The Quality of Life scale for Children (QoL-C)

SCHOLARONE™
Manuscripts

Title- The Quality of Life scale for Children (QoL-C)

Abstract

Purpose-There is a lack of valid and reliable generic measures of Health Related Quality of Life(HRQoL) for children under eight. This was a preliminary study to assess the psycho-metric properties of the newly formulated Quality of Life Scale for Children (QoL-C), which uses a pictorial response format.

Method- 335 primary school children completed the QoL-C on two occasions, two weeks apart. Children aged 4-7 were interviewed one-to-one whilst children aged 8-9 completed the measure as a class activity. Test re-test reliability, convergent validity and child-parent concordance were assessed.

Findings- Only one child refused to complete the QoL-C, which suggests the measure is user-friendly. Test re-test reliability was moderate for the measures total score (ICC=0.48, 95% CI 0.39, 0.57) but low to fair for individual items (K from 0.13 to 0.37). Internal consistency was moderate (α=0.42 time one, 0.53 time two). A small significant correlation was found between the QoL-C and Child Health Meter in the expected direction (r=-0.32) suggesting convergent validity. There was low concordance between the children's QoL-C responses and parents responses (r=0.19) to a parallel measure.

Research implications-Our results suggest that further development of this measure is needed. However,our findings indicate that one-to-one support increases the reliability of very young children's responses. The use of pictures, emoticons and minimal text used in the QoL-C should be investigated further.

Value- Low parent-child concordance underscores the importance that younger children get the opportunity to share their views about their HRQoL

Keywords – QoL-C, quality of life, children, measure, health,

Article Classification: Research paper

A growing interest in the field of Quality of Life (QoL) research has led to an expansion in the use of Health Related Quality of Life (HRQoL) measures within health services (Varni *et al*., 2007a). The consensus is that measures of general HRQoL should be based on several domains that are necessary for healthy functioning (Wallander *et al*., 2001); namely physical, mental and social wellbeing (World Health Organisation (WHO) 1948; National Institute for Clinical Excellence (NICE), 2010). The aim of generic HRQoL measures is to gain an accurate representation of an individual's perception of their current health state (Hurst *et al*., 1997). There are also disease specific measures that assess the impact of clinical interventions for particular chronic illnesses (Brazier and Longworth, 2011) . These measures provide data that can be used to evaluate interventions on clinical trials across different health conditions (Food and Drug Administration (FDA), 2006), inform government and organisation policies, and improve patient healthcare experiences (Varni *et al*., 2007b; Robinson, 1993). There is a tension between generic measures that allow comparison across multiple conditions, versus disease specific measures, that may be more sensitive to change in the particular condition being measured, but make meaningful comparison across different areas difficult (Brazier and Longworth, 2011).

While several HRQoL measures have been developed for adults, comparatively few measures are available for children (Varni *et al*., 2007a). In the most recently conducted review of child HRQoL measures it was found that of the 30 generic measures available, only 11 were designed as self-report for children under eight (Solans *et al*., 2008). There is also the DISABKIDS measure for children aged 4-7 (Chaplin *et al*.,2008) and the younger years PedsQL that is self-report for children aged 5-7 (Varni *et al*.,2007b) . There are some disease specific measures available for children, for example -measures for childhood Asthma (Raat *et al*.,2005a), skin diseases (Beattie and Lewis-Jones,2006) and diabetes (Varni *et al*.,2003b), but further examination of measures that focus on specific conditions is beyond the scope of this article (Eiser and Morse, 2001b).

 The lack of current generic measures for children to complete may be due to the difficulties associated with designing measures appropriate and relevant to -younger children (Schwab-Stone *et al*., 1994). HRQoL is a complex concept, and its translation into a simple questionnaire that very young children can reliably and validly complete, represents a considerable methodological challenge (Schwab- Stone *et al.,* 1994; Ravens-Sieberer *et al.,*2006). There have been previous attempts at developing self-report measures for younger children, which suggest that valid data can be collected from children under eight, with further developmental related difficulties below the age of five (Lawford *et al.*, 2001). However, the lower age limit of children's ability to self-report HRQoL has yet to be sufficiently tested and so is currently unknown (Everhart and Fiese, 2009; Varni *et al., *2007b).

Research into health perception suggests that children as young as three are able to effectively communicate their opinion about their health, provided that the measure is presented in a format that they understand (Rebok *et al.,* 2001); in a way that they are willing to engage with (Varni *et al.,* 2007b); and holds their attention (Everhart and Fiese, 2009; Siegal *et al.,* 1990). However, developmental issues can also fundamentally threaten

the reliability of a child's self-report. Longitudinal studies have found that children show a response shift, due to a change in the way children understand and respond to questions as they develop, rather than a change in the latent variable being measured (Sprangers and Schwartz, 1999). Developmental psychologists suggest that children aged 4-7 can apply a dichotomous thinking style, which can result in the choice of extreme scores despite the availability of other options (De Civita et al., 2005). It has been suggested that that if children are encouraged to base their responses on several examples they will be more likely to provide representative answers (De Civita et al., 2005). To reduce the influence of developmental constraints, questions should always be about normal functioning within the age group being studied, so that lack of ability due to immaturity is not misinterpreted as poor HRQoL (Wallander et al., 2001).

