

# A comparison of children and adolescent's self-report and parental report of the PedsQL among those with and without autism spectrum disorder

M. A. Stokes<sup>1</sup> · L. Kornienko<sup>1</sup> · A. M. Scheeren<sup>2,3</sup> ·  
H. M. Koot<sup>2,3</sup> · S. Begeer<sup>2,3</sup>

Accepted: 15 December 2016 / Published online: 22 December 2016  
© Springer International Publishing Switzerland 2016

## Abstract

**Purpose** Children and adolescents with autism spectrum disorders (ASD) are understood to experience a reduced quality of life compared to typically developing (TD) peers. The evidence to support this has largely been derived from proxy reports, in turn which have been evaluated by Cronbach's alpha and interrater reliability, neither of which demonstrate unidimensionality of scales, or that raters use the instruments consistently. To redress this, we undertook an evaluation of the Pediatric Quality of Life Inventory™ (PedsQL), a widely used measure of children's quality of life. Three questions were explored: (1). do TD children or adolescents and their parents use the PedsQL differently; (2). do children or adolescents with ASD and their parents use the PedsQL differently, and (3). do children or adolescents with ASD and TD children or adolescents use the PedsQL differently? By using the scales differently, we mean whether respondents endorse items differently contingent by group.

**Methods** We recruited 229 children and adolescents with ASD who had an IQ greater than 70, and one of their parents, as well as 74 TD children or adolescents and one of their parents. Children and adolescents with ASD (aged

6–20 years) were recruited from special primary and secondary schools in the Amsterdam region. Children and adolescents were included based on an independent clinical diagnosis established prior to recruitment according to DSM-IV-TR criteria by psychiatrists and/or psychologists, qualified to make the diagnosis. Children or adolescents and parents completed their respective version of the PedsQL.

**Results** Data were analysed for unidimensionality and for differential item functioning (DIF) across respondent for TD children and adolescents and their parents, for children and adolescents with ASD and their parents, and then last, children and adolescents with ASD were compared to TD children and adolescents for DIF. Following recoding the data, the unidimensional model was found to fit all groups. We found that parents of and TD children and adolescents do not use the PedsQL differently ( $\chi^2_{(46)} = 64.86, p = \text{ns}$ ), consistent with the literature that children and adolescents with ASD and TD children and adolescents use the PedsQL similarly ( $\chi^2_{(69)} = 92.22, p = \text{ns}$ ), though their score levels may differ. However, children and adolescents with ASD and their parents respond to the PedsQL differently ( $\chi^2_{(115)} = 190.22, p < 0.001$ ) and contingently upon features of the child or adolescent.

**Conclusions** We suggest this is due to children or adolescents with ASD being less forthcoming with their parents about their lives. This, however, will require additional research to confirm. Consequently, we conclude that parents of high-functioning children with ASD are unable to act as reliable proxies for their children with ASD.

**Keywords** Autism · ASD · PedsQL · Item response analysis · Child · Children · Parent · Proxy reporting · Differential item function

✉ M. A. Stokes  
mark.stokes@deakin.edu.au

<sup>1</sup> School of Psychology, Deakin University, 75 Pigdons Rd, Waurn Ponds, VIC 3216, Australia

<sup>2</sup> Section Clinical Developmental Psychology, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

<sup>3</sup> EMGO Institute for Health and Care Research, Amsterdam, The Netherlands

## Introduction

Autism spectrum disorders (ASD) are a group of developmental disabilities thought to affect about 1 in 68 children [9]. When compared to typically developing children (TD), many of them experience a markedly reduced well-being, measured as quality of life (QoL; [17]), not to mention the stress experienced by their parents [32, 38]. QoL is understood to involve both affective and cognitive elements [10, 14, 16]. Accurate measurement of QoL in children and adolescents with ASD is particularly important. This informs much of the utility of interventions. If interventions are not improving the well-being and quality of life of individuals with ASD, the value of these interventions would be questionable. Hence, it is important to establish accurate measurement of QoL among those with ASD and to establish the psychometrics of these QoL measurements. Yet exploration of QoL within populations with ASD has been limited. In the current study, we closely examine parent- and child-reported QoL in a large sample of children with and without ASD.

Recently two reviews have been published on QoL in autism. Ikeda et al. [25] undertook a systematic review of QoL assessments in children and adolescents with ASD and found that of 13 articles evaluating QoL, 11 evaluated QoL against either norms or TD control participants. Van Heijst and Geurts [54] reported a meta-analysis on quality of life in autism across the lifespan. They included 10 studies, overlapping with around 50% of the Ikeda et al. study. Thus, to date eleven papers have compared children and adolescents with ASD against norms or TD children and adolescents. Each found QoL was significantly reduced in those with ASD.

The above studies relied on parent report (4 studies), self-report (3 studies), or combined self- and parent-reported QoL (4 studies). The variable use of informants highlights the importance of analysing the value of self- and parent report, including possible measurement biases. To date, no studies have analysed these issues. This is surprising, because autism has been suggested to affect the child's ability to reflect on his or her own experiences, and there has been much debate over the use of self-report measures in ASD. For instance, parents were found to rate the QoL of their children with autism lower than the children themselves do [29]. It could thus be argued that individuals with ASD and their parents use QoL scales differently from typically developing groups. Though this may not be specific to ASD, Bastiaansen et al. [4] reported that proxy-reported (i.e. parents or clinicians) QoL was found to be lower than that reported by the children themselves for all psychiatric or developmental conditions assessed. Despite that this result was more general than

ASD, this still suggests proxy reports of QoL may be used differently than self-reported QoL and that this would then apply to those with an ASD. Given that QoL scores are based on subjective experience or parental estimations of these experiences, it is difficult to establish whether the underlying trait (subjective well-being) is indeed similar in those with ASD compared to TD groups. However, even in the absence of this information, we can disentangle whether parents and children differ in their probabilities of responding to QoL measures, using new techniques.

The Pediatric Quality of Life Inventory™ version 4.0 (PedsQL) is one of the most widely studied and cited assessments of child QoL reported in the literature, with more than 1100 citations to date (*cf.* [www.pedsq.org](http://www.pedsq.org)) covering many thousands of children, both with and without various health conditions, and including at least seven evaluations of the item structure of this instrument. The PedsQL consisted of 23 items evaluating aspects of a child or adolescents quality of life. These are divided into four subscales addressing physical functioning, emotional functioning, social functioning, and last, school functioning. All items are responded to on five-point Likert scales (0 = never a problem to 4 = almost always a problem). Various findings suggest that underlying factor structure of the PedsQL is multidimensional, consisting of four [44], five [37, 42, 60], or even six factors [1, 28], and various configurations of second order factor models [23]. The four subscales may be each treated as a separate dimension; however, the PedsQL™ 4 manual specifies that all items provide a single, unidimensional score of well-being, representing the sum of all items [56]. In the present study, we seek to evaluate whether parents and children are responding to the same unidimensional overall construct of quality of life.

To date, the PedsQL has been used in six studies of children and adolescents with ASD [31, 36, 49, 50, 52, 55]. Reliability has generally been assessed by Cronbach's  $\alpha$ , and most of the reported values fall within acceptable limits for Cronbach's  $\alpha$  (0.72–0.93). However, Cronbach's  $\alpha$  is an insufficient basis from which to conclude either general reliability or that the scale is unidimensional. Cronbach's  $\alpha$  reflects only the lower bound of reliability, not the actual reliability, is vulnerable to manipulation through the number of test items, reflects the average relatedness of items for a particular sample, and does not reflect the unidimensionality of a particular set of items even when scores are high [41, 48, 51].

In the present day, with ready access to much more powerful techniques, Cronbach's  $\alpha$  is an insufficient index to establish the reliability of an instrument. Item response analysis or the Rasch model both offer a more sophisticated and nuanced way to establish reliability of an instrument,

as well as its unidimensionality [12, 21]. For instance, using item response analysis, groups can be compared in their differential probability to endorse items; specifically, as well-being increases in an individual, it would be expected that they select or rate higher scores of well-being. If respondents in one group, for example parents, differ systematically from respondents in another group, for example children, the two groups would show a differential probability to endorse items. In these instances, differential item functioning (DIF) is used to measure differences between respondents' likelihood to respond based upon their group membership and their level of trait or ability. Item response analysis first evaluates each individual's ability and assigns all respondents to a class interval based upon their assessed trait level ability (i.e. overall score). Thus, all individuals are of known abilities (i.e. class interval) and known groups. All persons in the same class interval have the same underlying level of trait. Each class interval is relatively homogenous for the trait and can now be compared for differences in group membership (i.e. person traits). DIF analysis then tests the interaction of group and level of trait. Two kinds of DIF may be found. Uniform DIF arises where one or more groups have a consistent advantage relative to other groups—the test is therefore easier for one or more groups across all levels of ability. However, where the advantage is contingent upon both group and class interval (i.e. level of trait), this is non-uniform DIF. The presence of non-uniform DIF suggests modifications need to be made to the scale or instrument being assessed [57].

