

TITLE: Protocol for Generation of a Patient-Reported Outcome Measure of Quality of Life in Heart Valve Disease: The VALVQ

AUTHORS

Ariel Pons, Department of Medicine, Otago Medical School, University of Otago, Dunedin, New Zealand (Orcid ID: 0000-0003-0851-034X).

Prof. Gillian Whalley, HeartOtago, Department of Medicine, Dunedin School of Medicine, University of Otago, Dunedin, New Zealand (Orcid ID: 0000-0002-5713-6967)

Prof. Crispin Jenkinson, Harris Manchester College, University of Oxford, United Kingdom (Orcid ID: 0000-0002-3692-0779).

Dr. David Morley, Nuffield Department of Population Health, University of Oxford, United Kingdom (Orcid ID: 0000-0003-1911-5855).

Dr. Sean Coffey, HeartOtago, Department of Medicine, Dunedin School of Medicine, University of Otago, Dunedin, New Zealand (Orcid ID: 0000-0003-2312-4666).

Corresponding Author: Dr. Sean Coffey, Department of Medicine, Dunedin School of Medicine, PO Box 56, Dunedin 9054, New Zealand. Email: sean.coffey@otago.ac.nz

Word count: 3350

Key words: Quality of Life, Heart valve disease, Psychometrics, Questionnaire, Item response theory

ACKNOWLEDGEMENTS: This study was funded by the Department of Medicine, University of Otago, New Zealand.

ABSTRACT (236 words)

Background: There is an increasing prevalence of people worldwide with heart valve diseases (HVD), especially rheumatic heart disease, aortic stenosis, and mitral regurgitation, as well as people with a previous valve repair or replacement. Treatment decisions for HVD can be complex, making quality of life an important factor, but no questionnaire to measure quality of life across the lifespan of HVD exists. In this article, we describe the protocol for the development of such a questionnaire.

Methods and Results: The project will occur over four phases. First, people with HVD, family members and clinical experts will be interviewed to generate a list of questions ('items') that comprehensively describe participants' quality of life. In the second phase, this will be formatted into a questionnaire that is pilot tested for functionality. In the third phase, items will be selected according to item distributions, factor analysis and rotation, and item response theory using the Graded Response Model to generate a final questionnaire containing only the best-performing items, which will then be tested for validity. Validity assessments will be repeated after final questionnaire administration in a new sample in the fourth phase.

Conclusion: The article gives a template for development of a patient report outcome measure (PROM) in the health sciences. It is expected that the final questionnaire, called the VALVQ, will allow clinical trials to more sensitively assess quality of life changes across the spectrum and lifespan in HVD.

INTRODUCTION

Heart valve disease (HVD) consists of a variety of conditions where the valves within the heart lose their ability to correctly direct blood flow, usually due either to excessive leaking (regurgitation) or restriction to flow (stenosis). HVD contributes to considerable morbidity and mortality worldwide, with an increasing prevalence seen over the past few decades(1). There is a growing cohort of people who have previously had a valve repair or replacement (VRR), but their epidemiology is much less well studied.

Although any heart valve can have clinically significant disease, three conditions have the largest impact: rheumatic heart disease (RHD), aortic stenosis (AS), and degenerative mitral valve disease causing mitral regurgitation (MR).(1) Rheumatic heart disease commonly occurs in young and socioeconomically disadvantaged populations. It is due to an infection by Group A Streptococcus which triggers an immunologic reaction, causing regurgitation and/or stenosis predominantly of the mitral valve and less commonly regurgitation of the aortic valve.(2) Aortic stenosis typically occurs in the elderly due to calcification of the valve, but can occur in younger people with a congenital malformation. Degenerative MR can occur in both the young and the elderly due to a wide range of primary and secondary causes.(1)

