
Article:
Carlton, J. orcid.org/0000-0002-9373-7663 (2019) Comparison of the CAT-QoL and PedsQLTM instruments in measuring quality of life in amblyopia treatment: preliminary results. Strabismus, 27 (3). pp. 165-171. ISSN 0927-3972

https://doi.org/10.1080/09273972.2019.1618351

Comparison of the CAT-QoL and PedsQLTM instruments in measuring quality of life in amblyopia treatment: preliminary results

Corresponding Author

Jill Carlton

Health Economics and Decision Science (HEDS)

School of Health and Related Research (ScHARR)

University of Sheffield

Regent Court

30 Regent Street

Sheffield

S1 4DA

United Kingdom

j.carlton@sheffield.ac.uk

Tel: 0114 2220799

Fax: 0114 2724095

Keywords: Amblyopia Quality of Life Questionnaire Treatment

ABSTRACT

Background/Aims

The Child Amblyopia Treatment Questionnaire (CAT-QoL) is a patient-reported outcome (PROM) measure designed to assess the impact of amblyopia treatment. The aim of this study was to compare the psychometric properties of two PROMs; the CAT-QoL instrument and PedsQL™, a generic paediatric PROM. This work was part of a wider project to develop a condition-specific PROM for children with amblyopia.

Methods

342 participants were recruited in a UK multi-centre study. Quality of life data was collected through using the CAT-QoL and the PedsQL™ instruments. The psychometric performance of the CAT-QoL and PedsQL™ were examined in terms of acceptability, reliability, and validity.

Results

Both instruments demonstrated good reliability (CAT-QoL Cronbach's $\alpha = 0.793$; PedsQL™ $\alpha = 0.872$). The convergent validity of the CAT-QoL and PedsQL™ instruments was tested by comparing the instruments to each other. There was a moderate correlation between the PedsQL™ and the CAT-QoL scores, and this relationship was statistically significant ($r_s = -0.517$, $p < 0.000$). No statistical significance was found between the level of amblyopia severity and the mean PedsQL™ score ($p = 0.420$).

Conclusion

It was possible to assess the impact of amblyopia treatment using the CAT-QoL and PedsQL™ instruments. The preliminary findings from this are not conclusive, and it is not possible to advocate

the use of one questionnaire over another based upon psychometric performance demonstrated here. This may be due to the sample population, as there were limited numbers of participants with severe amblyopia. Both the CAT-QoL and PedsQL™ instruments were noted to have some issues with ceiling effects at an individual item level. The CAT-QoL and PedsQL™ were reliable (as determined by Cronbach's alpha). The PedsQL™ instrument was not able to discriminate between amblyopia severity groups (discriminant validity). Further research is required to formally assess the psychometric properties of the CAT-QoL questionnaire.

INTRODUCTION

The Child Amblyopia Treatment Questionnaire (CAT-QoL) is a disease-specific patient reported outcome (PROM) instrument designed to measure the impact of amblyopia treatment from the child's perspective. The CAT-QoL was developed using an iterative approach, including systematic literature review; focus groups with clinicians; semi-structured interviews with children with amblyopia; cognitive de-briefing and ranking exercises; and Rasch analysis.[1-5] The CAT-QoL was designed for children aged 4-7 years, and the content and format of the instrument reflects this by having a low task burden. The refined instrument consists of eight items, each with three response levels.

Psychometric validation is the process by which an instrument is assessed for reliability and validity through a series of defined tests on the population group for which it is intended.[6] The aim of this study is to explore the psychometric properties of the CAT-QoL instrument, to determine the ability of the instrument to measure the impact of amblyopia treatment from a child's perspective.

MATERIALS AND METHODS

CAT-QoL

Seven treatment-specific versions of the CAT-QoL were created (patch; drops; glasses; patch and drops; patch and glasses; glasses and drops; glasses, patch and drops), with each version worded slightly differently to reflect the type of treatment the child is undertaking. All items are scored on a 3-level response scale. Individual item responses are scored from 0 to 2 (least to worst) meaning the instrument has a range of 0-16. The summative score is then converted into a Rasch score (as shown in Table 1) where a greater score indicates a worse quality of life (or greater impact of treatment on the individual).

