Published in final edited form as:

Qual Life Res. 2016 December ; 25(12): 3191–3196. doi:10.1007/s11136-016-1344-0.

Applying Generalizability Theory Methods to Assess Continuity and Change on the Adolescent Quality of Life-Mental Health Scale (AQOL-MHS)

Ligia M. Chavez, Ph.D.¹, Pedro Garcia, M.S.¹, Nyrrma Ortiz, Ph.D.¹, and Patrick E. Shrout, Ph.D.²

Behavioral Sciences Research Institute, University of Puerto Rico, Medical Sciences Campus

²New York University, Department of Psychology

Abstract

Purpose—The Adolescent Quality of Life Mental Health Scale (AQOL-MHS) is designed to measure quality of life in clinical samples of Latino adolescents aged 12–18 years. Initial findings support its reliability, validity and conceptual model for use of its three domains (Emotional Regulation, Self-Concept, and Social Context). Our current study tests the usefulness of the AQOL-MHS for tracking changes in HRQOL during the course of service use.

Methods—Three waves of data were collected from 59 participants who were recruited from mental health clinics, where waves were spaced four months apart. Participants were receiving services at baseline assessment and were tracked for follow-up appointments regardless of treatment status. We analyze conventional reliability statistics for individual differences (e.g. Cronbach's alpha and test-retest correlations) and to estimate the reliability of change, we carried out a variance decomposition analysis.

Results—The psychometric analyses from Chavez et al. [1] were replicated with comparable results. A Generalizability Theory (GT) analysis revealed that the AQOL-MHS domains can measure change reliably. The reliability estimates varied from .65 to .78. Although there was reliable change at the individual level, on the average the AQOL-MHS means improved only slightly over time.

Conclusions—The reliability of change for all three scales in the AQOL-MHS was acceptable, but consistently lower than the Cronbach's alpha values for each wave. Future work will use this approach to adapt our current measure to provide a better reliability of within person change and thus broaden its applicability for prospective use.

Corresponding author: Ligia M. Chavez, Ph.D., University of Puerto Rico, Medical Sciences Campus, PO Box 365067, San Juan, PR 00936-5067. Phone: (787) 754-8624; Fax: (787) 767-5959 ligia.chavez@upr.edu.

Conflict of Interest: The authors 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.”

“This article does not contain any studies with animals performed by any of the authors.”

Informed consent: “Informed consent was obtained from all individual participants included in the study.”

Keywords

Quality of Life; Generalizability Theory; Mental Health; Adolescents

Introduction

The term “Quality of Life” (QOL) was first introduced in the mid to late 1970’s as a key term in medical indexes but its systematic study took off a decade later, primarily in the field of oncology. Quality of life (QOL) is considered a broad multidimensional concept that includes both physical and psychological functioning, subjective evaluations of both positive and negative aspects of life, as well, as encompassing both medical and non-medical aspects [2–4]. Recent years have seen a transformation from general to personalized medicine, and researchers have found utility in defining QOL specific to different disabilities and health problems that diverse populations experience. For example, there are now a number of instruments designed to measure specific domains such as asthma [5], cancer [6], diabetes, multidimensional fatigue [7] and craniofacial differences [8] among others. Such specific measures are often used to follow an individual’s change over time and with treatment; and for comparing sub-groups of patients.

In the spirit of focusing on specific populations and disorders, we developed a QOL instruments for adolescents with mental health problems, the Adolescent Quality of Life-Mental Health Scale (AQOL-MHS) [1, 9]. This instrument differs from other QOL instruments developed for children or adolescents (e.g. the KIDSCREEN [10], PedsQL [11], YQOL-R [3]) in that it is a specific instrument developed for clinical mental health populations. The AQOL-MHS was developed using a grounded theory approach where youths themselves defined what they believed to be important concepts with items that are self-reported (patient-centered), and attention was given that the items be both culturally and developmentally appropriate. The uniqueness of having a clinic population define the domains meant that they could discuss issues such as stigma and medication use in ways that were meaningful to them as patients.

