

Satisfaction Ratings of QOLPAV: Psychometric Properties Based on the Graded Response Model

Ssu-Kuang Chen · Fang-Ming Hwang · Sunny S. J. Lin

Accepted: 11 September 2011 / Published online: 9 October 2011
© Springer Science+Business Media B.V. 2011

Abstract A scale measuring quality of life (QOL) is important in adolescent research. Using the graded response model (GRM), this study evaluates the psychometric properties of the satisfaction ratings of the Quality of Life Profile Adolescent Version (QOLPAV). Data for 1,392 adolescents were used to check IRT assumptions such as unidimensionality and local item dependence (LID). The goodness of fit of the GRM to the data and the item characteristic curves were evaluated. The reliability and validity analyses included item/test information, Cronbach's α , and convergent and discriminant validity. Differential item functioning (DIF) procedures were also performed to detect item bias. The results provide evidence that the items sufficiently measured one single dimension. Few pairs of questions were flagged as LID due to content or wording similarity. Five items did not fit the GRM, and 4 were low in item discrimination. The findings also suggest that the assessment had appropriate reliability and validity. The DIF impact on the assessment score was considered minor. Because QOLPAV includes a respondent's perceived importance of various life aspects, a short form that only considers important life aspects in the overall QOL estimation for each respondent becomes feasible within the framework of IRT. Future studies focusing on the development of a QOL overall index using the items from QOLPAV are recommended.

Keywords Quality of life · The graded response model · Psychometric properties · Satisfaction ratings · Importance weights

S.-K. Chen · S. S. J. Lin
Institute of Education, National Chiao Tung University, 1001 Ta-Hsueh Rd,
Hsinchu 300, Taiwan, ROC
e-mail: Schen75025@gmail.com

S. S. J. Lin
e-mail: sunnylin.nctu@gmail.com

F.-M. Hwang (✉)
Department of Education, National Chiayi University, No. 85, Wenlong Village,
Minxiong Township, Chiayi County 621, Taiwan, ROC
e-mail: fmh@mail.ncyu.edu.tw

1 Introduction

Adolescence is a critical phase of a person's life and is associated with changes in one's body, relationships, and self-image. Quality of life (QOL) concerning an adolescent's overall well-being has received increasing attention in adolescent research (Goldbeck et al. 2007; Shek 2005; Snyder et al. 2010; Valois et al. 2004). One conceptual model of QOL that is adopted widely, particularly in health care research, is based on the World Health Organization's definition of health as incorporating aspects of physical, psychological, and social function (WHOQOL Group 1994). Within this framework, various life aspects are assessed to fit specific research needs.

In the past few decades, the majority of QOL instruments have either been disease-specific or have focused on the functional performance of an individual's daily life. However, generic instruments that examine individuals' perceptions to various aspects of their lives are also important and have drawn attention from researchers in various areas. Wallander et al. (2001) suggested that disease-specific or health-related QOL instruments are preferable when "disease impact" is of interest. Researchers should keep in mind that "disease impact" is not equivalent to QOL. According to Phillips (2006), a generic instrument usually considers all relevant domains that affect one's life and is thus more comprehensive than a disease-specific instrument. Generic instruments are particularly useful when comparisons are conducted across different groups of adolescents (Lee et al. 2009; Wallander et al. 2001).

Measuring QOL in adolescents presents special challenges, including how to define the most relevant domains in adolescents' lives and how to conceptualize adolescents' perceptions of those domains (Patrick et al. 2002). According to Wallander et al.'s (2001) recommendation, the Quality of Life Profile Adolescent Version (QOLPAV; Raphael et al. 1996) is one of 5 generic adolescent QOL scales that are worth considering for future work. The QOLPAV focuses on the holistic nature of adolescents' lives (Zekovic and Renwick 2003). Using the Centre of Health Promotion's (CHP) model of QOL (Renwick and Brown 1996), the QOLPAV emphasizes 3 domains of life: being, belonging, and becoming. QOLPAV includes everyone, which allows for realistic comparisons between people with different objective life opportunities (Phillips 2006). In this assessment, individuals are asked to rate each item for importance and satisfaction using a 5-point Likert scale. Importance serves as a weight for a score conversion from satisfaction to a QOL score for each item. The measurement instrument has shown good reliability (Cronbach's $\alpha > 0.80$) based on data from Canadian adolescents aged 14–20 years. The measurement properties of QOLPAV have also been examined in terms of factor structures and convergent validity (Raphael et al. 1996; Bradford et al. 2002; Meuleners et al. 2003; Meuleners and Lee 2005). In an exploratory factor analysis, 8 factors were found in QOLPAV (Bradford et al. 2002). A second-order confirmatory factor analysis suggested there is one overall QOL second-order factor, which is divided into factors of psychology, environment, social function, opportunity for growth, and health (Meuleners et al. 2003).

Previous analyses of QOLPAV items mainly focused on the classical testing model. Recently, there has been an increased use of item response theory (IRT) in polytomously scored scales, such as 5-point Likert scales (Gomez 2008; Hays et al. 2007). IRT allows researchers to evaluate the functioning of individual response options and how accurately each item measures across the satisfaction continuum. Item banks based on polytomous IRT models are beneficial because IRT allows people to answer different questions, but scored on the same metric/scale. This approach facilitates comparisons across different samples who only consider items that are relevant to their life satisfaction in terms of

importance. As far as we know, no research has been carried out on the evaluation of the IRT properties of QOLPAV. The present study aims to evaluate the psychometric properties of the satisfaction rating of QOLPAV (QOLPAV-SAT) based on a polytomous IRT model, that is, the graded response model (GRM; Samejima 1969). Once the items are evaluated, further development of QOLPAV-SAT can be implemented (e.g., an overall QOLPAV index, a short version of QOLPAV-SAT, and a computerized adaptive assessment).

1.1 The Graded Response Model

The GRM, introduced by Samejima (1969), belongs to the category of item response models for which the responses are on an ordinal scale, such as Likert-type scales. The response options are ordered, for example, with respect to the degree of agreement, satisfaction, domain trait, or success to the solution of a problem. The GRM allows each item to have a slope (discrimination) and k thresholds ($k = \text{the number of categories} - 1$), which represent the locations of the latent domain on the scale.