Pediatric Quality of Life Inventory (PedsQL™)

The PedsQL™ was developed to allow accurate and reliable reporting of child health.[7;8] A number of different formats exist, which include parent proxy-reporting and different age versions. The items

for each of the forms are essentially identical, and differ only in appropriate language and grammatical tense. The Young Child Report (aged 5-7) version was used in this study. It comprises 15-items that are reported on a 3-level response scale ("not at all", "sometimes" and "a lot"). Items are reversed-scored and transformed to a 0-100 scale, with a larger number indicating better health-related quality of life (HRQoL). The PedsQL™ has been widely tested and validated in both healthy individuals and patients.[7-9]

Patient Cohort

Data used in this study was collected from nine sites across England, United Kingdom (UK).[5] Inclusion criteria was that used during development of the descriptive system.[3;4] The study was approved by the National Health Service Research Ethics Committee for Airedale, UK, (REC Ref: 07/Q1201/5), and followed the tenets of the Declaration of Helsinki. Written parent/guardian consent was obtained prior to data collection. Each participant was asked to complete a version of the CAT-QoL and PedsQL™ questionnaires, issued by the clinician. Socio-demographic and clinical data was collected by the clinician.

Psychometric evaluation

Acceptability

Acceptability was assessed by calculating completion rates and missing data values. For the purpose of this study, the acceptable amount of overall missing data and individual missing data is ≤10%.[10] Floor and ceiling effects describe the amount of responses given at either end of the scale. A high percentage of floor or ceiling is suggestive that the content validity of the instrument is limited.[11] For the purpose of this study, a level of ≤20% was considered acceptable.[10;12]

Reliability

Reliability was assessed by calculating the internal consistency of the scale, defined by a Cronbach's alpha score. Values of ≥ 0.70 indicate that the instrument is reliable.[10;12-14]

Validity: Construct validity: convergent, discriminant validity and known-group differences

Convergent validity was assessed by calculating the Spearman's correlation coefficient (r_s) between the CAT-QoL and PedsQL™ instruments. A strong correlation of Spearman rank correlation coefficient is defined > 0.70; moderate 0.30 to 0.70; and weak < 0.30.[15] The convergent validity of the CAT-QoL and PedsQL™ instruments was tested by comparing the instruments to each other. The hypothesis was that the correlations would not be strong between the two instruments, as they measure different things. This was explored by examining the overall CAT-QoL and PedsQL™ scores.

Discriminant validity was assessed by performing a one way ANOVA to assess the statistically significant difference in instrument scores across amblyopia severity groups. SES values can be described to fall within the ranges of: "small" 0.2 – 0.5; "medium" 0.5 – 0.8; and "large" effect size > 0.8.[16] The hypothesis tested was amblyopia severity level would correlate with HRQoL scores. That is, the "severe" amblyopia group will show a worse QoL than the "moderate" group; the "moderate" group will show a worse QoL than the "mild" group; and so on. Subjects were categorised in terms of the interocular severity difference between the two eyes at the time of the questionnaire. The amblyopia severity groups chosen were that adopted by the PEDIG group in their multi-centre studies examining treatment outcomes for amblyopia.[17-19] These were; mild amblyopia 0 ≥ 0.3 logMAR; moderate amblyopia 0.31 ≥ 0.60 logMAR; and severe amblyopia > 0.61 logMAR.

All data was analysed using SPSS 19.0.

RESULTS

Patient Cohort

The socio-demographic details of the study sample have been reported.[5] 342 subjects participated in the study, however some subjects were excluded from analysis due to missing clinical data (n=11). A total of 331 subjects were included in the analysis. It should be noted that participants completed a 5- or 6-response level draft version of the CAT-QoL instrument containing 11-items. The data was recoded such that only the responses for the final 8-item version was included in the comparisons presented here.

Acceptability

Table 1 shows the completion rate of both instruments, and indicates there was an issue with missing data. The CAT-QoL shows a percentage of missing data that exceeds the acceptability criteria adopted for this study. However, some of the participants were issued the incorrect version of the questionnaire due to administrator error (n = 18). When these respondents were excluded from the analysis, the number of respondents with missing data decreased, and fell within the acceptability criteria. The percentage of floor and ceiling effects, and completion rate for each individual item on the CAT-QoL instrument were calculated (Table 2). Most items demonstrate high ceiling effects, with the exception of item 2 (*feeling/sensation on face*). These fell within accepted levels for each item, with the exception of item 8 (*playing with friends*). The floor effects and percentage of missing data were low for each item.