While the item structure of the PedsQL has previously been analysed (cf. [18, 27, 33]), none have published this to date with ASD, evaluated how individuals with ASD might use the scale differently from TD individuals, or how parents and children with ASD may differ in their use of the PedsQL, thereby establishing the differential item functioning. The presence of DIF in a scale renders its use with different informants as problematic and likely to lead to erroneous conclusions [12, 21]. Agreement between parents and children on measures of QoL has been found largely to vary by measure [20, 53], and the PedsQL is no exception. Studies evaluating the PedsQL suggest DIF is present between parents of, and their TD children [27], between parents of different genders evaluating their own healthy children [18], between children with special health care needs and unaffected children [24], and between healthy and unwell children from the general population [30, 33]. Further, it has been found that some levels on items are not the most likely choice for any level of trait, known as disordered thresholds [26]. Therefore, the model may not be unidimensional in its standard form.

We aimed to establish if the PedsQL was responded to by parents of and ASD or TD children similarly.

Consequently, further aims were to establish if DIF was present or not between parents and their children with ASD, between parents and their TD children, and between children with ASD and TD children. It was hypothesised that (1) parents of TD children when answering about their child and their children themselves would answer the PedsQL in similar ways; (2) parents of children with ASD when answering about their child and their children themselves would not differ on the PedsQL; and (3) the PedsQL was a unidimensional measure in children that assessed QoL regardless of the presence of a diagnosis of ASD.

## Method

### Participants

The sample consisted of 74 parents of TD children and the 74 children themselves, as well as 160 parents of children with ASD and 229 children with ASD. Gender distributions were equal between groups; there being 63 boys and 11 girls in the TD group ( $M = 85.1\%$ ,  $F = 14.9\%$ ) and 196 boys and 33 girls in the group with ASD ( $M = 85.6\%$ ,  $F = 14.4\%$ ;  $\chi^2_{(1)} = 0.009$ ,  $p = 0.923$ ). Children and adolescents with ASD were recruited from special primary and secondary schools in the Amsterdam region. Children were included based on a clinical diagnosis of Autism or Aspergers disorder established prior to recruitment according to DSM-IV-TR criteria [2] by psychiatrists and/or psychologists who were not involved in the current research project and who were qualified to make the diagnosis. The diagnostic process included psychiatric and neuropsychological examinations. The comparison group was recruited via public primary and secondary schools in the Amsterdam region. Parents confirmed the absence of ASD in the comparison group.

Data from all participants who successfully completed the Dutch version of the Peabody Picture Vocabulary Test-III [19] and who had a receptive verbal IQ of at least 70 or higher were included in the analyses. Data from these high-functioning participants with ASD (HFASD) were included when they met the clinical cut-off on the Social Responsiveness Scale [15, 45]. Consequently, the HFASD group consisted of 202 participants (173 boys; 29 girls) versus 68 participants in the comparison group (58 boys; 10 girls). Mean age of the final HFASD group was significantly higher compared to the comparison group ( $t_{(268)} = 4.06$ ,  $p < .001$ ,  $d = .50$ ; see Table 1). Receptive verbal IQ did not differ between the groups ( $t_{(268)} = 1.53$ ,  $p = .12$ ,  $d = .19$ ).

**Table 1** Descriptives for the ASD group and the typically developing comparison group

Child variables	Total ASD group ( <i>n</i> = 202)		Comparison group ( <i>n</i> = 68)		Group differences
	<i>M</i> (SD)	Range	<i>M</i> (SD)	Range	
Gender ( <i>n</i> boys/girls)	173/29	58/10	$\chi^2 = 0.00, n.s.$		
Age (years)	13.7 (3.1)	6.4–20.5	12.0 (2.8)	6.0–17.1	$t_{(268)} = 4.06, p < .01$
Receptive verbal IQ (PPVT)	104 (12.7)	72–132	107 (13.4)	77–144	$t_{(268)} = 1.53, n.s.$
Total ADOS 3/4	5.8 (4.1)	0–19	–	–	
Total SRS	73.4 (11.4)	47–106	49.6 (5.1)	40–59	$t_{(70.7)} = 26.55^\ddagger, p < .01$
Employment	Mother	Father	Mother	Father	Mother * $U_{(203)} = 3075, p < .05$
No profession	13	1	4	0	Father * $U_{(199)} = 3062, n.s.$
Elementary occupation	1	2	0	0	
Lower occupation	23	19	6	7	
Middle occupation	74	82	14	19	
Higher occupation	47	46	16	14	
Academic occupation	9	14	13	12	
Missing	62	65	21	22	
Education					Mother * $U_{(215)} = 3549, n.s.$
Elementary school	0	0	2	4	Father * $U_{(213)} = 3527, n.s.$
Lower vocational ed.	5	2	17	26	
Lower general ed.	7	5	14	10	
Secondary vocational ed.	4	15	46	42	
Secondary general ed.	8	7	21	28	
Higher professional ed.	10	11	52	39	
Academic ed.	18	11	16	25	
Otherwise	2	3	13	4	
Missing	20	20	48	51	
% living with both bio. parents	73.1%		84.6%		$Z = 1.92^\dagger, p < 0.05.$

\* Mann–Whitney U; † proportion-based Z test; ‡ *t* test based upon unequal variances

## Procedure

After receiving written permission from their parents, children and adolescents were invited to participate. All tests were administered by trained psychologists and master students and took place at the participants' schools. Parents reports were obtained via mail. Quality of life was evaluated by use of the age appropriate versions of the PedsQL™ version 4, autistic symptomatology was confirmed by formal diagnosis and report from psychologists independent of this study and an SRS score greater than 60. Normal intelligence was confirmed using the Peabody Picture Vocabulary Test.

## Measures

**Quality of Life** Subjective quality of life was assessed using the *Pediatric Quality of Life Inventory™ version 4.0* (PedsQL; Parent, Child, 23 items; see Table 2). The instructions ask the respondent to indicate how much of a

problem an item has been for the child during the past month. By formulating the instruction in this way, the informant is not asked to rate the presence of a certain behaviour, but if present, its impact on the child's everyday functioning. The items are scored on a five-point Likert scale (0, 1, 2, 3, 4). Four subscales and a total score can be computed, covering the following dimensions of QoL: (1) physical functioning (8 items, e.g., “hard to do sports” or “having hurts”); (2) emotional functioning (5 items, e.g., problems with “feeling angry” or “trouble sleeping”); (3) social functioning (5 items, e.g., “trouble getting along with peers” or “being teased”); and (4) school functioning (5 items, e.g., “trouble keeping up with schoolwork” or “missing school”). Good reliability and validity have been reported for the American and Dutch versions [5] of the PedsQL. The PedsQL has three self-report versions for children, 5–7, 8–12, and 13–18 years. The age appropriate version was used with each child.

*The Autism Diagnostic Observation Schedule* (ADOS; [34]) assesses autism across age, developmental level, and

**Table 2** PedsQL items

Item	
Physical functioning (problems with...)	
1	Walking more than one block
2	Running
3	Participating in sports activity or exercise
4	Lifting something heavy
5	Taking a bath or shower by him or herself
6	Doing chores around the house
7	Having hurts or aches
8	Low energy level
Emotional functioning (problems with...)	
9	Feeling afraid or scared
10	Feeling sad or blue
11	Feeling angry
12	Trouble sleeping
13	Worrying about what will happen to him or her
Social functioning (problems with...)	
14	Getting along with other teens
15	Other teens not wanting to be his or her friend
16	Getting teased by other teens
17	Not able to do things that other teens his or her age can do
18	Keeping up with other teens
School functioning (problems with...)	
19	Paying attention in class
20	Forgetting things
21	Keeping up with schoolwork
22	Missing school because of not feeling well
23	Missing school to go to the doctor or hospital

language skills by observing social and communication behaviours. During a semi-structured observation, the ADOS interviewer offers playful activities (e.g. reading a story book) and topics of discussion (e.g. peer problems) to assess the socio-communicative abilities of the participant. Each of the participant's behaviours is rated on a scale ranging from normal behaviour (0) to clearly deviant and autistic behaviour (2). An ADOS score of 7 or higher is indicative of the presence of an ASD [34, 35, 39].

