Guiding Residency Program Educational Goals Using Institutional Keyword Reports from the Ophthalmic Knowledge Assessment Program Examination

Isdin Oke, MD[1]   Steven D. Ness, MD[1]   Jean E. Ramsey, MD, MPH[1]   Nicole H. Siegel, MD[1]
Crandall E. Peeler, MD[1]

[1] Department of Ophthalmology, Boston Medical Center, Boston University School of Medicine, Boston, Massachusetts

Address for correspondence Isdin Oke, MD, Department of Ophthalmology, Boston Medical Center, 850 Harrison Ave 3rd Floor, Boston, MA 02118 (e-mail: isdinoke@gmail.com).

Abstract

Introduction Residency programs receive an institutional keyword report following the annual Ophthalmic Knowledge Assessment Program (OKAP) examination containing the raw number of incorrectly answered questions. Programs would benefit from a method to compare relative performance between subspecialty sections. We propose a technique of normalizing the keyword report to determine relative subspecialty strengths and weaknesses in trainee performance.

Methods We retrospectively reviewed our institutional keyword reports from 2017 to 2019. We normalized the percentage of correctly answered questions for each postgraduate year (PGY) level by dividing the percent of correctly answered questions for each subspecialty by the percent correct across all subsections for that PGY level. We repeated this calculation for each PGY level in each subsection for each calendar year of analysis.

Results There was a statistically significant difference in mean performance between the subspecialty sections ($p = 0.038$). We found above average performance in the Uveitis and Ocular Inflammation section (95% confidence interval [CI]: 1.02–1.18) and high variability of performance in the Clinical Optics section (95% CI: 0.76–1.34).

Discussion The OKAP institutional keyword reports are extremely valuable for residency program self-evaluation. Performance normalized for PGY level and test year can reveal insightful trends into the relative strengths and weaknesses of trainee knowledge and guide data-driven curriculum improvement.

Keywords
► Ophthalmic Knowledge Assessment Program
► institution keyword reports
► residency program
► didactics
► education
► curriculum

The Ophthalmic Knowledge Assessment Program (OKAP) is an annual examination completed by ophthalmology residents across North America and many other parts of the world.[1–3] Each participant receives a performance report several weeks after the examination.[4] Overall performance is classified by the cognitive domain and subspecialty section of each question.[5] The cognitive domains include three categories: *Recall, Interpretive,* and *Decision-Making/Clinical Management.* Subspecialty sections correspond to the 13 volumes of the Basic Clinical Science Course (BCSC), the comprehensive curriculum from the American Academy of Ophthalmology (AAO) from which questions for the OKAP are derived.[6] Scaled scores and percentile ranks are reported.[5] Scaled scores indicate how many standard deviations above or below average a resident performs compared with all test-takers that year, regardless of training level. Percentile ranks indicate the percentage of other examinees at the same training level who score below the resident.

Each residency Program Director also receives a similar cognitive domain and keyword report that summarizes the cumulative performance of residents within their program. The cognitive domain report includes a scaled score for the program as a whole. The keyword report shows only the raw number of questions answered incorrectly, broken down by postgraduate year (PGY) level and subspecialty. The OKAP User's Guide encourages residency programs to use this information to identify program-wide gaps in knowledge.[5] However, the performance between different trainee levels within a program varies with years of clinical experience, and the performance between different years of the test fluctuates with test difficulty. The keyword report does not provide an intuitive approach for assessing the relative performance between the subspecialty sections for a given residency program which can make it difficult to interpret. In this study, we propose a method to analyze the institutional keyword report to identify relative strengths and weaknesses in trainee exam performance between subspecialty sections and best guide future educational curriculum development.

Methods

In this study, we retrospectively reviewed Boston Medical Center's keyword reports from 2017 to 2019. We did not include reports from earlier years because the structure of the OKAP exam, including the number and naming of subspecialty sections, changed between the 2016 and 2017 test years. We also focused our review on this time period because it included the most recent test years without significant changes in our didactic curriculum. All 12 residents over the three PGY training levels completed the OKAP in 2018 and 2019, while 11 completed the test in 2017. This quality improvement project did not involve the use of patient information and did not require approval from our Institutional Review Board.