As there are fewer reliable and valid measures available for very young children, proxy reports are often used for children under eight (Eiser and Morse, 2001a). However, there is often a lack of concordance between proxy reports and the child's self-report (Davis et al., 2007). Disagreement may be due to children having a different perspective to their parents or not fully understanding the questions, but it is difficult to know whose view to prioritise (Theunissen et al., 1998). Pal (1996) has suggested that proxy reports should only be used to provide additional insight, rather than as a substitute for the child self-report. Furthermore NICE have recently proposed that measures to evaluate QoL should be available to all children (NICE, 2010).

Therefore, there is a need to test the lower limit of children's ability to report their HRQoL and there should be standardisation across ages if practitioners and researchers wish to collect longitudinal data or compare scores across all ages.

A current problem with the majority of child HRQoL measures is that they are time consuming to administer (Solans et al., 2008), so especially in younger children, responses are likely to become skewed by fatigue (Varni et al., 2007b). Additionally, few measures are truly completed independently by children under eight. For example in the younger years PedsQL, parents assist their child in understanding and completing the measure (Varni et al., 2007b). Whilst no bias is intended, the children's answers may in fact reflect their parent's views (Connolly and Johnson, 1999). Neutral interviewers may carry a lower risk of bias, but are expensive and impractical in some environments where QoL measures are completed. A further issue is that, many QoL measures ask children to remember their health over a number of weeks. Retrospective recall in younger children has been demonstrated to be inaccurate so the usage of a shorter time scale may increase reliability (Rebok et al., 2001).

The EQ-5D construct

One of the most commonly used adult measures of HRQoL in adults is the European Quality of life - 5 Dimensional survey (EQ-5D). The EQ-5D, which was devised by the EuroQoL Group (www. euroqol.org) and is the primary measure used by NICE in clinical evaluations (Brazier and Longworth, 2011; Kind et al., 2005). The EQ-5D is a general measure of HRQoL covering five domains of healthy functioning: mobility, self-care, usual activities, pain/discomfort and anxiety/depression (Dolan, 1997). Scores on each domain range from 1-3 and total scores range from 5-15, with lower scores implying better QoL. Research has indicated that the EQ-

5D has strong test re-test reliability, validity and produces comparable results across Europe (Brooks *et al.,* 2003).

There is a version of the EQ-5D for young people (EQ-5D-Y), which was successfully adapted from the adult measure for use with children aged 8-15 years. The EQ-5D-Y has less reliability and validity among 8-9 year olds, because the measure includes a considerable amount of text and complex concepts, such as mobility, that younger children may not understand (Jelsma and Ramma, 2010). During the adaptation from the adult EQ-5D, researchers emphasized a need to establish the lower age limit of the EQ-5D construct (Burström *et al.,* 2011b) as there were no children under the age of eight included in the validation studies (Burström *et al.,*2011b). The lack of an adapted version of the EQ-5D for children under eight means that this age range may be being under-represented in NICE clinical evaluations. The short format of the EQ-5D suggests it may be ideal to adapt for use with younger children.

Introducing the QoL-C

The Quality of life scale for Children (QoL-C) is a new self-report HRQoL measure designed for children aged 4-9 years. Based on the format of the EQ-5D in consultation with the EuroQoL group, the measure uses a single item for five domains and uses simple language, such as 'moving around' instead of 'mobility'. Relevant pictures of children are used alongside each question, as well as written examples to assist comprehension (Rebok *et al.,* 2001; Burström *et al.,* 2011a). The measure uses three faces (happy, neutral and sad) as response options providing a visual anchor to the options 'no problems', 'some problems' and 'a lot of problems' (Burström *et al.,* 2011a). Emoticon scales have been successfully used in other quality of life measures for this age range (Chaplin *et al.,*2008). The five questions come together to produce a total QoL score of 5-15, with lower scores implying better QoL. Children during the development process found the vertical visual analogue scale ranging from 1-100 used in the EQ-5D and EQ-5D-Y confusing, so as an alternative, the QoL-C uses a health meter, presented horizontally to resemble the number lines used to develop numeracy for 3-5 year olds (Department for Education and Employment, 1999). The Child Health Meter asks children to rate their general health from 0-10, with emoticons presented alongside to indicate 'worst' health (sad emoticon = 0), 'OK' health (neutral emoticon = 5), and 'best' health (happy emoticon= 10), (Badia *et al.,*1999).

The current study was an initial assessment of the psychometric properties and practical application of the QoL-C in a healthy sample of primary school children. We anticipated that the measure would be feasible and acceptable. We predicted there would be moderate correlations between the first and second completion of the QoL-C, and given the very young age of our sample, that test re-test reliability would be lower among the youngest children (Olson *et al.,* 2007). We predicted low correlations between child and parent reports as commonly reported in the literature (Klassen *et al.,* 2006) and furthermore we expected a significant correlation between the QoL-C and the Child Health Meter.

Method

Participants
Participants included 335 children (aged 4-9) and 147 parents from two mainstream primary schools in the South West of England. The exclusion criteria were: children with English as a second language, children who demonstrated a clear lack of understanding, and children who did not want to complete the measure

Measures

QoL-C
The QoL-C, as described previously, was developed by researchers at the University of Exeter Medical School. Advice was sought from children in two school council groups aged 6-11 to guide the adaption of the wording and layout of the adult EQ-5D measure to make it more child friendly. The measure asks the child to describe their QoL 'today' (see Appendix 1).