*The Social Responsiveness Scale* (SRS; [15]) measures the severity of autism spectrum symptoms as they occur in natural social settings, with a 65-item questionnaire completed by parent or teacher. Several studies have found evidence for good internal consistency, test–retest reliability, interrater reliability, construct validity, and convergent validity (with both the ADOS and ADI-R) of the SRS [11, 59].

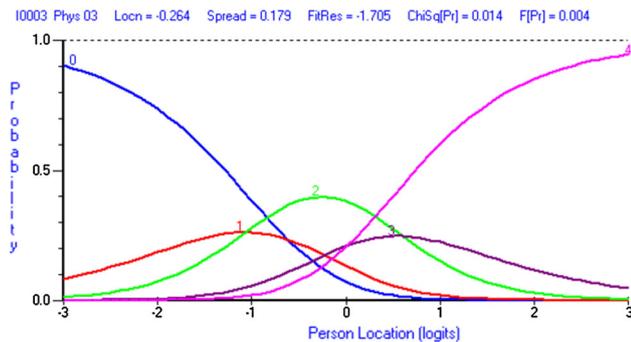
*The Peabody Picture Vocabulary Test* (PPVT; [19]) is designed as a test of receptive vocabulary achievement and verbal ability. The test consists of a series of pictures and is

suitable for a wide age range (2–90 years). The participant has to match an orally given word to a picture. The total score is converted to a verbal IQ. The reliability of the PPVT tested with split–split half and test–retest administration is excellent and the construct and content validity good. The validity of the PPVT is evidenced by strong correlations between PPVT scores and overall intelligence [6, 7].

### Data analyses

Data were analysed using RUMM2030 (2012; RUMM Laboratory Pty Ltd, Perth, Western Australia) and Winsteps (version 3.92.1, © John Linacre, 2016, [www.winsteps.com](http://www.winsteps.com)). In order to evaluate the hypotheses, five main analyses were undertaken. The first consisted of parents of TD children and the children themselves, contrasted for DIF by respondent (parent vs child), and with a second further analysis treating paired respondents as a repeated factor in a separate facet analysis. The third consisted of data for parents of children with ASD and the children themselves, contrasted for DIF by respondent (parent vs child), and the fourth by treating paired respondents as a repeated factor in a separate facet analysis. The fifth consisted of data for children (ASD and TD), contrasted for differential item functioning (DIF) by diagnosis. Given the number of analyses to be undertaken, Bonferroni adjustments were made to error rates where appropriate. There were five main analytic approaches undertaken each with two principle unrelated analyses (fit and information). Bonferroni, Sidak, Holm-Bonferroni, and False Discovery Rate (step 1) techniques each suggest setting  $\alpha = 0.01$  for this number of tests.

In all models, cases missing data were removed listwise. Thereafter, an unrestricted polytomous or partial credit (PC) model was assessed across four parameters (location, scale, kurtosis, skew; giving 92 parameters to fit for 23 items). Models were all fitted to a convergence criterion of 0.001 or smaller, unless otherwise stated. A PC model allows each item to have different thresholds (the balance point between choosing, for example, 0 or 1). In all analyses, the five levels or ratings were found to be disordered for most items (*cf.* Fig. 1). Disordered thresholds indicate that a category is never the most likely to be chosen or endorsed regardless of level of trait. Consequently, in each model, all items were rescaled to three categories (0 → 0, 1 → 1, 2 → 1, 3 → 1, and 4 → 2), collapsing the three intermediate categories into a single category. Thereafter, all cases with extreme data were removed from each model. This strategy retained the broad structure and all items of the PedsQL, while maximising the fit to the uni-dimensional measurement model. Following this, model fits were evaluated and a likelihood ratio test contrasting



**Fig. 1** Category probability curves indicating disordered thresholds for children with ASD for item 3 “it is hard for me to do sports activity or exercise”. Nowhere on curve 3 is a person more likely to choose response 3 than responses 2 or 4

the PC model to the rating model (where all thresholds are equal and are independent of the item).

## Results

### Parents compared to TD children

Prior to analysing the differences between children and parents of children with an ASD, it was first necessary to establish if there were differences between TD children and their parents on the PedsQL. Of the original 68 parents and their TD children, 2 children and 17 parents were removed due to missing data, giving a sample of 66 TD children and 51 parents. The likelihood ratio test was significant ( $\chi^2_{(21)} = 63.56, p < 0.001$ ), suggesting the unrestricted PC model contained more information than the rating model. Consequently, a PC model was used to analyse the data. The data were found to fit the Rasch model [ $\chi^2_{(46)} = 64.86, p = 0.03$ ; Infit 1.05 (0.58–1.78); Outfit 1.06 (0.62–1.39)].

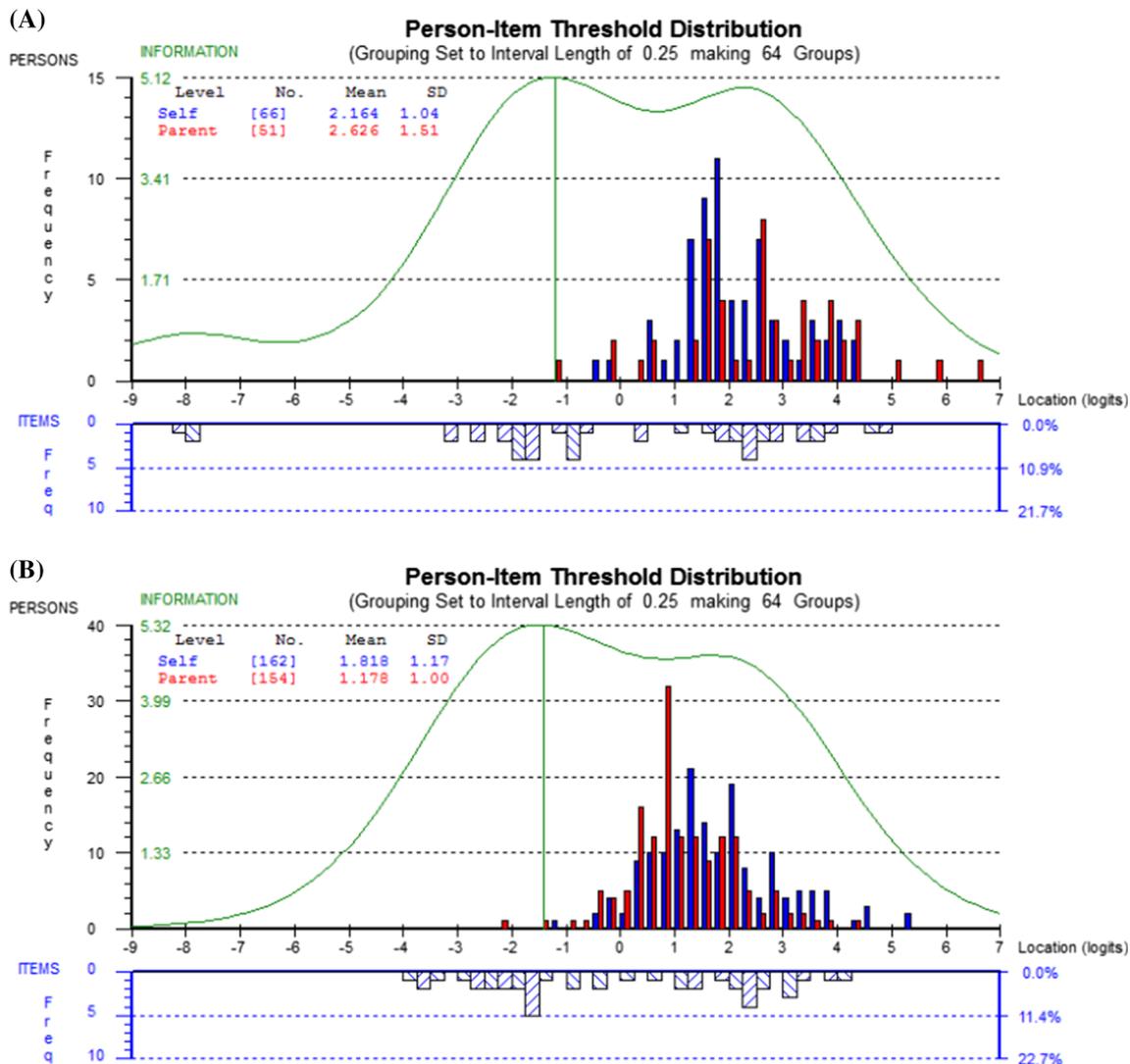
Reliability indices were strong (Cronbach’s  $\alpha = 0.858$ ; PSI = 0.851), while the mean fit residual for items (Residual Fit = 0.193; SD = 0.859) and persons (Residual Fit = -0.292; SD = 0.991) was close to expected values, with no items having excessive fit residuals. The mean person location on the latent continuum was 2.366 (SD = 1.280). The plot of person–item thresholds was examined (see Fig. 2a), with scores for parents and TD children not differing ( $F_{(1,121)} = 3.84, p > 0.05, \eta^2 = 0.03$ ). A facet analysis was undertaken on the rescaled data treating rater (child or parent) as the repeated factor ( $n = 47$  pairs), revealing both raters were unidimensional ( $\chi^2_{(92)} = 133.05, p = \text{ns}$ ). Parents of TD children consistently overestimated their TD child’s own assessment of QoL (see Fig. 3). At the individual level (comparing parent to child for the 23 items), 13 parents significantly differed from their child ( $p < 0.01$ ). In seven

