Published in final edited form as:

J Child Fam Stud. 2018 October ; 27(10): 3176–3184. doi:10.1007/s10826-018-1158-5.

Measurement Invariance of the Adolescent Quality of Life-Mental Health Scale (AQOL-MHS) across Gender, Age and Treatment Context

Ligia M. Chavez¹, Patrick E. Shrout², Pedro García¹, Erick Forno³, and Juan C. Celedón³

¹Behavioral Sciences Research Institute, University of Puerto Rico, Medical Sciences Campus

²New York University, Department of Psychology

³Division of Pulmonary Medicine, Allergy and Immunology, Children's Hospital of Pittsburgh of UPMC, University of Pittsburgh, Pittsburgh, PA

Introduction

Adolescence is a challenging life stage, particularly for youth with mental disorders. While medical services often focus on symptom reduction, improving quality of life is equally important. We previously developed the Adolescent Quality of Life-Mental Health Scale (AQOL-MHS), a Spanish language instrument that focuses on the QOL of adolescents (12–18 years old) with mental health problems. The AQOL-MHS was developed in Spanish with adolescents living on the island of Puerto Rico, with promising results (Chavez et al. 2012; 2014).

The AQOL-MHS was constructed to have good psychometric properties that include internal consistency and test-retest reliability. We also have evidence for convergent and discriminant validity, and results for known-groups' validity supported the ability of the instrument to identify differences between adolescents with a mental health diagnosis, global impairment in functioning, and clinic versus community samples. We have not yet determined, however, whether the AQOL-MHS items are interpreted in the same way by boys and girls, older and younger adolescents, and youth in mental health treatment versus those in the community. Because adolescence is a vital time of change, adjustment and development, it is possible

Corresponding author: LM Chavez, 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.

The content of this article is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflict of Interest: The authors declare that they have no conflict of interest.

Ethical approval: The University of Puerto Rico Medical Sciences Campus IRB approved this study. "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."

Author Contributions:

LC: designed and executed the study, conducted data analysis, and wrote the manuscript. PS: collaborated in conceptualization and study design, data analyses and writing of the manuscript. PG: conducted data management and data analysis. EF and JCC: participated in study design, data management, and drafting the manuscript.

that expressions of quality of life related to emotional regulation, self-concept and social-context could differ across age groups and contexts. Moreover, the relevance of the AQOL-MHS constructs to youth with mental health difficulties might lead youth who are receiving treatment to interpret them differently than those with no formal connections to the mental health treatment system.

When the data suggest that items are interpreted the same way, we say the items and the scales composed of those items are invariant (Meredith and Teresi 2006). If items are not interpreted the same way, then conventional methods of summing or averaging item responses can lead to biases in comparisons of groups and individuals within groups. In recent years, the issue of measurement invariance has gained increasing interest (Ang et al. 2016; Chung et al. 2016; Maes et al. 2016).

Invariance is tested by comparing the correlations of latent variables to item responses across different groups (using factor loadings in confirmatory factor analysis), and by comparing average item responses after adjusting for latent group differences. Gregorich (2006) provided a systematic guide for testing measurement invariance. The most basic tests of invariance requires that factor analyses of items within a measure yields the same number of factors and the same item clusters within factors (*dimensional* and *configural* invariance). The next level is called *metric* invariance; this means that the factor loadings for each item are the same across the groups being compared and therefore that the item has the same association with the latent variable in each group. Given metric invariance, one can test *strong factorial invariance* (also called *scalar invariance*), which means that the level of average item endorsement, after adjusting for any group differences on the latent variable, are the same over groups. If items do not relate to the same factors, then it is not possible to interpret differences in observed item responses. When scalar invariance is violated, but metric invariance holds, then it is possible to obtain a statistical estimate of the latent trait for group comparisons using factor analysis software, instead of using simple averages of item responses because this can lead to biases (Gregorich 2006).

