Reliability of Chinese web-based ocular surface disease index questionnaire in dry eye patients: a randomized, crossover study

Xin-Mei Zhang[1,2], Lan-Ting Yang[1,2], Qing Zhang[1,2], Qing-Xia Fan[1], Can Zhang[1], Yue You[1], Chen-Guang Zhang[1], Tie-Zhu Lin[1], Ling Xu[1], Salissou Moutari[3], Jonathan E. Moore[4,5], Emmanuel E. Pazo[1], Wei He[1]

[1]Department of Ophthalmology, He Eye Specialists Hospital, Shenyang 110034, Liaoning Province, China

[2]The Second Affiliated Hospital of Dalian Medical University, Dalian 116000, Liaoning Province, China

[3]School of Mathematics and Physics, Queens University Belfast, University Road, Belfast, Northern Ireland BT7 1NN, United Kingdom

[4]Cathedral Eye Clinic, 89-91 Academy Street, Belfast, Northern Ireland BT1 2LS, United Kingdom

[5]Biomedical Sciences Research Institute, University of Ulster, Coleraine, Northern Ireland BT52 1SA, United Kingdom

Co-first authors: Xin-Mei Zhang and Lan-Ting Yang

Correspondence to: Wei He and Emmanuel E. Pazo. Department of Ophthalmology, He Eye Specialist Hospital, No.128 North Huanghe Street, Shenyang 110034, Liaoning Province, China. ericpazo@gmail.com; hewei05@hsyk.com.cn

Received: 2020-09-25　　Accepted: 2021-01-22

Abstract

● **AIM:** To assess the reliability of web-based version of ocular surface disease index in Chinese (C-OSDI) on clinically diagnosed dry eye disease (DE) patients.

● **METHODS:** A total of 254 Chinese participants (51% male, 129/254; mean age: 27.90±9.06y) with DED completed paper- and web-based versions of C-OSDI questionnaires in a randomized crossover design. Ophthalmology examination and DED diagnosis were performed prior to the participants being invited to join the study. Participants were randomly designated to either group A (paper-based first and web-based second) or group B (web-based first and paper-based second). Final data analysis included participants that had successfully completed both versions of the C-OSDI. Demographic characteristics, test-retest reliability, and agreement of individual items, subscales, and total score were evaluated with intraclass correlation coefficients (ICC), Spearman rank correlation, Wilcoxon test and Rasch analysis.

● **RESULTS:** Reliability indexes were adequate, Pearson correlation was greater than 0.8 and ICCs range was 0.827 to 0.982; total C-OSDI score was not statistically different between the two versions. The values of mean-squares fit statistics were very low compared to 1, indicating that the responses to the items by the model had a high degree of predictability. While comparing the favorability 72% (182/254) of the participants preferred web-based assessment.

● **CONCLUSION:** Web-based C-OSDI is reliable in assessing DED and correlation with the paper-based version is significant in all subscales and overall total score. Web-based C-OSDI can be administered to assess individuals with DED as participants predominantly favored online assessment.

● **KEYWORDS:** dry eye disease; ocular surface disease index; Rasch analysis; test-retest reliability; web-based questionnaire

INTRODUCTION

The worldwide prevalence of dry eye disease (DED) is assessed to be anywhere from 5% to 50% and reported to be highly prevalent in China and globally[1-4]. It continues to accelerate due to factors such as increase in multimedia screen usage, ageing population, and environmental factors. Several objective clinical tests are available for evaluating DED and due to the inherent variability of clinical features in DED favors the use of subjective assessment[5-6]. Among various DED questionnaires, ocular surface disease index (OSDI) is one of the most popular DED assessment questionnaire

following its conception in 1997[7]. DED can lead to discomfort and impaired vision, along with decrease in quality of life and work productivity[8-9]. DED is generally managed by artificial tears, warm compresses, omega-3 fatty acid, anti-inflammatory drugs, tetracyclines, secretagogues, intense pulse light (IPL), cholinergics, lacrimal plug, systemic immunosuppressives, eye lid massage and expression, serum tears, amniotic membrane biologic corneal bandage lens to name a few[10-13].

Traditionally, survey data in research has been collected using paper questionnaire[14]. However, in recent years, this method is facing challenges as multiple reports have stated that the response rates have declined by approximately 1% per annum in various countries[15]. Recent smartphones offer advanced computing and communication capability. Smartphone, along with web-based health-related services, is transforming clinical research settings. Since 2011, World Health Organization (WHO) has recognized the use mobile phones and other electronic devices for medical and public health practices under the umbrella of mHealth[16]. WeChat (Tencent Holdings Ltd., China) a smartphone application has a large user base in Asia and offers a real-time platform for sharing information. Currently, web-based questionnaires utilizing WeChat platform is rapidly growing in field of telemedicine. Additionally, multiple studies have utilized, and validated questionnaires administered via WeChat-based for health-related research and clinical practice[17-19]. In the last decade, due to the increased adoption of the internet, researchers have adopted in using web-based data entry and direct e-mail for collecting data[20-21]. Since internet-based questionnaires are increasingly gaining popularity in survey research, it is imperative to test the instruments' reliability. While scholars have explored methods of validation, administration, real-world considerations, and reliability of electronic versions of patient response outcomes measures (PROM)[22-23] and a growing number of clinical researchers support the use of web-based survey methods and instruments in reducing the hurdles of logistic associated with large sample size survey research[24-25]. Gwaltney et al's[26] Meta-analysis suggests that there is an overall high level of agreement between paper and electronic versions of health-related questionnaires. The review included peer-review articles from the fields of allergies, asthma, alcoholism, cardiology, diabetes, diabetes, gastrointestinal disease, pain assessment, psychiatry, and rheumatology. On the other hand, a study from the European Organization for Research utilizing Treatment of Cancer Quality of Life Questionnaire-Core 30 questionnaire reported a minor, nevertheless statistically significant differences of 3 to 7 mean score points (on a scale of 100-point) related with various methods of questionnaire administration[27]. Clayton et al[28] while comparing the equivalence of web-based and paper-based subscale of OSDI in

DED patients with a sample size of 68 participants, primarily consisting of Caucasian (n=43) demographic found no statistically significant difference between the paper-based and web-based version. However, the rigorous reliability of OSDI in the Chinese language has not been assessed. Additionally, it has been documented that health-related questionnaires scores have the potential to be culturally biased or neutral[29-30]. Therefore, this current study aims to assess the reliability of web-based OSDI questionnaire in Chinese language (C-OSDI) for evaluating the ocular surface health of DED participants in comparison with the paper-based administration of C-OSDI.