Presentations of these valve diseases vary but symptoms classically involve shortness of breath on exertion (or, in late-stage disease, shortness of breath at rest), angina, and fatigue.(3) Left untreated, HVD gradually leads to pressure and volume overload of the left ventricle, which eventually causes potentially fatal heart failure.(4) The treatment for HVD is valve repair or replacement (VRR). There are a wide range of replacement/repair modalities, necessitating the importance of QOL as an outcome in HVD. However, currently QOL cannot be sensitively measured in HVD since there is no existing questionnaire designed to measure QOL in HVD; existing studies of QOL in HVD use questionnaires developed in patients with heart failure or angina,(5, 6) or 'generic' questionnaires designed to compare across different diseases.(7, 8) One questionnaire has recently been generated to measure QOL in AS across transcatheter replacement,(9) but an instrument to measure QOL throughout the lifespan of HVD and across broader forms is required. In this article, we describe the protocol for development of a questionnaire to measure QOL in HVD, which we call the Valve Quality of life (VALVQ).

METHODS

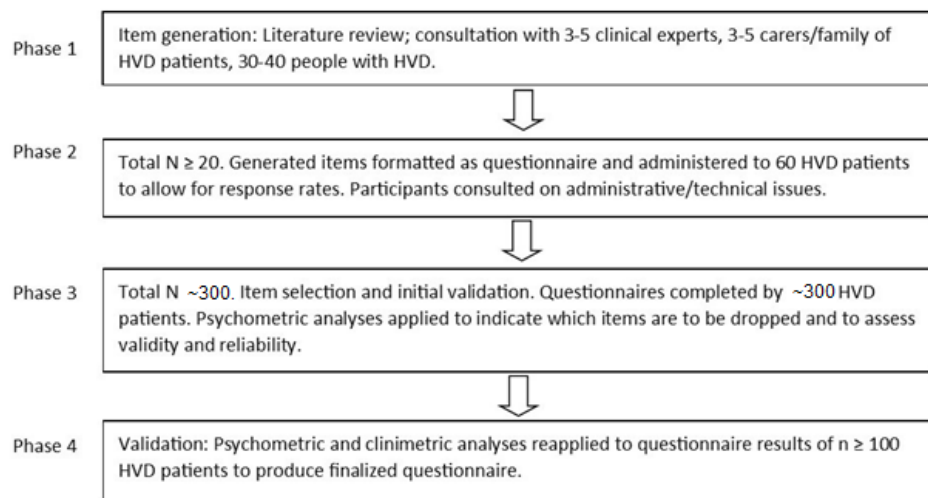
Standard psychometric practice in the development of PROMs was used, and in particular the evaluation of the instrument as described in the document “Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims” produced by the Food and Drug Administration, was followed.(10) The COREQ Checklist was used to guide reporting of this process, and is reported in the appendix.(11)

Questionnaire development

The fundamental process of robust questionnaire generation is to first generate a questionnaire where all factors reported by participants which reflect changes in the disease state of interest and could relate to the outcome of interest are included. Secondly, the questionnaire is then shortened into the minimum number of best-functioning items. Generating the initial saturated questionnaire allows the experience of those with the disease state to be recorded; researchers cannot know which factors are important and which are not to people living with the disease state, and so must not rely on assumption. Subsequently shortening the questionnaire is necessary to produce a practical questionnaire and for the questionnaire’s statistical performance.

Questionnaire development will be split into four phases (Figure 1: Overview of study design). Phase one consists of item generation, where individuals with HVD, their families/carers, clinical experts, and results of a literature review will be used to form an item list that can be expected to cover all significant indicators of QOL in a HVD cohort. Pilot testing for basic functionality forms phase two, while questionnaire testing and dropping items form phase three, with phase four being validation.

Figure 1: Overview of study design



Subject Inclusion and Exclusion

All participants will be aged over 18 years with a good knowledge of English and a diagnosis of one of the following forms of HVD: AS, MR, or RHD, or have undergone valve replacement/repair. The valve disease must be clinically significant which we define as moderate/severe AS or MR, any severity of clinically significant RHD, or any form of VRR. Participants will be excluded if they have a comorbidity more significant than their HVD, have recently been hospitalized, are pregnant (due to changes in heart function during pregnancy), are cognitively impaired, are deemed to be at risk of an impaired outcome from participating in this project, or do not consent to take part. All participants will be required to provide written informed consent. This project was approved by the New Zealand Health and Disabilities Ethics Committee (approval number 19/NTA/163).