Normalized Performance
To analyze the OKAP institutional keyword report, we sought to normalize the raw scores for each PGY level for each test year. For each PGY level and subspecialty section of the keyword report, we tallied the incorrect responses (►**Fig. 1**, red box) and calculated the percentage of correct answers. We then normalized the percentage of correctly answered questions by:

$$P = \frac{C}{\bar{C}}$$

Where P is the normalized performance such that $P = 1$ represents average performance across all subspecialties, $P > 1$ represents above average performance, and $P < 1$ represents below average performance. C is the percent of correctly answered questions by a PGY level for a given subspecialty and $\bar{C}$ (with a line on the top) is the mean percent correct across all subspecialties for that PGY level. We repeated this calculation for each PGY level and calendar year of analysis. We also calculated the breakdown of cognitive domains (►**Fig. 1**, blue box) and individual keywords (►**Fig. 1**, yellow box) for incorrectly answered questions as these metrics can be used to guide specific interventions if any outlying subspecialty sections are identified. We also combined the normalized performance scores for all PGY levels in all testing years to identify program-wide trends in subspecialty performance over the study period.

Statistics
Statistical analysis was performed using R Studio, version 1.2.1335 (RStudio, Inc., Boston). We performed one-way analysis of variance to assess for statistically significant difference in subspecialty performance. Post-hoc analysis was performed using the Tukey–Kramer method. We report 95% confidence intervals (CI). Statistical significance was defined as $p < 0.05$.

Results

Our institution's normalized performance in each subspecialty section for the 2017 to 2019 study period is shown in ►**Fig. 2**. There was a statistically significant difference in the normalized performance between all subspecialties ($p = 0.038$). We found above average performance in the Uveitis and Ocular Inflammation section (95% CI: 1.02–1.18) that was statistically significant ($p = 0.031$). Though performance in the remaining sections did not differ significantly from the mean, our analysis allowed us to visualize above average, average, and below

| Cognitive Level | Keyword/Topic | Number of residents who answered incorrectly | | | | |
| | | Training Level | | | | |
		1	2	3	4	TOTAL
	FUNDAMENTALS AND PRINCIPLES OF OPHTHALMOLOGY					
II	Orbit and Ocular Adnexa	2	3	4	0	9
I	Bowman Layer	4	3	2	0	9
I	Ocular Pharmacotherapeutics; Diagnostic Agents	3	2	1	0	6
II	Orbital Anatomy	3	0	3	0	6
I	Eye Development	2	0	3	0	5

Fig. 1 Annotated excerpt from Ophthalmic Knowledge Assessment Program institutional keyword report from the Fundamentals and Principles of Ophthalmology section illustrating the components involved in the analysis including number of incorrectly answered questions per postgraduate year (red box), cognitive domain category (blue box), and keyword topics (yellow box).
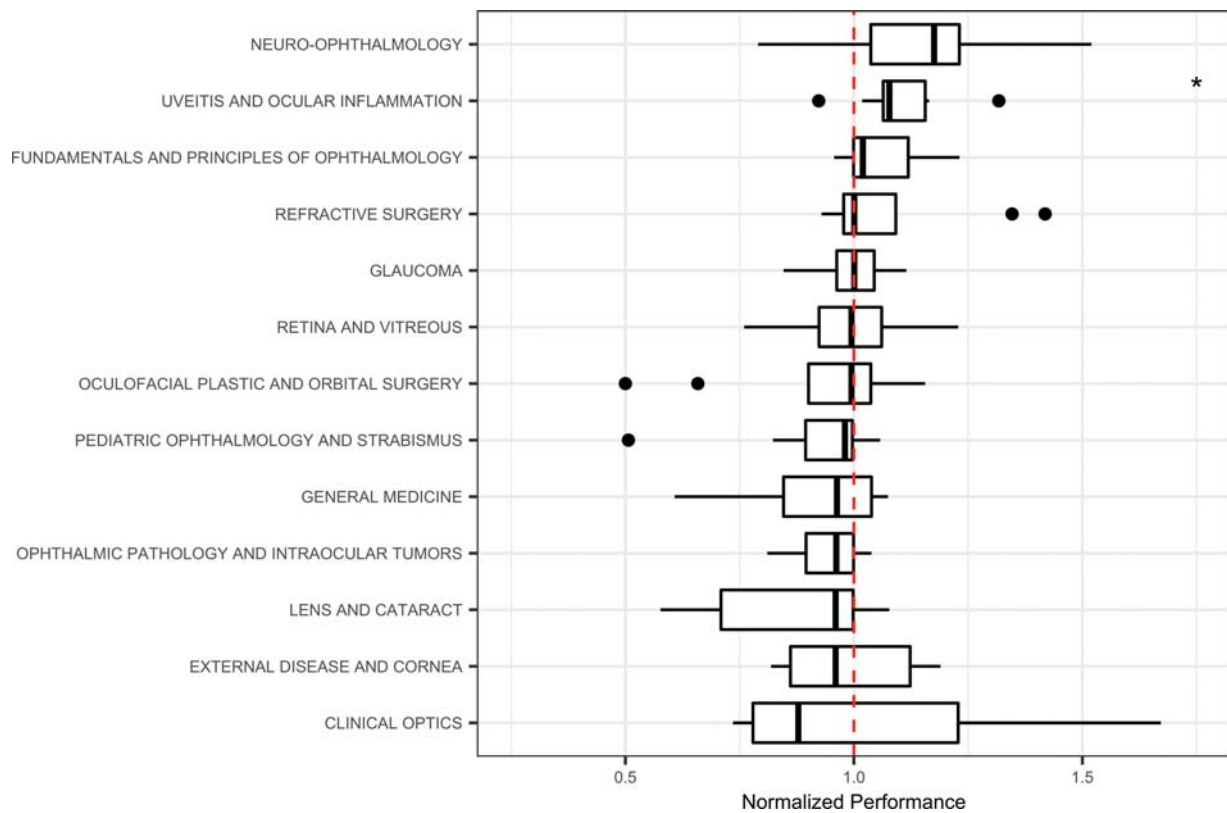
Fig. 2 Box plot comparing relative performance between different subspecialties in our residency program during the 2017, 2018, and 2019 exam years (*: $p < 0.05$). Black dots represent outliers (1.5x the interquartile range above the upper quartile and below the lower quartile)

