

promoting access to White Rose research papers



Universities of Leeds, Sheffield and York
<http://eprints.whiterose.ac.uk/>

This is an author produced version of a paper published in **Quality of Life Research**.

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/10941>

Published paper

Stevens, K. (2009) *Developing a descriptive system for a new preference-based measure of health-related quality of life for children*, *Quality of Life Research*, 18 (8), pp.1105-1113
<http://dx.doi.org/10.1007/s11136-009-9524-9>

1 **Running head:** New health descriptive system for children

2

3

4 **Developing a Descriptive System for a New Preference Based**
5 **Measure of Health Related Quality of Life for Children.**

6

7

8

9 Katherine Stevens, BSc, MSc, PhD

10

11 Health Economics and Decision Science (HEDS), ScHARR,

12

13 The University of Sheffield, UK.

14

15

16 **Address for correspondence:**

17 Katherine Stevens

18 Health Economics and Decision Science

19 School of Health and Related Research,

20 The University of Sheffield,

21 Regent Court, 30 Regent Street

22 Sheffield S1 4DA, UK

23 Telephone: 0114 222 0841

24 Fax: 0114 272 4095

25 Email: K.Stevens@Sheffield.ac.uk

26

27 **Abstract**

28

29 **Objectives**

30 The use of preference based measures (PBM) of health related quality of life
31 (HRQoL) is increasing in health care resource allocation decisions. Whilst there are
32 measures widely used for this purpose in adults, research in the paediatric field is
33 more limited. This paper reports on how the descriptive system for a new paediatric
34 generic PBM of HRQoL was developed from dimensions identified in previous
35 research.

36

37 **Methods**

38 Existing scales from the paediatric literature were reviewed for suitability and scales
39 were also developed empirically, based on qualitative interview data from children, by
40 taking adverbial phrases and confirming the ordinality by a ranking exercise with
41 children. The resulting scales were applied to the dimensions from the previous
42 research.

43

44 **Results**

45 No suitable scales were found in the paediatric literature, so the empirically derived
46 scales were used resulting in 7 different types. Children were successfully able to
47 rank these to determine the ordinality and they were applied to the dimensions.

48

49 **Conclusions**

50 This work has empirically developed a descriptive system for the dimensions of
51 HRQoL identified in previous research. Further research is needed to test the
52 descriptive system on a paediatric population and reduce the number of dimensions
53 to be amenable to health state valuation.

54

55 **Key Words:** Health care rationing, pediatrics, quality adjusted life years, quality of life

56

57 **Abbreviations**

58 PBM Preference based measure

59 HRQoL Health Related Quality of Life

60 QALYs Quality Adjusted Life Years

61

62

63 **Introduction**

64 The use of preference based measures (PBM) of health related quality of life
65 (HRQoL) is increasing in health care resource allocation decisions. In the United
66 Kingdom (UK) in particular, the National Institute for Health and Clinical Excellence
67 (NICE) specifies that for its reference case, a PBM measure be used to quantify the
68 benefits of an intervention [1]. PBMs allow the calculation of quality adjusted life
69 years (QALYs) by combining length of life with quality of life, which can be used in
70 economic evaluation as part of a decision making process. Whilst there are PBM
71 widely used for this purpose in adults, research in the paediatric field is more limited
72 [2].

73

74 Research by Stevens [3] reported on the first stage in the development of a new
75 generic paediatric PBM for children age 7 to 11 years, in order to start addressing
76 this gap. The paper reported on the process of identifying relevant dimensions of
77 health related quality of life (HRQoL) for inclusion in the new measure. They were
78 identified by undertaking qualitative interviews with children aged 7-11 years with a
79 wide range of acute and chronic health conditions, to find out how their health
80 affected their lives. The children were divided into two age groups according to their
81 school year (7-9 years and 9-11 years). Each group was sampled, interviewed and
82 analysed independently to explore whether these groups identified the same
83 dimensions and therefore shared a common HRQoL framework. The research found
84 that they did share a common framework as the dimensions identified were almost
85 identical, therefore a measure could be developed for the age group 7-11 years as a
86 whole. Eleven dimensions were identified from the interviews, covering social,
87 emotional and physical aspects of HRQoL. These dimensions are reproduced from
88 Stevens [3] in Table 1.

