



Standard Setting Report for the Part II-PAM Examination *February 2010*

Background

Throughout its history, the National Board of Examiners in Optometry (NBEEO) has strived to develop and administer examinations that reflect the contemporary practice of optometry. As a result, the exams have undergone several transformations, based on the results of systematic external and internal reviews. The most recent exam restructure follows the recommendations of major task force reports for enhancing the clinical relatedness and the integration of content within the exams. The empirical basis of the current exam structure was a practice analysis study commissioned by the NBEEO in 2004 (Soroka, Bennett, & NBEEO, 2006).

In the current restructure of the NBEEO examinations, Part I has been renamed “Applied Basic Science” (ABS) and consists of a blend of Basic Science and Clinical Science items, that is ‘old’ Part I and Part II. As implied by its name, the ABS exam will have the requirement of testing basic science that is related to the clinical practice of optometry and most items will have the requirement of referencing a clinical condition either explicitly or implicitly. The ABS examination is administered to student candidates in the spring of their third academic year. It consists of a total of 500 items which are administered in four 3 ½ hour sessions each containing 125 items. The first administration of the ABS exam was in March 2009.

Part II, renamed “Patient Assessment and Management” (PAM) is a lengthened version of the PAM section of the ‘old’ Part III. As the new Part II exam, PAM is lengthened to two 3 ½ hour sessions each consisting of 30 cases. The exam consists of a total of 350 items associated with 60 patient cases. The existing student eligibility requirements are the same for the new Part II, that is, being midway through the fourth academic year.

Part III, renamed “Clinical Skills,” is a lengthened version of the current Clinical Skills section of the old Part III. This performance test will comprise Part III in its entirety. The existing student eligibility requirements will remain the same for the new Part III, that is, being near the end of the fourth academic year, with the targeted administration remaining in the spring.

Description of Report

When a new certification examination is established or is greatly modified, as in the present case, it is necessary to set a cutoff score through a process known as *standard setting*. This report summarizes the results of the standard setting study for the restructured Part II-PAM examination, which will be offered beginning December 2009.

This report contains two major sections. The first section contains background information on the Part II-PAM examination and an overview of standard setting methods. The second section presents details of the standard setting study to establish the cut score for the PAM exam including the method used for the present study, the results, and an assessment of data quality. The report concludes with a brief summary and recommendation.

Overview of the Part II, PAM examination

The Part II-PAM examination consists of 60 simulated patient cases administered over two sessions, each 3 ½ hours in length. Each patient case begins with a scenario in which the patient history and clinical data are presented. These data include at least one visual (e.g., color ophthalmic photographs; contact lens fluorescein pattern; spectacle frame fitting problem; visual field plots; other instrumentation printouts). The scenarios are followed by either 5 or 6 related multiple-choice test items, for a total of 350 items. Each test item contains as many as 10 options, only one of which is correct.

Cases on the Part II-PAM exam are distributed between the two major content areas and among the 17 conditions listed on the NBEO content matrix. This matrix is the same for all Parts of the NBEO examinations. The distribution of cases is as follows:

Disease/Trauma (60% - 70% of cases):

- Lids / lashes / lacrimal system / ocular adnexa / orbit (2 – 4 cases)
- Conjunctiva / cornea / refractive surgery (6 – 8 cases)
- Lens / cataract / IOL / pre- and post-operative care (3 – 5 cases)
- Episclera / sclera / uvea (2 – 4 cases)
- Vitreous / retina (6 – 8 cases)
- Optic nerve / neuro-ophthalmic pathways (5 – 7 cases)
- Glaucoma (2 – 4 cases)
- Emergencies (2 – 4 cases)
- Systemic health (2 – 4 cases)

Refractive Status/Sensory Processes/Oculomotor Processes (30% – 40% of cases):

- Ametropia (3 – 5 cases)
- Ophthalmic optics (1 – 3 cases)
- Contact lenses (3 – 5 cases)
- Low vision (1 – 3 cases)
- Accommodative / vergence / oculomotor anomalies (2 – 4 cases)
- Amblyopia / strabismus (1 – 3 cases)
- Perceptual function / color vision (1 – 3 cases)
- Visual and human development (1 – 3 cases)