In the GRM, a two-stage process is used to obtain the probability that an individual will receive a given category score on a certain item. In the first stage, the cumulative probability that an individual with a certain domain level receives a given category score or higher on item i is expressed by:

$$p_{ix}^*(\theta) = \frac{\exp[Da_i(\theta - b_{ix})]}{1 + \exp[Da_i(\theta - b_{ix})]} \quad (1)$$

For Eq. 1, a_i is the discrimination parameter of item i , θ is the domain level, and b_{ix} is the category threshold associated with category x for item i . Note that a_i is a constant for each category on item i . The responses to item i are successive integers 0, 1, ..., m_i . For each item, an a_i term and a set of m_i category thresholds (b_{ix}) are estimated. The categories within an item are ordered in terms of location, and the steps within an item are completed in order. For simplicity, $D = 1.0$ is used to place parameters on a logistic metric.

The second stage involves obtaining the probability that an individual will respond in a given category. This category probability can be obtained by subtracting adjacent cumulative probabilities conditional on θ such that:

$$p_{ix}(\theta) = p_{ix}^*(\theta) - p_{i,x+1}^*(\theta) \quad (2)$$

The probability of responding in the lowest category (category 0) or higher ($p_{i0}^*(\theta)$) is defined as 1.0, and the probability of responding in category $m_i + 1$ or higher ($p_{i,m+1}^*(\theta)$) is defined as 0.0. Therefore, Eq. 2 can be used to obtain the probability of responding in either of the two extreme categories. The probability of responding in the lowest category is calculated by subtracting $p_{i1}^*(\theta)$ from 1.0, and the probability of responding in the highest category is calculated by subtracting 0.0 from $p_{im}^*(\theta)$.

2 Methods

2.1 Sample

Participants were drawn from the list of respondents to a national adolescent well-being survey. Regional clusters were classified according to official Taiwan territorial divisions

(northwest, midwest, southwest, and east/islands). The numbers and percentages of participating schools and students in each region were determined with educational statistics published by the Taiwan Ministry of Education. Schools were randomly selected in each region, and one class was randomly selected from each school. The final sample consisted of 602 middle school students (297 male, 305 female) and 790 high school students (457 male, 333 female) who completed questionnaires assessing their perceived importance and satisfaction ratings for a variety of life aspects.

2.2 Instrument

QOLPAV consists of 54 items and examines the QOL of 3 domains of adolescent functioning: *being*, *belonging*, and *becoming*. *Being* is further divided into physical being, psychological being, and spiritual being; *belonging* includes physical belonging, social belonging, and community belonging; and *becoming* describes practical becoming, leisure becoming (enjoyment), and growth becoming. Each sub-domain contains 6 items. Because grade 9 and grade 12 students in Taiwan are required to take high school and college entrance exams, respectively, a question addressing the opportunity of educational tracks in the next school level was added. Therefore, the total number of questions in the survey was 55.

Participants were asked to rate each item for importance and satisfaction using a 5-point Likert scale. The importance and satisfaction scores ranged from 1 (not at all important/no satisfaction at all) to 5 (extremely important/extremely satisfied). The QOL scores were computed based on Raphael et al.'s equation: $QOL = (\text{importance score}/3) \times (\text{satisfaction score} - 3)$.

The QOLPAV scale in this study was forward translated into Chinese by the authors. A backward translation was performed by a translator who had never seen the original survey questions. The backward translation was compared with the original, and the discrepancies between them were checked.

2.3 Item Analyses

The psychometric properties were first examined based on traditional item analyses (e.g., item response frequency and Cronbach's α). We checked the assumptions of unidimensionality and local item independence before using PARSCALE (Muraki and Bock 2003) to calibrate items within the framework of the GRM.

2.3.1 Unidimensionality

QOLPAV-SAT is designed so that a higher score represents a higher life satisfaction. To combine item information across items in the IRT models, the item responses must represent one dominant construct. Therefore, we utilized confirmatory factor analysis to examine whether the dataset was sufficiently unidimensional, as suggested by Hill et al. (2007). Because Likert-type scale scores are rarely normally distributed, categorical confirmatory factor analysis (CCFA) using diagonally weighted least squares (DWLS) in LISREL (Jöreskog and Sörbom 1993) was adopted to evaluate unidimensionality (Hill et al. 2007). Before the analysis, we first examined the data and used listwise deletion to eliminate missing data. The fit indices obtained from CCFA were compared with Browne and Cudeck (1993), McDonald and Ho (2002), and Hu and Bentler's (1999) cutoff criteria

for model-data-fit. The criteria included a comparative fit index (CFI) > 0.95 and a root mean square error of approximation (RMSEA) at < 0.08 (reasonable fit).

Because the scale was originally designed based on 3 domains (being, belonging, and becoming), an alternative model was adopted to reflect these domains and to evaluate whether a general factor exists in the measurement model (i.e., a bi-factor model). The results are helpful in determining whether separate subscales are necessary. This analysis was conducted by comparing factor loadings on the general factor from the bi-factor model with those on the single factor from the unidimensional model. Here, a large absolute difference indicates a distortion of dimensionality caused by a factor other than the general factor (Reise et al. 2007; Cook et al. 2009).

Alternative approaches for assessing sufficient unidimensionality were also performed, including the exploratory factor analysis (EFA) using polychoric inter-item correlations and the internal consistency reliability (Cronbach's α). The standard of unidimensionality in EFA was that the first factor accounts for at least 20% of the variance and that the ratio of the first and second eigenvalues is greater than 4.

2.3.2 Local Item Independence

Within the framework of IRT, the probabilities of each item response are driven by the latent construct; after accounting for the latent construct, there should be no relationship between the responses to different items. Therefore, residual correlations among items were examined after the dominant factor was removed in the unidimensional model, which can be obtained from LISREL. High residual correlations (absolute value > 0.2) were flagged and considered as a possible local item dependence (LID) (Reeve et al. 2007). Because large modification indices (MI) between items might be an indication of LID, MI was also utilized to identify the LID problems (Hill et al. 2007).

2.3.3 Evaluating Model-Data-Fit

Because the GRM was selected in this study, we examined how closely the actual data correspond to the GRM. The extent of closeness between the actual data and the predicted model can be indicated by model-data-fit indices, for which we used $S-G^2$ (Orlando and Thissen 2003) and $S-X^2$ (Orlando and Thissen 2000). These indices compare expected and observed frequencies of item category responses for various levels of the total observed score. IRTFIT (Bjorner et al. 2007) was used to calculate these two indices based on the calibrated item parameters obtained from PARSCALE. In addition, visual information of model-data-fit was obtained from the IRT categorical response curves and item information curves.

2.3.4 Reliability and Validity Analyses

In IRT, reliability is conceptualized as “information” of an item or a test and measurement precision, which can be different across the levels of the domain being measured. High information is associated with low standard error of measurement (SE), which suggests greater measurement precision. Therefore, IRT item/test information plots were used to inspect whether an item/test could provide accurate measurement at different domain levels.