89

90 Having identified these dimensions, the next stage in the development of a paediatric
91 PBM for children was to create a descriptive system based on these dimensions that
92 is suitable for use in economic evaluation. This paper reports on how this descriptive
93 system was developed. The aim was to begin to develop a descriptive system
94 suitable for health state valuation, based on the dimensions identified from the
95 previous interview work [3].

96

97 **Background**

98 Existing non preference based quality of life measures have generally taken an
99 approach to descriptive system development whereby a series of items or statements
100 are developed using focus groups, the literature or interviews. Work is then done to
101 develop order and scales for these items, or response options could be based on
102 Likert scale type responses [5]. These are then reduced or sorted into factors or
103 dimensions using psychometric techniques. Reduction of items is common as
104 generally long lists of items are generated which are too long to have each item in
105 the final questionnaire, hence testing is useful to identify redundant items (for
106 example if items are not used or are very similar to another item), incomprehensible
107 or ambiguous items and to test the internal consistency of a scale [5]. Factor analysis
108 or Rasch techniques can be used to do this and can also be used as complements
109 rather than alternatives [6].

110

111 The work by Stevens [3] took a different approach to the development of the
112 dimensions, in that the dimensions of paediatric health related quality of life were
113 determined directly from qualitative interviews and analysis. The qualitative work
114 provides supporting evidence as to why the dimensions arose and the terminology of
115 the dimensions is based on the terminology used in the interviews. There is very little
116 guidance in the literature about how to develop levels for dimensions directly. One
117 way could be to consider the use of standard response scales from the literature.

118

119 Most existing measures use categorical response scales for their items, including
120 those based on options relating to frequency (e.g. never, sometimes, often), the
121 intensity/severity of a dimension (e.g. a little, moderately, a lot), or the level of
122 agreement with something (strongly agree, disagree etc), also known as a Likert
123 scale. [5]

124

125 Existing generic PBM have taken different approaches when using scales. The EQ-
126 5D takes the severity approach, using three levels for each dimension, the Health
127 Utilities Index (HUI)2/3 has a mixture of both (severity and frequency) and the SF-36
128 (used to obtain the SF-6D) has a mixture of both, but is mainly a frequency based
129 approach [7]. The levels on the EQ-5D descriptive system, (a generic preference
130 based measure for adults), were developed to be ordinal and were developed using
131 an expert panel. The developers also recommend using severity based scales
132 although they do not justify why [8].

133

134 It can make a substantial difference to the descriptive system depending on the scale
135 used. For example a frequency based scale may not capture the range of how
136 something can affect a person, e.g. you can always be worrying, but only at a low
137 level, which is different to being extremely worried. Equally, a scale based on severity
138 may not adequately describe frequency. Another type of scale which is used in health
139 status measures is the level of agreement, which asks a respondent how much they
140 agree (or disagree) with a statement. This type of scale does not really make sense
141 for a preference based measure as you do not want a separate scale for each item
142 level. There is also a scale which asks you to indicate how much something bothers
143 you, however, again this is not suitable for a preference based measure as it is not
144 useful for societal valuation, but may be useful for individual clinical decision making.

145

146 The majority of scales used in existing paediatric measures are categorical response
147 type scales with a variety of response options and the vast majority are frequency
148 based rather than severity [9]. Most do not give any explanation as to how the levels
149 or scales were developed. Those with a shorter recall period, the 16D/17D and
150 HUI2/3 are statement based [10], [11].

151

152 There is not much empirical work in the paediatric field with regard to the use of
153 response options and children's ability to understand and use them across ages. [9]
154 Many existing measures use response options with between 3 and 7 points and there
155 is literature which has shown that the number of categories used by raters should be
156 in the region of between 5 and 7 as a maximum [4], [5]. Some measures use the
157 same number of response options for each question, and some use different
158 numbers of response options. The HUI2/3 and the 16D/17D use descriptive
159 statements instead, however these are still ordinal [10] [11]. There are also
160 developmental differences in children's ability to understand and respond to items on
161 a Likert scale. Eight year old children can accurately use a 5 or 7 point scale to rate
162 their health status whereas younger children tend to use more extreme responses.
163 Some instruments have used visual aids to help with this, for example the Child
164 Health and Illness Profile, which uses graduated circle sizes for the response options
165 [12].

166

167 Another important feature of descriptive system development is the recall period.
168 This is the time frame respondents are asked to think about when completing a
169 questionnaire. In existing paediatric generic measures, there is a whole mixture of
170 recall periods, from several weeks to the current day. More research is needed in this
171 area about what is appropriate for children and different health conditions [9], [13].