Approximately 100 – 120 items of the 350 items on the PAM exam are categorized as TMOD (Treatment and Management of Ocular Disease) items. A TMOD breakout score and pass-fail decision based on these items will be reported for state boards requirements. The content of TMOD items includes one or more of the following:

- Formulation of most appropriate disease diagnosis which will be treated and managed
- Clinical correlation of basic science principles related to disease diagnosis and treatment
- Selection of treatment/management, including systemic considerations
- Dose, form, schedule, and duration of treatment
- Contraindications and side effects of medication, including systemic considerations
- Follow-up and prognosis, including reassessment of diagnosis after initiating treatment
- Treatment and management of ocular emergencies and urgencies

Overview of Standard-Setting Methods

There are two basic approaches to setting performance standards for examinations, each appropriate for a given situation. The two methods are commonly referred to as norm-referenced standards and domain-referenced standards.

Norm-referenced Standards

A norm-referenced standard defines acceptable performance for a candidate in reference to all other candidates by comparing each candidate to the "norm" group that took the exam. This method is most appropriate when only a certain number of positions (jobs, student vacancies) are available to be filled, and the goal is to fill these positions with the most qualified individuals. Colleges and universities might use a norm-referenced standard in order to select only the top 50% of the applicants. Similarly, a business might want to select only 20 of 200 applicants for a position. In both of these instances, the standard would be set so that a predetermined percentage of applicants "passed" the exam. Norm-referenced standards can be established in less systematic ways, too. One common method is to set a passing score at one standard deviation below the mean (average) score of the group. Such a standard will pass approximately 84% of the candidates. Another method is to set the passing score at the group mean, which will pass about 50% of the candidates. Some licensure programs may determine that 75% or 80% of the candidates will pass, and then set the passing score to ensure that this outcome is achieved. These standards are arbitrary because there is no compelling reason for a licensing board to ensure that a fixed percentage of candidates fail.

Domain-Referenced Standards

The second method for establishing performance standards defines acceptable performance in terms of the level of knowledge required in order to ensure that only qualified individuals pass the exam. Licensure exams are developed according to content specifications; these specifications list the subject domains that the exam will cover, and also state the number of test items in each subject domain. Some *criterion level of knowledge* in each of the domains is presumed to be required for safe and effective practice. The goal of establishing a criterion-referenced standard is to identify how much knowledge is enough. In other words, the criterion-referenced standard specifies the percentage of a content domain that must be mastered in order to pass the examination. Not surprisingly, the terms domain-referenced and criterion-referenced are used interchangeably.

A domain-referenced standard specifies the number (or percentage) of test items that a candidate must answer correctly in order to pass an examination. Such standards can be established arbitrarily, or in a rational, systematic manner. For example, it would be possible for an authority to arbitrarily determine that candidates must get 80% of the test items correct in order to pass an exam. The standard would be arbitrary because it was established without carefully studying the occupation being licensed and without carefully evaluating the content and difficulty of the examination.

NBEO staff advocates a more systematic approach. This involves first having a panel of experts determine the knowledge and skills required for safe practice, and then carefully evaluating each item on the examination in relation to this criterion.

Review of Domain-Referenced Standard Setting Methods

Many methods for setting domain-referenced performance standards appear in the measurement literature. In general, methods are classified in terms of the types of materials the panelists will focus their judgments. Some methods require judges to review items and are referred to as "test (or item) centered;"

Other require judges to systematically look at candidate work; these are called “examinee centered.” A brief overview of different standard setting methods is provided below within the two broad classifications.

Test/item centered methods:

- In the **Angoff method** (1971), panelists review multiple-choice items and then provide an estimate of the probability of minimally competent candidates answering each item correctly. The ratings for each panelist are typically summed over items in the test, and these sums are averaged across panelists, to determine the panel performance standard.
- Angoff (1971) also referred to a simpler version of his method, in which the standard setting panelist provides an estimate of whether or not the just-qualified candidate would answer the item correctly. This method is referred to as the **yes/no method** (Impara & Plake, 1997).
- In the **Ebel method** (1972), panelists make item-by-item judgments to classify items along two dimensions: difficulty and relevance. Once items are classified into a matrix based on the combination of the dimensions, panelists provide a judgment in terms of expected percent correct as to how the borderline examinee will perform on the items contained within that cell.
- The **Nedelsky method** (1954) can only be used with multiple choice items since it requires a judgment about each possible wrong answer. The judge’s task is to look at each item and identify the wrong answer choices that the borderline examinee would recognize as incorrect. The reciprocal of the number of choices not ruled out represents the probability that the borderline candidate will answer the item correctly by random guessing after eliminating the choices he/she knows to be incorrect.
- The **bookmark method** (Lewis, Mitzel, & Green, 1996) requires panelists to review specially constructed booklets, termed “ordered item booklets,” in which test items are ordered according to difficulty (using a p-value or an IRT difficulty). The judges’ task is to place a bookmark between the two items in the ordered booklet that represent the content that borderline candidates should be likely to know and be able to do.