SUBJECTS AND METHODS

Ethical Approval　This study was reviewed and approved by the Institutional Review Board of He Eye Specialist Hospital, Shenyang, China and followed the tenets of the Declaration of Helsinki. All participants signed an informed consent after receiving a detailed explanation and possible consequences of participating the study. Data from the participants was collected between September 2019 to December 2019. In this randomized, crossover design study all participants completed both paper-based and web-based versions of the same C-OSDI questionnaire and were previously clinically diagnosed with DED. The C-OSDI questionnaire quantitatively measures the subjective symptoms of DED[31].

Two hundred and fifty-four consecutive consenting Chinese adults were enrolled into this prospective study. Diagnostic criteria: 1) At-least 1 of 6 symptoms of dryness, burning, sandiness, tiredness, discomfort of the eye and/or blurred vision with non-invasive tear break-up time (NITBUT) ≤10s[5]. 2) At-least 1 of 6 symptoms: dryness, burning, sandiness, tiredness, discomfort, and blurred vision accompanied by corneal fluorescein staining (CFS) score[32]. Inclusion criteria: full legal age, diagnosis of DED, follow study guidelines, read, and comprehend the questionnaire without help or support, complete the entire study protocol, and provide signed consent. Exclusion criteria: lacking the ability to give informed consent and participation in other studies (burden of participation), best-corrected visual acuity (BCVA) <20/20, previous ocular surgery or trauma, acute inflammation, blepharal dysraphism, history of blepharal and periorbital skin disease or allergies in the last 1-month, history of herpes zoster infection, rheumatic immune systemic diseases, pregnancy, breastfeeding, and use of photosensitive drugs/foods.

Experimental Design　The pretesting and pilot testing phase of the study consisted of evaluating usability, accessibility, and clarity of the web-based version of C-OSDI questionnaire by 3 ophthalmologist and 3 non-experts. This was conducted to assess the functionality of the web-based C-OSDI questionnaire, which was identical to the validated paper-based OSDI questionnaire.

Three hundred patents between 18 and 62 years of age voluntarily participated under the study. Participants were assessed for eligibility during their initial visit at the clinic and eligible participants were requested to enroll for the study. Participants enrolled for the study were asked to complete the questionnaires during their visit at the hospital under the observation of trained physicians. This study was designed as a 2-group (armed), prospective, crossover, randomized study. All participants were required to complete both versions of the same C-OSDI questionnaire (paper-based and web-based). Participants in group A, first completed the paper version followed by the web version on their personal smartphone. A 20-minute break was allotted between the paper and web sessions. While participants in group B filled out the web version followed by the paper version on their personal smartphone with 20min break between the two sessions (Figure 1). Taking into consideration that symptoms of DED can vary from day to day, and environmental conditions, both groups completed their both versions of the of the C-OSDI questionnaire on the same day with 20min break in between them. Additionally, we tried to mitigate the carry-over effect of the previous questionnaire with an interval break of 20min.

Randomization Participants were randomly enrolled to either group A or group B in a 1:1 ratio by a computer-generated randomization list with a specified seed and block size of 4. Prior to the administration of the questionnaires, written instructions were provided to all participants and was completed at the hospital under the supervision of three trained medical doctors (Fan QX, Zhang C, and You Y).

Questionnaire OSDI (Allergan Inc, Irvine, CA, USA) is a frequently used instrument to assess DED, which comprises of 12 items, and the final score range from 0 (no symptoms) to 100 (severe symptoms) points[7]. The 12 items of the questionnaire are sub-grouped into three subscales. Authors followed the guidelines for self-administered questionnaire design to reduce the risk of errors (Figure 2)[33]. Industry standard guidelines for translation were employed to achieve a scientifically accurate translation of the OSDI questionnaire from English to Chinese[34].

Clinical Assessment Full ophthalmic examination including BCVA (Snellen) at 4 m, corneal conjunctival examinations with slit lamp microscope and intraocular pressure (IOP) measurements were performed. Subjects were evaluated for DED before the administration of the C-OSDI questionnaires using the following assessments: NITBUT was measured using the Keratograph 5M (Oculus, Germany) and three times consecutively measurements were obtained. The median value was recorded used in the final analysis. Tear film lipid layer (TFLL) interferometry: DR-1 (Kowa, Nagoya, Japan) was performed to assess TFLL quality and graded from 1 to
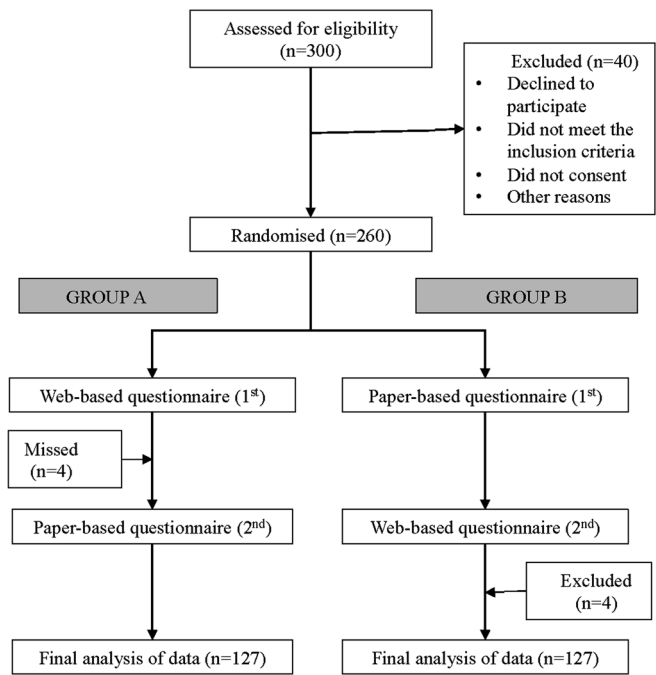

Figure 1 Study flow diagram.

Figure 2 Screenshot of web-based C-OSDI version.

5 according to Yokoi DE severity grading system[35]. CFS: the cornea after instilling fluorescein were evaluated using the

Efron system and was scored between 0 and 4[36]. Conjunctival hyperemia (CH) was assessed using the Keratograph 5M (Oculus, Germany). The redness scores (RS; accurate to 0.1 unit) generated by the device[37].