Subject Screening and Enrolment

Individuals with HVD will be identified through echocardiographic databases of hospitals in New Zealand. Individuals deemed to meet the criteria will be contacted and sent an information sheet and consent form. Screening of consecutively screened participants will be performed to provide a representative sample of the VRR population undergoing echo.

Questionnaire format

The questionnaire was formatted as semantic differential items with five-point scales, as this format better reflects QOL in this context. In contrast to the more common Likert scale, which has one statement with which respondents rate their agreement, a semantic differential presents paired bipolar opposing statements and the respondent answers by noting their position of relative agreement between the opposing statements. When measuring emotive concepts such as QOL, framing bias is a particular concern, but a Likert scale can only avoid framing bias by alternating the direction (i.e. positive vs negative wording) of its statements, which induces respondent error. The semantic differential format, with its opposing statements, has an inherently reduced risk of framing bias without an increase in error.(12)

Scales of QOL measures are typically either four or five-point scales. There are advantages and disadvantages to each: a four-point scale forces respondents to not tend towards the middle, but can cause participants to leave an increased number of questions unanswered.(13) A five-point scale does allow respondents to regress to the middle, but this reduces respondent burden. Since the semantic differential format chosen for this questionnaire already has a higher respondent burden, the questionnaire was created with a 5-point scale.

Statistical methods

A sequential list of statistical methods is described in the results section – here we provide more detail on the methods used. This project will use factor analysis and item response theory (IRT) in its analyses. Factor analysis is the identification of underlying factors driving the variance in groups of items. Since we do not know how many factors there are in QOL in HVD – models of QOL in disease states vary widely in the number of factors (14, 15) – we used an exploratory factor analysis (EFA).(16) This reports the number of factors identified, in contrast to a confirmatory factor analysis in which a number of factors are input. EFA has two steps: first, the number of factors are identified in factor analysis, and then factor rotation is used to determine which items are grouped to which factor. Factor analysis and rotation is required before IRT, since IRT can only be applied within a single factor.(17)

Item response theory aims to explore the level of ‘latent trait’ – the variable being measured, e.g. QOL – in respondents. The model we chose provides detail on the function of individual items by generating discrimination and difficulty parameters for each item. ‘Discrimination’ is an item’s ability to differentiate between respondents with different levels of the latent trait. ‘Difficulty’ is the level of latent trait required for a respondent to acquire higher scores on the item. Item parameters can be used not only for item selection when developing a questionnaire but also to investigate the functionality of the final questionnaire items. For example, parameters can assess whether responses are being driven by a factor other than the latent variable. This is done by assessing for differential item functioning (DIF). Item parameters can also be used to generate scoring algorithms for the resulting questionnaire that can reflect respondents’ latent trait more accurately and make more use of complex data than classical test methods.(18) Item Response Theory, however, assumes univariance; that there is only one factor in the data. As such, IRT can only be performed after factor analysis and rotation, and the analysis must be performed separately within each factor. Discrimination parameters are also used here: while a high discrimination parameter indicates good item function, an excessively high discrimination parameter (four times than that of others)(17) indicates the violation of unidimensionality. IRT also assumes monotonicity, local independence, and invariance. Monotonicity is the assumption that, as a respondent’s level of latent trait increases, so does their probability of acquiring a higher questionnaire score; local independence is that, for each respondent, their answers to different items will be independent of other items, but reflect their latent trait; invariance is that item parameters are valid regardless of respondents’ scores.(19-21)

There is, at present, no strict consensus on the sample size calculation required for IRT. Some authors suggest a two-parameter model with a polytomous outcome on typically requires at least 400 participants to produce useful output,(22) but others have completed valid and useful studies with smaller sample sizes.(23, 24) Effective assessment of the latent variable requires a well-distributed dataset; a smaller dataset may likely be sufficient if the data are of good quality rather than skewed by biases or floor-ceiling effects. As such, we aimed for 300-400 participants with a variety of forms of HVD.