Candidate centered methods:

- In the **borderline group method** (Livingston & Zieky, 1982), a group of borderline examinees is identified based on the ratings of teachers and/or other qualified judges. The test scores for these examinees are gathered and their median test score is typically used as the performance standard.
- In the **contrasting groups method** (Livingston & Zieky, 1982), raters are asked to identify one group of examinees whose member are clearly above a particular performance standard and another groups whose members are clearly below that standard. Then the test score distributions of these two groups are contrasted to select the performance standard.

PAM Standard Setting Study

Participants

The development of a realistic, defensible performance standard requires the participation of those who are knowledgeable about: (1) the tasks required to practice in the profession in a safe and effective manner, and (2) the content covered by the exam. In addition, participants should be familiar with the goal of a licensure program, and be able to consider the interests of both the candidates and the public.

In the summer of 2009, the NBEO scheduled the Part II standard setting to coincide with other exam development activities at the National Board office in Charlotte. During the fall of 2009, three groups participated in the standard setting for the Part II-PAM examination; each group devoted a full-day of meeting time to this activity. Each group consisted of clinical practitioners from a variety of geographical

and practice settings, as well as faculty members from various schools of optometry around the U.S. Other participants included four members of the NBEO Board of Directors and a representative from the Association of Regulatory Boards in Optometry. Three members of the NBEO staff attended the various meetings and participated in the discussions. The meetings were coordinated by the NBEO Director of Psychometrics.

A total of 24 participants took part in the Part II-PAM exam standard setting:*

1. The Part II Examination Council which serves in an advisory capacity and reviews the overall Part II exam for accuracy and balance of content during the 2009-2010 exam cycle. This group met August 5-8, 2009 and included the following individuals:

Stanley Hatch, O.D., M.P.H.	Plattsburgh, NY
David King, O.D.	China Grove, NC
Joseph Sullivan, O.D.	Kansas City, KS
Randy Vance, O.D.	Big Rapids, MI
Gary Williams, O.D.	Owego, NY
Janet Carter, O.D.	Board Liaison, Los Vegas, NV
Don Crouch, O.D.	Board Liaison, Storm Lake, IA

2. The second panel met October 3-6, 2009 and consisted of the following individuals:

Denise Goodwin, O.D.	Forest Grove, OR
James Hoekel, O.D.	Saint Louis, MO
Lynn Greenspan, O.D.	Philadelphia, PA
Edward Revelli, O.D.	Berkeley, CA
Linda Rouse, O.D.	Fort Lauderdale, FL
Muriel Schornack, O.D.	Rochester, MN
Stanley Woo, O.D., M.S.	Houston, TX
Nancy Peterson-Klein, O.D.	Board Liaison, Big Rapids, MI

3. The third panel met October 22-25, 2009 and consisted of the following individuals:

Christopher Barbour, O.D., Ph.D.	Hood River, OR
Richard Castillo, O.D., D.O.	Broken Arrow, OK
Julie DeKinder, O.D.	Saint Louis, MO
Barry Fisch, O.D.	Newton, MA
Mark Ostermeir, O.D.	Grants Pass, OR
Karen Peschke, O.D.	San Marcos, CA
Dawn Pewitt, O.D.	San Diego, CA
Tammy Than, O.D.	Birmingham, AL
Elizabeth Wyles, O.D.	Chicago, IL
Alan King, O.D.	Board Liaison, Dickinson, ND

* The National Board wishes to acknowledge the efforts of ALL of the individuals who contributed to the development of the Part II-PAM examination. It would have not been possible without the generosity and expertise of the many case authors, the Part II-PAM Committee and Council members.

