

OKAP™ International User's Guide



PO Box 7424

San Francisco, CA 94120-7424

Clinical Education Secretaries

Gregory L. Skuta, MD

Senior Secretary for Clinical Education

Louis B. Cantor, MD

Secretary for Ophthalmic Knowledge

Ophthalmic Knowledge Assessment Program Committee

J. Timothy Stout, MD, PhD Chair

Alice Song, MD

Andrew G. Lee, MD

Daniel A. Johnson, MD

Edmond H. Thall, MD

Evelyn A. Paysse, MD

Julia C. Lane Stevens, MD

Steven J. Gedde, MD

AAO Staff

Beth Wilson

Director of CME & Programs

Kathryn Peters

OKAP Program Manager

BACKGROUND ON THE OKAP™ AND OKAP™ INTERNATIONAL:

The American Academy of Ophthalmology (AAO) is an organization whose membership currently comprises 93% of US ophthalmologists, and is open to ophthalmology residents. In addition to US members, international ophthalmologists make up about 30% of AAO membership. The AAO provides advocacy, assistance with practice management, and clinical-education programs for the benefit of its members. The Ophthalmic Knowledge Assessment Program (OKAP™) examination is one of the AAO's clinical-education programs. The OKAP exam is a 250-item multiple choice test that is administered to ophthalmology residents in each year of training. It is designed to measure the ophthalmic knowledge of residents, relative to their peers, in order to facilitate the ongoing assessment of resident progress and program effectiveness. The OKAP™ International is the same test given to US residents in a prior year. Prometric, the AAO's testing administration vendor, provides assistance in the development of test items and with the psychometric analysis of item performance.

OKAP and OKAP International exam items are developed from a Content Outline that is based largely on the same topical areas as the AAO's Basic and Clinical Science Course (BCSC) books. The BCSC is a highly recommended curriculum of basic knowledge and new developments in ophthalmology, but it is not meant to be an exhaustive survey of the entire field. Therefore, preparation for the OKAP or OKAP International exam should also include other authoritative sources such as textbooks and scientific journals. The Content Outline is available upon request.

INTERPRETATION OF OKAP AND OKAP INTERNATIONAL RESULTS

Guidelines for understanding and interpreting OKAP and OKAP International results are provided in this section. Recommendations for using results to further resident education are given in the section following this one, called UTILIZATION OF OKAP AND OKAP INTERNATIONAL RESULTS.

OKAP International results are determined the same way as OKAP results, with the important distinction that international residents will also receive a percentile rank that incorporates scores from the United States, as well as their own institution. The information in this section applies to both the OKAP and OKAP International exams.

Each annual OKAP examination is carefully constructed to be a valid and reliable measure of ophthalmic knowledge. An OKAP score carries information about how well a resident has done relative to the other residents who have taken the exam. Such scores are "norm-referenced" in the sense that they refer to the norm of performance established by the other examinees—the same as being graded "on the curve".

Scores Are Confidential: The AAO releases OKAP results only to residents and their program directors, but scores also may be disseminated to other program faculty in support of their intended educational purpose. (Residents should be given the names of faculty members who have knowledge of their performance on the exam.) Residency programs must maintain all score reports.

Resident Score Report: Performance on the 250-item test as a whole, and on the 11 individual subtests, is reported as Scaled Scores and Percentile Ranks. These are simply two alternate ways of indicating how well one examinee performs relative to others. Developers of the OKAP have decided to use scaled scores to represent a resident's performance relative to all others taking the exam—regardless of training level—and to use percentile ranks to represent a resident's performance relative only to those at one's own level of training (Level 1 = First year, Level 2 = Second year, and Level 3 = Third and Fourth years). Scaled scores and percentile ranks are both derived directly from the "number-correct" or "raw" scores, which is simply equal to the number of items answered correctly.

The Scaled Scores transform a resident's raw scores on the OKAP, and all of its subtests, to a scale that indicates how many standard deviations the score falls above or below the average of all residents regardless of training level. To establish this common scale, common values are substituted for the means and standard deviations of the various raw-score distributions. Scaled scores computed for the test as a whole substitute 500 for the mean and 100 for the standard deviation. Scaled scores computed for the subtests substitute 50 for each subtest mean and 10 for each subtest standard deviation.