Internal consistency (Cronbach's α) and convergent validity were carried out in this study. Convergent validity was performed using the correlation between the QOLPAV total satisfaction scores and the life satisfaction score from the sum of 5 7-point Likert-type items (Raphael et al. 1996). Validity analyses were also conducted to see if the QOLPAV-SAT score distinguishes among subgroups that, theoretically, should differ in mean scores. We conducted analyses of variance (ANOVAs) to compare the mean satisfaction scores of those who reported different levels of perceived health (poor, fair, good, very good, or excellent). ANOVA was also performed to compare the mean satisfaction scores of those who reported different health conditions (chronic diseases, psychological problems, or healthy).

2.3.5 Differential Item Functioning

Differential item functioning (DIF) examines whether an item is biased or in favor of certain subgroups. After controlling the overall score of a domain, people from different groups should have equal probabilities of endorsing specific item responses. In the context of self-reported adolescent QOL, it is reasonable to investigate whether boys and girls who are equal in perceived QOLPAV-SAT score are equally likely to select one particular response to an item. Because QOLPAV is a generic QOL questionnaire that can be adopted by research for general or vulnerable populations, people in different objective health conditions are expected to respond equally to most of the life aspects, given that their overall QOLPAV-SAT scores are the same. Therefore, DIF was examined in terms of objective health condition and sex.

Because the items in QOLPAV-SAT were analyzed using IRT, we conducted an IRT-based DIF evaluation for two types of subgroup (boys vs. girls and healthy vs. those with chronic diseases). We used the IRT likelihood ratio approach (IRTLRDIF) by Thissen (2001) along with the effect size calculation provided by VisualDF (Meade 2010). The analysis was conducted based on the items that fit the GRM. The readers are referred to Thissen et al. (1993) for detailed IRTLRF techniques. The indices from VisualDF are composed of signed item difference in sample (SIDS), unsigned item difference in sample (UIDS), expected score standardized difference (ESSD), and expected test score standardized difference (ETSSD). ESSD and ETSSD can be explained using the guidelines of Cohen's d (Meade 2010). In VisualDF, we also checked DIF using visual information (i.e., plots of expected score across the trait continuum).

3 Results

3.1 Unidimensionality

The CCFA results of the one-factor model from LISREL showed an RMSEA of 0.067, indicating a reasonable fit of the unidimensional model. The CFI value of 0.96 was higher than the cutoff criterion of the model fit (CFI > 0.95). However, the bi-factor model with one general factor and 3 domain-specific factors had a slightly better fit (RMSEA = 0.054; CFI = 0.97). The factor loadings in the unidimensional model and those on the general factor in the bi-factor model are nearly identical (see Table 1). Unidimensionality was confirmed using EFA, in which the ratio of eigenvalues of the first and second factors was approximately 6.89:1 and the variance accounted for by the dominant factor was 32%. In general, the unidimensional and bi-factor results together

Table 1 Means, standard deviations, and factor loadings

Item	M	SD	CCFA	Bi-factor			
				1-Factor	G	Factor 1	Factor 2
<i>Being</i>							
1	3.41	0.88	0.32		0.29	0.19	
2	4.34	0.92	0.42		0.39	0.19	
3	3.30	0.92	0.50		0.45	0.35	
4	3.42	1.06	0.54		0.50	0.32	
5	3.66	0.99	0.56		0.51	0.39	
6	3.44	0.97	0.55		0.52	0.31	
7	2.86	1.16	0.45		0.40	0.35	
8	3.47	0.93	0.62		0.56	0.44	
9	3.86	1.08	0.61		0.55	0.42	
10	3.23	1.20	0.57		0.53	0.32	
11	3.53	1.04	0.49		0.46	0.24	
12	3.67	0.98	0.50		0.49	0.15	
13	3.74	1.03	0.58		0.54	0.32	
14	3.89	1.04	0.66		0.59	0.47	
15	3.69	1.08	0.66		0.59	0.44	
16	3.09	1.28	0.24		0.23	0.09	
17	3.93	0.88	0.53		0.52	0.11	
18	3.58	0.97	0.43		0.41	0.19	
<i>Belonging</i>							
19	3.72	1.05	0.34		0.32		0.29
20	3.52	1.10	0.40		0.38		0.41
21	2.90	1.24	0.45		0.43		0.39
22	3.76	0.99	0.54		0.50		0.65
23	3.54	1.01	0.51		0.48		0.63
24	3.57	1.05	0.52		0.50		0.46
25	3.81	0.89	0.63		0.64		0.11
26	3.58	0.93	0.58		0.59		0.06
27	3.84	1.07	0.57		0.58		0.15
28	3.10	1.16	0.38		0.40		-0.09
29	3.17	1.10	0.35		0.37		-0.11
30	4.14	0.89	0.62		0.64		-0.10
31	3.92	0.90	0.63		0.65		0.01
32	3.96	0.93	0.64		0.66		0.06
33	3.53	1.09	0.54		0.55		-0.05
34	3.04	1.19	0.45		0.47		-0.13
35	3.77	1.02	0.68		0.71		-0.08
36	3.73	1.00	0.54		0.57		-0.13
37	3.13	1.13	0.48		0.50		-0.07
<i>Becoming</i>							
38	3.44	0.97	0.53		0.54		0.04
39	4.05	0.85	0.64		0.64		0.08

Table 1 continued

Item	M	SD	CCFA		Bi-factor		
			1-Factor	G	Factor 1	Factor 2	Factor 3
40	3.45	1.06	0.62	0.64			0.02
41	3.65	0.95	0.67	0.68			0.07
42	3.60	0.89	0.67	0.68			0.06
43	3.17	1.17	0.42	0.43			0.04
44	3.59	1.03	0.56	0.54			0.24
45	4.30	0.83	0.59	0.54			0.48
46	4.16	0.88	0.51	0.46			0.52
47	4.00	1.00	0.61	0.54			0.64
48	4.02	0.97	0.64	0.56			0.61
49	3.34	1.06	0.51	0.48			0.28
50	3.66	0.97	0.68	0.68			0.10
51	3.82	0.93	0.68	0.68			0.13
52	3.84	0.90	0.70	0.69			0.16
53	3.45	1.06	0.66	0.67			0.12
54	3.50	1.03	0.68	0.68			0.14
55	3.67	1.02	0.71	0.72			0.09

with EFA outcomes identified one primary dimension in QOLPAV-SAT, and it is appropriate to have one single assessment score to represent the construct.