The completion rate of the PedsQL™ is lower than that of the CAT-QoL on an item level basis (Table 3). The amount of missing data exceeds the accepted level for all of the items. This may suggest that the questions are either redundant in this population or that the respondents failed to understand the questions. The ceiling effects of the PedsQL™ are greater for each item compared to the CAT-QoL.

Five items have ceiling effects that exceeded the accepted level for this study. The floor effects for each item of the PedsQL™ instrument were low.

Reliability

The internal consistency of the CAT-QoL was 0.793. The PedsQL™ instrument had a higher value, α = 0.872. The reliability of the CAT-QoL and PedsQL™ scales were investigated further by assessing the item-total correlations and the Cronbach's alpha with each item, should that item be deleted (Table 4). If any of the items within the CAT-QoL or PedsQL™ instruments were to be removed, this would decrease the reliability of the scales.

Validity: Construct validity: convergent, discriminant validity and known-group differences

There was a moderate correlation between the PedsQL™ and the CAT-QoL scores, and this relationship was statistically significant (r_s = -0.517, p < 0.000). There was no statistical significance between the level of amblyopia severity and the mean PedsQL™ score (p=0.420). The results reject the hypothesis that amblyopia severity level would correlate with HRQoL scores, but it should be noted that the trends are in the right direction and are linear. As the number of subjects within the "severe" amblyopia group was low, these were merged with the "moderate" amblyopia group. Despite the merging of categories, there was no statistical significance between the level of amblyopia severity and the mean PedsQL™ score (p=0.406). There was a slightly stronger relationship between the CAT-QoL score and amblyopia severity, compared to the PedsQL™ score and amblyopia severity (r_s = 0.183 and 0.132, respectively). There was a trend for increasing CAT-QoL score with amblyopia severity. However, there is decrease in mean CAT-QoL score for the "severe" group. For the PedsQL™ instrument, there is a much weaker relationship.

The means of the CAT-QoL scores and the amblyopia severity level were compared (Table 6). There are statistically significant differences between the amblyopia severity level and mean CAT-QoL scores. A medium effect size was found between "no amblyopia" (equal VA) and "mild" amblyopia severity groups. A small effect size was found between "mild" and "moderate" amblyopia severity groups; and a small effect size between "moderate" and "severe" amblyopia groups.

There was no statistical significance between the level of amblyopia severity and the mean PedsQL™ score ($p = 0.420$). The effect size between amblyopia severity categories was smaller for the PedsQL™ instrument than the CAT-QoL for each scenario.

DISCUSSION

There is an increasing call for transparency in paediatric PROM development and reporting, however assessing the psychometric properties of any PROM is difficult.[20] The FDA state that the measurement properties of an instrument should be evaluated; those of reliability, construct validity and the ability to detect change.[21] They advocate developers of PROM instruments to provide hypotheses when presenting data on construct validity.[21] However, there are no universally accepted performance criterion to apply, and there is no "gold standard" instrument to compare it against. Various indirect tests of performance have been developed in an attempt to demonstrate instrument validity, and different studies have adopted different levels of acceptability, reliability, and validity.

Acceptability

There are no universally accepted values of missing data limits, with ranges of <5-10% described as acceptable.[10;12] The CAT-QoL was associated with less missing data than the PedsQL™ instrument. However, it should be acknowledged that incorrect versions of the CAT-QoL instrument were issued to a small number of respondents in the validation study. This may be considered as a potential

weakness of the instrument. It would be desirable to have one version of the instrument that could be administered to any child, irrespective of their amblyopia treatment. However, this was not possible due to the nature of some of the items requiring expansion (e.g. *feeling of drops on your face (like stinging, or cold)*). No record was kept as to the order subjects were presented with the two instruments. Further investigation is required to assess whether completion rates and missing data values differ if the ordering of the two instruments are randomised.