Parent Questionnaires
Parents completed the PedsQL version 4.0 (Varni *et al.,* 2001) appropriate to the age of their child. The PedsQL is a widely used proxy measure which provides valid and reliable data for children aged 2-16 years (Varni *et al.,* 2007c; Varni *et al.,* 2003a; Varni *et al.,* 2001). It asks about the child's health over the last month. Scores range from 1 (very bad) to 5 (very good). This proxy report was selected to accompany the child data as it covers similar domains to the QoL-C. Parents also completed a five-point likert scale (Parent General Health Rating) to represent the child's general health, (1=very good to 5=very bad), which was used in the mental health of children and young people survey 2004 (Green *et al.,* 2005), and selected as a validation tool for the Child Health Meter.

Procedure
Following ethical approval from Exeter University School of Psychology Ethical committee, two schools were invited to take part in the study. Parents were informed about the study via a letter in the child's book bag, and had two weeks in which to choose to opt their child out of the study. Children, who were not opted out, completed the QoL-C measure at school, on two occasions, two weeks apart. This two week timescale was used so that the children would be familiar with the measure but would leave a long enough time to avoid recollection effects. Previous research has used timescales of 1-4 weeks between data collections with this age range (Varni *et al.,* 2007c; Jokovic *et al.,* 2004; Ravens-Sieberer and Bullinger,1998).

For children aged 4-7(n=224), the QoL-C and the Child Health Meter was administered in a one-to-one interview style by trained researchers using a standardised script (Appendix 2). A trained researcher would ask the child to use the emoticon scale to complete some verbal practice examples. If reasonable responses were given to these practice questions then the child completed the measure. If the child showed clear signs of not understanding then the child would be excluded and their responses would not be used .Children aged 8-9 years (n=110) completed the measure as a class activity . A researcher was introduced to the class who would then run through practice examples as a group discussion (Appendix 2). The

children then independently completed the measure but could ask for support from researchers.

Parents were sent the measures via the child's book bag after the researcher's first visit to the school and were instructed to return them to the school office for collection in two weeks.

Statistical analysis

Data were double entered to identify data entry errors. Analysis was performed using Stata version 12.1 (StataCorp, 2011). These data were skewed and were therefore analysed using non-parametric tests. Feasibility was assessed by the number of missing responses and refusals. The relationship between age and QoL-C scores was assessed using the Kruskall-Wallis test. Inter-rater effects were assessed by a comparison of QoL-C total scores from respondents completing the questionnaire with each rater. Test re-test reliability for the QoL-C total scores and the Child Health Meter was assessed through an Intraclass Correlation Coefficient (ICC) (Fleiss and Cohen, 1973; Green *et al.*, 2005) and the Bland-Altman. Test re-test reliability on each item of the QoL-C was assessed by a weighted Kappa. Internal consistency was assessed by calculating Cronbach's alpha. Spearman's rank correlation was used to calculate concordance and assess construct validity between the QoL-C and the Child Health Meter.

Results

Feasibility, Demographics and Distribution of QoL-C total scores

Only 16 (5%) parents opted their child out of the study. One parent was willing for their child to take part but did not wish to complete the parent measures, and only one child refused to take part. The measure took approximately seven minutes to administer. Children had a mean age of 6.5 years (SD= 1.6) and the number of children per year of age varied between 34 (age 4) and 75 (age 6) (see Table 1). Age and gender were evenly distributed across the two schools. There was no significant difference in total scores found between genders (all p-values >0.1).

Insert Table 1 here

The age distribution was similar between year groups except at the extremes, as each school year groups contains children whose age varies by 12 months. Thus, there were fewer four and nine year olds compared to those aged five to eight.

Insert Table 2 here

The response rate amongst children was very high at both time one and time two (91%, 95%) and the response rate was at the anticipated level for parents contacted via the book bag (40%, 45%).Most children and parents scored child QoL at the healthy ends of the scales.

Insert Table 3 here

As can be seen from Table 3, QoL-C scores did not differ across ages at either time point. Although there was evidence of a difference between ages for the Child Health Meter scores, however the median and inter-quartile range suggest that this difference is not meaningful; children tended to report at the healthy end of the scale, as would be predicted of a young, fit general population sample.

Insert Table 4 here

Reliability
Researchers completed between 14 and 37 one-to-one interviews among the younger children, and Table 4 suggests that there was little difference in responses obtained by individual researchers among the children who completed the measure as an interview, or that those interviewed responded differently to those who completed the measure as a class activity.

Insert Table 5 here

Test re-test reliability
The Intraclass correlation (ICC) for the QoL-C total scores was 0.48, 95% CI (0.39, 0.57), which suggested a moderate level of test re-test agreement. The 95% limits of agreement (-4.0, 4.6), indicated that we would expect to see a decrease of up to 4 or an increase up to 4.6 points in QoL-C total scores between the first and second time points. The ICC for the Child Health Meter score was 0.39, 95% CI (0.30, 0.49); the 95% limits of agreement (-6.3, 6.6), indicated that we can expect to find a decrease up to 6.3 or an increase up to 6.6 between visits. Both of these potential differences in score between time one and time two visits are quite large. Interestingly, we did not see marked differences in reliability by age (see Table 5), which may be because one-to-one administration increased the reliability of the youngest children, or that we lacked sufficient power to detect a significant difference.

Insert Table 6 here

Test re-test reliability for the individual items is illustrated in Table 6; kappa values ranged between 0.13 and 0.37; the question about 'moving around' had a particularly low kappa value of 0.13, which suggested a particularly poor level of agreement about this item.

Internal reliability

The Cronbach's alpha values (0.42 for time one and 0.53 time two) indicated moderate internal consistency of items included in the QoL-C.