172

173 Many of the existing paediatric instruments based on a frequency approach ask
174 questions about how often something has been the case over the past few weeks.
175 The evidence from the qualitative interviews undertaken in previous work by Stevens
176 [3] is that children are able to recall information about their health and understand
177 and describe it well, but often have difficulty remembering when they had a particular
178 health problem or when an event had occurred. The advantage of asking about
179 HRQoL today, is that you are focusing on a point in time and you also remove any
180 potential problems with recall bias as children are thinking about the present time.
181 The disadvantage is that this may miss important episodes in the context of a clinical
182 trial for example, particularly in episodic conditions.

183

184 The main constraint in designing a descriptive system for a preference based
185 measure is that the health states defined by the system should be amenable to
186 valuation. Ideally, each dimension needs to contain levels (response scales) that are
187 ordered within it to fit this criteria well. There are also constraints on the number of
188 dimensions that can be included due to limitations on people's ability to process
189 information. [4] This paper reports on how levels were developed for the dimensions
190 identified in previous work [3] to form a descriptive system amenable to valuation.

191

192 **Methods**

193 The first stage in developing the levels (response scales) for the dimensions was to
194 determine whether they should be frequency or severity based. To do this, the data
195 from the original qualitative work for developing the dimensions was used [3]. All the
196 interview transcripts were reviewed and adverbial phrases were extracted when the
197 children were describing the dimensions and the way in which something was
198 described, for example, 'it's a bit annoying' or 'it's quite annoying'. Phrases were
199 extracted for each dimension separately and this was used to determine whether the

200 dimension was about severity or frequency. In this way, the decision was based on
201 the data.

202

203 Once this had been determined, the next step was to develop the scales for each
204 dimension. Scales were developed based on the qualitative interview data from
205 children and using guidance from the methodological literature [5] together with what
206 is required for a PBM (i.e. ordinal levels within each dimension) [14]. The principals
207 from the literature are as follows:

208

- 209 • Items should be clear, relevant and understandable
- 210 • Scales will be developed with 5-7 levels with a view to reduction in further
211 testing
- 212 • Language should be kept simple
- 213 • Double barrelled questions will be avoided (asking two different things within
214 one question)
- 215 • Negatively worded items will be avoided, using positive wording styles instead
- 216 • Vague quantifiers will be avoided, although this can be very difficult in
217 practice.

218

219 In addition, the following approach was also followed due to using the qualitative data
220 and the constraints of a PBM:

221

- 222 • The qualitative interviews were used to guide the wording of the levels, by
223 analysing how the children described the problem, e.g. It hurts a bit, and it
224 hurts a lot
- 225 • Levels were ordinal, using an adjectival scale with discrete responses
- 226 • Language was based on the qualitative data

227

228 From the original qualitative work, there were alternative wording terms used to
229 describe the dimensions, for example pain and hurt. Where more than one term
230 existed, the alternative wordings were each developed into separate questions for
231 future testing work about which was the most appropriate.

232

233 Not all terms were used as alternatives, as sometimes words were used by the older
234 age group and so were more complex, for example miserable. As the measure was
235 being developed for the two age groups combined (as they were found to have a
236 common HRQoL framework in the earlier research) [3], where there was a choice
237 over wording, the wording used by the younger age group was selected.

238

239 The final questions developed were as follows. Worried and scared were developed
240 as separate questions and sad and upset were developed as separate questions.
241 Miserable is just a more sophisticated wording style by the older children and was
242 therefore not included. Unhappy was felt not to be a good term for use in a
243 questionnaire as it is negatively worded and so was not included. Annoyed,
244 frustrated and angry were all developed as separate questions. Hurt and pain were
245 developed as separate questions. School work and learning were referred to as the
246 same thing in the interviews, therefore the younger children's terminology was used
247 (i.e. school work). Daily routine was the same for both age groups so this was
248 developed into a question. Tired and weak were developed into questions as drowsy
249 and weary were not in common across age groups, and energy is the opposite
250 meaning. Joining in activities was the same for both age groups so this was
251 developed into a question. Sleep was the same for both age groups so this was
252 developed into a question. Finally, jealous and embarrassed were both developed
253 into questions.

254

255 This resulted in seventeen questions in total: Worrying; Sad; Weak; Angry; Pain;
256 Frustrated; Hurting; School Work; Upset; Tired; Annoyed; Scared; Sleep;
257 Embarrassed; Jealous; Daily Routine and Joining in activities.