Calculations of Questionnaire Scores Questionnaire responses of the paper questionnaires were manually transferred into a password protected electronic spreadsheet by three trained medical doctors (Fan QX, Zhang C, and You Y) and responses were automatically transcribed after the participant concluded the questionnaire and downloaded into a password protected electronic spreadsheet. All questionnaires were checked for completeness in-terms of per-item basis and incomplete questionnaires were no included in the final analysis. The total C-OSDI score was obtained by the following official guidelines[7]. Direct comparison of individual items, subscales and total scores were the primary aim of this study. Following the completion of both versions of the C-OSDI questionnaire, participants were requested to choose whether they preferred paper-based, web-based or both versions of the questionnaire.

Statistical Analysis The sample size for this crossover design comparisons of means between the groups was derived from the equation: $(1-\rho)/2$; where ρ is an estimate of the expected correlation between the two modes of administration[26]. All statistical analyses for this study were conducted using SPSS (IBM, version 25). Questionnaires with missing values (items not filled) were not included in the final analysis. Descriptive sociodemographic characteristics of patients was determined by analyzing the frequency distribution of the overall data. Reliability, internal consistency, discrepancy of responses and the rate of consistency between paper-based and web-based responses were assessed. Reliability for the 12 individual items as well as for the 3 subscales (ocular symptoms, vision-related, and environmental triggers) and the total C-OSDI score under the OSDI guidelines were all calculated[7]. Shapiro-Wilk test inferred that the paired samples were not normally distributed. Due the ordinal nature of the data, Wilcoxon test was utilized to detect possible statistically significant differences in the test of parallel forms of reliability between the 12 items, 3 subscales and the total C-OSDI score. The mean values of the paper-based and web-based measures were calculated, Spearman rank correlation coefficient (Spearman ρ) for each item, subscale, and total score was used to assess consistency. To assess test-retest reliability, intraclass correlation coefficient (ICC; two-way random-effects model) was used. In this study, $P<0.05$ (2-tailed) was considered statistically significant differences (alpha=0.05). The psychometric properties of C-OSDI questionnaire were analyzed utilizing Rasch analysis. Further information and background to Rasch analysis in ophthalmic research by McNeely *et al*[38] is recommended.

RESULTS

The final analysis included 254 DED participants diagnosed under the criteria put forth by dye eye workshop (DEWS) the patients were classified as DED[32]. Initially, 300 patients were assessed for eligibility, however, 46 participants were excluded (Figure 1). Kemainder in groups A (*n*=127) and B (*n*=127) completed their questionnaires consecutively. There were no significant differences in response behavior, sociodemographic status, or therapy setting between the participants in either groups, the two groups were pooled in the final analysis. Table 1 shows the sociodemographic and clinical characteristics of the study group. The study consisted of 129 (51%) male and 125 (49%) female outpatients with a mean age of 27.90±9.06y. The mean BCVA value for both eyes were -0.10±0.01 logMAR, mean IOP for both eyes were 14.09±1.40 mm Hg. A total of 258 paper and web-based questionnaires were available, out of which 254 had completed all 12 items and therefore only 254 were selected for the final data analysis. Additionally, in group A, four patients had failed to complete both version of the questionnaires (Figure 1). Parallel test-retest reliability for all paper-based and web-based scale scores were assessed for each item, subscale, and total score (Table 2). Item 11 was found to have the lowest level of agreement (Spearman ρ=0.806, ICC=0.824). In the present study, standard deviations (SD) for total C-OSDI score for paper-based was 12.78 and web-based was 12.43. To assess the effects of the type of questionnaire (web-based or paper-based) and the sequence of administration, random-effects 2-way ANOVA was used. Since the order of administration was balanced (50%; *n*=172), no interaction was found among the type and administration order of questionnaires. Similarly, there was no effect of the type of questionnaire or administration order. As shown in Table 2, reliability indexes were within the acceptable range, with Pearson correlations greatest for item 1 (0.965) and intraclass correlation ranging from 0.824 (item 11) to 0.989 (total C-OSDI score). Mean scores were significantly different for item 5 and subscale 1 score according to Wilcoxon signed rank tests for paired samples. However, the total C-OSDI score showed no significant difference. Table 2 also displays the Spearman rho correlation values between all individual items, subscales and total score. All 12 items, subscale and total score demonstrated a comparable correlation (>0.8). The distribution of web-based and paper-based total C-OSDI scores (0-100 points, where higher values reflect a worse state) are illustrated as box plots (Figure 3). The marginally higher total mean web-based C-OSDI score (29.87 points) *vs* paper-based (29.63 points) can be ascribed to the few outlies depicted in the boxplot and the difference was not found to be statistically significant (*P*=0.09; Table 3). The whisker of the web-based C-OSDI boxplot interquartile range was

**Table 1 Demographic and clinical information on study participants**

Variable	Overall group (*n*=254)
Age (y)	
Mean (SD)	27.90 (9.06)
IQR	20 to 23
Gender, *n* (%)	
Male	129 (50.8)
Female	125 (49.2)
Education level, *n* (%)	
Primary	0
Secondary	3 (1.2)
Tertiary	251 (98.8)
Visual acuity (logMAR)	
Mean, SD	-0.10 (0.01)
IQR	-1 to -1
Vision correction, *n* (%)	
Spectacles	108 (42.5)
Contact lens	79 (31.1)
Refractive surgery	28 (11)
None	39 (15.4)
NITBUT (s)	
Mean, SD	6.39 (2.00)
IQR	4.89 to 8.06
Corneal staining (Efron scale)	
Mean, SD	0.51 (0.74)
IQR	0-1
Tear lipid layer (DR-1 scale)	
Mean, SD	2.63 (0.70)
IQR	2 to 3
Conjunctival hyperaemia (Oculus scale)	
Mean, SD	1.06 (0.37)
IQR	0.80 to 1.30
IOP (mm Hg)	
Mean, SD	14.09 (1.40)
IQR	14 to 15

logMAR: Logarithm of the minimum angle of resolution; NITBUT: Non-invasive tear breakup time; IOP: Intraocular pressure; IQR: Interquartile range.

