Article

Development and Validation of the Singapore Thyroid Eye Disease Quality of Life Questionnaire

Melissa H. Y. Wong[1–4], Eva Fenwick[2,3], Ai Tee Aw[1,2], Ecosse L. Lamoureux[2,3], and Lay Leng Seah[1–4]

[1] Singapore National Eye Centre (SNEC), Singapore, Singapore
[2] Singapore Eye Research Institute, Singapore, Singapore
[3] Duke NUS Medical School, Singapore, Singapore
[4] Yong Loo Lin School of Medicine, Singapore, Singapore

Correspondence: Melissa H. Y. Wong, Singapore National Eye Centre (SNEC), Cataract and Comprehensive Ophthalmology, 11 Third Hospital Ave, Singapore, 168751, Singapore. e-mail: melissa.wong.h.y@singhealth.com.sg

Received: 5 February 2018
Accepted: 28 July 2018
Published: 1 October 2018

Keywords: quality of life; thyroid eye disease; Rasch analysis

Citation: Wong MHY, Fenwick E, Aw AT, Lamoureux EL, Seah LL. Development and validation of the Singapore Thyroid Eye Disease quality of life questionnaire. Trans Vis Sci Tech. 2018;7(5):14, https://doi.org/10.1167/tvst.7.5.14

Purpose: Current instruments to assess thyroid eye disease (TED) quality of life (QoL) were not developed using modern psychometric theory and may not be applicable to Asian populations. Therefore, we developed a psychometrically robust questionnaire, the Singapore Thyroid Eye Disease Quality of Life questionnaire (STED-QoL), for assessing QoL in Asian patients.

Methods: This cross-sectional study was conducted at the Singapore National Eye Centre between 2012 and 2015. In Phase 1, content for the questionnaire was developed using qualitative methods. A total of 20 patients participated in three different focus groups. Thematic analysis was conducted to identify relevant themes from which 12 items, rated on a 5-point Likert-type scale, were generated. In Phase 2, the pilot instrument was administered to 59 TED patients and psychometric assessment of the STED-QoL was conducted using Rasch analysis.

Results: After collapsing categories from five to four and deleting two misfitting items, we generated a 10-item STED-QoL befitting the Rasch model. The scale showed good criterion validity, with scores decreasing as severity of TED worsened: mild (1.78 logits), moderate (0.27 logits), and severe (0.92 logits). A 'Psychosocial' subscale also had adequate psychometric properties and psychosocial scores were significantly worse in those who underwent surgery for TED compared to those who had not (0.41 vs. 1.82 logits, $P = 0.021$).

Conclusions: The STED-QoL is a robust 10-item questionnaire specifically developed to measure the impact of TED on QoL and psychosocial well-being in an Asian population.

Translational Relevance: QoL assessment is important for holistic management of TED patients.

Introduction

Thyroid eye disease (TED) can be an incapacitating condition, with patients suffering a spectrum of clinical problems from mild dry eye symptoms to severe sight-threatening conditions, such as optic nerve compression and exposure keratopathy.[1] Facial disfigurement as a result of lid retraction, chemosis, squint, and proptosis often is a significant social embarrassment resulting in substantial psychologic burden.[2] Treatments include controlling and stabilizing the thyroid disease, followed by subsequent orbital decompression and then squint and eyelid surgery in those with moderate and severe disease. Patients often require multiple surgeries and must attend multiple follow-up appointments. Consequently, TED has been shown to have a substantial impact on patient quality of life (QoL), including reduced participation in activities of daily living and poorer emotional well-being.[4,5]

Using QoL questionnaires has become the standard approach for assessing patient-centered outcomes in healthcare.[2,3,5,6] Similarly, patient reported outcomes (PROs) now are required by regulatory agencies, such as the Food and Drug Administration and National Institute of Clinical Excellence, both in the United States, in clinical trials to assess the effectiveness of novel treatments from the patients' perspective. QoL is a broad concept that assesses the impact of an illness on patients' physical, mental, social, and functional health.[7] Disease-specific factors, such as symptoms, treatment burden, inconvenience, and task-specific difficulty, have resulted in the development of several condition-specific QoL questionnaires in Ophthalmology,[8–10] including two for TED, namely the 15-item Graves' Ophthalmopathy-QOL (GO-QOL) and the 3-item TED-QOL.[6,11] However, as yet there is no TED-specific questionnaire to study the QoL impact of TED in adults in Asia. While translations of current TED-specific questionnaires have been used in several studies,[12,13] they may lack cultural specificity as they were not developed locally. This is important given that racial and ethnic differences in the use of eye care services and perception of eye-related care have been demonstrated in epidemiologic studies in Singapore.[14,15] Moreover, differences in living conditions and cultural and environmental habits between Asian and Western populations may mean that the QoL impact of TED also is different in Asian populations.