In this project, oblique factor rotation was selected as it is assumed that factors underlying QOL will correlate with each other. Our IRT model was chosen on the basis of which model best theoretically matched the structure of our data, rather than numerical fit.(21) We therefore selected the Graded Response Model (GRM), a logistic model designed for polytomous outcomes, as it is specifically

designed for ordinal outcomes and especially scales that rate agreement,(35) and functions in practice at least as well as other polytomous models such as the partial credit model (PCM).(36) Moreover, while some consider the PCM to have more attractive statistical properties, especially for assessing latent variables in a population, our aim was to use IRT to select items – as the PCM does not generate a discrimination parameter, the GRM is superior for this task.(35) According to this selection by best theoretical structure, model fit was not calculated as it would not add value to our selection process.

To assess internal reliability, Cronbach's Alpha was selected. This was selected as there is no consensus on the best reliability coefficient, and while the more computationally complex Macdonald's Omega is being increasingly used, it has not been shown to be superior in practice to Cronbach's Alpha.(37)

RESULTS

Phase One: Item Generation

In phase one, a comprehensive list of items assessing all aspects of QOL in HVD will be generated through three sequential steps: Systematic literature review of QOL in HVD; semi-structured interviews of people with different presentations of HVD to determine their opinions of indicators of their QOL; semi-structured interviews of clinical experts of HVD (n=3-5) and family/carers of people with HVD (n=3-5) to determine their opinions of factors that indicate patient QOL in HVD until saturation is reached.

Transcription of interviews will be done by the research team. The conclusions will be used to generate a comprehensive list of items that can be reasonably expected to reflect all significant indicators of QOL in HVD. This will be formatted into a self-reported questionnaire which will be computerized and designed to be completed on a smartphone or other electronic device. A printed paper version will also be made available.

Phase Two: Pilot Testing

In order to identify any administration issues with the questionnaire, 20 questionnaires will be returned from participants with different forms of HVD. The questionnaire will also include a feedback form asking participants to identify any items they found difficult, unclear, insensitive, or irrelevant, and if there are any indicators of their QOL that have been missed. Items deemed clearly unsuitable in this phase may be dropped from the questionnaire, but the threshold to do so will be high as dropping items is to be done primarily in phase three.

Phase Three: Initial item Selection

Item selection will involve a sample of approximately 300-400 people with different forms of HVD. Participants will also be asked to complete concurrent measures of QOL (the SF-20 and KCCQ) for reliability and validity assessments.(25)

The following analyses will be performed sequentially to indicate which items are to be dropped. Items may be retained despite the results of these tests if they are important contributors to the descriptive capacity of the questionnaire according to the results of phase one.

- The standard deviation (SD) for each item will be calculated, and items with a very low SD may be dropped ('very low' SD being determined by a distribution plot).
- Floor and ceiling effects will be described to assess item performance as well as whether the 1-5 range item scale is suitable.
- Factor analysis and factor rotation will be used to identify questionnaire structure. Factor analysis will be exploratory, as the number of factors is unknown. Factor rotation will be oblique, as it is assumed that factors of QOL will have some correlation with each other rather than being completely independent, and varimax, since this is the most commonly used method.(26)

Phase Three: Item response theory analysis

After factor analysis, IRT will be used to compare items within each factor as follows:

- Discrimination parameters will be used to individually assess items; the distribution of discrimination parameters will be plotted to define low parameter values, and items with low discrimination parameters will be removed. Items with extremely high discrimination parameters (at least 4 times that of other items) will also be removed.
- Difficulty parameters within factors will be plotted to select between redundant items. Many items of the questionnaire will be different wordings of the same concept, as it is not known which wording will be found more relatable by respondents. Selecting between these 'redundant' items will be done by plotting difficulty parameters to see where the parameters are very similar. If two items have a similar difficulty parameter, the loss of one will not impact the ability of the questionnaire to measure the latent variable; if one item is a redundant item as described above and its differently-worded partner has a different difficulty parameter, then the item that has a similar difficulty parameter to any other item will be removed.