within the paper-based version. Bland-Altman chart (Figure 4) illustrates that the individual total C-OSDI scores of the two versions of the questionnaire are mostly close to one another. However, 13 out of 254 participants had their total C-OSDI scores beyond the SD on the web-based version. Figure 5 illustrates a positive correlation between total C-OSDI scores of the two questionnaires. Wilcoxon sign rank test was used to assess parallel reliability in single items, subscale, and total score of C-OSDI (Table 3). No systematic location difference was observed for continuous variables except for item 5 (poor vision) and subscale 1. However, most of the responses to the items had same response (ties) in both versions of
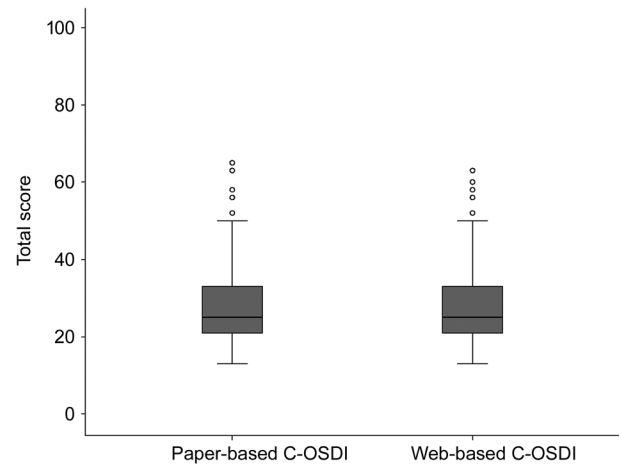
Figure 3 Boxplot distribution of web-based and paper-based C-OSDI total scores.
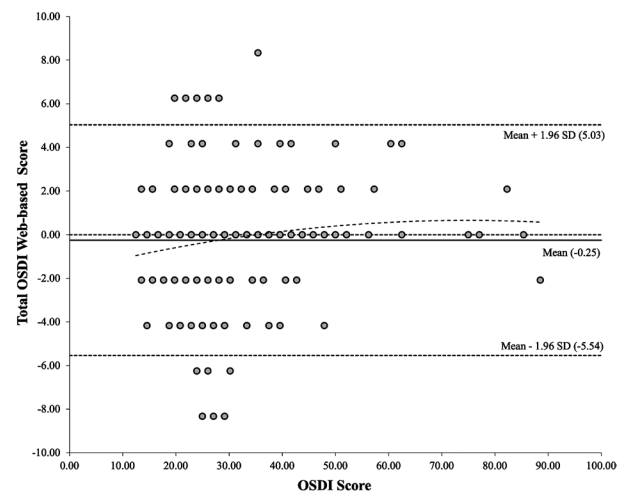
Figure 4 Bland-Altman analysis for clinical agreement between the web-based C-OSDI and paper-based C-OSDI final scores revealed a clinical difference (bias) of -0.25 units.
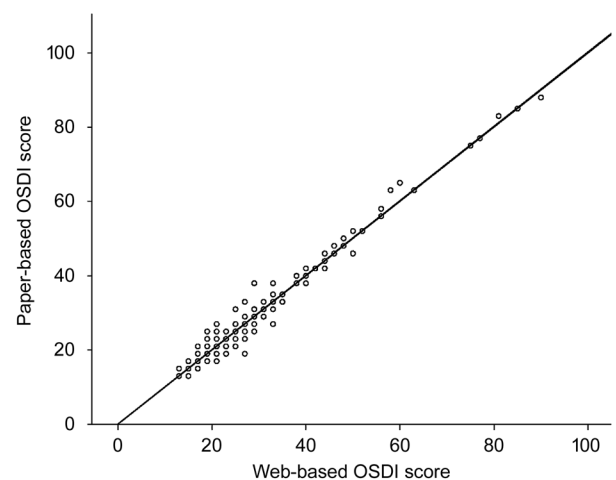
Figure 5 Correlation between web-based and paper-based OSDI total scores.

the questionnaires. These findings suggest a high parallel reliability. A moderately statistically significant difference could only be identified in subscale 1 (ocular symptoms).

Table 2 Parallel test-retest reliability of single items, subscale and total score

Items	Paper-based		Web-based		ICC (95%CI)	r (ρ)	P
	Mean (SD)	IQR	Mean (SD)	IQR			
Ocular symptoms (subscale 1)							
Item 1	1.09 (0.99)	0-1	1.08 (0.98)	0.75-1	0.982 (0.977-0.986)	0.965	<0.001[a]
Item 2	1.18 (0.92)	1-1	1.19 (0.91)	1-1	0.908 (0.884-0.927)	0.842	<0.001[a]
Item 3	1.14 (0.91)	1-1	1.13 (0.89)	1-1	0.946 (0.932-0.958)	0.911	<0.001[a]
Item 4	1.32 (0.97)	1-2	1.38 (0.97)	1-2	0.841 (0.80-0.873)	0.813	<0.001[a]
Item 5	1.26 (1.11)	0-2	1.41 (1.08)	1-2	0.85 (0.812-0.881)	0.825	<0.001[a]
Total subscale 1	12.49 (6.90)	8.33-16.67	12.90 (6.58)	8.33-16.67	0.951 (0.938-0.961)	0.92	<0.001[a]
Vision-related function (subscale 2)							
Item 6	1.18 (1.11)	0-2	1.19 (1.08)	0-2	0.959 (0.947-0.968)	0.931	<0.001[a]
Item 7	0.89 (0.96)	0-1	0.90 (0.96)	0-1	0.942 (0.926-0.954)	0.925	<0.001[a]
Item 8	1.32 (0.98)	1-2	1.28 (0.97)	1-2	0.827 (0.784-0.862)	0.815	<0.001[a]
Item 9	1.17 (0.95)	1-2	1.16 (0.92)	1-2	0.867 (0.833-0.895)	0.848	<0.001[a]
Total subscale 2	9.50 (5.83)	6.25-12.5	9.43 (5.71)	6.25-12.5	0.959 (0.947-0.968)	0.936	<0.001[a]
Environmental triggers (subscale 3)							
Item 10	1.16 (0.88)	1-1	1.11 (0.85)	1-1	0.854 (0.817-0.884)	0.82	<0.001[a]
Item 11	1.15 (0.85)	1-1	1.18 (0.81)	1-1	0.824 (0.78-0.859)	0.806	<0.001[a]
Item 12	1.32 (0.86)	1-2	1.30 (0.82)	1-2	0.855 (0.818-0.885)	0.854	<0.001[a]
Total subscale 3	7.57 (4.06)	4.17-10.42	7.49 (3.99)	4.17-10.42	0.927 (0.908-0.943)	0.882	<0.001[a]
Total OSDI score	29.63 (12.78)	21-35	29.87 (12.43)	21-33	0.989 (0.986-0.991)	0.935	<0.001[a]

OSDI: Ocular surface disease index. [a]Statistically significant difference.