Both the CAT-QoL and PedsQL™ instruments demonstrated acceptable levels of floor and ceiling effects. However, when we consider the instruments on an individual item basis, some items exceeded the accepted criteria (> 80%). On the CAT-QoL instrument, one item (*playing with friends*) was seen to have marginally greater ceiling effects than the accepted value (80.3%). The PedsQL™ instrument contained five items that fell outside of the accepted criteria. It could be argued that the criteria used for evaluating floor and ceiling effects at the scale level should be different to that used when assessing individual items. It is desirable for an instrument to be able to measure the full range across a spectrum. Difficulty can arise when an instrument is found to have high ceiling effects. This may result in the instrument not being able to detect an improvement in HRQoL over time when there has been an improvement in their overall condition (if their starting level of HRQoL is already very good).[22] However, the responsiveness of both the CAT-QoL and PedsQL™ instruments has yet to be tested in this population. Existing literature appeared to use the same criteria for floor and ceiling effects for both the scale and the item.[10;12] The presence of high ceiling effects may suggest an issue with the instrument itself, or it may also be due to the study sample. The study sample does have a large number of people with "mild" levels of amblyopia. The high ceiling effects found in the CAT-QoL and PedsQL™ scales may be linked to this. It should be noted that the data used in this analysis was that collected as part of the development of the CAT-QoL instrument. The presence of floor and ceiling

effects of the final 8-item CAT-QoL (three response level) instrument should be explored in an independent data set.

Reliability

The reliability of both the CAT-QoL and PedsQL™ instruments fell within acceptable levels. Item total correlations and the Cronbach's α with each item suggested that removing any item within either instrument would reduce the reliability of the overall scale. However, it is important to recognise that the data used in this study was the same as that of the development of the CAT-QoL instrument.[5] The reliability of the final 8-item (three response level) instrument should be explored in an independent data set.

Validity: Construct validity: convergent, discriminant validity and known-group differences

The CAT-QoL and PedsQL™ scores were found to moderately correlate. It was hypothesised that no (or weak) correlations would be found between the two instrument scores, as other generic PROMs have been found to be insensitive to particular medical conditions.[23-25] It should be recognised that the data used in this analysis was that collected as part of the development of the CAT-QoL instrument. The convergent validity of the final 8-item CAT-QoL (three response level) instrument should be explored in an independent data set. The findings in this study suggest that the PedsQL™ instrument may able to detect some of the HRQoL implications of amblyopia. However, a key component of instrument validity is the ability to detect differences between severity groups. Whilst this was assessed for the CAT-QoL instrument and the results described here, it should be noted that this was using the same dataset that was applied to refine the measure.[5] Independent assessment of psychometric performance has yet to be assessed for the 8-item CAT-QoL questionnaire, and further

data is required determine whether it is able to detect differences between severity groups. Using clinical indicators to assess the validity of a HRQoL instrument may not always be appropriate. Clinicians can postulate that a greater level of amblyopia severity results in a lower measure of HRQoL, but this may not actually be the case. It is not necessarily true that the worse the level of amblyopia, the lower the HRQoL score.

The study is not without limitations, and the results presented here are preliminary.[5] There are only small numbers of respondents in the "severe" category group. There are a number of reasons to account for this. The first is that of categorization: subjects were categorized into severity groups as used by the PEDIG studies.[17-19] However, this categorization is arbitrary, and may not universally accepted. Furthermore, the data collection period for this study was conducted over a short time period (approximately 4 months). Therefore, there was limited opportunity to collect data from respondents over a number of sequential visits. Such data could have been used to evaluate the test-retest ability, and properly assess responsiveness of both instruments. Further research is required to examine both test-retest reliability and responsiveness in subsequent validation surveys.

The overall purpose of the research was to develop a condition-specific PROM for children with amblyopia. The CAT-QoL was developed for children, by children, using children's data at every stage to inform the item content, response levels, language, and format of the measure itself. The research closely follows the guidance of the FDA, and has demonstrated that the CAT-QoL does capture the "patient's experience" in the target population. This "bottom-up" methodological approach has ensured high content and face validity of the instrument; and the format and language informed directly from children increases the scope of self-reporting.[1-5] The CAT-QoL is now ready for further studies to assess the reliability, validity and responsiveness of the instrument in an independent

sample. Once this has been established the CAT-QoL could be used in both clinical practice and research settings to calculate the impact of amblyopia treatment in the paediatric population, and offers an alternative to generic measures of paediatric HRQoL.