258

259 As described above, the qualitative data was used to develop levels (response
260 scales) for each of these 17 questions. In addition, the wording used tried to
261 incorporate the ways in which children had described the dimensions, for example for
262 worried, sad, angry, weak and embarrassed, children were often using the term 'feel'.
263 For hurt and pain, they were describing it in terms of it hurting or having pain.

264

265 Whilst the scales developed would be based on children's descriptions, the ordinality
266 of these scales needed to be confirmed. As children have been involved at every
267 stage of the development of this measure and the measure is intended for children, it
268 was important to verify the order of the scales with them.

269

270 The ordinality of the scales developed was tested by asking children to rank the
271 levels in order of their severity. Children were sampled from the same two schools
272 used in the original qualitative work. [3].

273

274 Levels (response scales) were created for each question by applying the scales
275 developed. These scales were applied to all seventeen questions: Worrying; Sad;
276 Weak; Angry; Pain; Frustrated; Hurting; School Work; Upset; Tired; Annoyed;
277 Scared; Sleep; Embarrassed; Jealous; Daily Routine; Joining in activities. Piloting of
278 the ranking work with children demonstrated that 17 ranking exercises was infeasible
279 for them to do in one sitting, and so a subset of the scales from the questions were
280 ranked, making sure each different type of scale developed was covered. This
281 assumes that the ordinality of the scale is independent of the item (question).

282

283 Cards were created for each question being tested, with each card displaying a level
284 and these were put together into a coloured envelope, one for each question/scale
285 being tested. Children were asked to choose an envelope, one at a time and asked
286 to rank the levels on the cards in order of severity (how bad they thought they were)
287 from best to worst. Ties were allowed. Where children ranked levels as equal they
288 were asked if they had a preference for the wording. The ranking work was first
289 piloted on 10 children aged 7-11 years (5 male and 5 female). They were able to
290 complete the tasks successfully and advised on the size of the cards, the font used
291 and the colours of the cards.

292

293 For the main study, 31 children were sampled from both schools involved in the
294 research and each child carried out the same number of ranking exercises. The aim
295 of the sampling was to get an equal balance across gender and all year groups and
296 to include both schools equally. The number of children included in the study was
297 based on what was possible given resource constraints, as there was only one
298 researcher undertaking this work, with a limited time period. Ethical approval and
299 consent from the parents of children in both schools had already been obtained when
300 the qualitative work was undertaken [3]. Children were sampled from those where
301 parents had given their consent for the researcher to approach the child to ask if they
302 would like to participate in the research. Children were approached one by one and
303 the study was explained to them with the aid of an information leaflet which they
304 could take and keep. The children had an opportunity to ask any questions they liked
305 before being asked if they would like to take part. If children consented to take part,
306 they were given the ranking tasks to do. All children carried the task out by
307 themselves with the researcher sat with them in the school library or the dining room.
308 The children's rankings for each of the sets were recorded by the researcher, along
309 with any comments on preferences for wording where levels were ranked equally.

310

311

312 **Analysis**

313 The rank data was analysed by looking at the mean ranking and variation (standard
314 deviation) and by using Kendall's coefficient of concordance test statistic. The
315 approach of looking at the mean ranking is similar to work undertaken by Keller et al
316 [15] as part of their work testing the equivalence of translations of widely used
317 response choice labels, where they looked at the mean response choice ratings by
318 country and language.

319

320 The Kendall statistic is between 0 and 1 and is a measure of the agreement between
321 rankings, 0 means there is no agreement between rankings. It measures the extent
322 to which ordering by each of two (or more) variables would arrange the observations
323 into the same numerical order [16].

324

325 The rank data was coded using the mid rank method [17], [18] as this is more
326 appropriate for this type of analysis and ensures that the sum of ranks is maintained.
327 That is, a rank of 1 was coded as 1, a rank of 2 was coded as 2 and where rankings
328 were tied, each tied ranking was given a value of the midpoint of the previous and
329 next ranks. For example, a ranking sequence where the second and third cards were
330 ranked equally was coded as 1,2.5,2.5,4,5.