Information plots will be generated for each item. These indicate at which level of the latent variable the item performs best – i.e. if it performs best at negative versus positive QOL, or at extremes of QOL versus around the median.(27)

A scoring algorithm will be generated by weighted sum scoring; that is, the difficulty parameter for each level of response to an item will be multiplied by the discrimination parameter for the item as a whole, and then scaled to produce a score of 0-100.(28)

Phase Three: Validation

Validity and reliability will be assessed once the final items have been selected.

- Differential item functioning will be used to assess the effect of gender and HVD status (native disease vs. replaced/repared valve)
- Test-retest reliability will be determined longitudinally using a second administration of the questionnaire 2-3 weeks later, using only participants whose QOL according to the concurrent QOL measures has not significantly changed.
- Internal reliability will be calculated cross-sectionally with Cronbach's Alpha.
- Local independence will be assessed by a correlation matrix of items' residual covariance.
- Concurrent criterion validity will be assessed cross-sectionally by comparison of this questionnaire with SF-20 and KCCQ scores.
- At each administration of the questionnaire, feasibility will be assessed by response and completion rates. A codebook of ambiguous, unclear, incomplete, or missing answers will be compiled to identify any items that may need alteration.
- 20 randomly selected participants and 3-5 clinical experts will be interviewed to determine their opinions on the questionnaire.
- Linguistic validity will be assessed through concept elaboration documents and translatability assessment. Concept elaboration documents deconstruct items into the simplest possible sentences, which are then assessed by the research team as to whether the interpretation was as intended. Translatability assessments examine whether the wording of items is free of culturally-specific terms and could be easily translated into another language (rather than the separate process of actually translating a questionnaire.)(29)

DISCUSSION

In this paper we describe an approach to developing a QOL measure that can be used across the spectrum and natural history of HVD. This approach can be used for development of other QOL measures. We initially focused on patients with AS, MR, RHD, and VRR, but envisage that the VALVQ can be used in other valve disease populations after appropriate testing.

While VRR reduces mortality in HVD, quality of life (QOL) is still an important outcome in HVD. Firstly, VRR has burdens of its own: replacement valves can be tissue valves, which have a limited life span and may need reoperation, or metallic valves, which last longer but require permanent anticoagulation.(30, 31) Secondly, there are many treatment options for HVD; intervention can be before or after overt symptoms develop, and can be performed with either surgical or transcatheter approaches, with a wide range of manufactured replacement valves available. QOL becomes an important guiding factor when there is no clear mortality advantage to a particular modality, and the improvements in expertise and quality of VRR ensures that most cases have multiple options which would be suitable, making QOL consideration an imperative. The literature reflects this, with an increasing number of studies assessing QOL after different valve treatment modalities.(32) Valve replacement also causes its own symptom burdens, such as the demand for long-term anticoagulation and blood tests when non-tissue valves are used, painful injections as part of RHD treatment, and recovery from surgical incisions. Finally, symptom burden is often not totally reduced after VRR,(33, 34) and clinical sequelae of living with HVD means that patients are often left with improved but still present heart failure symptoms.

Limitations

Phase one, in which the QOL of people with HVD is elicited, only uses interviews and does not use focus groups. The addition of focus groups would have allowed the researchers to elicit data that they could not do themselves since they, unlike focus group participants, do not share the lived experience of HVD.(38) However, phase one of this project occurred during nationwide lockdowns for COVID-19, which made focus groups unfeasible due to many participants being unfamiliar with online group meetings. Secondly, HVD often occurs in people who are elderly who often have multiple comorbidities, many of which result in similar limitations to QOL as HVD. This makes assessment of QOL due purely to HVD difficult. Equally, there is wide spectrum of severity of HVD, from asymptomatic mild disease to end-stage disease, making measurement of the QOL underlying questionnaire responses difficult at the

extremes of QOL. However, clinical decision-making at the extremes has a lesser requirement for quantification of QOL; treatment options have a lesser requirement of guidance by QOL. Assessment of the effect of clinical therapies on QOL is of more importance at intermediate levels of HVD.