In this article, we report psychometric analyses of new data to address the question of whether the AQOL-MHS items are interpreted the same way by youth recruited from mental health clinics and those recruited from a community sample, for males and females, and for younger and older youth. We also report analyses that are designed to select subsets of the original AQOL-MHS items that can provide more efficient estimates of each scale. Both goals require that we verify that items of each scale can be fit by a unidimensional latent variable model in the new data we report. In the second step of the analysis we compare the item characteristics across sample type, gender, and age while adjusting for group differences in the latent variable. Our overall goal is to provide users with a shortened version of the AQOL-MHS that is robust with regard to measurement artifacts when being used to assess a variety of youth.

Method

Participants

Participants came from two different populations, one recruited from mental health clinical settings and the other from a community sample. The clinical sample is composed of 59 participants 12–18 years old 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). These participants were recruited from mental health clinics in the San Juan Metropolitan Area of Puerto Rico, and represent a sample of convenience. All participants were receiving services at the time of the protocol assessment. Study exclusion criteria included cognitive impairment, history of severe brain injury, pervasive developmental disorder or evident sensory impairment.

The second sample includes children who were selected using a multi-stage probability sample design from randomly selected households in San Juan. A household was eligible if at least one resident was 6–14 years old. In households with multiple eligible children, one child was randomly selected by using Kish tables. Participants were originally sampled for a study of asthma in Puerto Ricans, and thus included children with asthma (defined as physician-diagnosed asthma and wheeze in the prior year) and children without asthma. The main recruitment tool was a screening questionnaire given to parents. Of the 1,111 children who were eligible for the 2009–2010 study, approximately 515 returned for a follow up study conducted between 2013 and 2017, when they were 9 to 20 years old. Only children aged 12 to 18 years were included in the current analysis ($n=295$). Analyses were conducted to determine if there were differences between children with and without asthma in the AQOL-MHS, and none were found.

Procedure

For both samples procedures consisted of after an initial visit to inform the service providers and their staff on the study. We also made routine visits to collect referrals in which information was reported on age, gender, diagnosis and severity. Parents were informed of the research project and those who expressed an interest in participating, were then enrolled for an initial interview. Parents signed a consent form and youth signed assent/consent forms (depending on the age). All participants received a research incentive for participating. The incentives offered were: \$25 for adolescents participating in the clinic sample and from \$20 to \$40 for the community sample, based on the location of the interview. The initial interview lasted around 1 hour for both samples.

Measure

The original AQOL-MHS [3] contained 31 self-report items that comprised three QOL scales: Emotional regulation (ER) (11 items), Self-concept (SC) (10 items), and Social context (SO) (10 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 ER items are all phrased as emotional dysregulation, but they are reverse coded so that larger numbers correspond to positive QOL. In the first revision of the AQOL-MHS,

the measure was shortened to 26 items (ER: 9 items; SC: 8 items; SO: 9 items) (Chavez et al. 2016). This was the version administered to the current participants. The AQOL-MHS was administered in Spanish.

Data Analyses

To achieve the goal of further shortening the AQOL-MHS and verifying a single factor structure, we carried out a conceptual item analysis that sought to identify items that were highly redundant and that could be burdensome for participants completing the measure. The first pair of redundant items was from the Self-Concept scale. Item 14 “It is important to believe in myself” and Item 20 “I value and accept myself as a person”. Item 20 was retained. The second pair of items was from the Social Context scale, Item 21 “*It is important for me to do well at school*” and Item 27 “*I feel it is important to get good grades*”. Item 21 was retained. In short, 2 items (14 and 27) were eliminated.

After removing redundant items, we fit one-factor confirmatory factor (CFA) models to each of the scales. We evaluated whether a single dimension was appropriate by examining the eigenvalues of the item correlation matrix, and computing three common fit statistics for structural equation methods (Muthén and Muthén 1998–2017). The fit statistics we used are the root mean square error of approximation (RMSEA), Tucker-Lewis fit index (TLI), and the comparative fit index (CFI). These were computed by Mplus software (Version 7.1; Muthén and Muthén 1998–2017 with robust maximum likelihood estimation (MLR)).