of these, the parent underestimated the child’s assessment and in six cases the parent overestimated the child’s score.

DIF analysis was undertaken comparing parents and their TD children. Using Bonferroni corrected tests, no items showed significantly different class intervals. Parents and children differed on three items (item 4—“hard to lift something heavy”:  $F_{(1,110)} = 36.97, p < .0007, \eta^2 = 0.25$ ; item 7—“hurt or ache”:  $F_{(1,110)} = 15.95, p < .0007, \eta^2 = 0.15$ , and item 19—“hard to concentrate”:  $F_{(1,110)} = 19.06, p < .0007, \eta^2 = 0.13$ ). There were no significant interactions of class interval by respondent. The three items with DIF were removed, resulting in a non-significant change in fit to the Rasch model; thus, the items were retained. Models were attempted with splitting out the items displaying DIF [57]; none resulted in model improvement. In summary, while parents of and TD children differed in their use of the PedsQL on some items, there were no concerning interactions between class interval and rater, meaning that while parents and children differed this did not depend upon the level of QoL in the child—both groups use the PedsQL in similar and consistent ways.

### Parents compared to children with ASD

Differences between parents of children with ASD and the children themselves were then explored. Of the 160 parents of and 202 children with ASD, 40 children with ASD and 6 of their parents were removed due to missing data. This left a sample of 162 children with ASD and 154 of their parents. The likelihood ratio test was significant ( $\chi^2_{(21)} = 205.61, p < 0.001$ ), suggesting the unrestricted PC model contained more information than the rating model. Consequently, a PC model was used to analyse the data. The data were restricted as detailed above, with all items rescored to have three categories; this did not result in a fitting model [ $\chi^2_{(115)} = 190.22, p < 0.001$ ; Infit 1.05 (0.59–1.79); Outfit 1.06 (0.61–1.55)]. One item, item 21 (“I have trouble keeping up with my schoolwork”), had an excessive fit residual (2.92 > 2.5), removal of which improved the fit, but only marginally ( $p < 0.01$ ), and so this item was retained for completeness. The failure of the test of fit alone is insufficient to conclude a lack of unidimensionality due to oversensitivity to sample size, which dictates the number of class intervals [46]. Further, given infit and outfit values were within acceptable ranges and that examination of fits revealed person and item fits were improved by retention of item 21; although classical reliability was slightly worsened, it was decided to retain the model with item 21 included. With item 21 retained, the reliability indices indicated the scale to be reliable in classical terms (Cronbach’s  $\alpha = 0.778$ ) and in Rasch model terms (PSI = 0.810). The mean fit residual for items (Residual Fit = -0.160; SD = 1.232) and persons



**Fig. 2** Person–item threshold distribution for all data for **a** parent of children with TD and the children themselves separated by respondent type; **b** parent of children with ASD and the children themselves

separated by respondent type (*red* = parents response, *blue* = children’s response). (Color figure online)

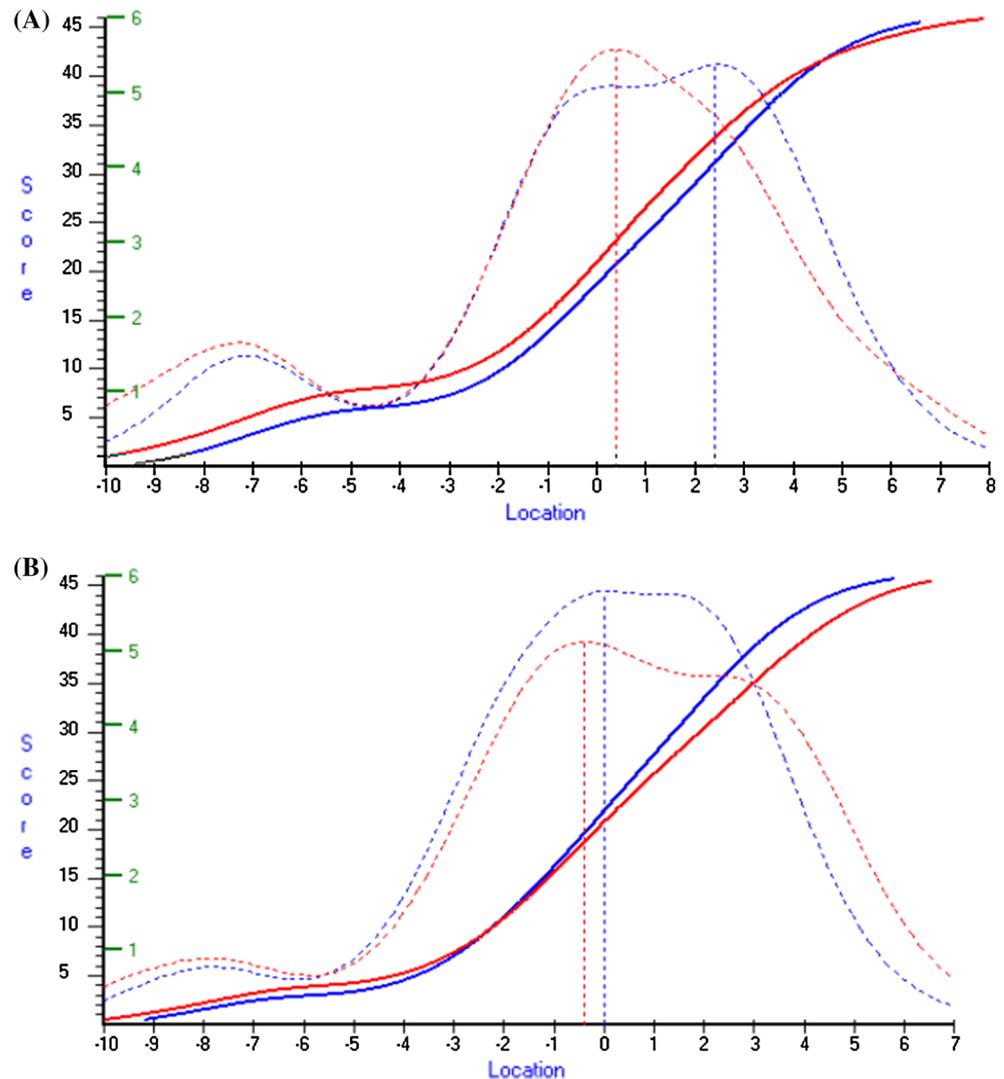
(Residual Fit =  $-.247$ ; SD = 1.113) was close to expected values. The mean person location on the latent continuum was 1.90 (SD = 0.990).