As noted above, there is no consensus on sample sizes, and so this paper cannot advise future researchers on selecting a sample size.

CONCLUSION

With the approach described here, we are confident that we can develop a QOL instrument for use in patients with HVD, both for individual patient care and for research. In particular, given the limitations of medical therapies in HVD, the VALVQ should allow more accurate measurement of participant-reported change in clinical trials.

STATEMENTS AND DECLARATIONS

This study was funded by the Department of Medicine, University of Otago, New Zealand. Each author contributed significantly to the development of this protocol. Ethics was granted by the New Zealand Health and Disability Ethics Committee (Ethics reference 19/NTA/163). All participants gave informed consent to participate in this study, and those who participated gave consent to publish. No authors of this study had competing interests. This project was funded by the Department of Medicine, University of Otago.

References

1. Coffey S, Roberts-Thomson R, Brown A, Carapetis J, Chen M, Enriquez-Sarano M, et al. Global epidemiology of valvular heart disease. *Nat Rev Cardiol.* 2021;18(12):853-64.
2. Passos LSA, Nunes MCP, Aikawa E. Rheumatic Heart Valve Disease Pathophysiology and Underlying Mechanisms. *Front Cardiovasc Med.* 2020;7:612716.
3. Carabello BA, Crawford FA. Valvular Heart Disease. *NEJM.* 1997;337(1):32-41.
4. Goody PR, Hosen MR, Christmann D, Niepmann ST, Zietzer A, Adam M, et al. Aortic Valve Stenosis. *ATVB.* 2020;40(4):885-900.
5. Green CP, Porter CB, Bresnahan DR, Spertus JA. Development and evaluation of the Kansas City Cardiomyopathy Questionnaire: a new health status measure for heart failure. *J Am Coll Cardiol.* 2000;35(5):1245-55.
6. Spertus JA, Winder JA, Dewhurst TA, Deyo RA, Prodzinski J, McDonnell M, et al. Development and evaluation of the Seattle Angina Questionnaire: a new functional status measure for coronary artery disease. *J Am Coll Cardiol.* 1995;25(2):333-41.
7. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care.* 1992;30(6):473-83.
8. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. *Ann Med.* 2001;33(5):337-43.
9. Styra R, Dimas M, Svitak K, Kapoor M, Osten M, Ouzounian M, et al. Toronto aortic stenosis quality of life questionnaire (TASQ): validation in TAVI patients. *BMC Cardiovasc. Disord.* 2020;20(1):209.
10. FDA. Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims [Available from: www.fda.gov/regulatory-information/search-fda-guidance-documents/patient-reported-outcome-measures-use-medical-product-development-support-labeling-claims].
11. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *Int J Qual Health Care.* 2007;19(6):349-57.
12. Friberg O, Martinussen M, Rosenvinge JH. Likert-based vs. semantic differential-based scorings of positive psychological constructs: A psychometric comparison of two versions of a scale measuring resilience. *Pers. Individ. Differ.* 2006;40(5):873-84.
13. Østerås N, Gulbrandsen P, Garratt A, Benth JS, Dahl FA, Natvig B, et al. A randomised comparison of a four- and a five-point scale version of the Norwegian Function Assessment Scale. *Health Qual. Life Outcomes.* 2008;6(1):14.
14. Hides L, Quinn C, Stoyanov S, Cockshaw W, Mitchell T, Kavanagh DJ. Is the mental wellbeing of young Australians best represented by a single, multidimensional or bifactor model? *Psychiatry Res.* 2016;241:1-7.
15. Longo Y, Coyne I, Joseph S, Gustavsson P. Support for a general factor of well-being. *Pers. and Individ. Differ.* 2016;100:68-72.
16. Finch WH. *Exploratory Factor Analysis.* Thousand Oaks, California. 2020. Available from: <https://methods.sagepub.com/book/exploratory-factor-analysis>.
17. Frank B, Baker S-HK. *The Basics of Item Response Theory Using R.* 1 ed: Springer Cham; 2017. 174 p.
18. Kim S-H, Cohen AS. Detection of Differential Item Functioning Under the Graded Response Model With the Likelihood Ratio Test. *Appl. Psychol. Meas.* 1998;22(4):345-55.
19. Baker FB, editor *The Basics of Item Response Theory.* Second Edition. 2001.