3.2 Local Item Independence

Based on the magnitude of the modification indices, the residual correlation, and the content of the items, more than 22 pairs of items were suspected of having LID (see Table 2). After allowing these locally dependent items to have correlated errors, the fit of the unidimensional model improved slightly ($CFI = 0.98$ and $RMSEA = 0.050$). The item parameters were examined after one item from the potential locally dependent (LD) item pairs was removed. The absolute changes of item parameter estimates obtained before and after removing one item from the LD item pairs ranged from 0.01 to 1.46. A change of 1.46 was found in item pairs 45 and 46, and a change of 0.99 was found in item pairs 22 and 23. Removing one item from item pairs 5 and 6, 14 and 15, 20 and 22, 28 and 29, or 31 and 32 also showed a slight change of item parameter estimates. We conceived these potential LD items as a result of presenting life aspects nearly identical to each other in terms of content, wording, and position in item sequence.

3.3 Item Calibration and Model-Data-Fit

IRTFIT (Bjorner et al. 2007) was used to test the model-data-fit. The $S-G^2$ and the $S-X^2$ statistics showed that items 1, 2, 16, 21, and 43 were misfit. The estimated item parameters and fit indices ($S-G^2$ and $S-X^2$) were presented in Table 3. The range of item discrimination was 0.62–1.84. Other than the misfit items, items 19, 20, 28, and 29 had item discrimination lower than 0.8, and their factor loadings on the unidimensional model

Table 2 Modification indices (MI) and residual correlations

Item pairs	MI	Res. corr.	Items	
4 and 5	602.83	0.26	4: My exercising and being fit	5: My physical health
4 and 6	186.52	0.21	4: My exercising and being fit	6: My nutrition and the food I eat
5 and 6	756.57	0.31	5: My physical health	6: My nutrition and the food I eat
7 and 8	145.63	0.22	7: Being free of worry and stress	8: How I feel about myself
8 and 9	207.67	0.22	8: How I feel about myself	9: Knowing who I am
13 and 14	325.51	0.26	13: Feeling part of things	14: Feeling that life has meaning
14 and 15	n/a	0.31	14: Feeling that life has meaning	15: Having hope for the future
17 and 38	126.94	0.23	17: Helping others	38: Doing volunteer work for others
20 and 22	128.37	0.24	20: The area of the country I live in	22: The house or apartment I live in
20 and 23	151.45	0.26	20: The area of the country I live in	23: The neighborhood I live in
21 and 22	201.32	0.25	21: The earth and its environment	22: The house of apartment I live in
22 and 23	3,076.55	0.43	22: The house or apartment I live in	23: The neighborhood I live in
22 and 24	205.10	0.25	22: The house or apartment I live in	24: The school I attend
23 and 24	214.58	0.27	23: The neighborhood I live in	24: The school I attend
28 and 29	220.69	0.35	28: Having a girlfriend/boyfriend	29: Having parties and things to go to
31 and 32	901.25	0.31	31: Being able to get medical/social services on my own	32: Getting good education
34 and 43	232.44	0.31	34: Having jobs available while still in school	43: The work I do at a job while still in school
40 and 42	131.24	0.17	40: Studying and doing homework	42: The things I do in school
41 and 42	313.41	0.25	41: The chores I do at home	42: The things I do in school
45 and 46	907.16	0.36	45: Having hobbies and personal interests	46: Indoor activities (e.g. TV, reading, etc.)
45 and 48	183.46	0.20	45: Having hobbies and personal interests	48: Participating in sports and recreation activities
50 and 51	853.87	0.21	50: Being successful at the things I do	51: Getting along better with others
Others	≤126			

(Table 1) were generally small (≤ 0.40). We conducted two more analyses by calibrating the item parameters (1) after removing the 5 misfit items and (2) after deleting the 5 misfit items and the 4 items with low discrimination. The item parameters were compared across

Table 3 Item parameter estimates and model-fit statistics

Item	a	b1	b2	b3	b4	df	$S-G^2$		$S-X^2$	
							Value	p	Value	p
<i>Being</i>										
1	0.63	-6.56	-3.39	0.70	3.52	153	187.73	0.03	177.85	0.08
2	0.35	-1.63	-0.96	0.71	2.99	163	1,511.81	0.00	1,213.93	0.00
3	1.01	-3.94	-1.68	0.63	2.75	148	159.21	0.25	148.41	0.48
4	1.06	-3.76	-1.44	0.29	1.82	153	135.43	0.84	128.99	0.92
5	1.16	-4.16	-1.91	-0.19	1.44	153	127.18	0.94	121.01	0.97
6	1.14	-3.87	-1.74	0.37	1.93	151	146.3	0.59	140.86	0.71
7	0.87	-2.35	-0.56	1.55	2.83	160	185.29	0.08	172.11	0.24
8	1.39	-3.14	-1.75	0.25	1.87	152	134.81	0.84	129.16	0.91
9	1.33	-3.20	-1.88	-0.38	0.72	149	119.84	0.96	118.12	0.97
10	1.16	-2.44	-0.84	0.53	1.71	162	138.89	0.91	135.08	0.94
11	1.04	-3.68	-1.80	0.09	1.73	153	159.83	0.34	157.61	0.38
12	1.03	-4.02	-2.25	-0.27	1.61	154	142.85	0.73	130.78	0.91
13	1.22	-3.39	-2.15	-0.15	1.10	150	144.65	0.61	139.37	0.72
14	1.48	-3.04	-1.84	-0.48	0.71	154	137.3	0.83	134.95	0.86
15	1.49	-2.84	-1.52	-0.14	1.00	156	126.85	0.96	122.64	0.98
16	0.41	-4.42	-1.95	1.67	3.89	165	219.07	0.00	193.43	0.06
17	1.10	-4.78	-3.16	-0.70	1.12	146	130.62	0.81	123.86	0.91
18	0.87	-4.74	-2.52	0.07	1.99	154	136.94	0.83	132.06	0.90
<i>Belonging</i>										
19	0.62	-5.58	-3.40	-0.58	1.88	156	149.5	0.63	147.59	0.67
20	0.76	-3.98	-2.33	-0.01	2.08	157	156.08	0.51	153.43	0.57
21	0.87	-2.30	-0.44	1.22	2.45	165	213.3	0.01	204.62	0.02
22	1.09	-3.80	-2.31	-0.42	1.35	150	148.78	0.51	140.16	0.71
23	1.06	-3.54	-1.97	0.01	1.81	157	154.63	0.54	143.03	0.78
24	1.06	-3.20	-1.96	-0.11	1.72	157	174.2	0.16	167.28	0.27
25	1.47	-3.57	-2.35	-0.41	1.25	142	126.56	0.82	118.99	0.92
26	1.33	-3.70	-2.03	0.13	1.60	149	133.71	0.81	128.90	0.88
27	1.25	-3.23	-1.90	-0.49	0.86	149	136.55	0.76	131.46	0.85
28	0.72	-3.26	-1.27	1.14	2.87	162	148.63	0.77	142.81	0.86
29	0.66	-3.94	-1.72	1.00	3.21	159	180.12	0.12	168.00	0.30
30	1.45	-3.76	-2.48	-1.04	0.49	139	114.58	0.94	112.07	0.95
31	1.49	-3.80	-2.52	-0.51	0.92	149	125.51	0.92	121.35	0.95
32	1.54	-3.31	-2.28	-0.67	0.80	143	102.21	0.99	94.90	0.99
33	1.13	-3.12	-1.68	0.08	1.51	154	139.81	0.79	131.14	0.91
34	0.88	-2.58	-0.92	1.08	2.40	160	173.65	0.22	170.37	0.27
35	1.71	-2.79	-1.66	-0.25	0.95	148	141.88	0.63	136.25	0.75
36	1.16	-3.58	-2.16	-0.29	1.29	156	160.03	0.40	148.08	0.66
37	0.96	-2.71	-1.07	0.91	2.31	157	186.19	0.06	175.77	0.15
<i>Becoming</i>										
38	1.11	-3.63	-1.87	0.38	1.97	149	150.05	0.46	143.19	0.62
39	1.51	-4.19	-2.64	-0.87	0.75	137	116.13	0.90	106.79	0.97