Table 3 Rank test-retest reliability of single items, subscale and total score (Wilcoxon rank test)

Parameters	Negative ranks	Positive ranks	Ties	P
Ocular symptoms (subscale 1)				
Item 1	5[a]	4[b]	245[c]	0.379
Item 2	15[a]	18[b]	221[c]	0.362
Item 3	11[a]	8[b]	235[c]	0.67
Item 4	19[a]	28[b]	207[c]	0.078
Item 5	16[a]	43[b]	195[c]	<0.001[d]
Subscale 1 score	41[a]	69[b]	144[c]	0.013[d]
Vision-related function (subscale 2)				
Item 6	11[a]	11[b]	232[c]	0.841
Item 7	13[a]	14[b]	227[c]	0.847
Item 8	23[a]	18[b]	213[c]	0.316
Item 9	18[a]	16[b]	220[c]	0.897
Subscale 2 score	45[a]	36[b]	173[c]	0.245
Environmental triggers (subscale 3)				
Item 10	27[a]	18[b]	209[c]	0.123
Item 11	26[a]	33[b]	195[c]	0.31
Item 12	20[a]	18[b]	216[c]	0.454
Subscale 3 score	52[a]	43[b]	159[c]	0.552
Total OSDI score	66[a]	82[b]	106[c]	0.09

OSDI: Ocular surface disease index. [a]Web-based score<paper-based score; [b]Web-based score>paper-based score; [c]Web-based score=paper-based score; [d]Statistically significant difference.

Additionally, the IQR of the "Item 5" for the paper-based and web-based questionnaires were also different (0-2 and 1-2 respectively). Although the web-based total mean score was slightly higher by 0.24 points but was not statistically significant different in comparison to the paper-based version. The most used metrics, in Rasch Analysis, to assess the randomness of the response to items are the mean-squares fit statistics (Outfit MNSQ and Infit MNSQ). The values of both Outfit MNSQ and Infit MNSQ are expected to be around 1, and any value far away from 1 suggests either a low or high degree of randomness in the response to the items, which could jeopardize the quality of the fitted model. Low values of Outfit MNSQ and Infit MNSQ highlight that the responses to the items are easily predictable, and this could result into the overfitting of the model. High values of Outfit MNSQ and Infit MNSQ points out that the responses to the items are very unpredictable, which could result into the misfitting of the model. The infits and outfit mean square statistics are below the 1.5 threshold and above the 0.5 threshold for all the items in both web-based and paper-based OSDI questionnaire. Therefore, all the items are relevant and capture the underlaying latent trait (Table 4). On the other hand, the items characteristic curves for all the all items in both web-based (Figure 6) and paper-based (Figure 7) OSDI questionnaire show that the rating 2 and 3 are redundant and only three
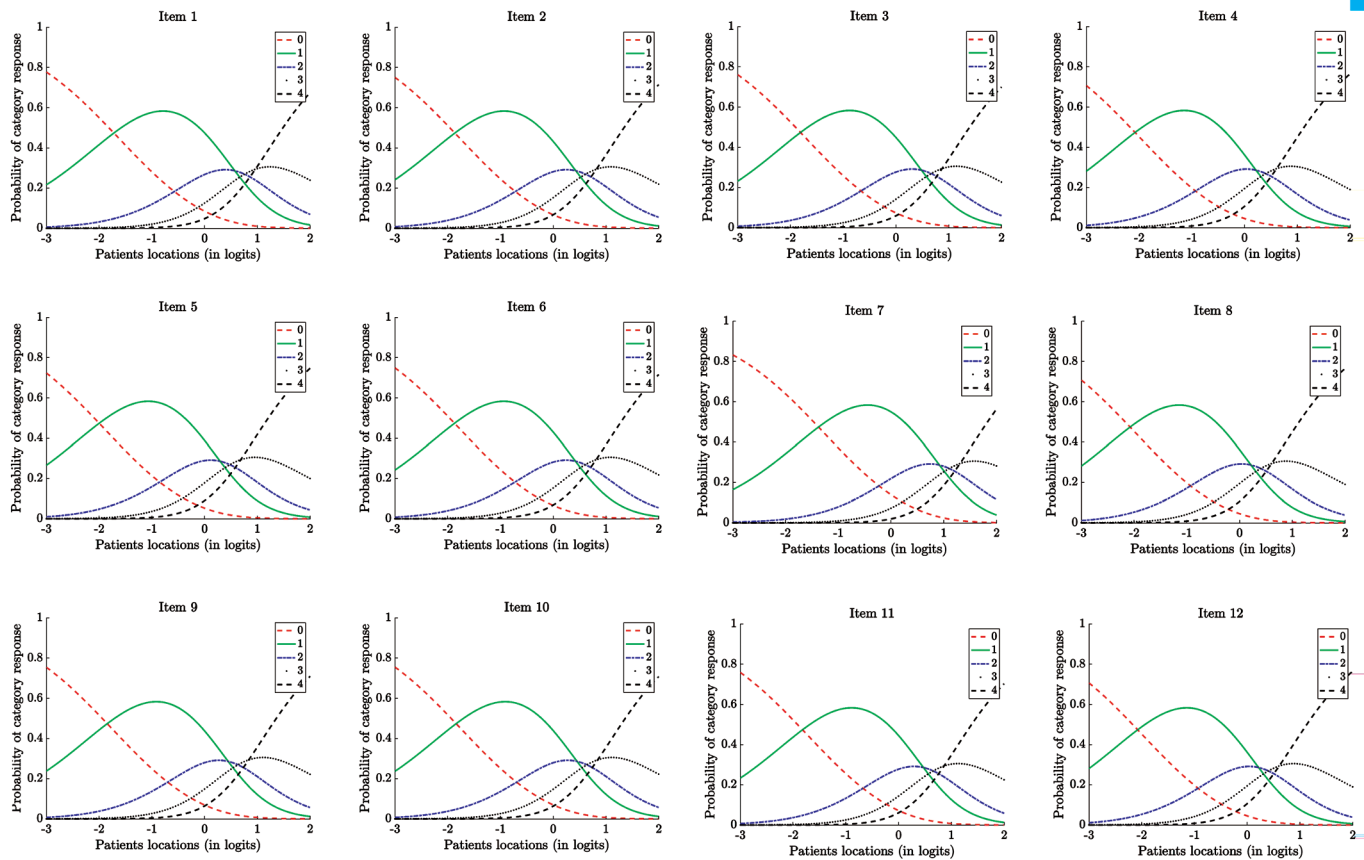
Figure 6 Items characteristic curves of web-based C-OSDI items.

rating scales, namely 0, 1, and 4 are enough to capture the underlaying latent trait.