The total scaled score now uses a process called equating that enables comparison of scaled scores from year to year. An advantage of the scaled score is that it makes it immediately clear where the group mean is located (500 for total test and 50 for subtests) as well as how far away from the mean the score is, expressed in standard deviation units (100 and 10 points represent a standard deviation on the total test and subtest respectively). The utility of transforming raw scores to scaled scores is two-fold: Unlike raw scores, scaled

scores carry information about how one has performed relative to one's peers, and allows comparison across subtests of varying length and difficulty.

As an example, if a resident's raw score for the whole 250-item test is 170, the actual mean of the raw scores is 150, and the actual standard deviation is 25, then that resident's raw score exceeds the mean by 20 items or $20/25 = 0.80$ standard deviation units. This raw score can be expressed in terms of the mean and standard deviation as follows: $150 + 0.80(25) = 170$. The corresponding scaled score is computed by substituting 500 for the mean and 100 for the standard deviation in this expression: $500 + 0.80(100) = 580$. Or if a resident's raw score on a particular subtest is 16, the actual mean is 19, and the actual standard deviation is 5, then that resident's score is 3 points or $3/5 = 0.60$ standard deviation units below the mean and the corresponding scaled score is $50 - 0.60(10) = 44$.

In general, about two-thirds of all residents will score within one standard deviation of the mean, so that their scaled scores will be between 400 and 600 for the test as a whole and between 40 and 60 on the subtests. Residents scoring within that range are part of the large, average group of residents who perform in a roughly equivalent manner. Residents scoring above or below this range are noteworthy because they differ from this overall average group.

A scaled score for the entire 250-item OKAP exam is a very reliable assessment of one's ophthalmic knowledge, relative to the other residents who also took the same year's version of the OKAP. Therefore, scaled scores from repeated administrations of the OKAP can be used to track one's progress over the entire term of residency. In 2006 when new scaling began, the average of the scaled scores for the entire test at the various levels of training, and for all levels combined, fell within the following ranges:

Level one residents	=	430-435
Level two residents	=	520-525
Level three residents	=	540-550
All levels combined	=	500

Note that the largest change in the scaled scores historically occurs between the first and second years of training, probably reflecting the fact that all of the content areas in the exam may not have been covered by the training-program curriculum until the second year.

Beginning in 2006 OKAP scoring used an equating plan in order to enable a scoring scale that could be compared across OKAP examinations taken in different years. Previous to 2006, inferences could only be based on a single year's results. The mean in 2006 was set to 500 as a reference point as indicated in the previous paragraph. For 2012, the performance of U.S. residents declined slightly from 2011, with the overall mean calculated as 541. This could be reflective of the transition from a pencil and paper exam to the now computer based exam. There was a large gap between the performance of level 1 and level 2 residents in keeping with the findings of past years.

	Yr 1	Yr 2	Yr 3	Yr 4	Total
2006	434.12	524.43	535.36	601.52	499.98
2007	468.24	556.08	567.75	609.55	532.56
2008	478.72	557.38	572.82	636.93	537.66
2009	473.47	549.34	562.90	640.29	530.31
2010	470.05	559.30	565.61	630.71	533.11
2011	485.72	563.01	578.77	624.38	544.43
2012	486.51	563.52	568.38	632.14	541.39

Percentile Ranks

The Percentile Ranks indicate the percentage of examinees, at a resident's own training level, who scored below that resident. If, for example, a second-year resident had a percentile rank of 60, it would mean that his or her raw score was higher than the raw scores of 60 percent of the second-year residents taking the exam in that year. Third- and fourth-year residents are grouped in the same training level for computation of percentile ranks. Like scaled scores, percentile ranks are provided for the test as a whole and for all subtests. No percentile ranks are reported for "fellows", those examinees who had already graduated prior to taking that year's OKAP.

Subtest Scores: Reviewing subtest scores can be helpful in identifying specific learning needs. The OKAP subtests parallel the 12 volumes of the AAO's Basic and Clinical Science Course (BCSC), and the 12 sections of the OKAP Content Outline. A subtest is comprised of the OKAP test items that pertain to that topic. Occasionally, items pertain to more than one topic and are included in more

than one subtest. For this reason, the sum of the items in the subtests is greater than the 250-item total for the whole exam.