Validity
A Spearman's Rank Correlation found a small but significant negative correlation in the predicted direction between the QoL-C total and the Child Health Meter score at time one

0.23, 95% limits of agreement CI (-0.34, -0.12) and time two -0.32, 95% CI (-0.42,-0.21), which indicated some consistency in children's responses to the two parts of the measure.

There was a small positive correlation 0.19, 95% limits of agreement CI (0.02-0.35), between the QoL-C total scores at time two and the parent PedsQL ratings, which is at the lower end, but not outside the range of agreement often seen between parents and children on measures of children's internal states (Davis *et al.,* 2007). There was also no evidence of a correlation between the Child Health Meter score and the Parent General Health Rating.

Discussion

Substantive findings
This study was a preliminary assessment of feasibility, reliability and validity of the QoL-C measure among children aged 4-9 years. Completion was brief for all children, including those assisted individually by a researcher. The measure demonstrated high levels of acceptability and feasibility, which suggests that the QoL-C offers an easy and practical method to open up conversations with children about different aspects of their general health that could be useful for practitioners working with children in a variety of contexts. Further research on a larger sample should empirically explore whether the theoretical influence of a dichotomous thinking style, is reduced by the use of emoticon, pictures and / or administration via one-to-one interview (Rebok *et al.,* 2001).

The results from this study suggest additional development is needed before the QoL-C is recommended as a research tool. The measure had moderate internal consistency and test re-test reliability -and validity is difficult to demonstrate with internal states, especially within the age range being studied (Harding, 2001). These results were not unprecedented in that the EQ-5D-Y another measure, that like the QoL-C was adapted from the EQ-5D, has previously reported moderate psychometric properties (Ravens-Sieberer *et al.,* 2010). The lack of an increase in test re-test reliability obtained with age was an unexpected finding, given previous research (Varni *et al.,* 2007b; Olson *et al.,* 2005), and importantly may indicate that individual support provided to the younger children successfully offset the impact of their young age. Our findings for the reliability of the Child Health Meter are particularly surprising given that visual analogue scales have been reported to be unreliable in children under six (Stanford *et al.,* 2006). It is possible that older children who completed the QoL-C independently felt pressured to complete the measure in the same time as their peers, and may have been more reluctant to ask for help. The inclusion of written examples as well as verbal ones offered a way of increasing understanding but we do not know if the children used them.

There was only limited confirmation of validity, with a significant but small correlation of the QoL-C total with the Child Health Meter (Raat *et al.,* 2005b), which may reflect that children consider aspects of their life experience differently compared with a general health question (Wille *et al.,* 2010). The comparison of the children's self-report and the PedsQL proxy report, as expected, yielded a small but significant correlation. Lack of parent child concordance is common in child research literature (Davis *et al.,* 2007). Whilst the domains assessed on the PedsQL were similar to the QoL-C's domains, it's possible that subtle

differences may have depressed correlations further. The PedsQL was the best available measure to validate against for our study as it has been successfully used in a variety of settings, but the poor parent-child agreement indicates the importance of further work to develop a robust measure for children in preference to the use of a proxy measure.

Strengths, limitations and directions for future research
This study benefitted from a moderate sample size for a developmental study, with a high response rate and consultation with children and experts during the development of the measure. The QoL-C was tested in the South West of England, with a predominantly white British population. Therefore, while the children who were consulted during the development of the measure preferred photographs to cartoons, in a more ethnically diverse population, providing photographs that all could identify with might be problematic. Further research could evaluate the impact of novel parts of the measure, such as the use of pictures, emoticons and simplified concepts. It is also important to investigate the impact of one-on-one administration on reliability as this could be key to the development of self-report measures for children under eight, though it raises costs and practical issues for large scale use. Whilst children were consulted during the development of this measure, consultation was limited by the resources available. Much more extensive qualitative research with different age groups should be conducted to inform further development of the measure.

Aside from psychometric and design issues, a key limitation of this study was that we were unable to ask the children *why* they had chosen a particular response. Whilst the usage of a one-to-one interview style and the practice questions served to check understanding, there is no guarantee that misconceptions were not occurring during the completion of the QoL-C (Rebok *et al.,* 2001). Further research should involve a qualitative element to explore the reasons given for responses and to further test/confirm children's level of understanding following the completion of the QoL-C measure. Adjustments of this nature to child QoL measures can seek to establish the lower limit of reliability for very young children.

The time scale of reporting used in the QoL-C may have influenced reliability. Whilst the 'today' time scale was used as children find it difficult to recall relevant examples over broader time scales (Rebok *et al.,* 2001), HRQoL may vary more if respondents are asked to report their health 'today' rather than 'over the past month'. Other measures use a broader time scale and report higher levels of reliability (De Civita *et al.,* 2005). The time scale used in the QoL-C was perhaps too narrow, turning the measure into more of an 'at the moment' assessment instead of a measure of a continual stable QoL trait (McGrath, 1990). Future QoL research for children should seek to develop optimum time scales for different age ranges as this will be more likely to capture an enduring picture of their QoL and enhance reliability (Rebok *et al.*, 2001).