The AQOL-MHS has three scales: Emotional regulation, Social context and Self-concept. Previous work [1] evaluated the reliability of the scale in two ways. We used both estimates of internal consistency (Cronbach’s Alpha) and test-retest reliability (intraclass correlations). Both the total scale and individual domain scales had reliability estimates that were .80 or better, indicating good to excellent reliability. We tested for convergent and discriminant validity and found that the AQOL-MHS had patterns of correlations that were consistent with our hypothesis driven predictions and theoretical assumptions. We also observed results for known-groups validity that indicated that the AQOL-MHS could differentiate adolescents at risk on variables such as having psychiatric diagnoses, functional impairment, having received mental health services, family functioning, having repeated a grade at school, and having experienced bullying within the last 30 days.

These results for AQOL-MHS focus on the psychometrics of individual differences cross sectionally. However, one important application of patient-centered measures is to track changes as the underlying health or mental health condition changes. Just because a measure

is reliable and valid for measuring individual differences at one time point does not mean that it will also be reliable and valid for measuring change [12–14]. The more a construct is trait-like, the higher the test-retest correlations of its measures, but the less likely that the measure will be sensitive to change. On the other hand, the more a construct is state-like, the more likely that its test-retest correlations will be lower than the reliability estimated at one point in time (e.g. with Cronbach's alpha). Trait and state variation can be considered simultaneously if a measure with multiple items is applied to the same persons at multiple points in time. Cranford et al. [12] showed how Generalizability Theory (GT) [15] can be used to decompose item response variance from longitudinal studies into trait and state components, and to use these components to estimate the reliability of change.

This article uses longitudinal data collected to apply the GT methods of Cranford et al. [12] to adolescent reports on AQOL-MHS items at three times over eight months. We compare the conventional reliability statistics for individual differences (e.g. Cronbach's alpha and test-retest correlations) to the GT estimates and evaluate the results from the perspective of researcher's who are interested in changes in QOL among adolescents with mental health problems.

Methods

Participants and procedure

Participants with one of five prevalent mental disorders: Attention Deficit Hyperactivity Disorder (ADHD), Conduct Disorder (CD), Oppositional Defiant Disorder (ODD), Generalized Anxiety Disorder (GAD) or Major Depressive Disorder (MDD), were recruited from 4 clinics of the San Juan Metropolitan area. These clinics included the APS Healthcare in San Juan, APS Healthcare in Bayamon, Pediatric Hospital, Child and Adolescent Clinic Medical Sciences Campus as well as from some private practices. The participants were assessed three times, four months apart. The inclusion criteria for the study were as follows: a) The youth was at least 12 years old, but not older than 18, b) The youth was a new client, in that he/she had received 3 or fewer sessions with a mental health professional in the current clinical setting, and c) The youth had not received mental health treatment at other settings in the 6 months prior to recruitment. In addition, study exclusion criteria include cognitive impairment, history of severe brain injury, pervasive developmental disorder, or evident sensory impairment.

After an initial visit to inform the mental health professional on the study we made routine visits to collect referrals in which reported age, sessions, diagnosis and severity. Parents were called to explain the research project and those who expressed interest an initial interview was coordinated with the interviewer for home visits. Parents signed a consent form in which they agreed to participate in the study for the course of a year.

All participants were receiving services at the time of the baseline assessment and were tracked for follow up appointments regardless of treatment status. Field work was completed at the end of January 2014, and data cleaning was completed in April 2014. All participants received a research incentive for participating. The incentives were incremental for each study visit. For the initial interview they were offered \$25.00, for follow-up 1, \$30.00 and

for follow-up 2, \$35.00. The initial interview lasted around 1 hour and the follow up interviews lasted around 45 minutes.