Results on the user preference survey were analyzed separately (Table 5). Of the 254 patient surveys that were completed, 9% (24/254) reported that they preferred the paper-based OSDI questionnaire, while 72% (182/254) preferred the web-based questionnaire and 19% (48/254) preferred both versions of the C-OSDI questionnaire. Regarding the version preference by participants, there were no significant associations found with age, gender, education level or level of DED severity. The median time to complete the paper-based C-OSDI was 109.5s, while for the web-based C-OSDI was 61s.

DISCUSSION

This study assessed the test-retest reliability of self-administered C-OSDI questionnaire *via* web-based user-interface. In accordance with the international guidelines, the validation of a web-based version must demonstrate equivalent measurement properties to its predecessor. This is can be measured by correlation and intraclass correlation. In general, reliability was found to be good for the web-based C-OSDI questionnaire as measured with ICC and Wilcoxon sign rank test. Spearman rho correlation analysis demonstrated that the mean differences were close to zero, implying high reliability of the web-based version of C-OSDI. Additionally, Rasch analysis revealed high degree of responses and predictability of the items.

Table 4 Infit and outfit mean square values for web-based and paper-based questionnaires

Items	Web-based OSDI		Paper-based OSDI	
	Infit MNSQ	Outfit MNSQ	Infit MNSQ	Outfit MNSQ
Item 1	1.06	1.04	1.02	0.99
Item 2	0.8	0.77	0.76	0.74
Item 3	0.91	0.91	0.92	0.93
Item 4	0.82	0.85	0.79	0.79
Item 5	1.38	1.34	1.47	1.4
Item 6	1.19	1.19	1.21	1.22
Item 7	1.29	1.26	1.29	1.26
Item 8	1.05	1.1	1.01	1.06
Item 9	1	1.02	0.93	0.96
Item 10	0.89	0.91	0.9	0.94
Item 11	0.76	0.8	0.84	0.89
Item 12	0.89	0.96	0.96	1.03

Patrick *et al*'s[39] Meta-analysis stated that an average correlation between paper-based and electronic administration was 0.90 without significant changes from various research relying on ICC or weighted kappa. Findings from our current study indicate that test-retest reliability, as measured by an ICC of the C-OSDI web-based version questionnaire achieved good (>0.80) results for subscales and total score. The Rasch analysis results suggests that all the 12 items contribute to capture the OSDI latent trait. Hence, they are all useful and
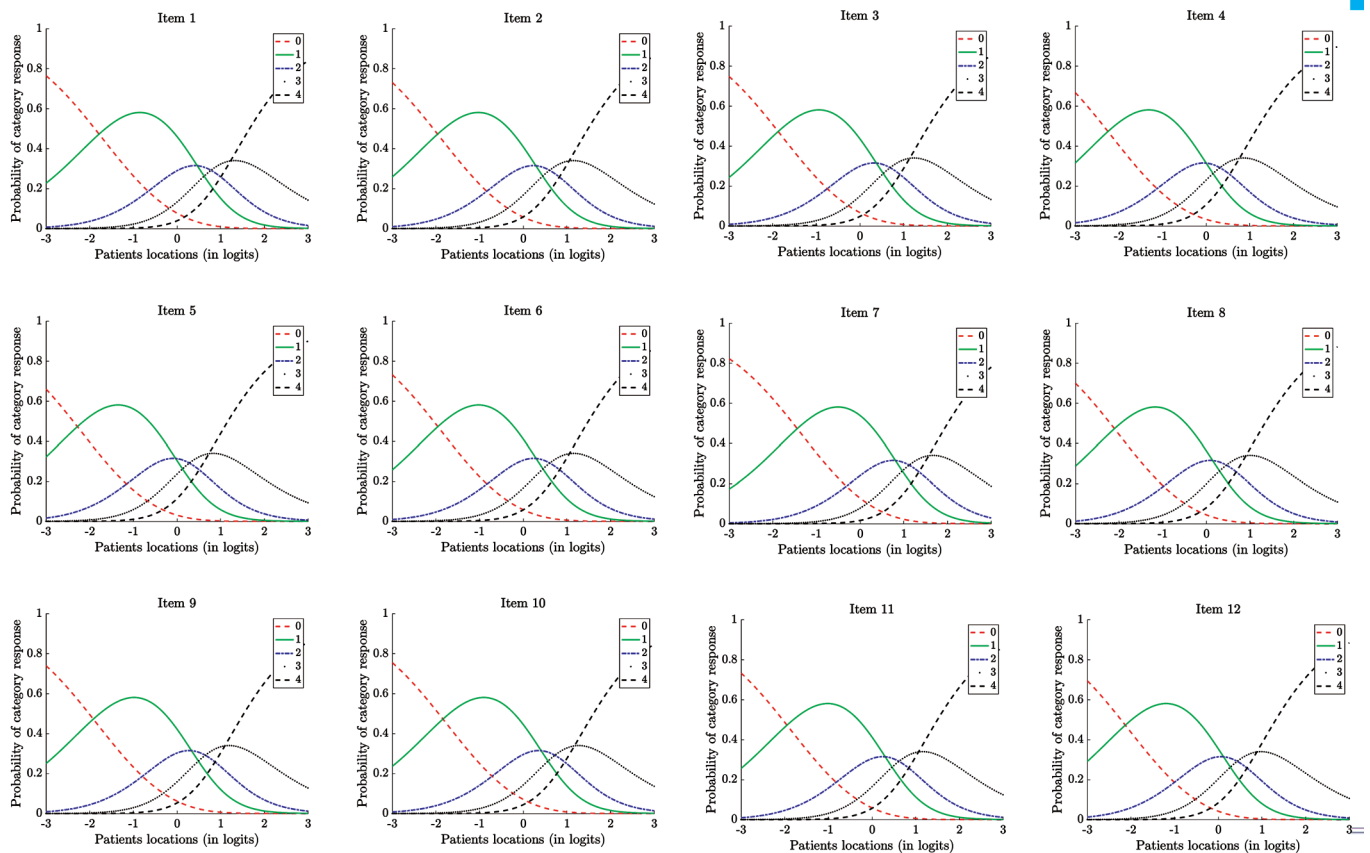
Figure 7 Items characteristic curves of paper-based C-OSDI items.