Furthermore, to our knowledge none of the current TED-specific QoL questionnaires has been validated using modern psychometric theory. While classical test theory (CTT) is useful in the early phases of instrument development, the benefits of item response theory, such as Rasch analysis, can provide a more robust and comprehensive psychometric evaluation than CTT.[16,17] The lack of a sophisticated TED-specific patient reported outcome measure restricts our understanding of the full impact of TED and related treatments on QoL.

We developed a TED-specific QoL questionnaire, the Singapore Thyroid Eye Disease Quality of Life questionnaire (STED-QoL), to assess the functional and psychosocial impact of TED from the patient's perspective. We described herein the content development phase of the STED-QoL and its psychometric evaluation using Rasch analysis.

Methods

Study Design and Population

This cross-sectional, two-phase, mixed method study was conducted between February 2012 and December 2015 at the Singapore National Eye Centre, a tertiary eye center in Singapore. The study adhered to the tenets of the Declaration of Helsinki and received ethical approval by the local institutional review board (Ref 2012/056/A).

Phase 1: Content Development using Qualitative Methods

Focus groups were conducted with TED patients to understand the breadth and depth of QoL issues associated with TED and its treatment that contributed to item generation. Patients were recruited from a single TED clinic, and were aged 21 years or older, had a diagnosis of any TED, and no significant hearing or cognitive impairment. Each focus group had six to eight participants and three focus groups were conducted using a thematic saturation technique.[18] Thematic saturation is when no new themes emerge from subsequent focus groups or interviews, suggesting that the topic has been covered comprehensively. Thematic saturation was determined in our study by analyzing detailed field notes taken during the focus group sessions to determine the emergent themes. During the last focus group, no new themes were deemed to have emerged and, therefore, no further focus groups were conducted. Care was taken to recruit patients across the spectrum of disease severity, age, sex, ethnicity, and socioeconomic background, so that the sample was representative of the local Singaporean population with TED.

A moderator's guide was developed from a comprehensive literature review and the questions focused on the effect of TED and associated visual issues on work, leisure activities, interaction with people, self-care, and psychologic well-being.[3,6,11–13] Focus groups were conducted by the senior researcher (SLL), questions were asked in an open-ended, neutral and sensitive fashion, and the participants were allowed to contribute their opinions freely. Nondirective probes were used to ensure that key areas were explored in an unbiased manner with all participants. A scribe also took notes on participants' responses and nonverbal cues. Focus groups were audiotaped and transcribed verbatim by an anonymous external transcriber. Each discussion lasted approximately two hours. The transcripts then were analyzed by two investigators (SLL and MW) who coded themes using an iterative process based on the constant comparative method,[19,20] which involves coding each transcript for relevant themes and iteratively comparing and contrasting these themes with subsequent transcripts, updating and refining the codes as each new transcript is analyzed. Any disagreement between the investigators was resolved by discussion. A total of 12 items across four QoL domains were generated, namely Activity Limitation (three items); Comfort (two items); Psychologic (four items); and Social (three items).

Phase 2: Psychometric Assessment of the STED-QoL Questionnaire using Rasch Analysis

The 12-item pilot questionnaire was interview-administered to patients aged ≥21 years with mild, moderate, and severe TED classified by their clinical activity score.[1] Patients also answered questions about their sociodemographic and medical histories. Information about treatment was self-reported by patients and confirmed using clinical case notes (Table 1). Rasch analysis was used to explore the psychometric properties of the STED-QoL questionnaire using Winsteps (version 3.91) software[20] and the Andrich rating scale model.[21] The Rasch model assumes that the probability of a given respondent affirming an item is a logistic function of the relative distance between the item's location (i.e., 'difficulty') and the respondent's location (i.e., 'impairment') on this linear scale.[22] The resulting person-measure calibrations are expressed in log of the odds units, or Logits.[23] During Rasch analysis, the observed pattern of responses is compared to the expected pattern using various fit statistics (outlined below).[24] While Rasch analysis generally does not require large numbers of participants,[25] certain Rasch parameters, such Differential Item Functioning as (DIF) are particularly affected by very small sample sizes.[26] Therefore, we did not assess DIF in this study.

