Rasch Analysis of the Ocular Surface Disease Index (OSDI)

Bradley E. Dougherty,¹ Jason J. Nichols,² and Kelly K. Nichols²

PURPOSE. The Ocular Surface Disease Index (OSDI) is a 12-item scale for the assessment of symptoms related to dry eye disease and their effect on vision. Its reliability and validity have been investigated within the classical test theory framework and, more recently, using Rasch analysis. The purpose of the present analysis was to more completely investigate the functioning of its response category structure, the validity of its three subscales, and the unidimensionality of the latent construct it is intended to assess.

METHODS. Responses to the OSDI from 172 females participating in the Dry Eye in Postmenopause (DEIM) study who had previously been diagnosed with dry eye or reported significant ocular irritation and dryness were analyzed. Response category structure and item fit statistics were evaluated for assessment of model fit. Person separation statistics were used to examine the validity of the subscales. Unidimensionality was assessed by principal component analysis (PCA) of model residuals.

RESULTS. The recommended five-category response structure resulted in disordered response thresholds. A four-category structure resulted in ordered thresholds. Item infit statistics were acceptable for all 12 items. Person separation with this category structure was adequate, with a person separation index of 2.16. None of the three subscales demonstrated adequate person separation. PCA showed one other significant factor onto which the three environmental items loaded significantly.

CONCLUSIONS. All items demonstrated acceptable fit to the model after collapsing categories to order the response thresholds. The original subscales did not prove valid, and there is some evidence of multidimensionality and poor targeting. (*Invest Ophthalmol Vis Sci.* 2011;52:8630–8635) DOI:10.1167/iovs.11-8027

The Ocular Surface Disease Index (OSDI) is perhaps the most frequently used survey instrument for the assessment of ocular surface disease severity in dry eye research. It has been used in a variety of studies, both in the assessment of the severity of dry eye associated with certain conditions^{1–3} and in the investigation of the efficacy of dry eye treatments.^{4–7} The instrument, introduced in 1997 by the Outcomes Research Group (Allergan Inc., Irvine, CA),⁸ consists of 12 items that assess symptoms, functional limitations, and environmental factors related to dry eye. Each

item has the same five-category Likert-type response option, and each of the three subscales has its own question type.

Initial investigations of the reliability of the OSDI were conducted using Classical Test Theory methods, that is, using the Cronbach's α statistic to assess the internal consistency of the items.⁹ There are several excellent references on Classical Test Theory and its use in survey research.^{10,11} Schiffman et al.⁹ found that Cronbach's α for the OSDI was 0.92, and a factor analysis revealed three subscales (symptoms, environmental triggers, and vision-related function). They reported acceptable test-retest repeatability, but found that OSDI scores did not correlate particularly well with clinical tests for dry eye. Specifically, correlations for all subjects between OSDI score and tear break-up time, Schirmer's test, lissamine green, and fluorescein staining in the worse eye ranged from -0.21 to $+0.19$ and none of the correlations was statistically significant.

The use of Classical Test Theory to score survey instruments and evaluate their reliability has been criticized for several reasons. One of these is the treatment of Likert-type survey data as continuous, rather than ordinal. Another is the assignment of equal weight to each survey item in the calculation of an overall score, when one could imagine situations in which items may require different levels of the underlying trait for endorsement. For these reasons, the common method of generating an overall score from an instrument—summing and averaging the ordinal-level responses—is open to criticism. Also, the approach of using Cronbach's α to assess reliability does not provide information regarding the behavior of individual survey items, only the instrument as a whole.

A family of models known collectively as Item Response Theory (IRT) provides an alternative approach to the scoring and evaluation of survey instruments. The models have roots in education and aptitude testing, and seek to quantify the amount of some latent trait using survey responses. Rasch analysis, also sometimes referred to as a one-parameter logistic model, is often considered as the simplest of the IRT models.^{12,13} Rasch analysis was initially developed for educational testing, but is now frequently used in healthcare research to evaluate and score survey instruments.^{14–18} It provides interval-level data from survey responses, including estimates of the amount of the latent trait displayed by each subject ("person measure") and the amount of the trait necessary to respond in a certain way to each item ("item measure").¹² Moreover, Rasch analysis also provides item fit statistics that indicate whether the individual item is contributing to the measurement of the latent trait. There are excellent descriptions of Rasch analysis and its use with survey instruments published elsewhere.^{12,19–21}

Other IRT models include parameters in addition to the item difficulty parameter used in Rasch analysis. For instance, some models include an item discrimination parameter that allows for differences in the slope of the curve describing response probabilities for individual items.²² Rasch models are restrictive in the sense that data that do not conform to the model are not considered to be consistent with measurement, and the curves describing response probabilities are not free to vary between items.