Unidimensionality was further evaluated by assessing the Principal Components (PrC) of the residuals. The first eigenvector accounted for 10.9%, while the second and third accounted for 9.0 and 8.4%, respectively. The location scores of all items with unrotated PrC factor loadings greater than  $\pm 0.3$  were contrasted to all items, and no significant difference was found ( $t_{(315)} = .32$ ,  $p = \text{ns}$ ), while the two scores were highly related ( $r = 0.87$ ; [43]). Thereafter, the location scores of items loading highly onto the first eigenvector were compared to those that did not load onto the first eigenvector and their location scores. No significant difference was found ( $t_{(315)} = .17$ ,  $p = \text{ns}$ ), and the scores were strongly correlated ( $r = 0.67$ ). Last items

loading strongly onto the first eigenvector were compared to those loading strongly onto the second eigenvector, with no significant difference found ( $t_{(315)} = .20$ ,  $p = \text{ns}$ ,  $r = 0.60$ ). Consequently, it would appear that while the scale is not entirely unidimensional, it functions as a unidimensional scale among TD respondents. Examination of the residual correlation matrix revealed some local dependence ( $r > \pm 0.3$ ). Of the 231 pairs, 7 pairs (3%) of items had correlations greater than the criteria ( $\pm .37 > r > \pm .30$ ; items 2 and 3; 14 and 15; 14 and 18; 15 and 16; 19 and 20; 19 and 21; 22 and 23; see Table 2).

The person–item threshold plot was examined (see Fig. 2b). A significant difference was found between parents of children with ASD and the children themselves ( $F_{(1,314)} = 27.09$ ,  $p < .001$ ,  $\eta^2 = 0.08$ ). In this broad analytic overview, parents rated the children lower than

**Fig. 3** Distribution of scores for parents and children.  
**a** Distribution of scores by location for TD children (*blue line*) and parents (*red line*).  
**b** Distribution of scores by location for ASD children (*blue line*) and parents (*red line*).  
 (Color figure online)



children with ASD rated themselves (parents  $M = 1.18$ ,  $SD = 1.00$ ; children with ASD  $M = 1.82$ ,  $SD = 1.17$ ). A facet analysis was undertaken treating rater (child or parent) as the repeated factor, with 134 paired cases. The likelihood ratio test was significant ( $\chi^2_{(44)} = 190.12$ ,  $p < 0.0027$ ), suggesting the PC model was more informative than the rating model. The model was found to fit the data, revealing the two levels of rater (self and parent) were unidimensional ( $\chi^2_{(92)} = 71.57$ ,  $p = ns$ ), meaning both groups utilised the PedsQL on the same dimension. However, parents of children with ASD tended to be less extreme than their child in estimation of their child's QoL. For children with low QoL, parents tended to slightly overestimate their child's own assessment of QoL and underestimate for children with above average QoL (see Fig. 3). At the individual level (comparing parent to child for the 23 items) in 27 cases, parents significantly differed from their child ( $p < 0.01$ ). In 14 of these, the parent

underestimated the child's assessment, and in 13 cases, the parent overestimated the child's score.

Thereafter, DIF analysis was undertaken. Using Bonferroni corrected tests, no item had a significant interaction of class interval by respondent, two items showed significantly different class intervals (item 14—"trouble getting along with peers":  $F_{(1,304)} = 4.84$ ,  $p < .0007$ ,  $\eta^2 = 0.02$ , and item 21—"trouble keeping up with schoolwork":  $F_{(1,304)} = 6.29$ ,  $p < .0007$ ,  $\eta^2 = 0.02$ ). Parents and children differed on three items (item 4—"hard to lift something heavy":  $F_{(1,297)} = 28.91$ ,  $p < .0007$ ,  $\eta^2 = 0.09$ ; item 11—"feel angry":  $F_{(1,297)} = 38.77$ ,  $p < .0007$ ,  $\eta^2 = 0.12$ , and item 23—"miss school—doctor appointment":  $F_{(1,297)} = 16.58$ ,  $p < .0007$ ,  $\eta^2 = 0.05$ ). Removal of the three items displaying DIF by respondent did not improve model fit and generated other issues. Further, removal did not remove the overall group differences noted above, but instead resulted in an enlarged group difference

( $F_{(1,307)} = 60.31$ ,  $p < .001$ ,  $\eta^2 = 0.16$ ). Therefore, as removal of these three items would not improve fit, nor reduce overall group differences, these three items were retained. In summary, parents of and children with ASD differed in their use of some items of the PedsQL. Moreover, while parent's assessment of their child's QoL appeared to depend upon the child's level of QoL, overall parents of children with ASD significantly underestimated their child's QoL.

### ASD children compared to TD children

Of the final sample of 270 participant children, 40 participants with ASD and 2 TD participants were removed due to missing data. This left 162 children with ASD and 66 TD children. The likelihood ratio test comparing to the rating model was significant ( $\chi^2_{(21)} = 114.02$ ,  $p < 0.001$ , convergence 0.01), suggesting the unrestricted partial credit (PC) model contained more information than the rating model. The data were found to fit the unidimensional PC model [ $\chi^2_{(69)} = 92.22$ ,  $p = 0.01$ ; Infit 1.06 (0.49–1.98); Outfit 1.03 (0.52–1.58)]. While infit was high, scores such as this are considered unproductive, but acceptable. Consequently, no items were observed with excessive fit residuals. Consequently, the PC model was retained. Indices suggested the scale to be reliable in classical terms (Cronbach's  $\alpha = 0.842$ ) and in Rasch model terms [Person Separation Index (PSI) = 0.806].

Given the two groups of children had different ages, it was decided to examine whether PedsQL scores differed by age group (5–7, 8–12, 12–18 based upon the PedsQL self-report forms) and then to establish if an age-matched data set differed from the main analysis. PedsQL scores were not found to differ based upon age group ( $F_{(2,264)} = 2.73$ ,  $p = .07$ ,  $\eta^2 = 0.02$ ). Child participants were then matched on age ( $n = 55$  per group), which revealed the data fit the PC model [ $\chi^2_{(46)} = 61.06$ ,  $p = 0.07$ ; Infit 1.06 (0.43–1.72); Outfit 1.04 (0.46–1.74)]. Indices also suggested the scale to be reliable in Rasch model terms (PSI = 0.75), although the Cronbach's alpha suggested a lower than desirable fit ( $\alpha = 0.70$ ). No issues related to DIF were found in the data. Last, groups were not found to differ ( $F_{(1,108)} = 1.73$ ,  $p = .19$ ,  $\eta^2 = 0.02$ ). As the differences in age show no effect and that the model based solely age-matched data showed no differences, we proceeded with the main analysis relying upon all the data for child participants.

The mean fit residual for items was  $-0.162$  (SD = 1.035), and person fit residuals were  $-0.293$  (SD = 1.191) which are close to the expected values of 0 and 1. Mean person locations on the latent continuum were 1.731 (SD = 1.026); mean item locations on the latent continuum were 0.00 (SD = 0.561). The plot of person-