Response Category Function

By inspecting the category probability curves (Fig. 1), we can assess whether the thresholds advance in the expected order. Category thresholds represent the point at which two adjacent categories have an equal probability of selection by participants. If thresholds are disordered, it may be necessary to collapse adjacent categories if they are semantically similar and other fit statistics improve.[22]

Precision

Scale precision is determined using person separation index (PSI) and person reliability (PR) coefficients, which indicate the capacity of the STED-QoL to discriminate between differing levels of participants' QoL. A scale should be able to distinguish between at least three levels of QoL, which is reflected by values of >2.0 and >0.8, respectively.[22]

Table 1. Sociodemographic and Clinical Characteristics of Patients who Responded to the STED-QoL Questionnaire

Categorical Variables	$N = 59$	%
Sex		
Male	20	33.9%
Clinical severity of disease		
Mild[a]	20	33.9%
Moderate[a]	20	33.9%
Severe[a]	19	32.2%
Ethnicity		
Chinese	51	86.4%
Indians	4	6.8%
Malays	3	5.1%
Other	1	1.7%
On thyroid medication (yes)	56	94.9%
Had surgery for thyroid eye disease (yes)	19	32.2%
Had intravenous methyl prednisolone (yes)	18	30.5%
Had orbital radiation (yes)	11	18.6%
Continuous variables		
Age, years	49.3 mean	12 SD
Disease duration, years	4.3 mean	3.8 SD

[a] Severity based on EUGOGO classification.[1]

Unidimensionality

It is important that scales are unidimensional; that is, they should measure a single underlying construct. Unidimensionality is assessed using Principal Components Analysis (PCA) of residuals, where the raw variance explained for the first dimension should exceed 50% and the unexplained variance by first dimension should be <2 eigenvalues.[27] Item 'misfit' also may provide evidence of multidimensionality, as it suggests that an item is measuring a different trait to QoL. Item 'fit' to the underlying construct is assessed through an infit MnSq statistic,[28] where a value of <0.7 indicates redundancy and >1.3 suggests measurement 'noise' in the responses.[28] When multidimensionality is evident, the standardized residual loadings for items are assessed to determine if certain items load together (>0.4) and, if so, whether they form a conceptually relevant second dimension.[22]

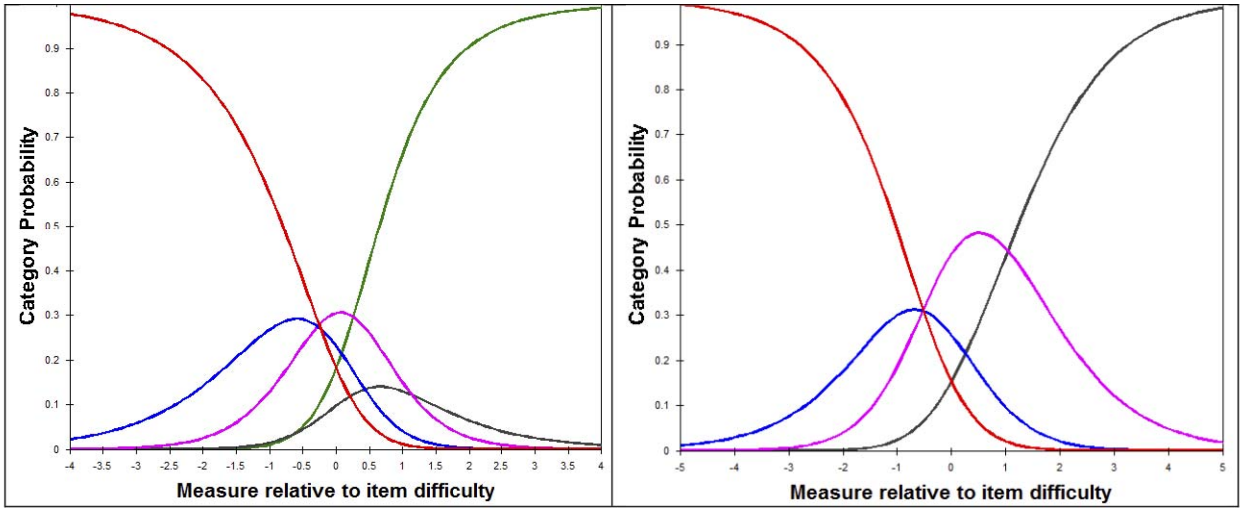

Figure 1. Category probability curves for the TED-QoL before (*left*) and after (*right*) modification.