Subtest Reliability: The main core topics within the 11 subspecialties are assessed in every OKAP exam, but because each yearly exam is comprised of a different set of test items, not every clinical point can be covered in every year. For example, because of the sampling of content, the Optics and Refraction subtest may give more emphasis to spectacle lenses one year and to contact lenses the next. Thus, in addition to reflecting the level of one's ophthalmic knowledge, one's score also might reflect variations in the sequence of topics presented in different residency programs.

Because subtests necessarily must contain fewer items than the test as a whole, subtest scores are somewhat less reliable and must be interpreted with more caution than a score based on all 250 items—particularly if the subtest is one of the shorter ones or if the score is extreme. For example, a resident who scores extremely well on a particular subtest when first taking the OKAP might show some decline on that subtest in a subsequent examination, simply as an artifact of lower subtest reliability. Though not as reliable as the score for all 250 items, subtest scores do indicate areas in which more study may be needed. Because scaled scores and percentiles are both derived from raw scores, they both have precisely the same reliability characteristics. If there are concerns about subtest performance, further evaluation is suggested in those areas. The 12 sections of the BCSC provide study questions to assist in this process.

Institutional Score Report: Each program receives an institutional report that indicates the average scaled scores (total and subtest) for residents in the program. It is useful to apply the scaled score interpretation for individuals to this report. That is, it is reasonable to assume that all programs whose average falls within the range of one standard deviation above or below the mean are experiencing similar outcomes on the content being measured by that score (i.e., 400-600 for the total test and 40-60 for subtests).

Programs with an average score falling more than one standard deviation above or below the mean should examine their resident profiles to determine if performance is uniformly above or below average, and how it compares with results from the previous year. In some cases, the performance of a single unusual resident may skew your program average, especially in smaller programs. Under such circumstances, it would be a good idea to re-compute the program's average after excluding that "outlying" value.

Keyword Report: Each resident receives a report that provides the keywords from the OKAP Content Outline for each item that he or she missed on the exam. Program directors also receive copies of these reports. The purpose of these reports is to help residents identify areas for further study, develop specific learning objectives in those areas, and devise learning plans for meeting those objectives.

Cognitive Levels: Mastery of clinical and scientific knowledge is tremendously important in the practice of ophthalmology, but knowledge alone is not sufficient. Residents must also learn to use their knowledge to interpret the clinical facts in a particular case and to solve the clinical problems that arise. OKAP test items are classified according to these three levels of cognitive functioning:

Level 1: **Recall** items measure an examinee's command of the facts, concepts, principles, and procedures on which the field of ophthalmology is based. But these items require more than rote recall (merely remembering a textbook definition). They also assess accuracy, depth, and comprehension; fine distinctions among subtly different concepts; and errors of commission and omission in route procedures.

Example of Level 1 item:

Which of the following medications is the treatment of choice of adult chlamydial conjunctivitis?

- > A. Systemic tetracycline
- B. Systemic ciprofloxacin
- C. Systemic sulfisoxazole
- D. Topical erythromycin

Level 2: **Interpretation** items likewise require recall of facts, but also require examinees to abstract meaning from those facts, recognize their implications, make inferences or extrapolations, explain or predict findings, or identify abnormalities in lab results and diagnostic images.

Example of Level 2 item:

Which of the following choices is the correct spherocylindrical notation for combining the cylinders +2.00 x 120° and

-2.25 x 30°?

- > A. +2.00 with -4.25 x 30°
- B. -0.25 with -2.25 x 120°
- C. +2.00 with -0.50 x 120°
- D. -4.25 with +2.00 x 30°

An examinee probably has not seen this particular situation and so cannot use recall alone to determine the correct response.

Level 3: Decision Making & Clinical Management items require examinees to apply their ability to recall relevant knowledge and draw correct conclusions to the complex process of making decisions about differential diagnosis in unusual circumstances, and setting priorities among competing clinical considerations in case management and treatment.

Example of Level 3 item:

A 60-year-old woman has a meibomian gland carcinoma of the left lower eyelid which extends through the orbital septum into the anterior orbit. Appropriate management should consist of:

- A. radiation to the orbit.
- B. systemic chemotherapy.
- > C. orbital exenteration.
- D. cryotherapy.

Unless an examinee has received explicit instruction about a patient in this particular circumstance, a problem-solving process is necessary to respond correctly.

UTILIZATION OF OKAP AND OKAP INTERNATIONAL RESULTS

The primary purpose of the OKAP and OKAP International is to support the clinical and basic-science education of residents by providing them a way to assess their ophthalmic knowledge relative to their peers and to identify individual learning objectives. Both tests are also designed to provide program directors with information that is useful in the formative evaluation of resident progress and program effectiveness.