Next, we evaluated measurement invariance following the steps laid out by Gregorich (2006) for testing configural, metric and scalar (strong factorial) invariance over the sample type, gender and age groups. Configural invariance was assessed by verifying that the single factor models fit in each of the groups being assessed. To test metric and scalar invariance models for each AQOL-MHS scale, we fit a series of nested CFA models in Mplus. The configural model, which we call Model 1, is a two group factor model with no parameters constrained to be the same across groups; this is essentially two parallel CFAs. The metric model (Model 2) is like the configural model, but the factor loadings are constrained to be the same and the factor variance is freed in the second group. Letting the number of items in each scale be designated as k , the Model 2 has $k-1$ additional degrees of freedom compared to Model 1. The scalar model (Model 3) was like Model 2, but k item intercepts were constrained to be the same across groups, and the factor mean was estimated in the second group; this model also had an extra $k-1$ degrees of freedom. The change in the fit of each successively constrained model can be tested using a test statistic of -2 times the differences in the log likelihood. Because robust ML estimation was used, this test statistic was scaled using the method of Satorra and Bentler (2001).

Results

Table 1 shows the demographic characteristics for both the community and clinic samples. The children in the two groups were comparable with regard to gender, context (community 45% vs clinical 52%) and age (community $M=14.52$, $SD=1.87$; clinical $M=14.90$, $SD=1.72$). The caretakers were also comparable in regard to gender, marital status and education.

However, compared to the clinical sample, the community sample has a higher number of caretakers who reported that they were economically living well.

Table 2 shows the mean and standard deviation for each of the 26 items in the community and clinical groups. All of the items were scored, so that higher values represent better QOL. In general, the average item responses were higher for the community than clinic respondents. When reviewing the descriptive statistics for each item by sample, we observed that Item 18 “I enjoy the support of others”, had very low variability in the clinical sample. We decided to eliminate Item 18 from further analyses.

Internal consistency (Cronbach’s alpha) for the AQOL-MHS scales is as follows: ER (.79 in clinical and .74 in community samples), SC (.85 clinical and .82 community), and SO (.79 clinical and .80 community). The intercorrelations among the three scales were positive and statistically significant with values of .25 (emotional and social), .63 (self and social), and .20 (emotional and self).

Confirmatory unidimensional latent structure

To verify that the psychometric characteristics previously found for the three scales are replicated in the current data, we carried out a series of confirmatory factor analyses that fit one factor models to the sets of items relevant to each scale. We found that the Emotional Regulation scale did not have a good fit for the one factor solution (RMSEA < 0.09; TLI > .82; CFI > .87). After carrying out an exploratory factor analysis for a two factor solution, we were able to identify Item 28 “*I am tired that people don’t understand me*”, as a major source of variance for a second factor. We decided to eliminate Item 28, and this made the one factor model fit better (RMSEA < 0.06; TLI > .90; CFI > .93). The factor loadings for the final Emotional Regulation analyses are shown in Table 2, along with loadings of the one factor solutions for Self Concept and Social Context scales.

Invariance Analyses

For each of the three AQOL-MHS scales, we carried out invariance analyses for sample type, gender and age. We first compared the community and clinical samples, and then merged these samples to test gender and age. To test age, we split the sample into younger (12–14 years, n=161) and older (15–18 years, n=134) groups. We fit the configural (Model 1), metric (Model 2) and scalar (Model 3) models, and compared the relative fit of each successively constrained model.

Emotional Regulation—Table 3 shows the fit statistics associated with the three models of the eight-item scale. The table has separate sections comparing groups defined by *sample type, gender and age*.