Table 5 User preference and time analysis

Questionnaire preference	Male (*n*)	Female (*n*)	Total (*n*, %)	Time of completion (s)	
				Mean (SD)	Median
Paper-based OSDI	13	11	24 (9)	134.24 (65.19)	109.5
Web-based OSDI	90	92	182 (72)	71.5 (34.50)	61
Both	26	22	48 (19)		

OSDI: Ocular surface disease index.

should be kept in the questionnaire. However, the items characteristic curves for the 12 items/questions, in both web-based and paper-based questionnaires, showed that only three rating scales are adequate, instead five. Additionally, good face validity was demonstrated as 72% of the respondents' preferred using the web-based version over the paper-based version when assessing their DED. Surprisingly in this current study, only 28% of the patients indicated that they favored the paper-based C-OSDI questionnaire. This could possibility be a biased indicator as the primary objective was to validate the web-based version of C-OSDI or that the transition to internet-based assessment is well accepted in China due to widescale mobile internet coverage in the last half decade. One of the important follow-ups of this study will be to continue a longitudinal follow-up of C-OSDI scores to assess whether web-based assessment can be of value in routine clinical DED care. The continuation of this study will allow longitudinal follow-up of electronically administered self-reported DED

scores and determine its value for clinicians and researchers. Screening test such as the OSDI questionnaire enables clinicians' early discovery of ocular surface alterations in a population allowing for prompt treatment, care, and monitoring. A screening test should have the benefits of being quick, easy to use, inexpensive and the ability to be administered by nonspecialized personnel. The OSDI one such popular tool clinical practice for DED. The present findings indicate that assessing DED using the web-based version was easy and reliable, and importantly, fulfils the criteria for migrating the need for a paper-based C-OSDI. However, migration to web-based C-OSDI could present some limitations since the variation of difficulty between some items were discovered. The mean of score of item and subscale 1 were found to be significantly different in the Wilcoxon rank test as the individual participants response were significantly different to the symptom severity. Therefore, this could further influence the final score[40]. Finally, it is possible that the

randomized crossover test re-test studies design can facilitate carryover effect. While interpreting the findings of our study, it must be taken into consideration that randomized crossover design test re-test study design suffers from internal validity, but the within-patient design offers better statistical power and reduces requirements for a large sample size. However, to compensate for carryover effect participants were given adequate time between test re-test. McNeely et al[38] suggest that Rasch analysis validated questionnaires such as the OSDI are centered on a single cohort and therefore at certain situations might not derive the most accurate assessment. However, further investigation is needed and will be carried out to validate these findings regarding the difficulty of items on Chinese version of OSDI. Although participants completed the web-based questionnaire in a shorter time than the paper-based questionnaire, it should be noted that patients that were administered the paper-based or web-based questionnaire first might have memorized their responses, however this could not be quantitatively assessed. The follow-up data to analyze the responsiveness of the web-based C-OSDI will be assessed in a forthcoming study.

To summarize, the web-based C-OSDI shows good reliability and could possibly mitigate the use of paper-based C-OSDI in assessing and monitoring individuals with DED. Additionally, good test-retest reliability suggests that web-based C-OSDI can be used for clinical studies that have a relatively moderate sample size.

ACKNOWLEDGEMENTS

REFERENCES

1 Stapleton F, Alves M, Bunya VY, Jalbert I, Lekhanont K, Malet F, Na KS, Schaumberg D, Uchino M, Vehof J, Viso E, Vitale S, Jones L. TFOS DEWS II epidemiology report. *Ocul Surf* 2017;15(3):334-365.

2 Song PG, Xia W, Wang ML, Chang XL, Wang JP, Jin S, Wang JW, Wei W, Rudan I. Variations of dry eye disease prevalence by age, sex and geographic characteristics in China: a systematic review and meta-analysis. *J Glob Health* 2018;8(2):020503.

3 Ma JF, Pazo EE, Zou ZH, Jin F. Prevalence of symptomatic dry eye in breast cancer patients undergoing systemic adjuvant treatment: a cross-sectional study. *Breast* 2020;53:164-171.

4 Tang YL, Cheng YL, Ren YP, Yu XN, Shentu XC. Metabolic syndrome risk factors and dry eye syndrome: a Meta-analysis. *Int J Ophthalmol* 2016;9(7):1038-1045.

5 Bron AJ, Abelson MB, Ousler G, Pearce E, Tomlinson A, Yokoi N, Smith JA, Begley C, Caffery B, Nichols K, Schaumberg D. Methodologies to diagnose and monitor dry eye disease: report of the Diagnostic Methodology Subcommittee of the International Dry Eye WorkShop (2007). *Ocul Surf* 2007;5(2):108-152.

6 Xu L, Zhang W, Zhu XY, Suo T, Fan XQ, Fu Y. Smoking and the risk of dry eye: a Meta-analysis. *Int J Ophthalmol* 2016;9(10):1480-1486.

7 Dougherty BE, Nichols JJ, Nichols KK. Rasch analysis of the ocular surface disease index (OSDI). *Invest Ophthalmol Vis Sci* 2011;52(12):8630-8635.

8 McDonald M, Patel DA, Keith MS, Snedecor SJ. Economic and humanistic burden of dry eye disease in Europe, north America, and Asia: a systematic literature review. *Ocul Surf* 2016;14(2):144-167.

9 Kobashi H, Kamiya K, Sambe T, Nakagawa R. Factors influencing subjective symptoms in dry eye disease. *Int J Ophthalmol* 2018;11(12):1926-1931.

10 Jones L, Downie LE, Korb D, Benitez-Del-Castillo JM, Dana R, Deng SX, Dong PN, Geerling G, Hida RY, Liu Y, Seo KY, Tauber J, Wakamatsu TH, Xu JJ, Wolffsohn JS, Craig JP. TFOS DEWS II management and therapy report. *Ocul Surf* 2017;15(3):575-628.

11 Fan QX, Pazo EE, You Y, Zhang C, Zhang CG, Xu L, He W. Subjective quality of vision in evaporative dry eye patients after intense pulsed light. *Photobiomodul Photomed Laser Surg* 2020;38(7):444-451.

12 Pazo EE, Huang H, Fan QX, Zhang C, Yue Y, Yang LT, Xu L, Moore JE, He W. Intense pulse light for treating post-LASIK refractory dry eye. *Photobiomodul Photomed Laser Surg* 2021;39(3):155-163.

13 Bhargava R, Kumar P, Kumar M, Mehra N, Mishra A. A randomized controlled trial of omega-3 fatty acids in dry eye syndrome. *Int J Ophthalmol* 2013;6(6):811-816.