Table 3 continued

Item	<i>a</i>	b1	b2	b3	b4	<i>df</i>	<i>S-G</i> ²		<i>S-X</i> ²	
							Value	<i>p</i>	Value	<i>p</i>
40	1.44	-2.68	-1.31	0.20	1.52	152	144.41	0.66	138.2	0.78
41	1.67	-3.07	-1.71	-0.04	1.31	143	114.1	0.96	110.11	0.98
42	1.70	-3.19	-1.83	0.00	1.53	142	128.99	0.78	138.80	0.56
43	0.86	-2.58	-1.47	0.78	2.42	156	228.72	0.00	219.33	0.00
44	1.21	-3.20	-1.92	0.07	1.41	147	138.82	0.67	138.67	0.68
45	1.32	-4.74	-3.05	-1.35	0.14	141	151.63	0.26	142.65	0.45
46	1.06	-5.02	-3.31	-1.27	0.49	149	136.56	0.76	127.03	0.90
47	1.32	-3.80	-2.18	-0.66	0.60	150	134.02	0.82	127.65	0.91
48	1.41	-3.80	-2.16	-0.68	0.55	147	124.58	0.91	122.85	0.93
49	1.04	-3.21	-1.54	0.51	1.96	157	167.84	0.26	161.52	0.39
50	1.64	-3.09	-1.76	-0.02	1.23	148	127.61	0.89	119.18	0.96
51	1.72	-3.14	-1.96	-0.38	1.05	142	124.77	0.85	116.06	0.95
52	1.78	-3.45	-2.07	-0.35	1.00	140	111.66	0.96	98.92	1.00
53	1.60	-2.62	-1.22	0.23	1.42	148	115.39	0.98	113.11	0.99
54	1.64	-2.65	-1.35	0.17	1.39	149	124.21	0.93	119.80	0.96
55	1.84	-2.66	-1.44	-0.12	1.09	145	131.78	0.78	129.94	0.81

these two analyses. The person parameter estimates were also evaluated in terms of the difference between two sets of person parameter estimates and their correlations. In general, the absolute difference between item parameter estimates ranged from 0 to 0.05. The person parameter estimates showed root mean square differences of 0.04 overall and a correlation of 0.99. We found that most of the discrepancies occurred at the extremes of trait continuum.

3.4 Reliability and Validity

The internal consistency of QOLPAV-SAT was high (Cronbach's $\alpha = 0.95$). The total information, with the exclusion of the 5 misfit items and the 4 items with low discrimination, is presented in Fig. 1. More information was provided on the left half of the domain scale. The total satisfaction score was significantly correlated with a life satisfaction score obtained from an external questionnaire (Pearson's $r = 0.48$). The validity of the QOLPAV-SAT can also be demonstrated using an ANOVA, which tests whether there are differences in satisfaction scores among various health conditions (see Table 4). The results indicate that adolescents in different perceived health conditions have different mean satisfaction scores ($F(4, 1,358) = 33.60, p < 0.001$). The post-hoc comparison (Games–Howell) showed that there was no difference between the mean satisfaction scores of those reporting poor health and those reporting fair health. However, the mean satisfaction scores for those indicating good, very good, or excellent health were significantly higher than those who perceived themselves to have poor health. The results did not show different mean satisfaction scores for healthy students, students with chronic diseases, and those with psychological problems. These results suggest that the perceived QOL was influenced by perceived health, but not by health conditions.

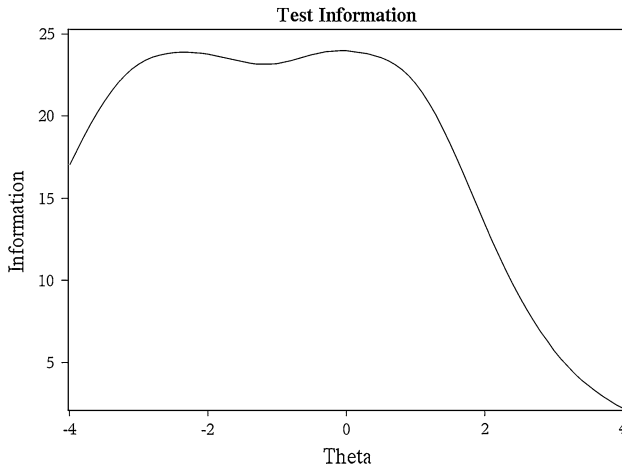


Fig. 1 Test information function obtained after removing the 5 misfit items and the 4 low discriminative items

Table 4 Descriptive statistics and ANOVA of QOLPVA total score in various health categories

	<i>N</i>	Mean	SD	<i>F</i>
Self-perceived health				33.60***
Poor	74	3.39	0.62	
Fair	677	3.47	0.51	
Good	251	3.62	0.44	
Very good	271	3.78	0.52	
Excellent	90	3.95	0.53	
Health problems				2.10
Chronic diseases (Asthma, diabetes, etc.)	287	3.57	0.53	
Psychological	22	3.82	0.56	
Healthy	1,058	3.59	0.53	