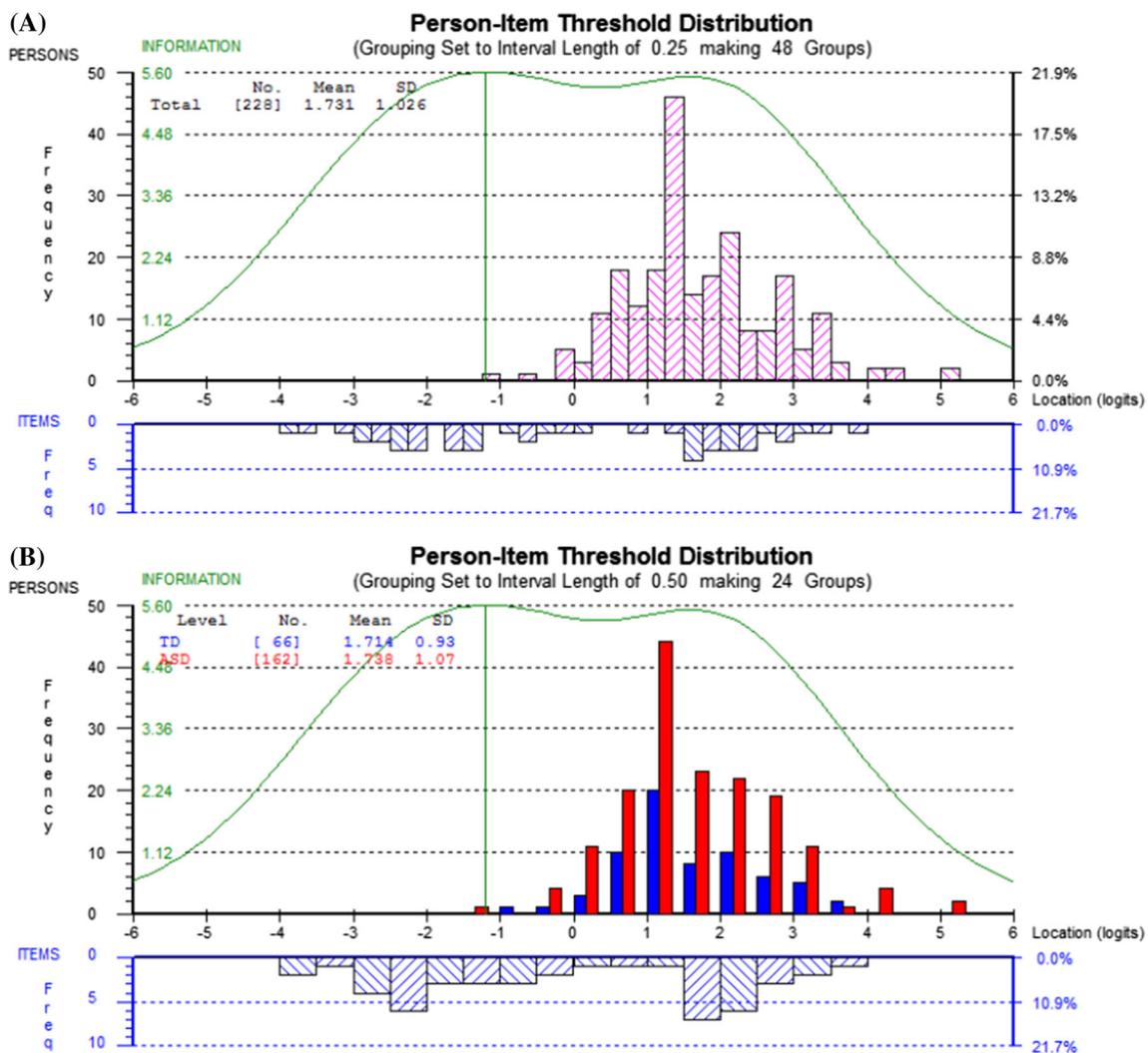
item thresholds was examined (see Fig. 4a), with no issues noted. Group differences by diagnosis were also examined (see Fig. 4b) and were not found to differ ( $F_{(1,226)} = 0.03$ ,  $p = 0.87$ ,  $\eta^2 < 0.01$ ). Subsequently, assessment of model structure and DIF was undertaken.

With 23 items in the data set evaluated for class interval (the band of ability or trait to which an individual belongs), diagnostic group, and the interaction of these, giving 69 tests, with  $\alpha = 0.05$ ,  $p$  was set to 0.0007. Examination of the results revealed no item differed by class interval, diagnostic group, or the interaction of these. In summary, it was found that children did not differ from each other by diagnostic group in how they assessed themselves on the PedsQL.

### Discussion

Children with ASD and TD children use the PedsQL in similar ways, and the overall PedsQL conformed to a unidimensional model. The PedsQL when scored as specified was not found to fit a unidimensional Rasch model and was replete with disordered thresholds on most items. However, when this was addressed by collapsing scoring (i.e.  $0 \rightarrow 0$ ,  $1 \rightarrow 1$ ,  $2 \rightarrow 1$ ,  $3 \rightarrow 1$ , and  $4 \rightarrow 2$ ), it was found that parents and TD children utilise the rescored PedsQL in similar ways, and a unidimensional model easily fitted their data. Further, while parents of and children with ASD also used the rescored PedsQL in similar ways, with the rescored scale showing conformity to a unidimensional model, parents of children with ASD estimated the level of QoL in their children in a manner dependent upon the child's level of QoL. Moreover, the rescored PedsQL displayed good levels of reliability in each of the three total scale assessments.

Others have previously reported that the PedsQL did not fit a unidimensional Rasch model [26, 30]. Disordered thresholds are where a particular response is never the most likely choice of a person at any given level of ability. Therefore, a modified model was analysed with three categories. Categories 1, 2, and 3 were collapsed into a single category, intermediate between the lowest and highest categories. Under this approach, we found all items in all groups functioned as expected. This model was found to fit the Rasch model, lacks any disordered thresholds, and was unidimensional when used by TD children, as well as their parents, and when used by children with an ASD. However, a unidimensional model for children with ASD and their parents was not obtained from model fit alone. It was necessary to infer from other indices of fit. The mean fit residual and person fit residuals were within acceptable limits. The item, "I have trouble keeping up with my schoolwork", item 21, had an excessive fit residual, but its



**Fig. 4** Person–item threshold distribution for **a** all data for child respondents and **b** separated by diagnostic group (*red* = HFASD, *blue* = TD). (Color figure online)

removal did not render the data fitting to the model, and fit indices were stronger with item 21 retained. Last, the lack of difference between parents of and children with ASD when analysed in the facet model revealed these groups treated the model unidimensionally. Analysing the data in this way tied each parent to their specific child. The success of the rescored model has implications for the scoring of individuals and suggests a modified scoring approach should be used in future where there are only three categories. Nonetheless, this result should be replicated before such a strategy is adopted, as this study was limited to a specific sample of children with ASD and a small sample of TD children, and their parents, and results differed by group.

The level of observed reliability in the rescored model, measured as Cronbach's  $\alpha$ , was good for the overall scale when measured among all participants. This was supported by person separation indices above 0.8 in all instances of

the rescored model. The PSI indicates the strength which the instrument is able to distinguish between individuals of different ability and ranges from 0 (low) to 1 (high; [12]). These results suggest that the PedsQL once rescored to three categories is reliable and can adequately discriminate between individuals of different ability.

We explored whether items were responded to similarly by parents and their children and by both groups of children. Others have found DIF in the PedsQL [18, 24, 27, 30, 33]. The present study did not find DIF in the way in which TD children and children with and ASD used the PedsQL. Nonetheless, parents of TD children and the children themselves differed on three items (4, 7, and 19) with effect sizes ranging from 12 to 25%. Parents of and children with ASD also differed on three items (4, 11, and 23) with effect sizes ranging from 5 to 12%. As the overall scales between TD children and their parents did not differ, it is reasonable to conclude these significant DIF

scores cancelled each other, but nonetheless, care should be taken when using these items with TD samples. However, as the parents of and children with ASD did differ at a group level in their use of the PedsQL, but not when each parent was tested against their child, this suggests the DIF may have contributed to the apparent group differences. However, removal of the three items displaying some respondent DIF did not remove the overall group difference, instead this exacerbated it. It should be noted that item 4 (“hard to lift something heavy”) differed between parents and their children for both the TD and ASD group. As children did not show DIF based upon diagnosis, it is likely that these differences in response to items may be related to the way parents interpret and respond.

The question of whether children with ASD have adequate insight to assess their own QoL can be addressed by facet analysis. It was found that TD children and their responding parents did not differ, with only about 3% of variance between them. A facet analysis treating respondent as the facet found no difference between respondents. Though not significant, parents of TD children consistently overestimated their child’s own estimate of QoL. By contrast, as a general trend, parents of children with ASD significantly underestimated their child’s QoL by 0.64 logits. However, the facet analysis revealed a more complex picture. This analysis paired each parent with their child. It was found that parents of and children with an ASD did not differ significantly in their scores. Children with ASD who had low self-reported QoL were rated as closer to average QoL by their parents, while children with ASD who had high QoL were rated by their parents also as closer to the average QoL for all children with ASD. In other words, parent estimates tended to be less extreme than their children’s responses. Thus, when rating their child’s QoL, parents differ from their child in how they perceive their child’s QoL. It is possible that children lack insight into themselves and that parents have better insight into their child. But this position is hard to parsimoniously reconcile with the data.

From the perspective of parsimony, parents must rely upon their observations of their child, who plausibly may conceal positive and negative aspects of their life. The facet analysis would suggest parents are not fully aware of their child’s QoL. This lack of shared knowledge has been noted among clinicians when dealing with children with ASD [3, 22]. Further, though in general, parents and children with ASD differ in how they responded to the PedsQL, when the target child was controlled for through the facet design, at the group level parents did not differ significantly from their matched child with ASD. This suggests, that for the most part, parents and children agree on their PedsQL rating. That there were a small number of

parents of children with ASD significantly over- or underestimating their children suggests that it is only in some instances that disagreement between parent and child occurs.