Procedure

Method

The study was conducted using the Yes/No method, which is similar to the original Angoff (1971) approach. This procedure is based on the notion that if it were possible to positively identify a large group of marginal candidates, they could be given the test and their test scores analyzed. Since this group has already been identified as “borderline,” it should be acceptable that some of this group would exceed the passing standard, while others would not. As a result, it would be reasonable to establish a “cut score” at their average score.

Since it is not practical to identify borderline candidates, give them a test, and then compute their average score, another plan must be used. This plan involves three steps: (a) asking a group of experts to identify a hypothetical marginally competent candidate, (b) asking them to decide whether or not that candidate would get each item correct; and (c) computing the average scores over all judges and over all of the items. The sum of the item rating for a given judge is an expected scores for a minimally qualified examinee. By asking the judges to make their determinations independently, it is reasonable to average their scores to derive a passing standard that represents, in some sense, a consensus agreement on the expected performance of minimally qualified candidates.

The choice of the Yes/No method for the Part II examination was based on a number of factors:

- The structure of the PAM exam, which is case-based, means that there is a certain amount of interdependence among the items associated with each scenario. A simple yes/no decision for each item allowed the panelists to take the relationships among the items into account as they made their judgments.
- The examination is very long and comprehensive with 60 cases and 350 items. In order to increase the probability that the judges would be able to complete the task of rating all of the items, a simple cognitive task (yes/no) was desirable.
- In the past, the NBEO has made extensive use of the Nedelsky method and many of the panelists in this study have participated in its use. The yes/no method offered the convenience of being an item/test centered technique and at the same time, it was a direct contrast with other methods that involve estimating item difficulties (e.g., Nedelsky, Angoff).
- Because the Part II-PAM exam has not yet been administered, it was not feasible to use an ordered item booklet, such as with the bookmark method.

Orientation

Each of the meetings ran for eight hours on August 5, October 5, and October 24 respectively. After a general orientation, participants discussed different approaches to standard setting and evaluated the advantages and disadvantages of both domain-referenced and norm-referenced standards. In addition, the various methods of setting domain-referenced standards were discussed along with a rationale for the use of the Yes/No method.

Approximately 3 hours was spent in training and orientation before the actual standard setting exercise. The primary goals of the the training were to:

- provide an overview of the PAM exam, in particular the case structure and the types of items within each case. The ABS groups were provided with an extensive overview of the Part II-PAM examination and a comparison with the original Part III-PAM.
- provide an overview to standard setting and to convey a sense of the importance of the process,

- discuss case/item difficulty and to teach judges to estimate the difficulty of item. Several cases from the April 2008 Part III-PAM exam were used to facilitate the training process.
- introduce the concept of minimal competence and to help judges arrive at their own personal definitions of borderline performance.

During the discussion of case/item difficulty, panelists were directed: *For each of the items in the case, indicate with 'Y' or 'N' whether or NOT a minimally proficient candidate **would** get the item correct.* Judgments were recorded, averaged, and then compared to the actual examinee performance. Participants discussed their individual judgments and shared the reasoning behind their decisions. Parts of this discussion addressed the factors that make the cases and/or test items easy or difficult, while other parts focused on candidate's level of preparation and education (e.g., amount/type of clinical experience of a 4th year optometry student).

The discussion of minimal competence (borderline performance) is an important part in the implementation of the Yes/No method. This conversation also focused on a group definition/perception of what constitutes 'entry-level.' To facilitate the discussion the marginally qualified candidate was contrasted with both expert and incompetent candidates. Expert candidates were characterized as having integrated knowledge; they are able to make connections between abstract concepts and concrete patient situations. In addition, experts have a wider breadth of knowledge which is often at a deeper level. The incompetent/unqualified candidate was characterized as someone who has a minimal amount of knowledge, is often unaware or unwilling to acknowledge their deficiencies and as a result, appears arrogant.

The marginally qualified candidate was characterized as someone with a 'thin' base of knowledge who wants to be told exactly how much they need to know. In addition, this individual knows just enough about the range of topics required in entry level optometry. This level of knowledge was characterized as 'bits and pieces' in one standard setting group and as 'fragmented' in another. The marginally qualified candidate does not have an integrated knowledge of optometry. In addition, he or she may not be a very curious person. Finally, the minimally qualified optometrist could be a very practical person who might have struggled to master basic concepts and who recognizes their limitations.