331

332 Where there was a very small difference between mean rankings, this was taken to
333 mean that only one statement was needed for the descriptive system. A difference in
334 mean ranking of less than 0.20 (chosen as a very low and conservative estimate)
335 was taken to be a small difference. Whilst a difference of 0.20 was an arbitrary
336 choice, this was chosen as the aim was to be conservative so that any removal of
337 levels due to redundancy was based on a clear overlap.

338

339 In order to choose between the statements, the variation and the preferences of
340 children for the wording was looked at, with the least amount of variation taking
341 priority.

342

343 **Results**

344 For every dimension, severity arose as the predominant characteristic. In a couple of
345 dimensions (worrying and angry/annoyed/frustrated) frequency arose in one case in
346 each. For worrying, this was a mixture of the two “I always get a bit worried”. For
347 angry/annoyed/frustrated, it was frequency “it’s always annoying”. For sleep, one
348 child described it in frequency terms “can’t get to sleep that often”. In the schoolwork,
349 activities and daily routine dimensions, children were describing how much they
350 could or couldn’t do something which again indicated a severity approach.

351

352 As the vast majority of dimensions and evidence within dimensions steered towards
353 a severity based approach, the dimension scales developed were based on this.

354

355 The adverbs and adverbial phrases used to describe the dimensions in the
356 qualitative data are listed below.

357

358	at all	a little bit	a bit	quite	quite a lot
359	much	a lot	very	very much	really

360

361 The only wording not included in this list was ‘kinda’, as this is a colloquial word and
362 was felt to be not appropriate to include.

363

364 Applying these phrases to the dimensions resulted in seven different types of scale,
365 some of which were very similar, but had subtle differences depending on how the

366 dimension fitted with the wording. There were therefore 7 unique scales to test in the
367 ranking work, and it was felt appropriate that each child should rank each one. Figure
368 1 gives the 7 scales tested and the dimensions (questions) to which each scale
369 applies.

370

371 All 31 children consented to take part in the ranking and all children completed all 7
372 ranking tasks. The characteristics of the sample are shown in Table 2. Table 3 shows
373 the mean rank order, standard deviation and difference in mean rank for each of the
374 7 scales.

375

376 Table 4 shows the Kendall coefficient for each scale which was very high for all
377 scales. The lowest was for scale 3 (school work). An agreement of 0.81 to 1.00 is
378 suggested to be almost perfect agreement for the Kappa statistic, which is another
379 statistical measure of agreement [19].

380

381 The difference in the mean rank order was very low for the statements highlighted in
382 bold in Table 3 (My sleep is very affected and My sleep is really affected had a
383 difference of 0.05, My school work is very affected and My school work is really
384 affected had a difference of 0.0. I feel very worried and I feel really worried had a
385 difference of 0.16).

386

387 As there was such a small difference between these mean rankings, it indicated that
388 only one statement was needed for the descriptive system. The preferences of
389 children when these statements were ranked equally are shown in Table 5. The
390 choice made over these three sets of statements where the difference in mean rank
391 order was low was as follows:

392

393 **(1) Sleep:** 'really' had a lower standard deviation and a smaller range (shown in
394 Table 3). The preferences of the children were equal. Therefore 'My sleep is
395 really affected' was chosen.

396

397 **(2) School work:** 'really' had a lower standard deviation and a smaller range
398 (shown in Table 3). 'Very' has one more vote. Therefore 'My school work is really
399 affected' was chosen.

400

401 **(3) Worried:** 'really' and 'very' have the same standard deviation and range
402 (shown in Table 3). 'Very' is preferred by one vote. Therefore 'I feel very worried'
403 was chosen.

404

405

406 The results of this ranking exercise were then applied to the scales on all questions
407 in order to form the draft descriptive system.

408

409

410 **Discussion**

411

412 A draft descriptive system has been developed from the dimensions formed from the
413 original qualitative work [3]. This descriptive system is based on the qualitative data
414 and is for both age groups combined. It contains 17 questions, some of which are
415 alternative wording for the same dimensions, as further testing is required to
416 determine the best wording. Instead of developing scales empirically, a scale could
417 have been used from the paediatric literature however the only severity based scale
418 in the literature for paediatric generic instruments is the scale from the KIDSCREEN
419 [20]. This scale is for children aged 8-18 years and uses the scale:

420

421 Not at all slightly moderately very extremely

422

423 The words slightly, moderately and extremely never appeared in the qualitative
424 interviews undertaken in the original qualitative research [3] and seem complex for
425 young children and so this was felt not to be a suitable option.