Targeting

Targeting measures how well item difficulty targets respondents' level of the underlying construct; for example, QoL. Targeting is calculated by determining the difference between the mean of item 'difficulty' (defined as 0 logits) and the mean of person 'ability.' Generally, the closer to 0 logits, the better the targeting; a difference of >1 logits indicates notable mistargeting.[28] Targeting also can be examined through visual inspection of the person-item map (Fig. 2), where the spread of items across the spectrum of participant ability level can be observed.

Criterion Validity

We assessed the criterion validity of the STED-QoL by testing its ability to discriminate between mild, moderate, and severe TED, number of medications (1 vs. >1), and whether patients had undergone surgery for TED (yes/no). Differences between groups were evaluated using analysis of variance (ANOVA). Post hoc Tukey pairwise comparisons were used to test for pairwise differences. Statistical analyses were undertaken using Stata version 14 (StataCorp, College Station, TX).

Results

Phase 1: Content Development using Qualitative Methods
Participant Characteristics

A total of 20 participants (mean age, 49 years; range, 25–74; Table 2) took part in the three focus groups. Group 1 had an equal number of male and female patients, while the focus groups 2 and 3 comprised only males and females, respectively. This strategy was adopted to ensure people of the same sex were comfortable discussing their problems among their own sex. Of the 20 participants, 10 (50%), 6 (30%), and 4 (20%) had mild, moderate, and severe TED, respectively. Following analysis of the focus group transcripts and a round table discussion with principal investigator and coinvestigators, 12 items across four QoL domains were generated, namely

Table 2. Sociodemographic and Clinical Characteristics of Focus Groups Participants ($n = 20$)

Characteristic	*n*	%
Age, years		
21–30	1	5
31–40	1	5
41–50	11	55
51–60	4	20
61–70	2	10
71–80	1	5
Sex		
Male	8	40
Education		
High school and below	5	25
Graduate and above	15	75
Severity of TED[a]		
Mild	10	50
Moderate	6	30
Severe	4	20

[a] Severity based on European Group on Graves' Orbitopathy (EUGOGO) classification.[1]

```
MEASURE PERSON - MAP - ITEM
         <more>|<rare>
    3    XXXX  +
               |
               |
               |
               |
            X  |
               |
               |
               |
               |
    2     XXX T+
               |
               |
          XXX  |
               |
          XXX  |
               |
              S|
               |
          XXX  |
    1       X  +T
               |
           XX  |  2_confident driving
               |  6_appearance
        XXXXX  |
          XXX M|S 8_avoid photos
          XXX  |
       XXXXXX  |
         XXXX  |  7_desire support   9_self-esteem
        XXXXX  |  1_daily activities
    0   XXXXX  +M
           XX  |  3_vary head position
            X S|  10_work
            X  |  11_social activates 5_dark glasses
               |
         XXXX  |S
               |  12_change appearance
               |
            X  |
               |  4_tape eyelids
   -1         T+T
         <less>|<freq>
```

Figure 2. Person–item map for the 12-item TED-QoL.

Activity Limitation (three items), Comfort (two items), Psychologic (four items), and Social (three items). The 12-item thyroid eye disease questionnaire is provided in Supplementary Table S1. The preceding item was "Because of your thyroid eye disease, how often do you..." and all items were rated on a 5-point Likert-type frequency scale, ranging from 'Not at all (4), Once in a while (3), Sometimes only (2), Most of the time (1), and All of the time (0).' Scoring was reversed for item 2 'do you feel confident to drive?'

Phase 2: Psychometric Properties of the STED-QoL Questionnaire

The pilot STED-QoL was administered to 59 patients with TED (mean age, 49.3 ± 12.0 years; 20 [33.9%] male), of whom 20 (33.9%), 20 (33.9%), and 19 (33.2%) had mild, moderate, and severe TED, respectively (Table 1). The response rate was 100%.