Measure

AQOL-MHS—The AQOL-MHS [1] is a self-report measure with 26 items that can be scored to produce three QOL scales: Emotional regulation (*ER*- 9 items), Self-concept (*SC*- 8 items), and Social context (*SO*- 9 items). Participants evaluate whether the item description applies to them using an 11 point scale that ranges from 0= "Completely disagree" to 10= "Completely agree". The responses to the *ER* scale were reverse coded so that larger numbers correspond to more positive QOL. The AQOL-MHS was measured at baseline and two additional follow-up visits every four months.

Statistical Analysis

We first replicated the psychometric results from Chavez et al. [1] by calculating Cronbach's alpha for each of the three scales at each time point, and by examining the two- and four-month correlations for each of the scales both between waves and across waves. We also conducted *t-tests* to assess change in QOL between the final assessment (Follow-up 2) and the baseline assessment, and estimated the unstandardized regression coefficient (b) and Cohen's effect size (d) [16] to provide information regarding the magnitude of the difference. Statistical significance was set at $p = .05$.

To estimate the reliability of change, we carried out analyses in the tradition of Generalizability Theory (GT) [15, 17, 18]. GT recognizes that variation in item responses can be due to multiple sources, and it provides a psychometric framework for comparing the variation due to the intended construct to variation that is considered to be noise. The framework is very general, but Cranford et al. [12] showed how it can be applied to analysis of change, where the sources of variation are components due to person (averaging across items and wave of interviews), item (averaging over persons and wave), time (averaging over persons and items) and also their two way interactions. To implement the Cranford approach we carried out a variance decomposition analysis using the VARCOMP procedure of SAS (Version 9.3) with restricted maximum likelihood (REML) estimation. The three way interaction of person by item by wave was used to estimate the underlying error variation.

According to Cranford et al. [12], the reliability of change (R_C) can be estimated using the Person*Wave and Error components of the variance decomposition. Their method assumes that all persons are asked to answer all items, and therefore that the items are fixed effects. The formula for calculating R_C is shown in equation 1. The impact of the error term is reduced by a factor of $1/m$, where m is the number of items in the scale.

$$R_C = \frac{\sigma^2_{\text{PERSON*WAVE}}}{\left[\sigma^2_{\text{PERSON*WAVE}} + \sigma^2_{\text{ERROR}} / m \right]} \quad \text{Eq. 1}$$

Results

Table 1 shows the means, standard deviations and alpha estimates for the three scales at each wave of data for 59 participants at baseline. The scale scores can range from zero to ten, and so the means indicate that for Social context and Self-concept the reported QOL in this sample is fairly high. In all three waves, the alphas for Emotional regulation were slightly smaller than the value of .82 reported by Chavez et al. (2014)[1]. However, the alphas for the other two scales were comparable to the values previously reported (Self-concept (.85) and Social context (.80)), and the alphas for all three scales were consistent across waves.

Also apparent in Table 1 is that QOL in all three domains did not change much over the three waves. Analyses of differences between baseline and follow-up 2 assessments indicated that the means for Emotional regulation increased subtly over time, ER $b=0.32$ $t(57)=2.07$, $p=.04$, $d=.31$, and for Self-concept a statistical trend $b=.19$; $t(57)=1.62$, $p=.11$, $d=.23$, while the Social context scale remained stable ($b=.08$; $t(57)=.68$, $p=.50$, $d=.11$).

Figures 1–3 show graphical representations of these changes in scores for the three AQOL-MHS scales. These figures include the mean change as heavy black lines and individual observed change as lighter lines. One can see from Figure 1 (first figure above on the left) that there is considerable variation in Emotional regulation at each time point and that a number of youth have consistently low or high levels of ER over time. However, there is also evidence of dramatic change across waves for some individuals. Figures 2 (Self-concept) and 3 (Social context) have less variation, and many youth who have high values at baseline remain high at follow-up 1 and 2. Like ER, there are some youth who have dramatic changes from one wave to the next.