Standard Setting Judgments

After panelists completed training, the December 2009 Part II-PAM examination was distributed. Participants also received the exam key (correct answers), and a form for recording their judgments. In order to account for fatigue effects among raters, panelists were asked to begin the case/item rating process at different points among the 60 cases. In other words, everyone did not start at Item 1 and proceed to Item 350.

The directions to panelists for the standard setting ratings were identical to the ones used in the training process:

*For each of the items in the case, indicate with 'Y' or 'N' whether or NOT a minimally proficient candidate **would** get the item correct.*

After the rating process, judgments were collected, entered into a spreadsheet, tabulated and summarized. Participants were not given the results of the tabulation and were not offered an opportunity to revise their estimates, as is common in some standard setting studies.

After the item rating activity, each judge was asked to complete a short survey. Among the questions on the survey was: "Given everything you know about the content and format of the exam (as well as the overall structure of the NBEO exams), the relevance of the test questions, and the level of knowledge

required for competent performance in optometry, what percentage of the questions should an examinee answer correctly in order to pass the exam? In short, what should the cut score be?” Responses to this question will be referred to as the *global cut score*. A second question of interest from the survey asked the judges to answer the following question: “Given your knowledge of the individuals taking this exam, what should the pass rate be? That is, what percentage of the examinees should pass?” Responses to this item will be referred to as the *global pass rate*.

Results

Case/Item Difficulty Ratings

This section summarizes the standard setting information obtained from the panelists. This includes the results of the Yes/No ratings, the global cut score judgments and the global pass rate judgments.

Table 1. Individual Judge Ratings by Exam Session

Judge	Session 1	Session 2	Total Test
1	0.866	0.821	0.844
2	0.687	0.639	0.663
3	0.802	0.824	0.813
4	0.689	0.735	0.712
5	0.827	0.776	0.801
6	0.552	0.647	0.600
7*	0.714	0.754	0.734
8	0.583	0.697	0.640
9	0.514	0.537	0.526
10	0.629	0.646	0.637
11	0.611	0.629	0.620
12	0.600	0.651	0.626
13	0.697	0.663	0.680
14	0.697	0.714	0.706
15	0.600	0.566	0.583
16	0.703	0.606	0.654
17	0.720	0.697	0.709
18	0.697	0.731	0.714
19	0.703	0.726	0.714
20	0.623	0.623	0.623
21	0.743	0.760	0.751
22	0.720	0.720	0.720
23	0.600	0.474	0.537
24	0.720	0.749	0.734
Mean	0.679	0.683	0.681

* incomplete data from this panelist

Table 1 presents the individual judges’ ratings for the Part II-PAM divided by exam session. The last row of the table presents the mean scores for the borderline candidate during each session of the PAM exam

over the entire group of judges. These values, 0.679 for session 1 and 0.683 for session 2, indicate that the judges rated the two ‘tests’ as roughly equal in difficulty. The judges would set as the overall cut score for the total test at 0.681 or 68.1% correct. Using this cut score, candidates correctly responding to 239 out of 350 items would pass the PAM exam.

Table 2 presents the information based on the global ratings described earlier. The values in the “Global Cut Score” row indicate that if this group of judges were to set a cut score based only on their general perception of the test as a whole, they would settle on a passing point of about 69.6% correct. The “Global Pass Rate” was calculated from responses to the question about what the pass rate for the Part II exam should be. This group of judges would find a pass rate of approximately 84% acceptable. These results should be considered supplemental information about the perceptions of panelists.

Table 2. Summary Judgments

	Mean	Std. Dev.
Global Cut Score:	0.696	0.058
Global Pass Rate:	0.839	0.062

Evaluation of Data Quality

Several analyses were conducted in order to evaluate the integrity and quality of the judge’s ratings. The simplest check was to evaluate the completeness of the data. The PAM Exam Council data were incomplete. This can be attributed to the timing of their participation in the standard setting study. The exam was not completely final (i.e., ready for the printer) when they participated in the case/item rating exercise. Some cases were edited further after they furnished their ratings (e.g., an item might have been eliminated because a case had too many). When the PAM Council’s item ratings were matched to the ratings provided by the other groups, 329 out of 350 items matched; that is, 94% of the PAM data was complete.