From the ¹The Ohio State University College of Optometry, Columbus, Ohio; and the ²University of Houston College of Optometry, Houston, Texas.

Supported in part by National Eye Institute of the National Institutes of Health Grants EY017951 (JJN, KKN), EY015519 (KKN, JJN), and 5T32EY013359 (BED).

Submitted for publication June 10, 2011; revised September 8, 2011; accepted September 16, 2011.

Disclosure: **B.E. Dougherty**, None; **J.J. Nichols**, None; **K.K. Nichols**, None

Corresponding author: Kelly K. Nichols, University of Houston College of Optometry, 505 J. Davis Armistead Building, Houston, TX 77204; knichols@optometry.uh.edu.

Other IRT models seek to describe the data as best as possible using extra parameters such as item discrimination. Massof²³ published a study in which he compared a Rasch model to a two-parameter logistic IRT model (the Muraki model) using data from visual functioning questionnaires. He demonstrated that the item discrimination parameter of the Muraki model was inversely proportional to the item fit statistics of the Rasch model.

There has been some recent work in the area of evaluation of dry eye survey instruments with Rasch analysis. Gothwal et al.¹⁴ examined the measurement properties of the McMonnies questionnaire using Rasch analysis. They found that person separation was inadequate for discriminating between more than two strata of dry eye severity and, therefore, the McMonnies questionnaire did not function as a valid measure to discriminate across disease severity.

Johnson and Murphy²⁴ developed the Ocular Comfort Index (OCI) to measure ocular surface disease symptoms using Rasch analysis. The instrument they developed has 12 items and a seven-category response structure. Person separation was good and all 12 of the final items had adequate Rasch fit statistics.

Simpson et al.²⁵ evaluated the Dry Eye Questionnaire, the McMonnies questionnaire, and the OSDI. One purpose of this study was to evaluate the Rasch item fit statistics of the instruments and use them to determine whether the surveys were unidimensional. For the OSDI, the authors found that all items had fit statistics within the acceptable range. Other aspects of the analysis, such as the functioning of the category structure and person separation statistics, were not reported.

Pesudovs and Noble²⁶ evaluated a single-item faces scale for measuring pain associated with severe ocular surface disease. They applied Rasch analysis to refine the category structure of the instrument. The study also used the scale to demonstrate the potential of Rasch analysis to increase sensitivity to changes after treatment for ocular surface disease, finding an increased effect size with Rasch analysis compared with conventional raw scores.²⁶

In light of this, we hypothesize that the application of Rasch analysis to responses to the OSDI from patients with dry eye disease might be beneficial in further understanding its psychometric properties. Thus, the purpose of this study was to investigate the OSDI using Rasch analysis in a sample of females 50 years of age and older who were participating in a study of dry eye in postmenopause and who had been previously diagnosed with dry eye or reported significant ocular dryness and irritation.

METHODS

Participants

The OSDI was administered to female participants in the Dry Eye in Postmenopause study at the College of Optometry at The Ohio State University. The OSDI scores of participants were included in the analysis if at least one of two criteria were met. The first of these was that the participant reported having been previously diagnosed with dry eye by an eye care provider. The second was that the participant answered "often" or "constantly" to *both* of the following questions: "How often do you experience eye dryness?" and "How often do your eyes feel irritated?" These questions were previously used by Schaumberg and colleagues²⁷⁻²⁹ for classification of patients by dry eye status in large-scale epidemiologic studies of the prevalence of dry eye. The mean age ($\pm$ SD) of participants was 63 ± 8 years. Potential participants were excluded from the study if they were <50 years of age, had worn contact lenses in the past 3 months, were taking eye drops for an ocular condition other than dry eye, had a history of any eye surgery other than secondary membrane removal after cataract extraction in the past year, or reported other significant ocular pathology. Informed consent was obtained from all participants, in accordance with The Health Insurance Portability and Accountability Act of 1996 (HIPAA) regulations and the Declaration of Helsinki.