Tables 2a and 2b show the correlations of the three AQOL-MHS scales within and between waves. In our previous study we found the two week test-retest reliability of the Emotional regulation, Self-concept and Social context scales to be .82, .90 and .83 [1]. The correlations of the baseline measures with the four month and eight month follow-up were smaller than these values, consistent with systematic change. For example, baseline Emotional regulation was correlated .55 and .37 with the four and eight month follow-up assessments, and baseline Self-concept was correlated .68 with the second wave of data and .42 with the third. The lowest level of correlations over waves was found for Social context. Baseline Social context was correlated .44 and .21 with four and eight month follow-up assessments.

The results of the GT analyses are shown in Table 3. The top part of the table shows how much of the item response variance can be attributed to between-person, time and item effects, as well as the two-way interactions. The largest portion of item response variance is associated with error, but this reflects error variation for individual items. The impact of the error variation is reduced when the eight or nine items are combined to form scale scores. Another large component of the variance for all three scales is the person effect (ranging from 10.5% to 21.2%), which reflects individual differences averaged over items and times. Consistent with Figures 1–3, little of the variation is due to overall change over time (averaging across items and persons). The variation associated with items ranges from 7.3% to 12.7%, but these differences in item severity do not have an impact on the measurement.

All persons answer all items, and thus the between person comparisons are not affected by this source of variation.

More important for the reliability of change is the person by time interaction, which ranges from 9.5% for Emotional regulation to 19.0% for Social context. This variation describes consistent changes of persons over the repeated measures (averaging over items). It is used in the numerator of Equation 1, which compares systematic change to total change variation. For Emotional regulation QOL the application of Equation 1 yields a reliability of change estimate of 0.65, and for both Social context and Self-concept the estimates are 0.78. These results indicate that for the AQOL-MHS, at least two thirds (.65) of the change variation revealed by the scales, is true within-person change in QOL over time.

Discussion

Using Generalizability Theory we were able to estimate the magnitude of systematic within-person change over time and to use these values to estimate the reliability of change (R_C). We found that the reliability of change for all three scales were acceptable, but consistently lower than the Cronbach's alpha values for each wave. The alpha values are appropriate for comparisons of subjects with regard to their QOL levels at a fixed point in time, but the R_C values are appropriate for comparing subjects with regard to their QOL change over time. Given that QOL is a primary outcome in many treatment studies and that there is an expectation that the treatment leads to QOL improvement, it is critical that the reliability of QOL change be formally examined.

Although all three AQOL-MHS scales can measure change over an eight month interval with adequate reliability, the results suggested that the Emotion Regulation scale was somewhat less reliable than the Social context and Self-concept scales (0.65 vs 0.78 and 0.78). From the variance decomposition we can determine that there was considerable person*wave variation in the Emotion Regulation scale, but there was even more variation in the error component. This scale includes items such as "I act crazy and out of control", and "When I get angry, my mind blocks up"; these items were phrased in terms of usual self, in other words they reflect personality traits that are relatively stable rather than recent emotional experiences. For example, the first of the above items might be reworded to be "I feel like I could act crazy and out of control". If adolescents are able to focus on the present state they might give a clearer signal about their changing QOL. Future plans include developing a state like version of the AQOL-MHS that could yield improved reliability of change by emphasizing state-level quality of life items rather than trait-level.

Our results are based on responses of clinical samples of adolescents in mental health care, and they will not necessarily generalize to other samples of youth. The extent to which quality of life changes over time, is likely to differ depending on the group of adolescents, for example those with no mental health problems or those with physical illness. Researchers who are interested in the reliability of change in other types of samples should conduct their own GT analyses.

The GT framework used in our analyses assumes homogeneity of items, similar to classical test theory of parallel measures. This assumption is like the one made implicitly when using Cronbach's alpha to estimate the reliability of measures at a single point in time. Insofar as some items differ in their relation to the underlying construct, the GT estimates of reliability might be conservative.