Finally the QoL-C was developed from a successful measure of adult HRQoL so the domains and items used within the QoL-C were limited in that they were constrained to those

originally considered relevant to adults (Dolan, 1997) . Whilst the adult domains have been carefully adapted to suit a younger audience, it may be that not all these domains are relevant or can be comprehended by this age range (Eiser and Morse, 2001b). Children may think of their QoL in an altogether different way to adults. It is an interesting observation within our study that the domain with lowest test re-test reliability was mobility. Although essential to assess functioning in a clinical population, anecdotal information from the interviews suggests that children frequently misunderstood this question. For example, data collection diaries, kept by researchers revealed children's queries included the relevance of the usage of a scooter to travel to school or the ability to hop. It is possible that a lack of problems with mobility in this group made the question harder to comprehend. There was a lack of discrimination within the healthy sample studied and this can be attributed to the usage of a sample that is medically well so any differences are likely to be minor, but also can be attributed to the generality of the measure itself. We did not have access to the health status of the children who participated in this study so we were unable to discern any difference between healthy and unwell children within the sample studied, though chronic ill health would be likely to be uncommon. The QoL-C has still to demonstrate whether it can discern differences in the QoL of children who have and do not have disability or chronic illness.

Conclusion

Overall, the results of this study are testimony to the challenges of developing a generic HRQoL measure for younger children. However the measure did prove easy to complete and generated moderate reliability. The findings also provided support for the use of neutral interviewers as an administration method. We recognise that this work is very preliminary and that a great deal of further testing and development is necessary before the QoL-C could be adopted for research, but hope that our work will provide a foundation for future research to optimise the ability of young children's voices to be heard in evaluations of interventions and services that are designed for them.

References

Badia, X., Monserrat,S., Roset, M., and Herdman, M. (1999), "Feasibility,
    validity and test–retest reliability of scaling methods for health states: The visual
    analogue scale and the time trade-off", *Quality of life research*, Vol.8
    No.4, pp. 330-310.

Beattie, P.E., and Lewis-Jones, M.S. (2006) "Comparative study of impairment of quality of
    life in children with skin disease and children with other chronic childhood diseases",
    British Journal of Dermatology, Vol .155 No.1, pp. 145-151

Brazier, J.E. and Longworth, L. (2011), "*NICE DSU Technical Support
    Document 8: An introduction to the measurement and valuation of
    health for NICE submissions*", Sheffield: *Decision Support Unit*. August,
    pp. 1-29.

Brooks, R., Rabin, R. and De Charro, F. (2003) *The Measurement and
    Valuation of Health Status Using EQ-5D: A European Perspective:
    Evidence from the Euroqol Biomed Research Programme*.Dordrecht:
    Kluwer Academic Publishers

Burström, K., Egmar, A. C., Lugnér, A., Eriksson, M. and Svartengren, M.
    (2011a), "A Swedish child-friendly pilot version of the EQ-5D
    instrument—the development process", *The European Journal of
    Public Health*, Vol. 21 No.2, pp. 171-177.

Burström, K., Svartengren, M. and Egmar, A. C. (2011b), "Testing a Swedish
    child-friendly pilot version of the EQ-5D instrument—initial results", *The
    European Journal of Public Health*,Vol. 21 No. 2, pp. 178-183.

Chaplin,J.E.,Koopman,H.M.,Schmidt,S. and The DISABKIDS Group (2008) "DISABKIDS
    Smiley Questionnaire: The TAKE 6 Assisted Health-Related Quality of Life Measure
    for 4 to 7-Year-Olds" *Clinical Psychology and Psychotherapy*, Vol 15, pp173–180

Connolly, M. A. and Johnson, J. A. (1999),"Measuring quality of life in
    paediatric patients",*Pharmaeconomics*, Vol.16 No. 6, pp. 605-625.

Davis, E., Nicolas, C., Waters, E., Cook, K., Gibbs, L., Gosch, A. and Ravens-
    Sieberer, U. (2007), "Parent-proxy and child self-reported health related
    quality of life: using qualitative methods to explain the discordance",
    Quality of Life Research, Vol.16 pp. 863–871.

De Civita, M., Regier, D., Alamgir, A. H., Anis, A. H., FitzGerald, M. J. and
    Marra, C. A. (2005),"Evaluating Health-Related Quality-of-Life Studies
    in Paediatric Populations Some Conceptual, Methodological and
    Developmental Considerations and Recent Applications",
    Pharmacoeconomics, Vol. 23 No.7, pp. 659-685.

Department for Education and Employment.(1999), "*The national numeracy
    strategy framework for teaching mathematics from reception to year 6*".
    Cambridge: Cambridge University Press.

Dolan, P. (1997) "Modelling Valuations For Euroqol Health States", *Medical
    Care,* Vol.35, pp. 1095-1108.

Eiser, C. and Morse, R. (2001a) "A review of measures of quality of life for
    children with chronic illness", *Archives of Disease in Childhood*, Vol.
    84, pp. 205-211.

Eiser,C. and Morse, R. (2001b) "Quality-of-life measures in chronic diseases of childhood",

Health Technology Assessment, Vol.5 No.4, pp.1-157

Everhart, R. S. and Fiese, B. H. (2009)," Development and initial validation of a pictorial Quality of life measure for young children with Asthma", *Journal of Paediatric Psychology*, Vol. 34 No.9, pp. 966-976.

Fleiss, J. L. and Cohen, J. (1973),"The Equivalence Of Weighted Kappa And The Intraclass Correlation Coefficient As Measures Of Reliability", *Educational and Psychological Measurement*, Vol.33 No.3, pp. 613-619.

Food and Drug Administration. (2006), "*Guidance for industry: patient reported outcome measures: Use in medical product development to support labelling claims*", Rockville: Centre for Drug Evaluation and Research, Food and Drug Administration.

Green, H., McGinnity, Á., Meltzer, H., Ford, T. and Goodman, R. (2005), *Mental health of children and young people in Great Britain, 2004.* New York: Palgrave Macmillan.