The 12-item STED-QoL initially displayed suboptimal fit to the Rasch model (Table 3). Although targeting was good (Fig. 2), the TED-QoL had disordered thresholds (Fig. 1) and poor precision (PSI = 1.54). There also was evidence of multidimensionality, with the raw variance explained <50% and a first contrast eigenvalue of 2.4, and three misfitting items (items 2, 4, and 5). Inspection of the standardized residual loadings for items revealed that items 6, 8, 9, and 12, all relating to psychosocial aspects of self-esteem, were loading together.

We undertook a series of amendments to improve STED-QoL's fit statistics. First, we collapsed categories 3 'once a while' and 2 'sometimes only' as thresholds were disordered and it is likely participants were unable to distinguish between these two conceptually similar response options (Fig. 1). Following this, precision improved to 1.83 although categories remained slightly disordered; however, given the small sample size and loss of information that would result from collapsing to a 3-category scale we chose not to collapse categories further. For item 2, however, the 4-category scale still was highly disordered and, therefore, we further collapsed categories to a 3-category scale, which resolved the disordering and markedly improved precision. Following this, items 4 and then 5 were iteratively removed due to substantial misfit, which further increased precision (PSI = 1.94) and reduced evidence of multidimensionality. This resulted in a 10-item STED-QoL questionnaire (Supplementary Table S2) with satisfactory psychometric properties (Table 3).

As the item loadings were suggestive of a psychosocial element of the scale, we also explored the possibility of a 6-item TED psychosocial subscale (items 6–9, 11, and 12) with a 4-category response scale. The psychosocial scale was unidimensional and had ordered thresholds; however, precision and targeting were suboptimal (PSI = 1.82, and difference between person and item means 1.34), and item 7 displayed misfit (infit MnSq 1.46). As fit statistics may improve in a larger sample, modifications to the subscale were not attempted at this stage. The 10-item STED-QoL showed good criterion validity, with scores decreasing markedly as severity of TED worsened from mild to moderate (Table 4: mild (1.78 logits), moderate (0.27 logits), and severe (0.92 logits). The *P* value for trend for decreasing STED-QoL scores was significant (*P* = 0.012), but was driven by the decrease from mild to moderate TED (*P* =

Table 3. Psychometric Properties of the 10-Item TED-QoL Questionnaire and 6-Item Psychosocial Subscale

Parameters	Rasch Model	TED-QoL ($n = 59$)[a]	TED-QoL: Revised ($n = 59$)	TED-Psychosocial ($n = 59$)
Items		1-12	1-3, 6-12	6-9, 11 & 12
Disordered thresholds	No	**Yes**	**Slightly**	No
Person separation index	>2.0	**1.54**	**1.94**	**1.82**
Person reliability	>0.8	**0.70**	**0.79**	**0.77**
PCA, variance by 1st factor	>50%	**45.5%**	**49.7%**	59.5%
PCA, Eigenvalue for 1st contrast & % unexplained variance in 1st contrast	<3.0, <5.0%	**2.4**[c]	**2.3**	1.6
Item fit (infit MnSq)	<1.3	**Item 4 (1.67)** **Item 5 (1.41)** **Item 2 (1.45)**	None	**Item 7 (1.46)**
Targeting, difference between person & item means	<1.0 logits	0.73	1.00	**1.36**

Bold values indicate misfit to the Rasch model.

[a] Ceiling effect: 4 (6.67%).

[b] Ceiling effect: 11 (18.3%).

[c] Items 6, 8, 9, 12 loaded substantively (>0.4) all relating psychosocial issues.

0.009). A similar decline in STED Psychosocial subscale scores also was observed as TED worsened ($P = 0.015$, Table 4). STED Psychosocial scores also were significantly worse in those who had undergone surgery for TED compared to those who had not (0.41 vs. 1.82 logits, $P = 0.021$). While a similar trend for undergoing surgery was observed for STED-QoL overall, it was not statistically significant (Table 4).

Table 4. Criterion Validity of 10-Item STED-QoL Questionnaire and STED-Psychosocial Subscale

	STED-QoL		STED Psychosocial	
	Mean (logits)	*P* Value	Mean (logits)	*P* Value
Severity				
Mild	1.78	0.012[a]	2.30	0.015[b]
Moderate	0.27		0.33	
Severe	0.92		1.43	
Surgery				
Yes	0.44	0.070	0.41	0.021
No	1.28		1.82	
Number of medications				
One	0.64	0.070	NS	N/A
More than one	1.41		NS	

[a] Overall *P* value for trend. Significant difference found only between mild vs. moderate ($P = 0.008$); mild vs. severe ($P = 0.197$); moderate vs. severe ($P = 0.392$) using post hoc Tukey pairwise comparisons.