We believe there is a need for specific QOL measure that can track significant changes over time, and that such a measure would be increasingly relevant for the adolescent clinical population. Our results provide validation support for the classic psychometric properties of the AQOL-MHS and also back the use of accompanying standard approaches with the utility of applying generalizability theory to the estimation of reliability of change over time. To our knowledge this is the first study to apply a GT framework to assess the reliability of change in a QOL measure. By using this approach we will be able to adapt our current measure to provide a better reliability of within person change and thus broaden its applicability for future use.

Acknowledgments

This study was supported by the National Institute of Health Research Grant #SC1 HD060888 and #SC1 GM109326. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

We would like to thank the Pediatric Hospital of the University of Puerto Rico Medical Sciences Campus, the APS Health Care System of Puerto Rico and Karen Mir, Ph.D. for their valuable support with the recruitment of participants.

References

1. Chavez LM, Ramírez R, Garcia P, Canino G, Mir K, Ortiz N, et al. Measurement properties of the Adolescent Quality of Life Mental Health Scale (AQOL-MHS). *Quality of Life Research*. 2014; 23(4):1327–1335. [PubMed: 24241819]
2. Chen H, Cohen P, Kasen S, Johnson JG, Berenson K, Gordon K. Impact of adolescent mental disorders and physical illnesses on quality of life 17 years later. *Archives of Pediatrics & Adolescent Medicine*. 2006; 160(1):93–99. [PubMed: 16389217]
3. Patrick DL, Edwards TC, Topolski TD. Adolescent quality of life, part II: initial validation of a new instrument. *Journal of adolescence*. 2002; 25(3):287–300. [PubMed: 12128039]
4. Rabkin, J.; Wagner, G.; Griffin, KW. Quality of life measures. In: Rush, JA.; Pincus, HA.; First, MB., et al., editors. *Handbook of Psychiatric Measures*. Task Force for the Handbook of Psychiatric Measures. Washington, D.C.: American Psychiatric Association; 2000. p. 135-150.
5. Osman L, Silverman M. Measuring quality of life for young children with asthma and their families. *European Respiratory Journal Supplement*. 1996; 21:35s–41s.
6. Cataudella D, Morley TE, Nesin A, Fernandez CV, Johnston DL, Sung L, et al. Development of a quality of life instrument for children with advanced cancer: the pediatric advanced care quality of life scale (PAC-QoL). *Pediatric Blood Cancer*. 2014; 61(10):1840–1845. [PubMed: 24947134]
7. Varni JW, Limbers CA, Newman DA. Using factor analysis to confirm the validity of children's self-reported health-related quality of life across different modes of administration. *Clinical Trials*. 2009; 6(2):185–195. [PubMed: 19342471]
8. Edwards TC, Patrick DL, Topolski TD, Aspinall CL, Mouradian WE, Speltz ML. Approaches to craniofacial-specific quality of life assessment in adolescents. *Cleft Palate Craniofac Journal*. 2005; 42(1):19–24.