average performance across the other subspecialties. Sections with above average performance included Neuro-Ophthalmology (95% CI: 0.99–1.32) and Fundamentals of Ophthalmology (95% CI: 0.99–1.14). Sections with average performance included Refractive Surgery (95% CI: 0.94–1.22), Glaucoma (95% CI: 0.93–1.07), Retina and Vitreous (95% CI: 0.90–1.10), and Oculofacial Plastic and Orbital Surgery (95% CI: 0.76–1.09). Sections with below average performance included Pediatric Ophthalmology (95% CI: 0.79–1.05), General Medicine (95% CI: 0.79–1.04), External Disease and Cornea (95% CI: 0.88–1.10), Ophthalmic Pathology and Intraocular Tumors (95% CI: 0.88–1.00), and Lens and Cataract (95% CI: 0.72–1.02). The Clinical Optics section (95% CI: 0.76–1.34) was found to have both the lowest median performance and the largest range in performance.

The cognitive domain distribution for incorrectly answered questions in each subsection is shown in ►**Fig. 3**. The section with greatest percentage of incorrect answers in the *Recall* domain was Fundamentals of Ophthalmology (70.5%). The Ophthalmic Pathology and Intraocular Tumors section had the highest rate of incorrect answers in the *Interpretive* domain (52.9%) and the Lens and Cataract Section had the highest rate of incorrect answers in the *Decision-Making/Clinical Management domain* (34.0%).

Discussion

Institutional keyword reports contain valuable information on OKAP exam performance of trainees within a residency program. Understanding performance patterns can allow programs to design data-driven curriculum changes to address relative weaknesses in specific subspecialty knowledge. Similarly, an appreciation of why certain subspecialties consistently rank well within a program may reveal educational practices worth exploring and applying to other subspecialties. While our specific calculations for relative performance are not generalizable to other institutions, the technique may be universally applied to provide residency programs with institution specific insight.