14 Campbell N, Ali F, Finlay AY, Salek SS. Equivalence of electronic and paper-based patient-reported outcome measures. *Qual Life Res* 2015;24(8):1949-1961.

15 Morton LM, Cahill J, Hartge P. Reporting participation in epidemiologic studies: a survey of practice. *Am J Epidemiol* 2006;163(3):197-203.

16 Logan AG, Jassal SV. Building a stronger care loop through mHealth technology. *JAMA Intern Med* 2018;178(6):809-811.

17 Wang YH, Zhao LM, Liu ZY, Li XM. Perioperative management by WeChat under the haze of COVID-19. *Int J Ophthalmol* 2020;13(7):1161-1163.

18 Yu Q, Xu L, Li L, Zhi M, Gu Y, Wang X, Guo H, Li Y, Fan Y, Yang B, Xue M, Lv M, Xu D, Zhang H, Li Y, Song Y, Deng Q, Huang X, Zhong J, Hu W, Zhu Y, Wang X, Cai J, Chen Y. Internet and WeChat used by patients with Crohn's disease in China: a multi-center questionnaire survey. *BMC Gastroenterol* 2019;19(1):97.

19 Yu A, Tu RX, Shao X, Pan AP, Zhou KJ, Huang JH. A comprehensive Chinese experience against SARS-CoV-2 in ophthalmology. *Eye Vis (Lond)* 2020;7:19.

20 De Castro A, Macías JA. SUSApp: a mobile app for measuring and comparing questionnaire-based usability assessments. In: ACM International Conference Proceeding Series. *Association for Computing Machinery* 2016.

21 Schleyer TK, Forrest JL. Methods for the design and administration of web-based surveys. *J Am Med Inform Assoc* 2000;7(4):416-425.

22 McMaster HS, LeardMann CA, Speigle S, Dillman DA, Millennium Cohort Family Study Team. An experimental comparison of web-push vs. paper-only survey procedures for conducting an in-depth health survey of military spouses. *BMC Med Res Methodol* 2017;17(1):73.

23 Blumenberg C, Menezes AMB, Gonçalves H, Assunção MCF, Wehrmeister FC, Barros FC, Barros AJD. The role of questionnaire length and reminders frequency on response rates to a web-based epidemiologic study: a randomised trial. *Int J Soc Res Methodol* 2019;22(6):625-635.

24 Braekman E, Berete F, Charafeddine R, Demarest S, Drieskens S, Gisle L, Molenberghs G, Tafforeau J, Van der Heyden J, Van Hal G. Measurement agreement of the self-administered questionnaire of the Belgian Health Interview Survey: paper-and-pencil versus web-based mode. *PLoS One* 2018;13(5):e0197434.

25 Terluin B, Brouwers EPM, Marchand MAG, de Vet HCW. Assessing the equivalence of web-based and paper-and-pencil questionnaires using differential item and test functioning (DIF and DTF) analysis: a case of the Four-Dimensional Symptom Questionnaire (4DSQ). *Qual Life Res* 2018;27(5):1191-1200.

26 Gwaltney CJ, Shields AL, Shiffman S. Equivalence of electronic and paper-and-pencil administration of patient-reported outcome measures: a meta-analytic review. *Value Heal* 2008;11(2):322-333.

27 Cheung YB, Goh C, Thumboo J, Khoo KS, Wee J. Quality of life scores differed according to mode of administration in a review of three major oncology questionnaires. *J Clin Epidemiol* 2006;59(2):185-191.

28 Clayton JA, Eydelman M, Vitale S, Manukyan Z, Kramm R, Datiles M 3rd, Temple A, Murphy E, Kim J, Hilmantel G, Rorer E, Hammel K, Ferris F 3rd. Web-based versus paper administration of common ophthalmic questionnaires: comparison of subscale scores. *Ophthalmology* 2013;120(10):2151-2159.

29 Lewis G, Araya RI. Is the General Health Questionnaire (12 item) a culturally biased measure of psychiatric disorder? *Soc Psychiatry Psychiatr Epidemiol* 1995;30(1):20-25.

30 Xue WW, Zhang P, Zou HD. Responsiveness and minimal clinically important difference of the Chinese version of the low vision quality of life questionnaire after cataract surgery. *Int J Ophthalmol* 2019;12(3):504-509.

31 Asiedu K, Kyei S, Mensah SN, Ocansey S, Abu LS, Kyere EA. Ocular surface disease index (OSDI) versus the standard patient evaluation of eye dryness (SPEED):a study of a nonclinical sample. *Cornea* 2016;35(2):175-180.

32 Craig JP, Nichols KK, Akpek EK, Caffery B, Dua HS, Joo CK, Liu ZG, Nelson JD, Nichols JJ, Tsubota K, Stapleton F. TFOS DEWS II definition and classification report. *Ocul Surf* 2017;15(3):276-283.

33 Jenkins CR, Dillman DA. Towards a theory of self-administered questionnaire design. *Bureau of the Census* 1995;165-196.

34 Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *J Clin Epidemiol* 1993;46(12):1417-1432.

35 Hosaka E, Kawamorita T, Ogasawara Y, Nakayama N, Uozato H, Shimizu K, Dogru M, Tsubota K, Goto E. Interferometry in the evaluation of precorneal tear film thickness in dry eye. *Am J Ophthalmol* 2011;151(1):18-23.e1.

36 Efron N, Morgan PB, Katsara SS. Validation of grading scales for contact lens complications. *Ophthalmic Physiol Opt* 2001;21(1):17-29.

37 Best N, Drury L, Wolffsohn JS. Clinical evaluation of the oculus keratograph. *Cont Lens Anterior Eye* 2012;35(4):171-174.

38 McNeely RN, Moutari S, Arba-Mosquera S, Verma S, Moore JE. An alternative application of Rasch analysis to assess data from ophthalmic patient-reported outcome instruments. *PLoS One* 2018;13(6):e0197503.

39 Patrick DL, Burke LB, Gwaltney CJ, Leidy NK, Martin ML, Molsen E, Ring L. Content validity-establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO Good Research Practices Task Force report: part 2-assessing respondent understanding. *Value Health* 2011;14(8):978-988.

40 Sullivan BD, Crews LA, Messmer EM, Foulks GN, Nichols KK, Baenninger P, Geerling G, Figueiredo F, Lemp MA. Correlations between commonly used objective signs and symptoms for the diagnosis of dry eye disease: clinical implications. *Acta Ophthalmol* 2014:161-166.