):353–357. doi:10.4300/jgme-d-15-00408.1.
5. Yemane L, Omoruyi E. Underrepresented in medicine in graduate medical education: historical trends, bias, and recruitment practices. *Curr Probl Pediatr Adolesc Health Care*. Oct 2021;51(10):101088. doi:10.1016/j.cppeds.2021.101088.
6. Spellacy WN, Carlan SJ, McCarthy JM, et al. Prediction of ABOG written examination performance from the third-year CREOG in-training examination results. *J Reprod Med*. 2006;51(8):621–622.
7. McCrary HC, Colbert-Getz JM, Poss WB, et al. A systematic review of the relationship between in-training examination scores and specialty board examination scores. *J Grad Med Educ*. 2021;13(1):43–57. doi:10.4300/jgme-d-20-00111.1.
8. American Urological Association. In-Service Exam (ISE)/Oncology Knowledge Assessment Test (OKAT). 2023. Accessed October 19, 2023. https://auau.auanet.org/content/in-service-exam

Copyright © 2024 American Urogynecologic Society. Unauthorized reproduction of this article is prohibited.