[b] Overall *P* value for trend. Significant difference found only between mild vs. moderate ($P = 0.014$); mild vs. severe ($P = 0.383$); moderate vs. severe ($P = 0.277$) using post hoc Tukey pairwise comparisons.

Discussion

We developed and validated a concise and psychometrically robust instrument, the STED-QoL questionnaire, to quantify the impact of TED and associated treatment from the patient's perspective in Singapore. The purpose of developing this questionnaire was to provide clinicians with the means to better understand the psychosocial and functional impact of TED on their patients, enabling them to provide better and more holistic disease management. The STED-QoL questionnaire can be implemented at several points along the clinical course of the disease; that is, at baseline, and during and after treatment to determine if QoL is improving for the patient as their disease severity lessens.

The original 12-item STED-QoL initially had several psychometric issues. We had to delete two items, which may have been due to unclear wording leading to poor understanding by participants. Similarly, we had to collapse response categories 2 and 3, which were likely too similar in meaning and were confusing to participants. This highlights the importance of careful writing of item content and response options,[29,30] as well as the importance of using Rasch analysis to detect and amend such psychometric issues, as this information is more difficult to access using CTT. Although measurement precision was slightly under the required value (PSI 1.94), this is likely related to the small sample size. Despite being a 10-item questionnaire, it was relatively well targeted to patient level of impairment. The psychosocial subscale has the potential to be useful for researchers focusing on the area of mental health; however, further psychometric testing in a larger sample is required to confirm whether this subscale functions as a valid stand-alone construct. We also explored whether a 4-item functioning questionnaire was possible; however, its psychometric properties were suboptimal (data not shown) probably due to item insufficiency.

An interesting note is that despite different sociocultural backgrounds, the items were fairly similar in our questionnaire compared to the Go-QOL questionnaire, which was developed for a Dutch population.[10] In particular, facial disfigurement and the need for camouflage, social interaction and ability to perform daily activities were common items. This suggests that the QoL impact of TED may be similar across cultures and some of the QoL challenges are universally evident. In terms of differences, our questionnaire was rated on a 4-point Likert-type scale (after collapsing 5 to 4 categories), unlike the Go-QOL questionnaire, which only had 3 responses.[31] The advantage of having more response options is that there is more coverage of the latent trait under measurement, which usually results in more precise measurement. Empirical evidence suggests that 4 to 5 response options are optimal in terms of coverage and function. Moreover, as mentioned earlier, the Go-QoL and 3-item TED-QoL were validated using CTT methods alone, whereas in our study we used Rasch analysis and CTT methods, enabling a thorough exploration of the reliability and validity of our STED-QoL. The 3-item TED-QoL that was developed by Fayers and Dolman,[5] though a much faster questionnaire to complete, may not be comprehensive enough for the physician to understand the full range of QOL issues experienced by the patient. Future work should focus on exploring the psychometric properties of STED-QoL in a larger sample and assessing convergent and divergent validity via correlations with the Go-QoL and 3-item TED-QoL instrument, as well as determine test–retest reliability.

Our results showed worse STED-QoL scores in patients with more severe disease as well as in those who had undergone surgery, although our results were unadjusted for confounders, such as age, sex, and sociodemographic and clinical factors. Similar findings have been reported by Delfino et al.,[32] who used the Spanish translated version of the Go-QOL questionnaire in a cohort of 71 patients (56 GO patients and 15 controls) and found lower QOL scores in patients with more severe GO.

The strengths of our questionnaire are the inclusion of a qualitative method to guide item development, use of Rasch analysis to validate the psychometric properties of STED-QoL, and a 100% response rate. This response rate reflected that our study sample was representative of the TED population in our country.[33] Use of the Rasch analysis in our study makes it more robust. With the increasing realization for a patient-centered health care system, patient-reported outcomes are of paramount importance. The Rasch model is the only item response theory model in which a person is characterized totally by the total score across the questions and items.[17]

While we agree that efforts to move vision-related PROs towards third generation item banks and critically appraised topics (CATs) are the ultimate goals in ophthalmology, we believe that short form paper-pencil questionnaires, such as the STEDQoL, are still important in the current research climate. Item banks and CAT are expensive to produce and, as TED affects a relatively small population, it may not be feasible to develop a TED-specific CAT. Similarly, as the scoring and reporting of CAT PRO data still are being streamlined, short forms remain useful for clinicians looking to gain quick and reliable measurement of a disease in the clinic.