*** $p < .001$

3.5 Differential Item Functioning

Two group differences of interest in the study were sex and objective health condition. The DIF analysis proceeded after removing the 5 misfit items and the 4 low discriminative items. As shown in Table 5, 19 items were flagged as DIF items in terms of sex difference based on the effect size statistics from VisualDF ($ESSD > 0.2$). Five items functioned differently for healthy adolescents and those with chronic diseases. There were items not detected as DIF in terms of ESSD, but they showed a slightly differential functioning across subgroups according to the visual plots from VisualDF (see Figs. 2, 3). In general, our results suggest that beyond the construct of life satisfaction, there were factors associated with sex that also affected the rating on these items. It is conceivable that most of the items in favor of boys were related to physical health, sports, and being fit. However, girls tended to give higher ratings for the items that are associated with characteristics such as helping others and studying. With respect to objective health condition, our study suggests

Table 5 IRTLRDIF and VidualDF results (based on 46 items)

Item	SIDS	UIDS	ESSD	Favor	Item contents
<i>Boys versus girls</i>					
3	-0.17	0.17	0.44	Boy	My appearance-how I look
4	-0.45	0.45	0.94	Boy	My excising and being fit
5	-0.11	0.11	0.24	Boy	My physical health
7	-0.21	0.21	0.50	Boy	Being free of worry and stress
8	-0.19	0.19	0.40	Boy	How I feel about myself
11	-0.13	0.13	0.28	Boy	Thinking and acting independently
13	-0.13	0.13	0.27	Boy	Feeling part of things
17	0.12	0.12	0.30	Girl	Help others
18	-0.10	0.10	0.26	*	My own ideas of right and wrong
22	0.10	0.10	0.23	*	The house or apartment I live in
24	0.09	0.09	0.22	Girl	The school I attend
34	0.11	0.11	0.28	Girl	Having jobs available while still in school
38	0.15	0.16	0.36	Girl \times boy**	Doing volunteer work for others
39	0.20	0.20	0.44	Girl	Looking after my appearance and hygiene
40	0.16	0.16	0.27	Girl	Studying and doing homework
42	0.13	0.13	0.26	Girl	The things I do in school
46	0.11	0.11	0.27	Girl \times boy	Indoor activities (e.g., TV, reading, etc.)
48	-0.22	0.22	0.41	Boy	Participating in sports and recreation activities
49	0.10	0.10	0.22	Girl	Visiting and spending time with others
<i>Healthy versus chronic diseases</i>					
5	-0.26	0.26	0.57	Healthy	My physical health
10	0.19	0.19	0.36	Chr. disease	Knowing where I am going
11	0.12	0.12	0.29	Chr. disease	Thinking and acting independently
22	0.10	0.10	0.25	*	The house or apartment I live in
34	0.13	0.14	0.32	Chr. disease \times healthy	Having jobs available while still in school

* Visual plots show no DIF

** Nonuniform DIF

that item 5 (“My physical health”) favored healthy persons, whereas adolescents with chronic disease seemed to express confidence in “Knowing where I am going” and “Thinking and acting independently.”

The expected test score standardized difference (ETSSD) from VisualDF was -0.014 for sex and -0.004 for objective health-condition. These values imply that the impact of differential functioning at the scale level is small in terms of the effect size taxonomies of Meade (2010).

4 Discussion

Our intent has been to provide a psychometric evaluation of the items in QOLPAV based on a polytomous IRT model. The finding suggests that the majority of items in terms of satisfaction ratings fit the GRM. This conclusion is supported by the results of the

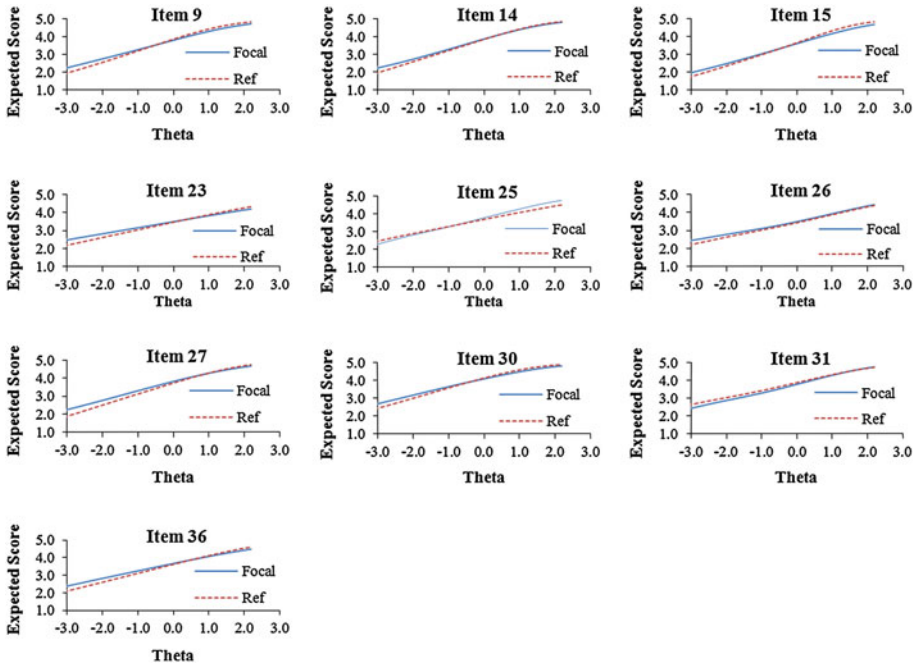


Fig. 2 DIF plots for sex (*focal* girls and *ref* boys)