The following information can be applied to both the OKAP and OKAP International.

OKAP test results may be used as a measure of competency in ophthalmic knowledge in comparison with all residents and within a residency year peer group. However, OKAP test results should never be used as the sole gauge of a resident's overall ophthalmic proficiency. Consideration of other relevant factors such as clinical acumen, surgical skill, patient rapport, and case-management ability always must enter into the process of resident evaluation.

To make best use of OKAP results, residents are encouraged to follow the six steps set forth below. Those in their second year of training, or beyond, should process their results in this way for every OKAP they take and evaluate their progress from one year to the next.

1. **Prioritize subspecialty areas for further study** (Note that all subspecialty areas will require further study throughout residency. The following suggestions are not meant to exclude any subspecialties from further study, merely to identify where to begin.):
 - Rank order the OKAP subtest results, from low to high, by both scale score and percentile.
 - Group subtest scaled scores according to whether they are below 40, between 40 and 60, or above 60.
 - Group subtest percentile ranks according to whether they are below the 25th, between the 25th and 75th, or above the 75th.
 - Because scaled scores and percentiles have been used to represent two different comparison groups, the priorities established may not always be identical for both.
2. **Identify specific learning objectives:**
 - Review Keyword Report.

- List specific topics of questions missed under each subtest.
 - Within each subspecialty, target as specific learning objectives the major headings in the Content Outline under which these topics are listed.
3. **Identify educational resources for meeting learning objectives:**
 - Use the Master Index, Section Indexes, and lists of educational resources in the BCSC to begin identifying medical textbooks, AAO resources, and references to the scientific literature for each area.
 - Continue this process with the bibliographies and reference sections in the resources identified from the BCSC.
 4. **Structure a learning plan:**
 - Begin with the areas shown to need the most study in step 1.
 - Specify how and when the resources from step 3 will be used to meet the objectives specified in step 2.
 5. **Review the learning plan with program director:**
 - Discuss the logic underlying the learning plan.
 - Make changes and additions to the learning plan recommended by program director.
 - Obtain program director's recommendation for the most appropriate faculty members to help carry out the learning plan.
 6. **Work on the learning plan with appropriate faculty** (Divulging specific OKAP results to faculty is optional but not necessary.):
 - Discuss the learning plan.
 - Identify additional related topics and materials to study.
 - Meet regularly to discuss progress.
 - Regularly inform the program director about progress made in accomplishing steps called for in the learning plan.

Clearly, one-on-one dialogue between residents and faculty is an indispensable part of clinical education and should not be limited to discussion of the OKAP. But regular faculty review of progress on OKAP learning plans could be useful as a starting point for a more comprehensive assessment of a resident's changing educational needs. In fact, assigning a faculty member to help a lower-scoring resident create and complete an OKAP learning plan is likely to provide a much more substantial benefit than the all-too-common practice of simply requiring that resident to take a review course.

In addition to the assessment of residents, OKAP is also useful in supplying information about the residency program itself. Program directors should examine the rank orderings among subtests for all of their residents, to see if any interesting patterns are present. Attempting to understand why certain subtests consistently rank well in the program may point up educational practices and techniques that would be worth discussing with the entire faculty.

FREQUENTLY ASKED QUESTIONS ABOUT OKAP INTERNATIONAL

1. *Can I get a copy of my OKAP International scores?*

Requests for scores must be addressed directly to the candidate's residency program, who must maintain these scores. If the candidate's residency program cannot supply an OKAP score, copies from the 2012 administration can be requested from the AAO for a USD \$25 fee per score year requested.

2. *How are OKAP International exams scored?*

The "number-correct" or "raw" score for the OKAP International, from which scaled scores and percentiles are computed, is simply the number of test items answered correctly. No passing score exists for the OKAP International.

3. *Does any of the information provided in the score report indicate the percentage of test items answered correctly?*

No. As with any exam of this type, slight variations in difficulty exist from year to year, and among subtests in the same year. Because percent-correct scores are confounded by these arbitrary differences in difficulty, they cannot really be compared over time or across subtests. For scaled scores and percentiles, however, differences in difficulty are removed by norming each test on its own level of performance. Because they reflect the one's rank among their peers, scaled scores and percentiles are ideal for making comparisons over time and across subtests. (It is important to remember that percentiles indicate the percent of examinees with lower scores, not the percentage of items answered correctly.)