426

427 The dimensions contain levels (response scales) which are based on severity which
428 was determined empirically from the qualitative data. The original interviews
429 contained a good mix of acute and chronic conditions such as sickness, fever, flu,
430 pneumonia, hearing problems, vision problems, asthma, weak wrists and ankles,
431 eczema, hyperactive fits and abnormal muscle growth. Children with these problems
432 all described the dimensions mainly in terms of severity, whether they had acute or
433 chronic conditions.

434

435 The ranking exercise worked well with children and they were successfully able to
436 complete the tasks with a 100% completion rate. The ordering of the statements
437 resulting from the analysis made sense at face value and there was very good
438 agreement in the rankings by children. Whilst the sample size was quite low in this
439 study, the high agreement in rankings gives confidence in the results produced.

440

441 The advantages of this new descriptive system being developed with children are
442 that the language is appropriate to children of this age group, which will aid self
443 completion and the content validity is likely to be increased. Whilst this has been
444 developed with children age 7-11 years, it may be the case that the descriptive
445 system is also suitable for other paediatric age groups, however this cannot be
446 determined without further empirical testing.

447

448 In comparison with the only other existing paediatric generic preference based
449 measure, the HUI2, all the dimensions in the new measure are based on severity,
450 whereas the HUI2 contains a mixture of severity and frequency based items. Both
451 measures are statement based however (rather than having an item and then a
452 standard response scale). This makes the descriptive systems more amenable to
453 valuation as a health state can be formed from these statements, whereas the
454 language may be clumsy with a standard response scale as the item and response
455 scale are separate.

456

457 The spacing of the scales is not necessarily even, however they do not have to be
458 equally spaced as ultimately this will be a preference based instrument and those
459 levels that are too close will drop out in future testing work. It is also likely that there
460 are too many levels as whilst the principle was to aim for 5-7 levels, a few of the
461 scales have more than this number (sleep and school work with 9 and 8 respectively)
462 however in scale development it is usual to start with too many levels and then
463 reduce these down. These issues will be addressed in future work.

464

465

466 **Conclusion**

467 This work has empirically developed a descriptive system for the dimensions of
468 HRQoL identified in the original interview work. As the methods were based on using
469 the data from children, the content validity should be increased. 17 questions are
470 contained within the descriptive system, some of which are alternative wordings for
471 the same dimension. Further research is needed to test these alternative wordings
472 on a paediatric population and to test the psychometric performance of this
473 descriptive system. In addition, due to the constraints of PBMs, the number of
474 dimensions will need to be reduced to be amenable to valuation. Further research is
475 required to do this.

476 **Acknowledgements**

477 This research was funded as part of an MRC Special Training Fellowship in Health
478 Services and Health of the Public Research.

479

480 Thanks go to The Children and Young People's Directorate at Sheffield City Council,
481 the staff of Firs Hill Community Primary School and Hunters Bar Junior School, to the
482 parents who gave their consent, to the staff at the schools and to all the children who
483 took part in the research.

484 **References**

- 485 1. NICE (National Institute for Clinical Excellence). (April 2004). Guide to the
486 Methods of Technology Appraisal.
487
- 488 2. McCabe, C. (May 2003). Estimating preference weights for a paediatric health
489 state classification (HUI2) and a comparison of methods. Ph.D.Thesis. The University
490 of Sheffield.
491
- 492 3. Stevens, K. J. (2008) Working with children to develop dimensions for a
493 preference based generic paediatric health related quality of life measure. Health
494 Economics and Decision Science Discussion Paper 08/04. Available from
495 <http://www.shef.ac.uk/scharr/sections/heds/discussion.html>
496 Accessed 14/08/2008
497
- 498 4. Miller, George A., "The Magical Number Seven, Plus or Minus Two: Some Limits
499 on our Capacity for Processing Information." Psychological Review, Vol. 63, No. 2,
500 March, 1956, pp 81-97.
501
- 502 5. Streiner, D.L. & Norman, G.R.. Health Measurement Scales. (1995). A Practical
503 Guide to their Development and Use. 2nd ed. Oxford University Press. Oxford.
504
- 505 6. Tennant. A., McKenna, S.P. & Hagell, P. (2004). Application of Rasch analysis in
506 the development and application of quality of life instruments. Value in Health, 7
507 (Supplement 1), S22-S26.
508
- 509 7. Brazier, J.E., Ratcliffe, J., Salomon, J. & Tsuchiya, A. (2007). Measuring and
510 Valuing Health Benefits for Economic Evaluation. Oxford University Press.
511