Rasch Analysis

The OSDI response structure contains five options that relate to the frequency of the effects of ocular surface disease: "none of the time," "some of the time," "half of the time," "most of the time," and "all the time." There are three question types: "Have you experienced any of the following during the last week?" (items 1-5); "Have problems with your eyes limited you in performing any of the following during the last week?" (items 6-9); and "Have your eyes felt uncomfortable in any of the following situations during the last week?" (items 10-12). Rasch analysis was performed with a commercial software knowledgebase (WINSTEPS version 3.69; Winsteps, Chicago, IL), using a three-level Andrich rating scale model.³⁰ For the response structure to be valid, the category thresholds, or the point on the logit scale of ability at which a subject is equally likely to choose between two adjacent categories, should be ordered. That is, these threshold person measures should increase in order with the categories so that subjects with increasing amounts of the trait of interest have increasing probabilities of selecting higher categories.³¹ If category thresholds proved disordered, categories were combined to obtain ordered thresholds. Once ordered category thresholds were established, instrument and item-level statistics were analyzed. Published guidelines regarding acceptable item fit and other Rasch analysis statistics were used to guide the analysis.²⁰

Item infit mean square statistics were used to determine whether individual items provided useful information for measurement of ocular surface disease severity. The infit mean square is an information-weighted fit statistic that compares observed data with model expectations. Items with infit values outside of 0.7-1.3 were eliminated one at a time, beginning with the most misfitting item, and the analysis was repeated until no items misfit.

The ability of the instrument to discriminate between participants was assessed using the person separation statistic. Person separation is a ratio of the variance explained by the measures to the total variance (including error variance).¹² A value of 2.0 was considered the minimum acceptable value and corresponds to the ability to differentiate between three levels of a trait. Person separation was also used to evaluate the validity of the three subscales, with the same minimum acceptable criterion.

If an instrument is used to report a single measure, it should assess only one latent trait. Principal component analysis (PCA) of Rasch residuals (performed using WINSTEPS version 3.69) was used to assess unidimensionality. If an instrument is unidimensional, then PCA of the model residuals should reveal no structure in those residuals.³² Significant loading onto other factors in the analysis is indicative of multidimensionality. Factors with eigenvalues (an indicator of the proportion of the total variation explained by an individual factor) >2.0 were considered to be evidence of significant multidimensionality.³³

RESULTS

Response Category Functioning

Category thresholds with the OSDI five-category response structure were shown to be disordered (Fig. 1). A four-category response structure, in which the categories "half of the time" and "most of the time" were combined, had ordered thresholds and fairly equal widths over which each category was the most likely response, which is desirable²⁶ (Fig. 2). A different four-category structure in which the categories "all the time" and "most of the time" were combined was also tested, but it did not result in ordered thresholds. The four-category structure that combines "half of the time" and "most of the time" was used for the rest of the analyses.

Item Statistics

Of the 172 female participants who completed the OSDI, 7 responded "none of the time" to all 12 items. Data from

FIGURE 1. Category probability curves for the five-category instrument showing underutilization of category 3 and disordered thresholds. Curves for items 1–5, 6–9, and 10–12 are shown in (a), (b), and (c), respectively. *Blue*: none of the time; *red*: some of the time; *green*: half of the time; *purple*: most of the time; *light blue*: all the time.

these subjects were not included in the analysis, leaving 165 subjects who were included. When the four-category response structure was used, all 12 items had infit mean square statistics within the acceptable range of 0.7 to 1.3. The item measures and infit statistics are shown in Table 1.

Person Separation

The person separation index for the 12-item instrument was 2.16, which indicates that the OSDI can adequately discriminate between patients.

Subscales

The person separation indices for each of the three subscales are shown in Table 2. None of the subscales met the criterion of a person separation index of at least 2.0, which indicates that none of the subscales adequately differentiated between different levels of the targeted constructs.

Unidimensionality

Principal component analysis of the standardized model residuals indicated that there was one additional factor onto which items were loading significantly. The first contrast had an

FIGURE 2. Category probability curves for the four-category instrument showing ordered thresholds. Curves for items 1–5, 6–9, and 10–12 are shown in (a), (b), and (c), respectively. *Blue*: none of the time; *red*: some of the time; *green*: half of the time or most of the time; *purple*: all the time.