9. Chavez L, Mir K, Canino G. Starting from scratch: the development of the Adolescent Quality of Life-Mental Health Scale (AQOL-MHS). [Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't]. *Culture, Medicine & Psychiatry*. 2012; 36(3):465–479.
10. KIDSCREEN Group. The KIDSCREEN questionnaires— Quality of life questionnaires for children and adolescents— Handbook. Lengerich: Papst Science Publisher; 2006.
11. Varni JW, Seid M, Kurtin PS. PedsQL 4.0: reliability and validity of the Pediatric Quality of Life Inventory version 4.0 generic core scales in healthy and patient populations. *Medical Care*. 2001; 39(8):800–812. [PubMed: 11468499]
12. Cranford JA, Shrout PE, Iida M, Rafaeli E, Yip T, Bolger N. A procedure for evaluating sensitivity to within-person change: can mood measures in diary studies detect change reliably? *Personality and Social Psychology Bulletin*. 2006; 32(7):917–929. [PubMed: 16738025]
13. Cronbach L, Furby L. How we should measure 'change'- or should we? *Psychological Bulletin*. 1970; 74:68–80.
14. Shrout, PE.; Lane, SP. Psychometrics. In: Mehl, MR.; Conner, TS., editors. *Handbook of Research Methods for Studying Daily Life*. New York: Guilford; 2012. p. 302-320.
15. Cronbach, L.J.; Gleser, G.C.; Nanda, H.; Rajaratnam, N. *The Dependability of Behavioral Measurements*. New York: Wiley; 1972.
16. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd. New Jersey: Lawrence Erlbaum Associates; 1988. ISBN 0-8058-0283-5, retrieved 20 April, 2016
17. Brennan, R.L. *Generalizability Theory*. New York: Springer-Verlag; 2001.
18. Shavelson, R.J.; Webb, N.M. Generalizability theory. In: Green, J.L.; Camilli, G.; Elmore, P.B., editors. *Complementary Methods for Research in Education*. 3rd. Washington, DC: AERA; 2005.

Table 1

Descriptive Statistics for the AQOL-MHS at each Measurement Time and Assessment of Internal Consistency

AQOL-MHS Baseline n=59	Mean	Std. Dev.	Cronbach's Alpha
Emotional regulation	5.76	2.28	0.79
Self-concept	8.14	1.78	0.85
Social context	8.83	1.42	0.86
Follow-up 1 n=58			
Emotional regulation	5.84	2.01	0.74
Self-concept	8.52	1.37	0.84
Social context	9.11	1.02	0.81
Follow-up 2 n=58			
Emotional regulation	6.42	1.92	0.73
Self-concept	8.52	1.58	0.89
Social context	8.99	1.40	0.85

Table 2

a. AQOL-MHS Scales Correlation coefficients within Wave

	ER	SC	SO
Emotional regulation	1.00	0.26	0.22
Baseline			
Self-concept		1.00	0.61
Social context			1.00
F1			
Emotional regulation	1.00	0.29	0.26
Self-concept		1.00	0.62
Social context			1.00
F2			
Emotional regulation	1.00	0.35	0.32
Self-concept		1.00	0.76
Social context			1.00

b. AQOL-MHS Scales Correlation coefficients between Waves

	Follow Up 1 (F1)			Follow Up 2 (F2)		
	ER	SC	SO	ER	SC	SO
ER	0.55	0.27	0.28	0.37	0.17	0.18
Baseline						
SC	0.26	0.68	0.49	0.21	0.42	0.26
SO	0.18	0.35	0.44	0.07	0.05	0.21
F1						
ER				0.59	0.19	0.21
SC				0.30	0.60	0.39
SO				0.37	0.33	0.44

Note: ER = Emotional regulation; SC = Self concept; SO = Social context

Table 3

Variance Decomposition of Items in the Three AQOL-MHS Scales

	Emotional Regulation		Self-Concept		Social Context	
	9		8		9	
Number of items	9		8		9	
Number of waves	3		3		3	
Variance Component	Estimate	%	Estimate	%	Estimate	%
Var(Person)	1.797	12.0%	1.245	21.2%	0.481	10.5%
Var(Time)	0.086	0.6%	0.029	0.5%	0.0	0.0%
Var(item)	1.900	12.7%	0.688	11.7%	0.334	7.3%
Var(Person*Time)	1.423	9.5%	0.892	15.2%	0.871	19.0%
Var(Person*item)	2.945	19.6%	0.989	16.8%	0.687	15.0%
Var(Time*item)	0.027	0.2%	0.0	0.0%	0.023	0.5%
Var(Error)	6.835	45.5%	2.027	34.5%	2.182	47.7%
GT reliability estimate Within Per Change (Rc)	0.652		0.779		0.782	