- 512 8. Kind, P., Brooks, R. & Rabin, R. (Eds). (2005). EQ-5D concepts and methods, a
513 developmental history. Chapter 3. Springer.
514
- 515 9. Eiser, C. & Morse, R. (2001). Quality-of-life measures in chronic diseases of
516 childhood. Health Technology Assessment. [5;4].
517
- 518 10. Apajasalo, M., Sintonen, H., Holmberg, C., et al. (1996). Quality of life in early
519 adolescence: A sixteen-dimensional health-related measure (16D). Quality of Life
520 Research. 5, 205-211.
521
- 522 11. Health Utilities Index. <http://healthutilities.biz/>
523 Accessed 14/07/2008
524
- 525 12. Riley, A. W., Forrest, C.B., Rebok, G.W., et al. (2004). The Child Report Form of
526 the CHIP-Child Edition: reliability and validity. Medical Care.. 42(3), 221-231.
527
- 528 13. Matza, L. S., Swensen, A. R., Flood, E. M., et al. (2004). Assessment of Health
529 Related Quality of Life in Children: A Review of Conceptual, Methodological, and
530 Regulatory Issues. Value in Health..7(1) 79-92.
531
- 532 14. Brazier, J.E., Deverill, M., Green, C., et al. (1999). A review of the use of health
533 status measures in economic evaluation. Health Technology Assessment. 3[9].
534
- 535 15. Keller, S. D., Ware, J. E., Gandek, B., et al (1998). Testing the Equivalence of
536 Translations of Widely Used Response Choice Labels: Results from the IQOLA
537 Project. Journal of Clinical Epidemiology. 51 (11), 933-944.
538

539 16. Bland, M. & Peacock, J. (2001). Methods based on rank order. In Statistical
540 questions in evidence-based medicine. Oxford Medical Publications. Oxford
541 University Press.
542

543 17. Argyrous, G. (2006) .Rank-order tests for two or more samples. In Statistics for
544 Research with a guide to SPSS. 2nd edition. Sage Publications. London.
545

546 18. Hinton, P. R. (1995). Statistics Explained. A Guide for Social Science Students.
547 Routledge.
548

549 19. Landis, J.R. & Koch, G. G. (1977). The measurement of observer agreement for
550 categorical data. Biometrics.. 33: 159-174.
551

552 20. MAPI research trust. Patient-Reported Outcome and Quality of Life Instruments
553 Database. Available from: www.proqolid.org
554 Accessed 13/04/2006
555
556
557
558
559
560
561

562 **Figure 1: Scales tested (and applicable wording for questions)**

563 Scale 1 (Worrying, Sad, Weak, Angry, Frustrated, Upset, Tired, Annoyed, Scared,
564 Embarrassed, Jealous)

565 I don't feel worried

566 I feel a little bit worried

567 I feel a bit worried

568 I feel quite worried

569 I feel very worried

570 I feel really worried

571

572 Scale 2 (Pain)

573 I don't have any pain

574 I have a little bit of pain

575 I have a bit of pain

576 I have quite a lot of pain

577 I have a lot of pain

578 I am really in pain

579

580 Scale 3 (Daily routine)

581 I have no problems with my daily routine

582 I have a few problems with my daily routine

583 I have some problems with my daily routine

584 I have many problems with my daily routine

585 I can't do my daily routine

586

587 Scale 4 (Hurting)

588 It doesn't hurt

589 It hurts a little bit

590 It hurts a bit

591 It hurts quite a bit

592 It hurts quite a lot

593 It hurts a lot

594 It really hurts

595

596 Scale 5 (Joining in activities)

597 I can join in with any of the activities that I want to

598 I can join in with most of the activities that I want to

599 I can join in with some of the activities that I want to I can join in with a few of the

600 activities that I want to

601 I can join in with none of the activities that I want to

602

603 Scale 6 (Sleep)

604 My sleep is not affected

605 My sleep is a little bit affected

606 My sleep is a bit affected

607 My sleep is quite affected

608 My sleep is affected quite a lot

609 My sleep is really affected

610 My sleep is very affected

611 My sleep is affected a lot

612 I can't sleep at all

613

614 Scale 7 (School Work)

615 My school work is not affected

616 My school work is a little bit affected

617 My school work is a bit affected

- 618 My school work is quite affected
- 619 My school work is affected quite a lot
- 620 My school work is really affected
- 621 My school work is very affected
- 622 I can't do my school work
- 623
- 624
- 625
- 626