Similarly, while other vision-specific questionnaires have undoubtedly been used to assess QoL relating to TED, it is always better, in our view, to have a disease-specific instrument because generic instruments may not be sensitive to disease-specific issues, particularly in the case of TED, which has some specific psychosocial issues relating to appearance.

There are a few limitations in our study. Despite being the largest tertiary referral center in the country, we could only recruit 59 TED patients because the prevalence of this disease is low. This small sample size could account for the lack of statistical significance in some criterion validity results and explain some of our suboptimal psychometric results; however, our results are promising and may show improvement in a larger sample. There also is an element of selection bias as this was not a multicenter study. There is potential to validate this 10-item questionnaire in another population with larger prevalence. In addition, the 12-items were developed from qualitative interviews with patients, and cognitive interviewing was not performed. This may have improved the wording of the items and reduced the misfit we observed with items 4 and 5.

Also, we only recruited Chinese TED patients for the focus groups. The Malay and Indian patients may have different issues from the Chinese population with regards to their disease. It would be interesting to validate this questionnaire in future in these two groups of patients in Singapore as well as other parts of Asia.

Another limitation is that we were unable to perform DIF analyses on our data due to the small sample size. If DIF were present for any item for population subgroups, such as age, sex, or disease severity (despite having similar underlying levels of impairment), this would affect the validity of the instrument. Future work in a larger sample is required to explore the presence of DIF for this new instrument.

Lastly, in the surgical group who had poorer scores, it would have been useful to know what their scores were pre- and postoperatively.

In conclusion, our 10-item Singapore STED-QOL questionnaire is a psychometrically robust questionnaire that can quantify the impact of TED and associated treatment from the patient's perspective in Singapore.

Acknowledgments

Supported by a Singapore National Eye Centre (SNEC) Research Health Endowment Fund.

Disclosure: **M.H.Y. Wong**, None; **E. Fenwick**, None; **A.T. Aw**, None, **E.L. Lamoureux**, None; **L. Leng Seah**, None

References

1. Bartalena L, Baldeschi L, Dickinson A, et al. Consensus statement of the European Group on Graves' orbitopathy (EUGOGO) on management of GO. *Eur J Endocr*. 2008;158:273–285.
2. Terwee CB, Dekker FW, Prummel MF, et al. Graves' ophthalmopathy through the eyes of the patient: a state of the art on health-related quality of life assessment. *Orbit*. 2001;20:281–290.
3. Lee H, Roh HS, Yoon JS, Lee SY. Assessment of quality of life and depression in Korean patients with Graves' ophthalmopathy. *Korean J Ophthalmol*. 2010;24:65–72.
4. Ittermann T, Völzke H, Baumeister SE, Appel K, Grabe HJ. Diagnosed thyroid disorders are associated with depression and anxiety. *Soc Psychiatry Psychiatr Epidemiol*. 2015;50:1417–1425.
5. Wickwar S, McBain H, Ezra DG, Newman SP, et al. The psychosocial and clinical outcomes of orbital decompression surgery for thyroid eye disease and predictors of change in quality of life. *Ophthalmology*. 2015;1222:2568–2576.
6. Fayers T, Dolman PJ. Validity and reliability of the TED-QOL: a new three-item questionnaire to assess quality of life in thyroid eye disease. *Br J Ophthalmol*. 2011;95:1670–1674.
7. Post MWM. Definitions of quality of life: what has happened and how to move on. *Top Spinal Cord Inj Rehabil*. 2014;20:167–180.
8. Zhu B, Ma Y, Lin S, Zou H. Vision-related quality of life and visual outcomes from cataract surgery in patients with vision-threatening diabetic retinopathy: a prospective observational study. *Health Qual Life Outcomes*. 2017;15:175.
9. Karadeniz Ugurlu S, Kocakaya Altundal AE, Altin Ekin M. Comparison of vision-related quality of life in primary open-angle glaucoma and dry-type age-related macular degeneration. *Eye (Lond)*. 2017;31:395–405.
10. Kim YS, Yi MY, Hong YJ, Park KH. The impact of visual symptoms on the quality of life of patients with early to moderate glaucoma. *Int Ophthalmol*. 2018;38:1531–1539.
11. Terwee CB, Gerding MN, Dekker FW, Prummel MF, Wiersinga WM. Development of a disease specific quality of life questionnaire for patients with Graves' ophthalmopathy: the GO-QOL. *Br J Ophthalmol*. 1998;82:773–779.
12. Ponto KA, Hommel G, Pitz S, Elflein H, Pfeiffer N, Kahaly GJ. Quality of life in a German Graves' orbitopathy population. *Am J Ophthalmol*. 2011;152:483–490.
13. Park JJ, Sullivan TJ, Mortimer RH, Wagenaar M, Perry-Keene DA. Assessing quality of life in Australian patients with Graves' ophthalmopathy. *Br J Ophthalmol*. 2004;88:75–78.