Acknowledgements

Data for this study was collected at the following sites: Bradford Teaching Hospitals NHS Foundation Trust; Doncaster and Bassetlaw NHS Foundation Trust; Harrogate and District NHS Foundation Trust; Maidstone and Tunbridge Wells NHS Trust; Medway NHS Foundation Trust; Sandwell and West Birmingham Hospitals NHS Trust; Sheffield Children's NHS Foundation Trust; Sheffield Teaching Hospitals NHS Foundation Trust; The Mid Yorkshire Hospitals NHS Trust. The author gratefully acknowledges the support from orthoptists and clinicians at each of the collaborating sites.

This work is produced by the author under the terms of Personal Development Award research training fellowship issued by the NIHR. The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, The National Institute for Health Research or the Department of Health.

Declaration of Interest

The authors report no conflict of interest.

References

1. Carlton J, Kaltenthaler E. Amblyopia and quality of life: a systematic review. *Eye*. 2011; 25(4):403-13.
2. Carlton J. Clinicians' perspectives of health-related quality of life (HRQoL) implications of amblyopia: a qualitative study. *Br Ir Orthopt J*. 2011;8:18-23.
3. Carlton J. Identifying potential themes for the Child Amblyopia Treatment Questionnaire. Optom Vis Sci. 2013;90:867-73.
4. Carlton J. Developing the draft descriptive system for the Child Amblyopia Treatment Questionnaire (CAT-QoL): a mixed methods study. *Health Qual Life Outcomes.* 2013;11:174.
5. Carlton J. Refinement of the Child Amblyopia Treatment Questionnaire (CAT-QoL) using Rasch analysis. *Strabismus.* 2019: doi: 10.1080/09273972.2019.1601743
6. Bowling A. The philosophy, theory and practice of research. In: *Research Methods in Health: Investigating health and health services*. 3rd ed. Maidenhead: Open University Press; 2009:125-82.
7. Varni JW, Seid M, Rode CA. The PedsQL: Measurement model for the Pediatric Quality of Life Inventory. *Med Care*. 1999; 37(2):126-39.
8. Varni JW, Seid M, Kurtin PS. Reliability and validity of the Pediatric Quality of Life Inventory Version 4.0 Generic Core Scales in health and patient populations. *Med Care*. 2001; 39(8):800-12.
9. Varni JW, Burwinkle TM, Seid M, Skarr D. The PedsQL 4.0 as a pediatric population health measure: Feasibility, reliability and validity. *Ambul Pediatr*. 2003; 3(6):329-41.
10. Hilari K, Lamping DL, Smith SC, Northcott S, Lamb A, Marshall J. Psychometric properties of the Stroke and Aphasia Quality of Life Scale (SAQOL-39) in a generic stroke population. *Clin Rehabil*. 2009; 23(6):544-57.

11. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker, Bouter LM, de Vet HC. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007; 60(1):34-42.
12. Lamping DL, Schroter S, Marquis P, Marrel A, Duprant-Lomon I, Sagnier PP. The Community-Acquired Pneumonia symptom quesitionnaire. *Chest*. 2002; 122(3):920-9.
13. Streiner DL, Norman GR. *Health measurement scales. A practical guide to their development and use.* New York: Oxford University Press; 2003.
14. Smith SC, Lamping DL, Banerjee S, Harwood RH, Foley B, Smith P, Cook JC, Murray J, Prince M, Levin E, Mann A, Knapp M. Development of a new measure of health-related quality of life for people with dementia: DEMQOL. *Psychol Med*. 2007; 37(5):737-46.
15. Tabachnick BG, Fidell LS. *Using Multivariate Statistics*. 4th ed. Boston: Allyn and Bacon; 2001.
16. Cohen K. A power primer. *Psychol Bull*. 1992; 112(1):155-9.
17. Pediatric Eye Disease Investigator Group. The clinical profile of moderate amblyopia in children younger than 7 years. *Arch Ophthalmol.* 2002;120(3):281-7.
18. Repka MX, Beck RW, Holmes JM, Birch EE, Chandler DL, et al. A randomized trial of patching regimens for treatment of moderate amblyopia in children. *Arch Ophthalmol*. 2003;121(5):603-11.
19. Pediatric Eye Disease Investigator Writing Committee, Rutstein RP, Quinn GE, Lazar EL, Beck RW, Bonsall DJ, et al. A randomized trial comparing Bangerter filters and patching for the treatment of moderate amblyopia in children. *Ophthalmol*. 2010;117(5):998-1004.
20. Tadic V, Hogan A, Sobti N, Knowles RL, Rahi JS. Patient-reported outcome measures (PROMs) in paediatric ophthalmology: a systematic review. Br J Ophthalmol. 2013;97(11):1369-81.
21. US Department of Health and Human Services Food and Drug Administration (FDA). Guidance for industry: patient-reported outcome measures: use in medicinal product development to support labeling claims. 2009.