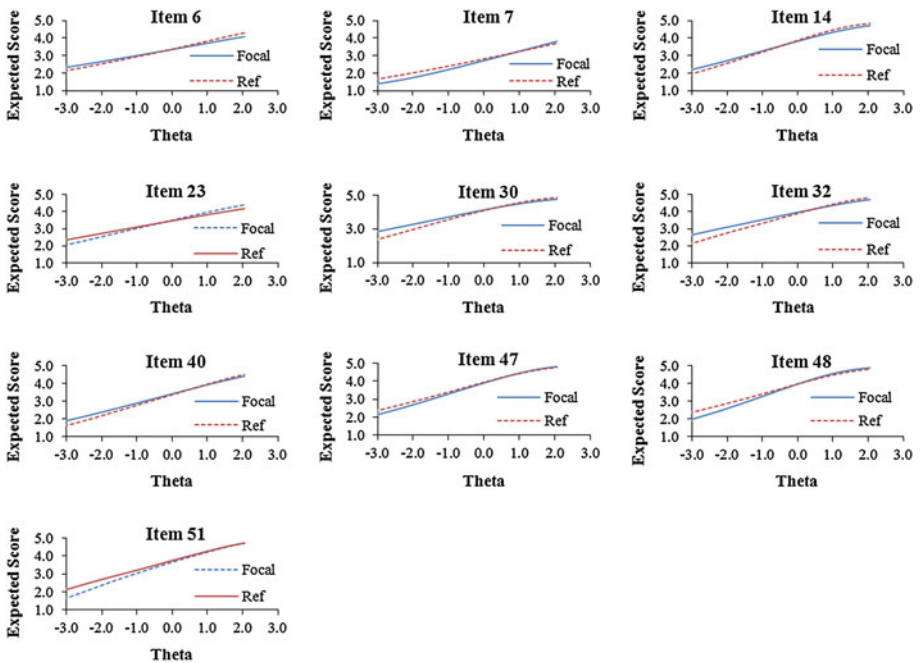


Fig. 3 DIF plots for objective disease condition (*focal* chronic disease and *ref* healthy)

examination in (1) unidimensionality, (2) local item independence, and (3) model-data-fit. In addition, the findings of reliability, validity, and differential functioning indicate that QOLPAV-SAT is a reliable, valid, and unbiased overall scale.

With respect to the IRT assumption evaluation, the 55 items analyzed in the present study represent various aspects of adolescent QOL. Items measuring such complex constructs are rarely strictly unidimensional. However, our results suggest that the QOLPAV-SAT measures one dominant/general construct, which not only meets the unidimensionality assumption of IRT but also supports the use of one summary index for overall adolescent QOL (Amtmann et al. 2010). Strong support for one homogeneous scale is also found from the high internal consistency estimates of the questionnaire.

To facilitate the use of IRT, it is necessary to avoid including both of the questions that are locally dependent on the score estimation. Among the 55 items, we suspect that several item pairs have potential LID problems. Removing one item of the LID pair from the scoring will result in more accurate estimates (Reeve et al. 2007). To determine which item should be retained, the principle of item selection in test development can be adopted. That is, the item with higher item information on the domain scale can be selected. However, we suggest an alternative method of item selection: the item of the LID pair receiving the higher importance rating is included. If both items in the pair are equally weighted in terms of importance, one of them can be arbitrarily selected.

Overall, the items provide good coverage of lower levels of quality of life satisfaction. The item category thresholds are concentrated on the lower end of the scale, and in turn, more information is provided toward the left of the continuum because the general adolescent population tends to be satisfied with various life aspects, which results in the non-normality of category responses that is usually observed in self-reported questionnaires. Figure 1 shows the test information function after the 5 misfit items and the 4 low discriminative items were removed, which indicates that the questionnaire can more accurately estimate the perceived QOL for those who are less satisfied with various life aspects.

The present study provides evidence that QOLPAV-SAT has high internal consistency and is able to distinguish between respondents who perceived themselves to be in poor versus good health. The DIF findings indicated possibilities of item bias in terms of objective health condition and sex. Because of the small differential functioning effect size of the scale score across subgroups, we determined that they have trivial impacts on the overall QOL score. Therefore, any significant QOLPAV-SAT score difference between samples with different health conditions is due to group differences rather than scale bias. However, caution should be taken when computerized adaptive testing or shorter versions of this scale are implemented because the DIF impact at scale level is unknown if the assessment version is different than the one analyzed in the present study. Because the scope of the study did not allow for investigating DIF using various approaches, some other promising techniques were not presented here. Further DIF investigations are needed for QOLPAV-SAT.

Wallander et al. (2001) noted that importance weights are crucial elements of QOL measurement. Trauer and Mackinnon (2001) also argued that the importance of an aspect of life in a satisfaction scale has explanatory value and should not be neglected. Nevertheless, the multiplicative method designed originally in QOLPAV is not psychometrically valid. If an aspect of life is not at all or not very important to an individual, then the satisfaction rating should not contribute to the overall QOL index. A scoring scheme that only considers the items rated “somewhat important,” “very important,” and “extremely important” may provide accurate and theoretically valid results. Future studies should

consider taking importance weights into account in the overall QOL index within the framework of IRT.

Acknowledgments This research was funded by the ROC National Academy of Educational Research (NAER-95-24-A-2-01-00-2-14) and the ROC National Science Council through three projects: “The Net and Taiwan Adolescents’ Physical and Mental Development” (NSC 97-2631-S-009-001), “Internet Use and Well-being of Taiwan Students: Longitudinal Analysis and the Extension of Fact Bank” (NSC-100-2631-S-009-001), and “Internet Use and Its Socio-cultural Impacts on Taiwan Adolescent Students” (NSC 97-2631-S-415-003).