Sample Type: The baseline or configural model (Model 1) for *sample type* had fit statistics of RMSEA=.06 (0.03, 0.08), CFI=.94, and TLI=.91. These show adequate fit compared to conventional standards for a good fitting model (RMSEA < 0.08; TLI > .95; CFI > .95). The metric model (Model 2) constrained loadings to be the same across samples and it did not have a significantly worse fit ($X^2_{\text{diff}}(7) = 5.08, p < .65$). The global fit statistics for Model 2

showed a slight improvement and remained acceptable. Model 3 tested scalar invariance by constraining item intercepts (i.e., item means) to be equal for both groups. Full scalar invariance was not well supported by the data; Model 3 had a marginally worse fit than Model 2 ($X^2_{\text{diff}}(7) = 12.91, p < .07$). Examination of fit diagnostics revealed the presence of differential item functioning (DIF) for Item 4, (“*When I was angry I took it out on something*”). Participants from the community endorsed this item more than expected given their inferred level of emotion regulation. We created Model 4, a partial scalar invariance model, which removed Item 4 from the list of items that had constrained item intercepts. This model did not have a worse fit than Model 2 ($X^2_{\text{diff}}(6) = 7.23, P = 0.30$).

Gender: Model 1, the baseline or configural model for *gender*, had fit statistics of RMSEA=.07 (0.04, 0.09), CFI=.92, and TLI=.88, and we interpret these to have marginally adequate fit. Model 2, the metric invariance model for *gender*, which constrained loadings to be the same across boys and girls, did not have a worse fit ($X^2_{\text{diff}}(7) = 5.68, p < .58$). In fact, the global fit statistics for Model 2 were better than those for Model 1. Full scalar invariance was not supported by the data; Model 3 had a statistically worse fit than Model 2 ($X^2_{\text{diff}}(7) = 24.07, P < .001$). As in the analysis of sample type invariance, Item 4 did not have invariant intercepts in males and females. Even after adjusting for emotion regulation, boys tended to endorse Item 4 more than girls. In addition, Item 1 (“*I have acted out of control.*”) is endorsed more by boys than girls with equal levels of emotion regulation. We created a partial scalar invariance model (Model 4), allowing these two items to vary in intercept level. This model fit better than Model 3, and was only marginally a worse fit than Model 2 ($X^2_{\text{diff}}(5) = 9.39, P = 0.09$).

Age: When Model 1 was fit to the younger and older groups (11–14, 15–18) the fit was adequate by conventional standards: RMSEA=.06 (0.03, 0.09), CFI=.94, TLI=.91. The metric invariance model (Model 2) did not fit worse than Model 1 ($X^2_{\text{diff}}(7) = 10.15, P = 0.18$), and the scalar model (Model 3) did not fit worse than Model 2 ($X^2_{\text{diff}}(7) = 11.23, P = 0.13$). These results suggest that both full metric and scalar invariance are a tenable assumption when comparing older and younger youth.

Self-Concept

Sample Type: We originally conducted the invariance analyses using a seven item Self-Concept scale. The configural model for *sample type* had fit statistics of RMSEA=.05 (0.00, 0.08), CFI=.96, and TLI=.94, indicating good fit (see Model 1a in Table 4). Model 2a, which constrained the loadings to be the same across clinic and community samples had worse fit than Model 1a ($X^2_{\text{diff}}(6) = 14.38, P = 0.02$). Examination of fit diagnostics revealed the presence of non-uniform DIF for Item 17 “I stand by my decisions”. Although it had a high loading in the community, it had a low loading in the clinic sample. Because youth in the clinic comprise our target population, we decided to remove Item 17 from further consideration. Hence we proceeded using a six item Self-Concept scale.