Last, TD children and children with ASD were compared. These two groups were not found to differ in their PedsQL scores or to differ in the way they used any particular item. Hence, it is reasonable to conclude that for the most part, children with ASD do have sufficient insight to respond to an instrument measuring the QoL. So are parents able to suitably act as proxies? From the facet analysis provided, it would appear that some parents of children with ASD tend to describe their child’s QoL closer to the mean of all children than the child themselves does. This may not be surprising, given that children with ASD may not readily be forthcoming with the details of their daily lives [3]. In addition, we need to consider that parents may compare their children with ASD to other TD children when responding to the QoL questions, while such social comparison may be more difficult for children with ASD. Still, for the most part, parents and children agreed, and so, parents are reasonable proxies for children, but given that children are able to answer questions pertaining to QoL themselves, it seems reasonable to rely upon the child where possible (also see [58], for the importance of the child’s view).

These results contradict findings by some and support findings by others. Interestingly, in measures of social insight and of insight into cognitive processes, various authors report children with ASD do not display insight [13, 40]. In the case of Mehzabin and Stokes, the conclusion was based upon comparison between parents’ responses and of self-responses to questions about sexual behaviour. Broadbent and Stokes found a lack of insight into cognitive processes and did evaluate this explicitly. However, the contradiction between their conclusion and the data herein reveals that there may be different levels of insight into cognitive and emotional processes within individuals with ASD. The current data support findings by Schriber et al. [47], Sheldrick et al. [49], and Berthoz and Hill [8]. These authors each report establishing suitable level of insight for the evaluation of emotional processes among individuals with ASD.

The present study is not without limitations. While the overall sample size was sound, the sample of TD children and their parents was considerably smaller than the sample with ASD. In addition, though all children were diagnosed with an ASD by a suitably qualified and trained independent professional using an extensive clinical assessment, not all participants obtained the ADOS threshold for ASD, though the SRS confirmed the diagnosis for all ASD participants.

## Conclusions

From the data presented herein, it is apparent that the PedsQL is a reliable measure that TD children and their parents can use, as can children with ASD, although their parents may not be as reliable reporters as the children themselves. The present results call into question the standard scoring structure for the PedsQL and suggest it may be wisest to restructure this into 3 categories rather than the typical 5. The most important finding here though is that children with ASD are able to adequately report on their own QoL.

**Acknowledgements** This study received 139,000 EURO funding from Stichting Fonds NutsOhra (Grant Number SNO-T-0701-116).

### Compliance with ethical standards

**Conflict of interest** Mark Stokes, Lara Kornienko, Anke Scheeren, Hans M. Koot and Sander Begeer declare that they have no conflict of interest.

**Ethical approval** All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

## References

- Ainuddin, H. A., Loh, S. Y., Chinna, K., Low, W. Y., & Roslani, A. C. (2013). Psychometric properties of the self-report Malay version of the Pediatric Quality of Life (PedsQLTM) 4.0 Generic Core Scales among multiethnic Malaysian adolescents. *Journal of Child Health Care, 19*, 229–238. doi:10.1177/1367493513504834.
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders—Text revision* (4th ed.). Washington, DC: Author.
- Attwood, T. (2007). *The complete guide to Asperger's syndrome*. London: Jessica Kingsley Publishers.
- Bastiaansen, D., Koot, H. M., Bongers, I. L., Varni, J. W., & Verhulst, F. C. (2004). Measuring quality of life in children referred for psychiatric problems: Psychometric properties of the PedsQL. *Quality of Life Research, 13*, 489–495.
- Bastiaansen, D., Koot, H. M., Ferdinand, R. F., & Verhulst, F. C. (2004). Quality of life in children with psychiatric disorders: Self-, parent, and clinician report. *Journal of the American Academy of Child and Adolescent Psychiatry, 43*(2), 221–230.
- Bee, H., & Boyd, D. (2004). *The developing child* (10th ed.). Boston, MA: Allyn and Bacon.
- Bell, N. L., Lassiter, K. S., Matthews, T. D., & Hutchinson, M. B. (2001). Comparison of the peabody picture vocabulary test-third edition and Wechsler adult intelligence scale-third edition with university students. *Journal of Clinical Psychology, 57*(3), 417–422.
- Berthoz, S., & Hill, E. L. (2005). The validity of using self-reports to assess emotion regulation abilities in adults with autism spectrum disorder. *European Psychiatry, 20*, 291–298.
- Biao, J. (2014). Prevalence of autism spectrum disorder among children aged 8 years—Autism and developmental disabilities monitoring network, 11 sites, United States, 2010. *Morbidity and Mortality Weekly Report. Surveillance Summaries, 63*(2), 1–21.
- Blore, J. D., Stokes, M. A., Mellor, D., Firth, L., & Cummins, R. (2010). Comparing multiple discrepancies theory and affective models of subjective wellbeing. *Social Indicators Research, 100*(1), 1–16. doi:10.1007/s11205-010-9599-2.
- Bölte, S., Poustka, F., & Constantino, J. N. (2008). Assessing autistic traits: Cross-cultural validation of the social responsiveness scale (SRS). *Autism Research, 1*(6), 354–363. doi:10.1002/aur.49.
- Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). New York: Routledge.
- Broadbent, J., & Stokes, M. (2013). Removal of negative feedback enhances WCST performance for individuals with Asperger's Syndrome. *Research in Autism Spectrum Disorders, 7*(6), 785–792.
- Campbell, A., Converse, P. E., & Rodgers, W. L. (1976). *The quality of American life: Perceptions, evaluations and satisfactions*. New York: Russell Sage Foundation.
- Constantino, J. N. & Gruber, C. P. (2005) *Social responsiveness scale (SRS)*. Western Psychological Services.
- Diener, E., & Diener, M. (1996). Most people are happy. *Psychological Science, 7*(3), 181–185.
- de Vries, M., & Geurts, H. (2015). Influence of autism traits and executive functioning on quality of life in children with an autism spectrum disorder. *Journal of Autism and Developmental Disorders, 45*(9), 2734–2743. doi:10.1007/s10803-015-2438-1.
- Doostfateme, M., Ayatollahi, S. M., & Jafari, P. (2015). Testing parent dyad interchangeability in the parent proxy-report of PedsQL<sup>TM</sup> 4.0: A differential item functioning analysis. *Quality of Life Research, 24*(8), 1939–1947. doi:10.1007/s11136-015-0931-9.
- Dunn, L. M., & Dunn, L. M. (2004). *Peabody picture vocabulary test (PPVT)-III-NL*. Amsterdam: Harcourt Test.
- Eiser, C., & Morse, R. (2001). Can parents rate their child's health-related quality of life? Results of a systematic review. *Quality of Life Research, 10*(4), 347–357. doi:10.1023/A:1012253723272.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Hobson, P. (2005). Autism and emotion. In F. R. Volkmar, R. Paul, A. Klin, & D. Cohen (Eds.), *Handbook of autism and pervasive developmental disorders* (3rd ed.). Hoboken, NJ: Wiley.
- Hoffman, S., Lambert, M. C., Nelson, T. D., Trout, A. L., Epstein, M. H., & Pick, R. (2013). Confirmatory factor analysis of the PedsQL among youth in a residential treatment setting. *Quality of Life Research, 22*, 2151–2157. doi:10.1007/s11136-012-0327-z.
- Huang, I. C., Leite, W. L., Shearer, P., Seid, M., Revicki, D. A., & Shenkman, E. A. (2011). Differential item functioning in quality of life measure between children with and without special health-care needs. *Value in Health, 14*, 872–883.
- Ikeda, E., Hinckson, E., & Krägeloh, C. (2014). Assessment of quality of life in children and youth with autism spectrum disorder: A critical review. *Quality of Life Research, 23*, 1069–1085. doi:10.1007/s11136-013-0591-6.
- Jafari, P., Bagheri, Z., Ayatollahi, S. M., & Soltani, Z. (2012). Using Rasch rating scale model to reassess the psychometric properties of the Persian version of the PedsQL<sup>TM</sup> 4.0 Generic Core Scales in school children. *Health and Quality of Life Outcomes, 10*(27), 1–11.
- Jafari, P., Bagheri, Z., Hashemi, S. Z., & Shalileh, K. (2013). Assessing whether parents and children perceive the meaning of the items in the PedsQL<sup>TM</sup> 4.0 quality of life instrument