TABLE 1. Item Measures and Infit Mean Square Fit Statistics for the 12 OSDI Items Using a Four-Category Response Structure, Collapsing “Half of the Time” and “Most of the Time”

Item	Item Measure (SE)	Infit Mean Square
1. Light Sensitivity	−0.75 (0.12)	1.29
2. Gritty	−0.81 (0.12)	1.10
3. Painful/Sore	0.42 (0.13)	1.18
4. Blurred Vision	0.27 (0.13)	0.93
5. Poor Vision	0.55 (0.13)	1.12
6. Reading	0.26 (0.13)	1.02
7. Driving at Night	0.12 (0.13)	1.14
8. Computer/ATM	0.47 (0.14)	0.71
9. Television	1.10 (0.15)	0.77
10. Windy	−0.69 (0.12)	0.97
11. Low Humidity	−0.78 (0.13)	0.80
12. Air Conditioning	−0.17 (0.13)	0.85

eigenvalue of 2.6 (11.1% of the total variance), which is more than can be attributed to random data. Items that loaded significantly (>0.4) onto this factor included the three environmental triggers items (“windy conditions,” “low humidity,” and “air conditioned”) and one other item (“gritty”). The second contrast had an eigenvalue of 1.6, or 6.6% of the total variance.

Because of the evidence of multidimensionality, we investigated whether a shorter instrument that does not contain the environmental triggers items might function as a valid instrument on its own. To investigate this question, we performed an analysis using items 1 to 9 with only a two-level Andrich rating scale model. The person separation index for this 9-item instrument was 1.82, which does not meet the criteria for adequate discrimination.

DISCUSSION

Our analyses indicate that the response category structure recommended for the OSDI responses currently is not ideal and can be optimized using Rasch analysis. We found that the categories should be collapsed to get them to work properly. Specifically, we found that combining the categories “half of the time” and “most of the time” was necessary. Once this change to the category structure was made, the categories functioned better.

Regarding the fit of the items to the Rasch model, we found results similar to Simpson et al.²⁵ The fit of the items was generally good, with fit statistics falling within the recommended range of Pesudovs et al.²⁰ for all items.

The person separation index for the OSDI was acceptable, at 2.16. This demonstrates that the full 12-item OSDI is a useful instrument for discriminating between people with varying levels of ocular surface disease. We also found that none of the three subscales had adequate person separation indices to function acceptably on its own.

Unidimensionality and Rasch analysis in general have been previously described for the OSDI only once, by Simpson et

al.²⁵ Our study explored additional aspects of the Rasch analysis and explored unidimensionality in another way. We found that the instrument does not meet the standard of unidimensionality when tested using PCA of the model residuals. This is an important requirement for the use of summary scoring, in that a summary score implies that all the items assess the same construct.

Previous analyses of the unidimensionality of the OSDI and the Ocular Comfort Index (OCI) were performed using item fit statistics but not PCA. Although the fit of the items to the model is one indicator of the unidimensionality of an instrument, PCA is another useful tool for the detection of multiple dimensions and may reveal evidence of multidimensionality not detected with item fit statistics alone.²¹ Our PCA indicates that there is evidence of multidimensionality in the OSDI. Specifically, the first contrast of the analysis showed unexplained variance of 2.6 eigenvalue units. Additionally, an analysis of the remaining nine items of the OSDI (not including the environmental triggers items) showed that they do not have adequate person separation to function as a separate scale.

The presence of multidimensionality in survey instruments is problematic, in that if more than one latent trait is being assessed by an instrument it becomes impossible to interpret a single score from that instrument as a measure of any one trait. We are not aware of any survey instrument specific to ocular surface disease that has been demonstrated to be unidimensional using PCA. Because dry eye is a multifactorial disease, investigators may have a desire to investigate the multiple aspects of the disease, such as symptoms and effects on visual functioning. One approach to managing this problem is to use multiple subscales, each of which is capable of assessing a single trait of interest in a valid manner. This would require subscales that have adequate discriminative ability, have items with acceptable fit statistics, and that are unidimensional. This approach would also require that scores from individual subscales, each of which is an indicator of a distinct latent trait related to ocular surface disease, not be combined into a single score for a larger instrument.

The need for more work in the area of patient-reported outcome measures in the area of dry eye and ocular surface disease was recently highlighted in the report on meibomian gland dysfunction from the International Workshop on Meibomian Gland Dysfunction.^{34,35} Future work in instrument development should seek to create unidimensional scales, rather than multidimensional scales that capture multiple aspects of the disease and report a single, difficult to interpret, score.