14. Wong T. Cataract extraction rates among Chinese, Malays, and Indians in Singapore: a population-based analysis. *Arch Ophthalmol*. 2001;119:727–732.
15. Wong T, Foster P, Seah S, Chew P. Rates of hospital admissions for primary angle closure glaucoma among Chinese, Malays, and Indians in Singapore. *Br J Ophthalmol*. 2000;84:990–992.
16. Petrillo J, Cano SJ, McLeod LD, Coon CD. Using classical test theory, item response theory, and Rasch measurement theory to evaluate patient-reported outcome measures: a comparison of worked examples. *Value Health*. 2015;18: 25–34.
17. Cappelleri JC, Jason Lundy J, Hays RD. Overview of classical test theory and item response theory for the quantitative assessment of items in developing patient-reported outcomes measures. *Clin Ther*. 2014;36:648–662.
18. Rice P, Ezzy D. *Qualitative Research Methods: A Health Focus*. Oxford: Oxford University Press 1999.
19. Maykut P, Morehouse R. *Beginning Qualitative Research: Techniques and Procedures for Developing Grounded Theory*, 2nd ed. Thousand Oaks, California: Sage, 1994.
20. Linacre JM. *A User's Guide to Winsteps/Ministeps Rasch-Model Programs*. Chicago, IL: MESA Press; 2005.
21. Andrich D. Rating formulation for ordered response categories. *Psychometrica*. 1978;43: 561–573.
22. Bond TG, Fox CM. *Applying the Rasch Model: Fundamental Measurement in the human Sciences*. London: Lawrence Erlbaum Associates, 2001.
23. Prieto L, Alonso J, Lamarca R. Classical test theory versus rasch analysis for quality of life questionnaire reduction. *Health Qual Life Outcomes*. 2003;1:27.
24. Vanalphen A, Halfens R, Hasman A, Imbos T. Likert or Rasch - nothing is more applicable than good theory. *J Adv Nurs*. 1994;20:196–201.
25. Linacre JM. Sample size and item calibration stability. *Rasch Measurement Trans*. 1994;7:328.
26. Scott NW, Fayers PM, Aaronson NK, et al. A simulation study provided sample size guidance for differential item functioning (DIF) studies using short scales. *J Clin Epidemiol*. 2009;62:288–295.
27. Pesudovs K, Burr JM, Harley C, Elliott DB. The development, assessment, and selection of questionnaires. *Optom Vis Sci*. 2007;84:663–674.
28. Linacre JM. What do infit and outfit, mean-square and standardized mean? *Rasch Measurement Trans*. 2002;16:878.
29. Boynton PM, Greenhalgh T. Selecting, designing, and developing your questionnaire. *BMJ*. 2004; 328:1312–1315.
30. Murray P. Fundamental issues in questionnaire design. *Accid Emerg Nurs*. 1999;7:148–153.
31. Khadka J, Gothwal VK, McAlinden C, Lamoureux EL, Pesudovs K. The importance of rating scales in measuring patient-reported outcomes. *Health Qual Life Outcomes*. 2012;10:80.
32. Delfino LC, Zunino A, Sapia V, Croome MD, Ilera V, Gauna AT. Related quality of life questionnaire specific to dysthyroid ophthalmopathy evaluated in a population of patients with Graves' disease. *Arch Endocrinol Metab*. 2017;61: 374–381.
33. Fincham JE. Response rates and responsiveness for surveys, standards, and the Journal. *Am J Pharm Educ*. 2008;72:43.