4. *Will the percentile and scaled score computed for the entire test equal the average of the corresponding subtest values?*

Not usually. One reason is that some items are counted in more than one subtest, depending on content, but all items can count only once in the total test. Another reason is that the specific numerical relationship between percentiles and raw scores, and between scaled scores and raw scores, can differ for every subtest. For example, correctly answering one more question might raise the percentile rank by 10 on one subtest, but only by 5 on another. Similarly, it is common for an identical change in the raw-scores for two subtests to result in different changes in the associated scaled scores. For percentiles, the change associated with a given raw score change can vary even within a single subtest, depending on where the raw score is located along the range from low to high. This occurs because scores tend to aggregate in a large group near the average. Adding one more item correct in that part of the range will raise the associated raw score above a greater percentage of examinees than would be the case nearer the ends of the range, resulting in a larger percentile increase, by definition. These scale inconsistencies means that neither the percentile nor the scaled score for the total test can be derived from the corresponding subtest values.

5. *Should I be concerned if my percentiles and scaled scores do not increase from year to year?*

More so if your scaled scores do not increase. The clear expectation is that a resident's ophthalmic knowledge will increase from year to year, and that OKAP International scores will reflect that increase. Because percentiles are used to represent OKAP International performance relative to one's own level of training, a percentile rank that stays the same over time simply means that your ophthalmic knowledge has increased at the same rate as the other OKAP International examinees at your level of training. But because scaled scores represent performance relative to all residents regardless of training level, a scaled scores that does not go up means that the OKAP International did not detect the expected increase in ophthalmic knowledge.

6. *Can my OKAP International scores validly be used in reaching decisions about my academic standing within my program?*

Yes, but only in conjunction with other information. The OKAP International is a valid and reliable instrument for measuring the ophthalmic knowledge of residents. But it is unwise to decide such an important question on the basis of a single use of any measuring device. Furthermore, there are many other factors contributing to resident competence that a test of ophthalmic knowledge does not assess.

7. *Who makes up the questions for the OKAP and OKAP International exams? Is there training involved?*

Test items for the OKAP and OKAP International are the responsibility of the Liaison Committees of the American Academy of Ophthalmology (AAO) and the American Board of Ophthalmology (ABO). The nineteen content expert members of these committees write exam items and the Chair decides which are suitable for the OKAP exam. Extensive item-writing training is required to generate valid and reliable test items that meet the committees' extremely high standards and is provided to all new members.

8. *What percentage of US ophthalmology residency programs participates in the OKAP exam? Is it only available for ACGME certified programs?*

The residents in all of the 117 US residency programs in ophthalmology, accredited by the Accreditation Council for Graduate Medical Education (ACGME), currently take the OKAP examination.

9. *How does Prometric obtain information about the training level to which an examinee belongs?*

Because Prometric does not link examinees from year to year, Prometric must depend on the training level the examinees select during their exam. Examinees often make mistakes in the training-level information they provide. One's training level is defined as one's year of ophthalmology residency (not PGY). If you are in your first year of ophthalmology residency, you are at training level 1, and so on.

10. *What can be done if a resident's percentile ranks are based on the wrong training level?*

If your percentile values are computed for an incorrect training level, ask your program director to write a letter to the AAO, verifying your correct training level and requesting that new percentiles be computed for you.

11. *Are any test items excluded from scoring due to aberrant item performance?*

Sometimes. Although all items have been extensively reviewed by content experts before administration, a few may not perform as expected. These items are sent to the chair of the AAO's OKAP Committee, who determines (sometimes with the help of other content experts) whether the answer key needs to be changed. In the past, a few of these items have been scored correct for all examinees.

Notice to Readers of the OKAP International User's Guide

The OKAP Committee of the American AAO of Ophthalmology has developed this *Guide* as an aid to residents and program faculty. We are very interested in obtaining as much feedback as possible from all users, and would greatly appreciate hearing your thoughts on how the *Guide* could be improved. Please communicate all comments and suggestions to:

Kathryn Peters , OKAP Program Manager
Clinical Education Division
American Academy of Ophthalmology
PO Box 7424
San Francisco, CA 94120-7424

Phone: +1 (415) 447-0201
Fax: +1 (415) 561-8557
E-mail: kpeters@aao.org