The targeting of the OSDI—how well the difficulty of the items matches the ability of the subjects taking the survey—was not ideal. This is shown in Figure 3, which indicates that many of the participants had an ability level higher than the level of most or all the items contained in the instrument. The average person measure for the participants in this study, all of whom reported previous dry eye diagnoses or significant ocular irritation, was −1.51. Ideally, the average item measure (set to 0 in the analysis) would be close to the average person measure and the range of ability covered by the set of items

TABLE 2. Rasch Summary Statistics for the OSDI and Its Subscales Using a Four-Category Response Structure, Collapsing “Half of the Time” and “Most of the Time”

Item Set	Mean Rasch Person Measure (SE)	Mean Rasch Item Measure (SE)	Person Separation Index	Item Separation Index	Mean Item Infit Mean Square
OSDI	−1.51 (0.12)	0.00 (0.18)	2.16	4.31	0.99
Subscale 1	−1.42 (0.11)	0.00 (0.31)	1.28	4.49	1.00
Subscale 2	−1.64 (0.17)	0.00 (0.35)	1.29	2.99	0.99
Subscale 3	−1.74 (0.27)	0.00 (0.49)	1.50	3.07	0.97

FIGURE 3. Person-item chart, which demonstrates nonideal targeting. The x-axis is the measure in logits, and the y-axis is the number of subjects or items located at that measurement location. Ideally, the average person measure would be closer to the average item measure, which is set to 0. The subjects represented by the bars to the left of 0 are not well-targeted by the instrument.

would be wide enough to adequately assess all the subjects. Johnson and Murphy²⁴ reported similar targeting for the OCI. However, in that study participants had not necessarily been diagnosed with dry eye or reported significant symptoms as had participants in this study, and it is important to consider differences in the participants when considering the targeting of instruments.

All the participants in this study were female and >50 years of age, thus limiting the ability to analyze whether there is differential functioning of items based on age or sex. However, dry eye and ocular surface disease are important concerns for postmenopausal females and information regarding the usefulness of survey instruments in this population is of great importance. Future studies should investigate differential item functioning. This would necessitate the inclusion of patients with a wider age distribution than that of the present study and of both males and females.

In conclusion, all items of the Ocular Surface Disease Index showed acceptable fit to a Rasch measurement model and adequate between-patient discrimination. However, there is evidence from principal components analysis that it is not unidimensional. Moreover, it is not ideally targeted for patients with dry eye disease. Future studies in patient-reported out-

come measures for dry eye and ocular surface disease should address these issues.

References

1. Gurdal C, Sarac O, Genc I, Kirimlioglu H, Takmaz T, Can I. Ocular surface and dry eye in Graves' disease. *Curr Eye Res.* 2011;36:8-13.
2. Rossi GC, Tinelli C, Pasinetti GM, Milano G, Bianchi PE. Dry eye syndrome-related quality of life in glaucoma patients. *Eur J Ophthalmol.* 2009;19:572-579.
3. Adatia FA, Michaeli-Cohen A, Naor J, Caffery B, Bookman A, Slovic A. Correlation between corneal sensitivity, subjective dry eye symptoms and corneal staining in Sjogren's syndrome. *Can J Ophthalmol.* 2004;39:767-771.
4. Monaco G, Cacioppo V, Consonni D, Troiano P. Effects of osmo-protection on symptoms, ocular surface damage, and tear film modifications caused by glaucoma therapy. *Eur J Ophthalmol.* 2011;21:243-250.
5. Luchs JJ, Nelinson DS, Macy JL. Efficacy of hydroxypropyl cellulose ophthalmic inserts (LACRISERT) in subsets of patients with dry eye syndrome: findings from a patient registry. *Cornea.* 2010;29:1417-1427.
6. Kim TH, Kang JW, Kim KH, et al. Acupuncture for dry eye: a multicentre randomised controlled trial with active comparison