22. Buck D, Clarke MP, Powell C, Tiffin P, Drewett RF. Use of the PedsQL in childhood intermittent exotropia: estimates of feasibility, internal consistency, reliability and parent-child agreement. *Qual Life Res*. 2012; 21(4):727-36.
23. Harper R, Brazier JE, Waterhouse JC, Walters SK, Jones NMB, Howard P. A comparison of outcome measures for patients with chronic obstructive pulmonary disease (COPD) in an outpatient setting. *Thorax*. 1997; 52(10):879-87.
24. Kobelt G, Kirchperger I, Malone-Lee J. Quality of life aspects of the over active bladder and the effect of treatment with tolterodine. *BJU Int*. 1999; 83(6):583-90.
25. Espallargues M, Czoski-Murray C, Bansback N, Carlton J, Lewis G, Hughes L, Brand C, Brazier J. The impact of age-related macular degeneration on health state utility values. *Invest Ophthalmol Vis Sci*. 2005; 46(11):4016-23.

Table 1 Rescoring of CAT-QoL instrument

CAT-QoL Raw Score	Person Scores	Interval Level Equivalences	Rounded Interval Level Equivalent Score
0	-3.60	-0.00000020	0.0
1	-2.65	2.14689245	2.1
2	-1.98	3.66101674	3.7
3	-1.51	4.72316363	4.7
4	-1.13	5.58192069	5.6
5	-0.81	6.30508453	6.3
6	-0.52	6.96045176	7.0
7	-0.25	7.57062125	7.6
8	0.01	8.15819187	8.2
9	0.27	8.74576249	8.7
10	0.55	9.37853085	9.4
11	0.84	10.03389808	10.0
12	1.16	10.75706192	10.8
13	1.53	11.59322011	11.6
14	1.98	12.61016926	12.6
15	2.61	14.03389807	14.0
16	3.48	15.99999976	16.0

It should be noted that this conversion chart can only be used when there is no missing data from an individual. It can only be used when complete data is present. For example, if an individual scored 14

(raw data score) this would be the equivalent of 12.6 on the re-scored measure. The final CAT-QoL scores range from 0-16, where a greater score indicates a worse quality of life (or greater impact of treatment on the individual).

Table 2 Descriptive and acceptability statistics for the CAT-QoL (Rasch scores) and PedsQL™ (n=331)

	N with completed questionnaire (%)	% with some missing data	Mean Score (SD)	Floor %	Ceiling %
CAT-QoL	292 (88.2)	**11.8†**	4.14 (2.87)	0.7	17.5
CAT-QoL correct version	286 (91.4)	8.6	4.14 (2.87)	0.7	17.5
PedsQL™	286 (86.4)	**13.6†**	83.90 (17.00)	0.3	18.5

† falls above accepted levels for this study (> 10%)

Table 3 Descriptive and acceptability statistics for individual items of the CAT-QoL (n=313)

Item	N	Mean Item Score (SD)	Missing data %	Floor %	Ceiling %
1 (*sad*)	308	0.49 (0.68)	1.6	10.4	61.7
2 (*feeling on face*)	306	0.69 (0.64)	2.2	9.5	40.8
3 (*hurt*)	309	0.44 (0.58)	1.3	4.2	60.5
4 (*doing schoolwork*)	300	0.38 (0.63)	4.2	6.3	68.3
5 (*other children*)	307	0.31 (0.57)	1.9	5.2	74.3
6 (*doing things*)	303	0.42 (0.68)	3.2	10.9	69.3
7 (*worried*)	307	0.31 (0.57)	1.9	5.2	74.3
8 (*playing friends*)	304	0.22 (0.48)	2.9	2.6	**80.3***