Since a key task of participants in a standard-setting study is to estimate the difficulty of the exam for a marginally proficient candidate, it is useful to examine the range of scores assigned by judges. Although participants are encouraged to have different performance expectations for borderline candidates, they should not be drastically different. The data presented in Table 1 indicates that the most lenient panelist would set a passing score of 52.6% correct, while the most stringent would set it at 84.4% correct. This 31.8 percent range is a bit large but this variation may be attributable to the dichotomous nature of the Yes/No method. Neither of these ‘extreme’ scores are outliers in the distribution of exam ratings; the set of ratings in Table 1 contains several scores in the 50s and several in the 80s. The relatively large number of raters in this study (24) makes the results more stable and eliminating these scores from the analysis has virtually no effect on the mean, or the recommended cut score.

Finally, the data were evaluated to determine the extent to which these results would be expected to generalize if a different group, if the study were conducted with another set of equally qualified panelists rating the same set of items. An analysis based on the variance components and generalizability theory was conducted to estimate the margin of error. This analysis was based on 23 judges with the most complete data (329 items). As a result, the index of dependability (ϕ) was 0.89, a very high value. The margin of error is ± 1 point. If the study were to be conducted with another group of participants, we could reasonably expect the passing score to be within 1 point (on the percent correct scale) of the value obtained from this study.

Summary of Report

The present study resulted in a cut score of 68.1% correct which the NBEO Board of Directors adopted at its December 2009 meeting. This score was applied to the scored items on the December 2009 PAM examination resulting in a cutscore of 233.6 out of 343 scored items. Consistent with accepted scoring practices, a raw score of 234 was the lowest passing score for the December 2009 PAM administration.

The PAM standard setting study focused on the initial, targeted, December 2009 examination. Given the scope of the Part II-PAM exam and the allotted amount of meeting time, it was not practical to obtain standard setting ratings for the April 2010 administration. In developing the December and April PAM administrations, extreme care was taken to make them parallel in terms of content and complexity (difficulty). However, since they contain different patient cases and items, the exams are not identical in terms of difficulty. As a result, percentage correct scores from the two exams will not reflect the same level of knowledge. To treat the scores as if they are comparable would be misleading to score users and unfair to test-takers (especially those who took the exam with potentially more challenging cases and items).

In order to carry the established cut score forward, a number of cases from the December exam will be administered again in April 2010. The repeated cases were selected as a mini-test in order to establish an equating link between the two examinations. This “anchor-test non-equivalent groups design” assumes that the mini-test represents the entire exam in terms of both content and difficulty (Kolen & Brennan, 1995). In that way, the established standard can be maintained.

Going forward, the use of mini-tests will not be necessary. As the item bank of cases is developed and calibrated, the Rasch item response theory model (1960) will be used to equate different examinations and to maintain the passing standard.

References

- Angoff, W.H. (1971). Scales, norms, and equivalent scores. In R.L. Thorndike (Ed.) *Educational measurement* (pp. 508-600). Washington, DC: American Council on Education.
- Ebel, R.L. (1972). *Essentials of educational measurement*. Englewood Cliffs, NJ: Prentice Hall.
- Impara, J.C. & Plake, B.S. (1997). Standard setting: An alternative approach. *Journal of Educational Measurement*, 34, 353-366.
- Kolen, M.J. & Brennan, R.L. (1995). *Test equating Methods and practices*. New York, NY: Springer-Verlag.
- Lewis, D.M., Mitzel, H.C., & Green, D.R. (1996 June). Standard setting: A bookmark approach. In D.R. Green (Chair), *IRT-based standard setting procedures utilizing behavioral anchoring*. Symposium conducted at the Council of Chief State School Officers National Conference on Large-Scale Assessment, Phoenix, AZ.
- Livingston, S.A., & Zieky, M. J. (1982). *Passing scores*. Princeton, NJ: Educational Test Service.
- Nedelsky, L. (1954). Absolute grading standards for objective tests. *Educational and Psychological Measurement*, 14, 3-19.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research.
- Soroka, M., Krumholz, D., Bennett, A., & National Board of Examiners Conditions Domain Task Force, (2006). The practice of optometry: National Board of Examiners in Optometry survey of optometric patients. *Optometry and Vision Science*, 83 (9), 625-634.