- intervention (artificial tear drop) using a mixed method approach protocol. *Trials*. 2010;11:Art. 107.
7. Gurdal C, Genc I, Sarac O, Gonul I, Takmaz T, Can I. Topical cyclosporine in thyroid orbitopathy-related dry eye: clinical findings, conjunctival epithelial apoptosis, and MMP-9 expression. *Curr Eye Res*. 2010;35:771-777.
 8. Walt J, Rowe M, Stern K. Evaluating the functional impact of dry eye: the Ocular Surface Disease Index (Abstract). *Drug Inf J*. 1997;31:1436.
 9. Schiffman RM, Christianson MD, Jacobsen G, Hirsch JD, Reis BL. Reliability and validity of the Ocular Surface Disease Index. *Arch Ophthalmol*. 2000;118:615-621.
 10. Salkind NJ. *Encyclopedia of Research Design*. Thousand Oaks, CA: SAGE Publications; 2010:3 vols.
 11. Crocker L, Algina J. *Introduction to Classical and Modern Test Theory*. 1st ed. Stamford, CT: Wadsworth Publishing; 1986:527.
 12. Bond TG, Fox CM. *Applying the Rasch Model. Fundamental Measurements in the Human Sciences*. 2nd ed. Mahwah, NJ: Erlbaum; 2007.
 13. Rasch G. *Probabilistic Models for Some Intelligence and Achievement Tests*. Copenhagen: Danish Institute for Educational Research; 1960.
 14. Gothwal VK, Pesudovs K, Wright TA, McMonnies CW. McMonnies questionnaire: enhancing screening for dry eye syndromes with Rasch analysis. *Invest Ophthalmol Vis Sci*. 2010;51:1401-1407.
 15. Garamendi E, Pesudovs K, Stevens MJ, Elliott DB. The Refractive Status and Vision Profile: evaluation of psychometric properties and comparison of Rasch and summated Likert-scaling. *Vision Res*. 2006;46:1375-1383.
 16. Massof RW, Fletcher DC. Evaluation of the NEI visual functioning questionnaire as an interval measure of visual ability in low vision. *Vision Res*. 2001;41:397-413.
 17. Pesudovs K, Garamendi E, Keeves JP, Elliott DB. The Activities of Daily Vision Scale for cataract surgery outcomes: re-evaluating validity with Rasch analysis. *Invest Ophthalmol Vis Sci*. 2003;44:2892-2899.
 18. Stelmack JA, Szlyk JP, Stelmack TR, et al. Psychometric properties of the Veterans Affairs Low-Vision Visual Functioning Questionnaire. *Invest Ophthalmol Vis Sci*. 2004;45:3919-3928.
 19. Massof RW. The measurement of vision disability. *Optom Vis Sci*. 2002;79:516-552.
 20. Pesudovs K, Burr JM, Harley C, Elliott DB. The development, assessment, and selection of questionnaires. *Optom Vis Sci*. 2007;84:663-674.
 21. Wright BD, Stone MH. *Measurement Essentials*. 2nd ed. Wilmington, DE: Wide Range, Inc.; 1999.
 22. Edwards MC. An introduction to item response theory using the Need for Cognition Scale. *Soc Person Psychol Compass*. 2009;3:507-529.
 23. Massof RW. Application of stochastic measurement models to visual function rating scale questionnaires. *Ophthalmic Epidemiol*. 2005;12:103-124.
 24. Johnson ME, Murphy PJ. Measurement of ocular surface irritation on a linear interval scale with the ocular comfort index. *Invest Ophthalmol Vis Sci*. 2007;48:4451-4458.
 25. Simpson TL, Situ P, Jones LW, Fonn D. Dry eye symptoms assessed by four questionnaires. *Optom Vis Sci*. 2008;85:692-699.
 26. Pesudovs K, Noble BA. Improving subjective scaling of pain using Rasch analysis. *J Pain*. 2005;6:630-636.
 27. Schaumberg DA, Dana R, Buring JE, Sullivan DA. Prevalence of dry eye disease among US men: estimates from the Physicians' Health Studies. *Arch Ophthalmol*. 2009;127:763-768.
 28. Schaumberg DA, Sullivan DA, Buring JE, Dana MR. Prevalence of dry eye syndrome among US women. *Am J Ophthalmol*. 2003;136:318-326.
 29. Gulati A, Sullivan R, Buring JE, Sullivan DA, Dana R, Schaumberg DA. Validation and repeatability of a short questionnaire for dry eye syndrome. *Am J Ophthalmol*. 2006;142:125-131.
 30. Andrich D. A rating formulation for ordered response categories. *Psychometrika*. 1978;43:561-573.
 31. Linacre JM. Optimizing rating scale category effectiveness. *J Appl Meas*. 2002;3:85-106.
 32. Linacre JM. Structure in Rasch residuals: why principal components analysis (PCA)? (Abstract). *Rasch Measurement Trans*. 1998;12:636.
 33. Linacre JM. *A User's Guide to WINSTEPS*. Chicago: MESA Press; 2009.
 34. Asbell PA, Stapleton FJ, Wickstrom K, et al. The International Workshop on Meibomian Gland Dysfunction: report of the Clinical Trials Subcommittee. *Invest Ophthalmol Vis Sci*. 2011;52:2065-2085.
 35. Nichols KK, Foulks GN, Bron AJ, et al. The International Workshop on Meibomian Gland Dysfunction: executive summary. *Invest Ophthalmol Vis Sci*. 2011;52:1922-1929.