* exceeds accepted levels for this study (≥ 80%)

Table 4 Descriptive and acceptability statistics for individual items of the PedsQL™

	PedsQL™				
Item	**N**	**Mean Item Score (SD)**	**Missing data %**	**Floor %**	**Ceiling %**
1 (*hard to walk*)	287	92.16 (23.27)	**13.3†**	4.2	**88.5***
2 (*hard to run*)	284	87.32 (27.52)	**14.2†**	5.6	**80.3***
3 (*hard to play sports or exercise*)	279	82.26 (30.56)	**15.7†**	7.2	71.7
4 (*hard to pick up big things*)	281	85.94 (29.70)	**15.1†**	7.5	79.4
5 (*hard to do chores*)	283	91.52 (23.04)	**14.5†**	3.5	**86.6***
6 (*feel scared*)	283	88.87 (25.08)	**14.5†**	3.9	**81.6***
7 (*feel sad*)	286	85.66 (26.90)	**13.6†**	4.2	75.5
8 (*feel mad*)	285	83.86 (30.01)	**13.9†**	7.0	74.7
9 (*worry about what will happen to you*)	284	84.86 (29.12)	**14.2†**	6.3	76.1
10 (*hard to get along with other kids*)	282	89.54 (23.61)	**14.8†**	2.8	**81.9***
11 (*other kids say they don't want to play with you*)	284	77.82 (31.19)	**14.2†**	7.0	62.7
12 (*other kids tease you*)	283	83.22 (27.79)	**14.5†**	4.2	70.7
13 (*hard to pay attention at school*)	280	78.57 (33.14)	**15.4†**	9.6	66.8
14 (*forget things*)	280	75.00 (31.68)	**15.4†**	7.5	57.5
15 (*hard to keep up with schoolwork*)	278	80.94 (31.16)	**16.0†**	7.6	69.4

† exceeds accepted levels for this study (> 10%) * exceed accepted levels for this study (≥ 80%)

Table 5 Item-total correlations and corrected Cronbach's α for the CAT-QoL instrument

	Scale mean if item deleted	**Scale variance if item deleted**	**Corrected Item-total correlation**	**Cronbach's α if item deleted**
1 (*sad*)	2.73	6.750	0.588	0.755
2 (*feeling on face*)	2.51	7.170	0.480	0.774
3 (*hurt*)	2.75	7.310	0.511	0.769
4 (*doing schoolwork*)	2.84	7.310	0.484	0.773
5 (*other children*)	2.87	7.444	0.457	0.777
6 (*doing things*)	2.80	6.776	0.582	0.756
7 (*worried*)	2.90	7.414	0.490	0.772
8 (*playing friends*)	2.99	7.923	0.409	0.784

Table 6 Discriminant validity of the PedsQL™ and CAT-QoL (Rasch scores) using PEDIG amblyopia classification

	PedsQL™				**CAT-QoL**			
	N	**Mean**	**SD**	**Effect Size**	**N**	**Mean**	**SD**	**Effect Size**
Equal VA	21	85.98	15.13		23	2.72	2.51	
				0.14				0.55
Mild	195	84.48	17.39		197	4.11	2.98	
				0.12				0.26
Moderate	59	82.44	15.55		64	4.87	2.41	
				0.26				0.43
Severe	10	76.56	21.84		7	3.84	2.97	
Total	285	83.89	17.04		291	4.16	2.86	
				p = 0.420				
Equal VA	21	85.98	15.13		23	2.72	2.51	
				0.10				0.55
Mild	195	84.48	17.40		197	4.11	2.98	
				0.17				0.22
Merged Moderate*	69	81.59	16.54		71	4.77	2.47	
Total	285	83.89	17.03		291	4.16	2.86	
				p = 0.406				P = 0.010†

* merged group

† p<0.05 in test of difference between adjacent severity