Harding,L.(2001)"Children's quality of life assessments: A review of generic and health related quality of life measures completed by children and adolescents" *Clinical Psychology & Psychotherapy*,Vol. 8, No. 2, pp. 79–96,

Hurst, N. P., Kind, P., Ruta, D., Hunter, M. and Stubbings, A. (1997), "Measuring health related quality of life in rheumatoid arthritis: validity, responsiveness and reliability of EuroQol (EQ-5D)", *British Journal of Rheumatology,* Vol. 36, pp. 551-559.

Jelsma, J. and Ramma, L. (2010), "How do children at special schools and their parents perceive their HRQoL compared to children at open schools?" *Health Quality of Life Outcomes*, Vol. 8 No. 72, pp. 1-7.

Jokovic,A.,Locker,D.,Tompson,B.,and Guyatt,G.(2004) "Questionnaire for Measuring Oral Health-related Quality of Life in Eight- to Ten-year-old Children", *Pediatric Dentistry* ,Vol.26, No.6, pp.512-518

Kind, P., Brooks, R. and Rabin, R. (2005), *EQ-5D Concepts And Methods: A Developmental History,* Dordrecht: Springer.

Klassen, A. F., Miller, A. and Fine, S. (2006), "Agreement between parent and child report of quality of life in children with attention-deficit/hyperactivity disorder", *Child: Care, Health and Development*, Vol. 32. No. 4, pp. 397-406.

Lawford,J.,Volavka,N.,and Eiser,C.(2001) "A generic measure of Quality of Life for children aged 3-8 years: results of two preliminary studies", *Developmental Neuro-rehabilitation*, Vol.4, No.4, pp.197-207

McGrath, P. A. (1990), *Pain in children: Nature, assessment, and treatment.* New York: Guilford.

National Institute of Clinical Excellence.(2010), "Promoting the quality of life of looked-after children and young people: NICE public health guidance 28", Vol. 89. Manchester: NICE.

Olson, L. M., Radecki, L., Frintner, M. P., Weiss, K. B., Korfmacher, P. and Siegel, R. M. (2007), "At What Age Can Children Report Dependably on Their Asthma Health Status?", *Pediatrics*, Vol. 119 No.1, pp. 93-103.

Pal, D.K. (1996),"Quality of life assessment in children: a review of conceptual and methodological issues in multidimensional health status measures"

Journal of Epidemiology and Community Health, Vol. 50, pp. 391-396.

Raat,H., Bueving,J.C., De Jongste,J.C., Grol,M.H., Juniper,J.C., and Van Der Wouden,J.C.(2005) "Responsiveness, longitudinal- and cross-sectional construct validity of the Pediatric Asthma Quality of Life Questionnaire (PAQLQ) in Dutch children with asthma", *Quality of Life Research* ,Vol. 14, No. 1, pp. 265-272

Raat, H., Botterweck, A. M., Landgraf, J. M., Hoogeveen, W. C. and Essink-Bot, M. (2005),"Reliability and validity of the short form of the child health questionnaire for parents (CHQ-PF28) in large random school based and general population samples",*Journal of Epidemiology and Community Health*, Vol. 59, pp. 75–82.

Ravens-Sieberer,U.,Wille,N.,Badia,X.,Bonsel,G., Burström,K.,Cavrini,G., Devlin,N.,Egmar, A.,Gusi,N.,Herdman,M.,Jelsma,J.,Kind,P.,Olivares,P.R.,Scalone,L.,and Greiner,W. (2010), "Feasibility, reliability, and validity of the EQ-5D-Y: results from a multinational study", *Quality of Life Research*,Vol. 19, No. 6, pp 887-897,

Ravens-Sieberer,U. and Bullinger,M(1998) "Assessing health-related quality of life in chronically ill children with the German KINDL: first psychometric and content analytical results", *Quality of Life Research*, Vol.7, No.5, pp.399-407

Ravens-Sieberer,U.,Erhart, M., Wille, N., Wetzel, R., Nickel, J., and Bullinger, M.(2006), "Generic Health-Related Quality-of-Life Assessment in Children and Adolescents", *PharmacoEconomics*, Vol. 24 No.12 , pp. 1199-1220.

Rebok, G., Riley, A., Forrest, C., Starfeild, B., Green, B., Robertson, J. and Tambor, E. 2001),"Elementary aged school-aged children's reports of their health: A cognitive interviewing study", *Quality of life research*, Vol. 10, pp. 59-70.

Robinson, R. (1993),"Cost -effectiveness analysis", *British Medical Journal*, Vol. 307, pp. 859-862.

Schwab-Stone, M., Fallon, T., Briggs, M. and Crowther, B. (1994),"Reliability of diagnostic reporting for children aged 6-11 years: a test-retest study of the diagnostic interview Schedule for Children-Revised", *American Journal of Psychiatry*, Vol. 151 No. 7, pp. 1048-1054.

Siegal, M., Pat, J and Eiser, C. (1990),"A re-examination of children's conceptions of contagion", *Psychology & Health*, Vol. 4 No. 2, pp. 159-165.

Solans, M., Pane, S., Estrada, M., Serra- Sutton, V., Berra, S., Herdman, M., Alonso, J. and Rajmil, L. (2008),"Health-Related Quality of Life Measurement in Children and Adolescents: A Systematic Review of Generic and Disease-Specific Instruments", *Value in Health*, Vol. 11 No. 4, pp. 742-764.