20. Nguyen T, Han H-R, Kim M, Chan K. An Introduction to Item Response Theory for Patient-Reported Outcome Measurement. *The patient*. 2014;7.
21. Samejima F. ESTIMATION OF LATENT ABILITY USING A RESPONSE PATTERN OF GRADED SCORES1. *ETS Research Bulletin Series*. 1968;1968(1):i-169.
22. Fayers P MD. *Quality of Life: The assessment, analysis and interpretation of patient-reported outcomes*. 2nd ed. Chichester: John Wiley & Sons; 2007.
23. Zhao NW, Mason JM, Blum AM, Kim EK, Young VN, Rosen CA, et al. Using Item-Response Theory to Improve Interpretation of the Trans Woman Voice Questionnaire. *Laryngoscope*. 2022.
24. Choi M, Park CG, Hong S. Psychometric evaluation of the Korean version of PROMIS self-efficacy for managing symptoms item bank: Item response theory. *Asian Nurs Res (Korean Soc Nurs Sci)*. 2022.
25. Hays R. *The Medical Outcomes Study (MOS) Measures of Patient Adherence*. 1994.
26. Browne MW. An Overview of Analytic Rotation in Exploratory Factor Analysis. *Multivariate Behav. Res*. 2001;36(1):111-50.
27. Bonifay W. *Multidimensional Item Response Theory*. Thousand Oaks, California. 2020. Available from: <https://methods.sagepub.com/book/multidimensional-item-response-theory>.
28. Kim J, Wilson M. Polytomous Item Explanatory Item Response Theory Models. *Educ Psychol Meas*. 2020;80(4):726-55.
29. Rust J, Golombok S, Kosinski M, Golombok S. *Modern Psychometrics : The Science of Psychological Assessment*. Florence, UNITED KINGDOM: Taylor & Francis Group. 2009.
30. Wunderlich NC, Dalvi B, Ho SY, Kux H, Siegel RJ. Rheumatic Mitral Valve Stenosis: Diagnosis and Treatment Options. *Curr Cardiol Rep*. 2019;21(3):14.
31. Boskovski MT, Gleason TG. Current Therapeutic Options in Aortic Stenosis. *Circ Res*. 2021;128(9):1398-417.
32. Pons A, Whalley G, Sneddon K, Williams M, Coffey S. A systematic review of non-procedural contributors to quality of life in heart valve disease. *Health Sci. Rev*. 2022;4:100050.
33. Sponga S, Perron J, Dagenais F, Mohammadi S, Baillot R, Doyle D, et al. Impact of residual regurgitation after aortic valve replacement. *Eur. J. Cardiothorac. Surg*. 2012;42(3):486-92.
34. Shingu Y, Iwano H, Murakami T, Katoh N, Ooka T, Katoh H, et al. Risk factors for residual mitral regurgitation after aortic valve replacement in patients with severe aortic valve stenosis and moderate mitral regurgitation. *Gen Thorac Cardiovasc Surg*. 2019;67(10):849-54.
35. van der Linden WJ, Hambleton RK. *Handbook of Modern Item Response Theory*. New York, NY, UNITED STATES: Springer New York. 1996.
36. De Ayala RJ, Dodd BG, Koch WR. A Comparison of the Partial Credit and Graded Response Models in Computerized Adaptive Testing. *Appl. Meas. Educ*. 1992;5(1):17-34.
37. Deng L, Chan W. Testing the Difference Between Reliability Coefficients Alpha and Omega. *Educ Psychol Meas*. 2017;77(2):185-203.
38. Guest G, Namey E, Taylor J, Eley N, McKenna K. Comparing focus groups and individual interviews: findings from a randomized study. *Int. J. Soc. Res. Methodol*. 2017;20(6):693-708.