- consistently: A differential item functioning analysis. *Global Journal of Health Science*, 5, 80–88.
28. Kaneko, M., Sato, I., Soejima, T., & Kamibeppu, K. (2014). Health-related quality of life in young adults in education, employment, or training: Development of the Japanese version of Pediatric Quality of Life Inventory (PedsQL) Generic Core Scales Young Adult Version. *Quality of Life Research*, 23(7), 2121–2131. doi:10.1007/s11136-014-0644-5.
  29. Kamp-Becker, I., Schröder, J., Muehlan, H., Remschmidt, H., Becker, K., & Bachmann, C. J. (2011). Health-related quality of life in children and adolescents with autism spectrum disorder. *Zeitschrift für Kinder-und Jugendpsychiatrie und Psychotherapie*, 39(2), 123–131. doi:10.1024/1422-4917/a000098.
  30. Kook, S. H., & Varni, J. W. (2008). Validation of the Korean version of the Pediatric Quality of Life Inventory™ 4.0 (PedsQL™) Generic Core Scales in school children and adolescents using the Rasch model. *Health and Quality of Life Outcomes*, 6(41), 1–15.
  31. Kuhlthau, K., Orlich, F., Hall, T., Sikora, D., Kovacs, E., Delahaye, J., et al. (2010). Health-related quality of life in children with autism spectrum disorders: Results from the Autism Treatment Network. *Journal of Autism and Developmental Disorders*, 40(6), 721–729. doi:10.1007/s10803-009-0921-2.
  32. Lai, W. W., Goh, T. J., Oei, T. P., & Sung, M. (2015). Coping and well-being in parents of children with autism spectrum disorders (ASD). *Journal of Autism and Developmental Disorders*, 45(8), 2582–2593. doi:10.1007/s10803-015-2430-9.
  33. Langer, M. M., Hill, C. D., Thissen, D., Burwinkle, T. M., Varni, J. W., & DeWalt, D. A. (2008). Item response theory detected differential item functioning between healthy and ill children in quality-of-life measures. *Journal of Clinical Epidemiology*, 61, 268–276.
  34. Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., et al. (2000). The autism diagnostic observation schedule-generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders*, 30, 205–223.
  35. Lord, C., Rutter, M., DiLavore, P. C., & Risi, S. (2008). *Autism diagnostic observation schedule manual*. Los Angeles, CA: Western Psychological Services.
  36. Limbers, C., Heffer, R., & Varni, J. W. (2009). Health-related quality of life and cognitive functioning from the perspective of parents of school-aged children with Asperger's Syndrome utilizing the PedsQL. *Journal of Autism and Developmental Disorders*, 39(11), 1529–1541. doi:10.1007/s10803-009-0777-5.
  37. Limbers, C. A., Newman, D. A., & Varni, J. W. (2009). Factorial invariance of child self-report across race/ethnicity groups: A multigroup confirmatory factor analysis approach utilizing the PedsQL 4.0 Generic Core Scales. *Annals of Epidemiology*, 19, 575–581. doi:10.1016/j.annepidem.2009.04.004.
  38. McStay, R. L., Dissanayake, C., Scheeren, A., Koot, H. M., & Begeer, S. (2014). Parenting stress and autism: The role of age, autism severity, quality of life and problem behaviour of children and adolescents with autism. *Autism*, 18(5), 502–510. doi:10.1177/1362361313485163.
  39. Molloy, C. A., Murray, D. S., Akers, R., Mitchell, T., & Manning-Courtney, P. (2011). Use of the autism diagnostic observation schedule (ADOS) in a clinical setting. *Autism*, 15(2), 143–162. doi:10.1177/1362361310379241.
  40. Mehzabin, P., & Stokes, M. A. (2010). Self-assessed sexuality in young adults with high functioning autism. *Research in Autism Spectrum Disorders*, 5, 614–621.
  41. Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.
  42. Pakpour, A. H., Zeidi, I. M., Hashemi, F., Saffari, M., & Burri, A. (2012). Health-related quality of life in young adult patients with rheumatoid arthritis in Iran: Reliability and validity of the Persian translation of the PedsQL™ 4.0 Generic Core Scales Young Adult Version. *Clinical Rheumatology*, 32, 15–22. doi:10.1007/s10067-012-2084-3.
  43. Pallant, J. F., & Tennant, A. (2007). An introduction to the Rasch measurement model: An example using the Hospital Anxiety and Depression Scale (HADS). *British Journal of Clinical Psychology*, 46, 1–18.
  44. Petersen, S., Hägglöf, B., Stenlund, H., & Bergström, E. (2009). Psychometric properties of the Swedish PedsQL, Pediatric Quality of Life Inventory 4.0 generic core scales. *Acta Paediatrica*, 98, 1504–1512. doi:10.1111/j.1651-2227.2009.01360.x.
  45. Roeyers, H., Thijs, M., Druart, C., De Schryver, M., & Schittekatte, M. (2011). *SRS, Screeningslijst voor autismespectrumstoornissen*. Amsterdam: Hogrefe.
  46. RUMM Laboratory P/L. (2013). *Interpreting RUMM2030* (5th ed.). Perth: RUMM Laboratory.
  47. Schriber, R. A., Robins, R. W., & Solomon, M. (2014). Personality and self-insight in individuals with autism spectrum disorder. *Journal of Personality and Social Psychology*, 106(1), 112–130. doi:10.1037/a0034950.
  48. Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, 74(1), 107–120. doi:10.1007/S11336-008-9101-0.
  49. Sheldrick, R. C., Neger, E. N., Shipman, D., & Perrin, E. C. (2012). Quality of life of adolescents with autism spectrum disorders: Concordance among adolescents' self-reports, parents' reports, and parents' proxy reports. *Quality of Life Research*, 21, 53–57. doi:10.1007/s11136-011-9916-5.
  50. Shipman, D. L., Sheldrick, R. C., & Perrin, E. C. (2011). Quality of life in adolescents with autism spectrum disorders: Reliability and validity of self-reports. *Journal of Developmental Behavioral Pediatrics*, 32(2), 85–89.
  51. Ten Berge, J. M. F., & Sočan, G. (2004). The greatest lower bound to the reliability of a test and the hypothesis of unidimensionality. *Psychometrika*, 69, 613–625.
  52. Thomas, S., Sciberras, E., Lycett, K., Papadopoulos, N., & Rinehart, N. (2015). Physical functioning, emotional, and behavioral problems in children with ADHD and comorbid ASD: A cross-sectional study. *Journal of Attention Disorders*. doi:10.1177/1087054715587096.
  53. Upton, P., Lawford, J., & Eiser, C. (2008). Parent-child agreement across child health-related quality of life instruments: A review of the literature. *Quality of Life Research*, 17(6), 895–913. doi:10.1007/s11136-008-9350-5.
  54. van Heijst, B. F. C., & Geurts, H. M. (2015). Quality of life in autism across the lifespan: A meta-analysis. *Autism*, 19(2), 158–167.
  55. Varni, J. W., Handen, B. L., Corey-Lisle, P. K., Guo, Z., Manos, G., Ammerman, D. K., et al. (2012). Effect of aripiprazole 2 to 15 mg/d on health-related quality of life in the treatment of irritability associated with autistic disorder in children: A post hoc analysis of two controlled trials. *Clinical Therapeutics*, 34(4), 980–992.
  56. Varni, J. W. (2014). *Scaling and scoring of the Pediatric Quality of Life Inventory™*. Lyon: Mapi research trust.
  57. Walker, C. (2011). What's the DIF? Why differential item functioning analyses are an important part of instrument development and validation. *Journal of Psychoeducational Assessment*, 29, 364–376.
  58. Wallander, J. L., & Koot, H. M. (2016). Quality of life in children: A critical examination of concepts, approaches, issues, and future directions. *Clinical Psychology Review*, 45, 131–143. doi:10.1016/j.cpr.2015.11.007.
  59. Wigham, S., McConachie, H., Tandos, J., Le Couteur, A. S., & Gateshead Millennium Study Core Team. (2012). The reliability

- and validity of the Social Responsiveness Scale in a UK general child population. *Research in Developmental Disabilities*, 33(3), 944–950. doi:[10.1016/j.ridd.2011.12.017](https://doi.org/10.1016/j.ridd.2011.12.017).
60. Yeung, N. C., Lau, J. T., Yu, X. N., Chu, Y., Shing, M. M., Leung, T. F., et al. (2013). Psychometric properties of the Chinese version of the Pediatric Quality Of Life Inventory 4.0 Generic Core scales among pediatric cancer patients. *Cancer Nursing*, 36, 463–473. doi:[10.1097/NCC.0b013e31827028c8](https://doi.org/10.1097/NCC.0b013e31827028c8).