Sprangers, M. A. G. and Schwartz, C. E. (1999),"Integrating response shift into health-related quality of life research: a theoretical model", *Social Science Health & Medicine*, Vol. 48 No. 11, pp. 1507-1515.

Stanford, E. A., Chambers, C. T. and Craig K. D. (2006), "The role of developmental factors in predicting young children's use of a self-report scale for pain", *Pain*, Vol. 120 No. 1, pp. 16–23.

StataCorp. (2011), *Stata Statistical Software: Release 12*. College Station: StataCorp LP.

Theunissen, N. C. M., Vogels, T. G. C., Koopman, H. M., Verrips, G. H. W., Zwinderman, K. A. H., Verloove-Vanhorick, S. P. and Wit, J. M. (1998),"The Proxy Problem: Child Report versus Parent Report in Health-Related Quality of Life Research", *Quality of Life Research*,Vol. 7, pp.387-397.

Varni, J. W., Limbers, C. A., and Burwinkle, T. M. (2007a),"Literature Review: Health Related Quality of Life in Paediatric Oncology: Hearing the Voices of Children", *Journal of Paediatric Psychology*, Vol. 32 No. 9, pp. 1151-1163.

Varni, J. W., Limbers, C. A. and Burwinkle, T. M. (2007b), "How young can children reliably and validly self-report their health-related quality of life?: An analysis of 8,591 children across age subgroups with the PedsQL™ 4.0 Generic Core Scales", *Health and Quality of life outcomes*, Vol. 5 No. 1, pp. 1477-7525.

Varni, J. W., Limbers, C. A., and Burwinkle, T. M. (2007c),"Parent proxy-report of their children's health-related quality of life: an analysis of 13,878 parents' reliability and validity across age subgroups using the PedsQL™ 4.0 Generic Core Scales", *Health and Quality of Life Outcomes,* Vol. 5, No. 2 pp. 1477-7525.

Varni, J. W., Seid, M., and Kurtin, P. S. (2001),"PedsQL 4.0: reliability and validity of the Pediatric Quality of Life Inventory version 4.0 generic core scales in healthy and patient populations", *Medical Care*, Vol. 39 No. 8, pp. 800-812.

Varni, J. W., Burwinkle, T. M., Seid, M., and Skarr, D. (2003),"The PedsQL™ 4.0 as a pediatric population health measure: Feasibility, reliability, and validity", *Ambulatory Pediatrics*, Vol. 3 No. 6, pp. 329-341.

Varni,J.W., Burwinkle,T.M., Jacobs,J.R.,Gottschalk,M., Kaufman,F., and Jones,K.L.(2003)"The PedsQL™ in Type 1 and Type 2 Diabetes :Reliability and validity of the Pediatric Quality of Life Inventory™ Generic Core Scales and Type 1 Diabetes Module", *Diabetes Care*, Vol. 26 No. 3, pp. 631-637

Wallander, J. L., Schmitt, M. and Koot, H. M. (2001),"Quality of Life Measurement in Children and Adolescents: Issues, Instruments and Applications", *Journal of Clinical Psychology*, Vol. 57 No. 4, pp. 571–585.

Wille. N., Badia, X., Bonsel, G., Burstrom, K., Cavrini, G., Devlin, N., Egmar, A., Wolfgang, G., Gusi, N., Herdman, M., Jelsma, J., Kind, P., Scalone, L. and Ravens-Sieberer. (2010),"Development of the EQ-5D-Y: a child friendly version of the EQ-5D", *Quality of Life Research*, Vol. 19 No. 6, pp. 875-886.

World Health Organisation. (1948), "Constitution of the world health organisation: basic document", Geneva: WHO.

www.euroqol.org

14

Tables for the Quality of Life Scale for Children (QoL-C) resubmission article

Table 1 Percentage age distribution of children who participated at either time point

Age	N	Percentage distribution
4	34	10.2
5	64	19.2
6	75	22.5
7	51	15.3
8	67	20.1
9	43	12.9
	Total	100.0% (N=334)

N is number of respondents per age

Note age is out of 334 not 335 – there is one child who has missing data for age.

Table 2 *Median and Interquartile range of scores on all measures*

QOL measure	N	Median (IQR)
QoL-C total T1[1]	310	7 (6 to 8)
QoL-C total T2[1]	309	6 (5 to 8)
Child Health Meter T1[2]	311	10 (8 to 10)
Child Health Meter T2[2]	307	10 (7 to 10)
Parent General Health rating[3]	146	1 (1 to 2)
Parent PedsQL rating[4]	136	16 (9 to 22)

T1 and T2 indicate time point one and two, n: number of respondents in sample of total sample, IQR: inter-quartile range

[1] Individual items are scored 1 (no problems), 2 (some problems) and 3 (a lot of problems) – total score is between 5 and 15

[2] Scale 0-10

[3] Individual items scored 0 (never), 1 (almost never), 2 (sometimes), 3 (often) and 4 (almost always) – total score is between 0 and 92

[4] Scored 1 (very good), 2 (good), 3 (fair), 4 (bad) and 5 (very bad)

Table 3 *Median and IQR – Kruskal–Wallis test for age*

QOL measure	4	5	6	7	8	9	p
QoL-C total T1	7 (6 to 9)	6 (5 to 8)	7 (6 to 9)	7 (6 to 9)	7 (6 to 8)	7 (6 to 7)	0.23
QoL-C total T2	6 (5 to 10)	5 (5 to 8)	7 (5 to 8)	6 (5 to 8)	6 (5 to 8)	6 (5 to 7)	0.22
Child HM T1	10 (10 to 10)	10 (10 to 10)	10 (8 to 10)	9 (7 to 10)	9 (7 to 10)	9 (8 to 10)	0.003
Child HM T2	10 (7 to 10)	10 (10 to 10)	10 (10 to 10)	10 (7 to 10)	9 (6 to 10)	9 (6 to 10)	<0.001

HM: Health Meter, T1 and T2 indicate time point one and two, *indicates significance above 0.005.