The primary benefit of this information is that it allows residency programs to design educational initiatives to meet medical knowledge-based ophthalmology milestones.[7] For example, the relative quantity and distribution of subspecialty didactics through the academic year could be adjusted based on an annual assessment of the keyword report. Using our institution's reports, we were able to identify below average performance in the Clinical Optics section (►**Fig. 2**). Certain exam sections, Clinical Optics in particular, require the memorization of formulas that are not otherwise used routinely in a clinical setting. Preparation efforts for these sections may benefit from additional review sessions closer to the date of the exam. Similarly, sections with a strong emphasis on the cognitive domain *Recall* may benefit from increased didactic sessions through the academic year with greater focus on the BCSC curriculum from which test questions are derived. In contrast, the cognitive domains *Interpretive* and *Decision-Making/Clinical Management* may benefit most from increased educational
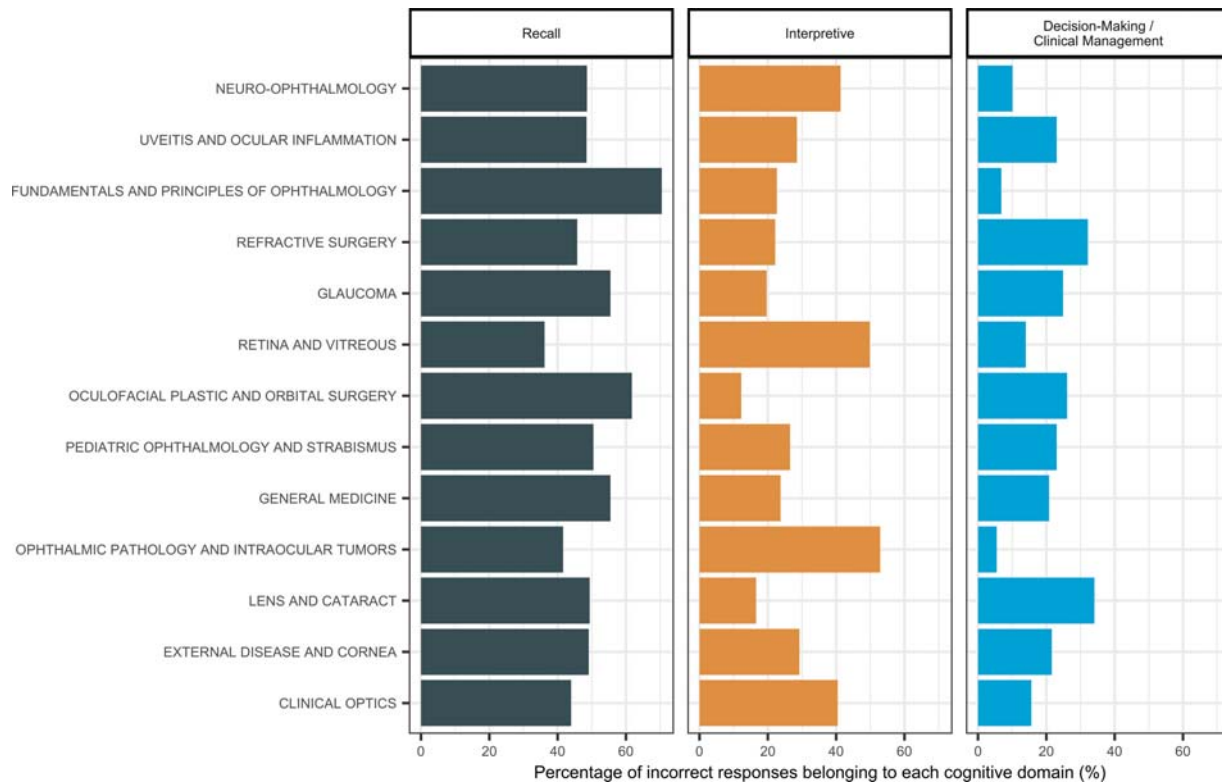
Fig. 3 Bar plots showing the cognitive domain of incorrectly answered questions in each subspecialty section. Percent of incorrect responses is calculated for each subspecialty as the number of incorrect responses in a given domain divided by total number of incorrect responses. Cognitive domains include *Recall (I)*, *Interpretive (II)*, and *Decision-Making/Clinical Management (III)*.

initiatives in a clinical setting. Potential interventions include adjusting resident rotation schedules to optimize subspecialty service exposure to address any relative weaknesses identified by this analysis. The specific keywords (►**Fig. 1**, yellow box) provide an excellent starting point for specific subjects that could be covered during a potential intervention.

There are many advantages to analyzing OKAP performance using a normalized approach. First, the method involves retrospective analysis of the institutional keyword reports that each residency program participating in the OKAP receives annually. Second, normalization across PGY level and test year allows programs to compare performance of all residents within an institution without the bias of years of clinical training or variability in test difficulty from year to year. Third, this approach allows for further subgroup analysis into specific test years or PGY levels. Access to this information can alert a program and allow for earlier intervention with targeted didactics or clinical rotations. In addition, analyzing keyword reports before and after an educational intervention can provide an objective way to quantify the impact of the intervention. Finally, the anonymity of the report analysis is an important benefit not to be overlooked. Not only can this method be performed without risk of loss of confidentiality of individual test scores but also the normalized performance of a residency program can be compared between institutions without revealing raw program performance. Sharing of this information may be particularly helpful in the design of interinstitutional didactic curricula.