627 **Table 1: Dimensions of Health Related Quality of Life [3]**

628

	(7-9 years)	(9-11 years)
1	Worried Scared	Worried
2	Sad Upset	Sad Upset Unhappy Miserable
3	Annoyed Frustrated	Annoyed Frustrated Angry
4	Hurt Pain	Hurt Pain
5	School work	Learning
6	Daily Routine	Daily Routine
7	Tired Weak Drowsy	Tired Weak Energy Weary
8	Joining in activities that want to	Joining in activities that want to
9	Sleep	Sleep
10	Jealous	
11		Embarrassed

629

630

631 **Table 2: Characteristics of the sample**

632

Characteristic	N
Hunter's Bar Junior School	16
Firs Hill Community Primary School	15
Male	15
Female	16
Y3 (age 7-8 years)	8
Y4 (age 8-9 years)	8
Y5 (age 9-10 years)	8
Y6 (age 10-11 years)	7
White	17
Mixed/dual heritage	2
Asian or Asian British	12
Black or Black British	0
Chinese	0

633

634

635 **Table 3: Mean rank order, standard deviation (SD) and difference in mean rank,**
636 **for each set of statements**
637

Level	Mean rank order	SD	Difference
I can join in with any of the activities that I want to	1.10	0.30	0.92
I can join in with most of the activities that I want to	2.02	0.49	1.06
I can join in with some of the activities that I want to	3.08	0.43	0.73
I can join in with a few of the activities that I want to	3.81	0.46	1.19
I can join in with none of the activities that I want to	5.00	0.00	
My sleep is not affected	1.00	0.00	1.52
My sleep is a little bit affected	2.52	0.71	0.26
My sleep is a bit affected	2.77	0.59	1.05
My sleep is quite affected	3.82	0.75	1.26
My sleep is affected quite a lot	5.08	0.50	1.23
My sleep is affected a lot	6.31	0.69	0.92
My sleep is very affected	7.23	0.92	0.05
My sleep is really affected	7.27	0.76	1.73
I can't sleep at all	9.00	0.00	
My school work is not affected	1.19	1.08	1.32
My school work is a little bit affected	2.52	0.70	0.32
My school work is a bit affected	2.84	0.66	1.02
My school work is quite affected	3.85	0.83	1.16
My school work is affected quite a lot	5.02	0.70	1.27
My school work is very affected	6.29	1.08	0.00
My school work is really affected	6.29	0.69	1.71
I can't do my school work	8.00	0.00	
I don't feel worried	1.00	0.00	1.27
I feel a little bit worried	2.27	0.48	0.73

I feel a bit worried	3.00	0.55	0.73
I feel quite worried	3.73	0.60	1.69
I feel very worried	5.42	0.45	0.16
I feel really worried	5.58	0.45	
I don't have any pain	1.00	0.00	1.29
I have a little bit of pain	2.29	0.42	0.42
I have a bit of pain	2.71	0.42	1.58
I have quite a lot of pain	4.29	0.48	0.79
I have a lot of pain	5.08	0.59	0.55
I am really in pain	5.63	0.66	
I have no problems with my daily routine	1.00	0.00	1.27
I have a few problems with my daily routine	2.27	0.40	0.45
I have some problems with my daily routine	2.73	0.40	1.31
I have many problems with my daily routine	4.03	0.18	0.94
I can't do my daily routine	4.97	0.18	
It doesn't hurt	1.00	0.00	1.34
It hurts a little bit	2.34	0.57	0.55
It hurts a bit	2.89	0.59	0.89
It hurts quite a bit	3.77	0.48	1.52
It hurts quite a lot	5.29	0.51	0.66
It hurts a lot	5.95	0.57	0.81
It really hurts	6.76	0.56	

638

639

640

Table 4: Kendall Coefficient

641

Set	Kendall Coefficient
1	0.925
2	0.939
3	0.880
4	0.918
5	0.914
6	0.954
7	0.933

642

643

644

Table 5: Preference of children when statements were ranked equally

645

	Statement	Children's Preference (n preferring each statement)
1	My sleep is very affected	1
	My sleep is really affected	1
2	My school work is very affected	3
	My school work is really affected	2
3	I feel very worried	3
	I feel really worried	2

646

647