Table 4 *Median score on the Qol-C total scores per rater for children aged 4-7 at time one and time two and the median scores for children not interviewed aged 8-9*

		Time one				Time two	
Rater	N	Median(IQR)	Range[5]	Rater	N	Median(IQR)	RangeError! Bookmark not defined.
1	30	6 (5 to 7)	5 to 10	1	29	6 (5 to 8)	5 to 11
2	37	8 (6 to 9)	5 to 12	2	35	7 (5 to 10)	5 to 13
3	34	7 (6 to 8)	5 to 11	3	31	6 (5 to 8)	5 to 10
4	26	7 (5 to 10)	5 to 12	4	24	5.5 (5 to 8)	5 to 10
5	17	7 (6 to 8)	5 to 12	5	17	6 (5 to 7)	5 to 10
6	14	6 (5 to 6)	5 to 9	6	21	6 (5 to 7.5)	5 to 12
None	152	7 (6 to 8)	5 to 11	None	161	6 (5 to 8)	5 to 12

n is number of respondents in sample of total sample, IQR: inter-quartile range, NI: not interviewed

Table 5 *Intraclass Correlation Coefficient by age*

QOL measure	4	5	6	7	8	9
QoL-C total	0.51 (0.24 to 0.77)	0.55 (0.37 to 0.71)	0.28 (0.11 to 0.54)	0.46 (0.26 to 0.68)	0.61 (0.47 to 0.76)	0.24 (0.04 to 0.62)
CHM	0.44 (0.19 to 0.72)	0.74 (0.62 to 0.84)	--[6]	0.20 (0.04 to 0.55)	0.26 (0.08 to 0.54)	0.41 (0.19 to 0.66)

CHM: child health meter 1. Too few responses to report

---

[5] Scores can range from 5 to 15

7 Scores can range from 5 to 15

[6] could not get an estimate of ICC for age 6 VAS scores, xtreg, xtmixed gave answer 0.004, loneway (fixed effects 0.38 (0.28 to 0.48) and similar for xtreg with fixed effects.

Table 6 *(Weighted) Kappa Coefficient for each domain*

Question	(weighted) kappa and 95% CI[7]
Usual activities	0.34 (0.18,0.50)
Pain	0.37 (0.25,0.48)
Self care	0.27 (0.14,0.43)
Mobility	0.13 (-0.04,0.35)
Sad	0.35 (0.23,0.45)

CI: confidence interval

---

[7] Confidence intervals obtained using bootstrap (N=1000), bias-corrected estimates (also have percentile estimate option)
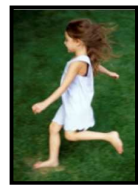
Appendix 1: The QoL-C measure

Name: ....   Age:................

EQ–5D–C

Do you have any problems with...?

Please put a circle around the cartoon face that shows your health today.

	No problems	Some problems	A lot of problems

Moving (e.g. walking around)

Looking after myself

(e.g. washing or dressing myself)

Doing usual activities

(e.g. going to school, playing, hobbies, doing things with family/friends)

Having pain (being sore)

Feeling worried, sad or unhappy

For Review Only

The Child Health Meter

Office Use: Baseline/Follow up
Child ID: Class: Year Group: School: Age: Date: /   /

Child Health Meter

We would like to know how you are feeling today.

The line is numbered from 0–10.
0 is the <u>worst</u> health you could imagine.
10 is the <u>best</u> health you could imagine.

Circle the number you think shows how good or bad your health is <u>today</u>.

0   1   2   3   4   5   6   7   8   9   10

Version

Appendix 2: Standardised script used with children in class and 1:1

Provisional Script for to introduce study in class

Note: We do not intend this script to be read verbatim. There is room to change the exact wording if necessary. The structure should remain the same though, and the finalised examples of happy/neutral/sad responses should not be changed between classes. Each child MUST be debriefed.

Introduction:

"My name is etc. Today, we'd like you to help us with a special project. We're going to ask you some questions about how you are and if you do or do not have problems with some different things. The answers you give us will be used to help other children like you.

"There are no right or wrong answers; we just want to know how you feel. It is very important you answer the questions on your own, your friends might have completely different ideas to you, but we really want to hear what you yourself have to say.

"To answer the questions on the first side of the page you need to circle the face you think shows if you have problems with some things. If you don't have any problems, circle the happy face; if you have some problems, circle the OK face and if you have a lot problems, circle the sad face. Let's do some examples now:

"Do you have any problems clapping your hands?"

"Do you have any problems tying your shoelaces?"

"Do you have any questions? If you don't understand something then please put your hand up and we will come and help you! Finally, whatever you say on the sheets is a secret. The only people who will know the answers are you and us. Your teachers and your parents won't see your answers. If your answers make us worried that you're in danger or being hurt, or that your friends are in danger and being hurt, we might have to talk about them to some special grownups."

Debriefing

"Well done, you have been so helpful (praise of some kind). Do you have any questions?

Version 1:2/11/11

"If you have any worries after today, talk to your teacher or mum or dad. Thank you for helping us, well done (hand them a sticker)."

Version 1:2/11/11