There are also several limitations of this approach and reasons to carefully interpret the results. First, since the number of categories and subsection names in the OKAP exam changed between 2016 and 2017, we are not able to combine and collectively analyze keyword reports from before 2016 with reports from 2017 onwards. Second, smaller residency programs may have increased difficulty detecting patterns given greater fluctuation in individual performance associated with fewer trainees. Wide confidence intervals due to the presence of outliers could result in a subspecialty area with high variability in performance. Variability may be seen in subspecialties with high testing uncertainty characterized by an increased percentage of guessed answer choices in the multiple-choice exam. Both high- and low-scoring outliers can affect the interpretation of mean program performance and thus programs may consider further subgroup analysis and recomputing program averages after excluding certain outliers. Third, many factors besides institutional didactic strength are involved in test-taking performance including individual test-taking abilities and residents with English as a second language. There is also some degree of overlap between the cognitive domains defined in the OKAP user manual.[5] *Recall* questions measure an examinee's command of facts, concepts, and principles procedures, *Interpretive* questions measure abstraction of facts to identify implication, make inferences and predictions, and *Decision-Making/Clinical Management* questions measure problem solving ability in recalling relevant knowledge to make appropriate decisions about diagnosis and treatment. Not all subspecialty sections have an equal distribution of questions from these

three domains, which must also be taken into consideration when comparing the relative performance in each section. Finally, normalized performance is institution specific and does not reflect performance compared with the national average. Absence of difference between the subspecialty sections could correspond to either stellar performance or need for improvement across all categories and therefore should be interpreted in the context of the cumulative score report.

Residency programs can take advantage of the valuable cumulative data of their trainees to set program educational objectives and guide curriculum changes just as individual participants can use the performance report of the annual exam to guide their future study goals and plans. Performance on the OKAP examination has been associated with performance on the American Board of Ophthalmology licensing examinations, and OKAP scores are frequently used as criteria in fellowship applications.[8–11] We hope this method will serve as a valuable tool to for residency program self-evaluation and data-driven curriculum improvement to maximize resident success and ensure a broad, well-rounded curriculum.

Conflict of Interest
None declared.

References

1 Liesegang TJ. New directions for the Ophthalmic Knowledge Assessment Program (OKAP) examination. Ophthalmology 1994;101(01): 194–198
2 O'Day DM. Examinations and ophthalmic education. Ophthalmology 2012;119(10):1947–1948
3 Zafar S, Wang X, Srikumaran D, et al. Resident and program characteristics that impact performance on the Ophthalmic Knowledge Assessment Program (OKAP). BMC Med Educ 2019; 19(01):190
4 Chen PP. Resident OKAP performance. Ophthalmology 2012;119 (03):656
5 Rapuano C, Stout T. Ophthalmic Knowledge Assessment Program Committee. OKAP User's Guide . Available at: https://www.aao. org/okap-exam. 2019. Accessed September 15, 2020
6 Oke I, Siegel NH, Peeler CE, et al. Completing the basic and clinical science course as a first-year ophthalmology resident. J Acad Ophthalmol 2019;11:e54–e58
7 Arnold A. Developing the educational milestones for ophthalmology. J Grad Med Educ 2014;6(01, Suppl 1):144–145
8 Lee AG, Oetting TA, Blomquist PH, et al. A multicenter analysis of the ophthalmic knowledge assessment program and American Board of Ophthalmology written qualifying examination performance. Ophthalmology 2012;119(10):1949–1953
9 Carey A, Drucker M. Standardized Training Examinations among Ophthalmology Residents and the American Board of Ophthalmology Written Qualifying Examination First Attempt: The Morsani College of Medicine Experience. Journal of Academic Ophthalmology 2014;07:e008–e012
10 Kempton JE, Shields MB, Afshari NA, Dou W, Adelman RA. Fellow selection criteria. Ophthalmology 2009;116(05):1020-–1020.e4
11 Johnson GA, Bloom JN, Szczotka-Flynn L, Zauner D, Tomsak RL. A comparative study of resident performance on standardized training examinations and the American Board of Ophthalmology Written Examination. Ophthalmology 2010;117(12): 2435–2439