References

- Amtmann, D., Cook, K. F., Jensen, M. P., Chen, W.-H., Choi, S., Revicki, D., et al. (2010). Development of a PROMIS item bank to measure pain interference. *Pain, 150*, 173–182. doi:[10.1016/j.pain.2010.04.025](https://doi.org/10.1016/j.pain.2010.04.025).
- Bjorner, J. B., Smith, K. J., Stone, C., & Sun, X. W. (2007). *IRTFIT: A macro for item fit and local dependence tests under IRT models [Computer program]*. Lincoln, RI: Quality Metric.
- Bradford, R., Rutherford, D. L., & John, A. (2002). Quality of life in young people: ratings and factor structure of the quality of life profile-adolescent version. *Journal of adolescence, 25*, 261–274. doi:[10.1006/yjado/469](https://doi.org/10.1006/yjado/469).
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Beverly Hills, CA: Sage.
- Cook, K. F., Kallen, M. A., & Amtmann, D. (2009). Having a fit: impact of number of items and distribution of data on traditional criteria for assessing IRT’s unidimensionality assumption. [Article]. *Quality of Life Research, 18*(4), 447–460. doi:[10.1007/s11136-009-9464-4](https://doi.org/10.1007/s11136-009-9464-4).
- Goldbeck, L., Schmitz, T. G., Besier, T., Herschbach, P., & Henrich, G. (2007). Life satisfaction decreases during adolescence. *Quality of Life Research, 16*, 969–979. doi:[10.1007/s11136-007-9205-5](https://doi.org/10.1007/s11136-007-9205-5).
- Gomez, R. (2008). Parent ratings of the ADHD items of the disruptive behavior rating scale: Analyses of their IRT properties based on the generalized partial credit model. *Personality and Individual Differences, 45*, 181–186.
- Hays, R. D., Liu, H., Spritzer, K., & Cella, D. (2007). Item response theory analyses of physical functioning items in the medical outcomes study. *Medical Care, 45*(1), S32–S38.
- Hill, C. D., Edwards, M. C., Thissen, D., Langer, M. M., Wirth, R. J., Burwinkle, T. M., et al. (2007). Practical issues in the application of item response theory: A demonstration using items from the pediatric quality of life inventory (PedsQL) 4.0 generic core scales. *Medical Care, 45*(1), S39–S47. doi:[10.1097/01.mlr.0000259879.05499.eb](https://doi.org/10.1097/01.mlr.0000259879.05499.eb).
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Model, 6*, 1–55.
- Jöreskog, K., & Sörbom, D. (1993). *LISREL8: Structural equation modeling with the command language*. Hillsdale, NJ: Erlbaum.
- Lee, A. H., Meuleners, L. B., & Fraser, M. L. (2009). Adolescent quality of life in Australia. In V. R. Preedy & R. R. Watson (Eds.), *Handbook of disease burdens and quality of life measures* (pp. 2538–2554). New York: Springer.
- McDonald, R. P., & Ho, M.-H. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods, 7*, 64–82.
- Meade, A. W. (2010). A taxonomy of effect size measures for the differential functioning of items and scales. *Journal of Applied Psychology, 95*, 728–743.
- Meuleners, L. B., & Lee, A. H. (2005). Quality of life profile—Adolescents: Assessing the relationship of covariates to scale scores using structural equation modeling. *Quality of Life Research, 14*, 1057–1063. doi:[10.1007/s11136-004-2573-1](https://doi.org/10.1007/s11136-004-2573-1).
- Meuleners, L. B., Lee, A. H., Binns, C. W., & Lower, A. (2003). Quality of life for adolescents: Assessing measurement properties using structural equation modeling. *Quality of Life Research, 12*, 283–290.
- Muraki, E., & Bock, R. D. (2003). *PARSCALE for windows (version 4.0) [computer software]*. Lincolnwood, IL: Scientific Software International.
- Orlando, M., & Thissen, D. (2000). Likelihood-based item-fit indices for dichotomous item response theory models. *Applied Psychological Measurement, 24*, 50–64. doi:[10.1177/01466216000241003](https://doi.org/10.1177/01466216000241003).

- Orlando, M., & Thissen, D. (2003). Further examination of the performance of S-X², an item fit index for dichotomous item response theory models. *Applied Psychological Measurement, 27*, 289–298. doi:[10.1177/0146621603027004004](https://doi.org/10.1177/0146621603027004004).
- Patrick, D. L., Edwards, T. C., & Topolski, T. D. (2002). Adolescent quality of life, part II: Initial validation of a new instrument. *Journal of Adolescence, 25*, 287–300. doi:[10.1006/yjado/471](https://doi.org/10.1006/yjado/471).
- Phillips, D. (2006). *Quality of life: Concept, policy and practice*. New York, NY: Routledge.
- Raphael, D., Rukholm, E. B., Hill-Bailey, P., & Donato, E. (1996). The quality of life profile-adolescent version: Background, description, and initial validation. *Journal of Adolescent Health, 19*, 366–375.
- Reeve, B. B., Hays, R. D., Bjorner, J. B., Cook, K. F., Crane, P. K., Teresi, J. A., et al. (2007). Psychometric evaluation and calibration of health-related quality of life item banks: Plans for the patient-reported outcomes measurement information system (PROMIS). *Medical Care, 45*(5 Suppl. 1), 22–31. doi:[10.1097/01.mlr.0000250483.85507.04](https://doi.org/10.1097/01.mlr.0000250483.85507.04).
- Reise, S. P., Morizot, J., & Hays, R. D. (2007). The role of bifactor model in resolving dimensionality issues in health outcome measures. *Quality of Life Research, 16*, 19–31. doi:[10.1007/s11136-007-9183-7](https://doi.org/10.1007/s11136-007-9183-7).
- Renwick, R., & Brown, I. (1996). The centre for health promotion's conceptual approach to quality of life: being, belonging, and becoming. In R. Renwick, J. Brown, & M. Nagler (Eds.), *Quality of life in health promotion and rehabilitation: Conceptual approaches, issues and applications*. Thousand Oaks, CA: Sage Publications.
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement, 17*, 34.
- Shek, D. T. L. (2005). Economic stress, emotional quality of life, and problem behavior in Chinese adolescents with and without economic disadvantage. *Social Indicators Research, 71*, 363–383. doi:[10.1007/1-4020-3602-7_12](https://doi.org/10.1007/1-4020-3602-7_12).
- Snyder, A. R., Martinez, J. C., Bay, R. C., Parsons, J. T., Sauers, E. L., & McLeod, T. C. V. (2010). Health-related quality of life differs between adolescent athletes and adolescent nonathletes. *Journal of Sport Rehabilitation, 19*, 237–248.
- Thissen, D. (2001). *Software for the computation of the statistics involved in item response theory likelihood-ratio tests for differential item functioning [computer software]*. Chapel Hill: University of North Carolina.
- Thissen, D., Steinberg, L., & Wainer, H. (1993). Detection of differential item functioning using the parameters of item response models. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 67–113). Hillsdale, NJ: Erlbaum.
- Trauer, T., & Mackinnon, A. (2001). Why are we weighting? The role of importance ratings in quality of life measurement. *Quality of Life Research, 10*, 579–585. doi:[10.1023/A:1013159414364](https://doi.org/10.1023/A:1013159414364).
- Valois, R. F., Zullig, K. J., Huebner, E. S., & Drane, J. W. (2004). Life satisfaction and suicide among high school adolescents. *Social Indicators Research, 66*, 81–105. doi:[10.1023/B:SOCI.0000007499.19430.2f](https://doi.org/10.1023/B:SOCI.0000007499.19430.2f).
- Wallander, J. L., Schmitt, M., & Koot, H. M. (2001). Quality of life measurement in children and adolescents: Issues, instruments, and applications. *Journal of Clinical Psychology, 57*, 571–585. doi:[10.1002/jclp.1029](https://doi.org/10.1002/jclp.1029).
- WHOQOL Group. (1994). The development of the WHO quality of life assessment instrument (the WHOQOL). In J. Orley & W. Kuyken (Eds.), *Quality of life Assessment: International perspectives* (pp. 41–60). Berlin: Springer.
- Zekovic, B., & Renwick, R. (2003). Quality of life for children and adolescents with developmental disabilities: review of conceptual and methodological issues relevant to public policy. *Disability and Society, 18*, 19–34.