The configural model 1 for *sample type* had fit statistics of RMSEA=.06 (0.02, 0.10), CFI=.96, and TLI=.93, demonstrating reasonable fit (See Table 4). Model 2, which constrained the loadings to be the same across the samples did not fit worse than Model 1 ($X^2_{\text{diff}}(5) = 5.88$,

$P=0.32$). Moreover, Model 3, which represents scalar invariance, fit nearly as well as Model 2 ($\chi^2_{\text{diff}}(5) = 9.98$, $P=0.08$). Because the comparison of Models 2 and 3 yielded marginal significance, we looked for a specific item among the six that had a notably different intercept, but we did not find one that stood out. The variation in the intercepts appeared to be distributed across the six items.

Gender: The configural model for *gender* showed adequate fit of configural (RMSEA=.06 (0.02, 0.10), CFI=.93, and TLI=.90), and Model 2 (metric invariance) did not have a worse fit ($\chi^2_{\text{diff}}(5) = 6.74$, $P=0.24$). However, the scalar model showed a decreased fit ($\chi^2_{\text{diff}}(5) = 25.52$, $P<0.01$). Item 29 (“*I have thought that things will work out for me in the future*”) was endorsed more often by females than males, even after adjusting for self-concept. Once the intercept for this item was released, the fit was not worse than Model 2 ($\chi^2_{\text{diff}}(4) = 6.57$, $P=0.16$).

Age: The configural model for age groups (12–14 years, 15–18 years) had adequate fit by conventional standards: RMSEA=.06 (0.02, 0.10), CFI=.96, TLI=.93. The further constraints imposed by the metric and scalar invariance models did not result in statistically worse fit (Table 4). This suggests that we can assume full metric and scalar invariance.

Social Context

Sample Type: The configural model for sample type has a good fit (RMSEA=.05 (0.00, 0.08), CFI=.96, and TLI=.94), and the fit for Model 2, the metric model, was not worse ($\chi^2_{\text{diff}}(6) = 3.19$, $P=0.78$) (Table 5). However, the scalar model (Model 3) showed a decreased fit relative to Model 2 ($\chi^2_{\text{diff}}(6) = 35.58$, $P<0.01$). This seemed to be due to Item 12, “*I have gotten along with others*”, which was endorsed more often by clinical participants than community after adjusting for any differences in social context between the groups. Model 4 allowed the intercept for Item 12 to be different between the groups, and this partial scalar invariance model did not fit worse than Model 2 ($\chi^2_{\text{diff}}(5) = 8.16$, $P=0.15$).

Gender: The configural model for gender had a good fit (RMSEA=.05 (0.00, 0.08), CFI=.96, and TLI=.94), and again the metric invariance model did not worsen the fit ($\chi^2_{\text{diff}}(6) = 6.98$, $P=0.32$). The Scalar invariance model, however, did have significantly worse fit ($\chi^2_{\text{diff}}(6) = 12.58$, $P=0.05$) and the difference seemed to occur for Item 9, “*I have enjoyed sharing and doing things with my family*”. This was endorsed more often than expected by males compared to females. Model 4, the partial scalar invariance model, allowed Item 9 to have a group specific intercept and its fit was not worse than Model 2 ($\chi^2_{\text{diff}}(5) = 7.14$, $P=0.21$).

Age: The configural model for age fit well (RMSEA=.04 (0.00, 0.07), CFI=.97, and TLI=.96), and again the metric model did not have worse fit ($\chi^2_{\text{diff}}(6) = 3.33$, $P=0.77$). Again the scalar invariance model had worse fit ($\chi^2_{\text{diff}}(6) = 14.14$, $P=0.03$), and the difference seemed to be due to Item 24, “*It is important for me to have the support of my teachers*”. After adjusting for social context, younger children tended to endorse this more highly than expected. The partial invariance model, Model 4, had an improved fit relative to Model 3 and it was not a worse fit than Model 2 ($\chi^2_{\text{diff}}(5) = 7.40$, $P=0.19$).

Summary of invariance analyses—Of the eight Emotion Regulation items, two showed some indication of differential item function (Items 1 and 4) for comparisons of clinic vs. community participants and boys and girls, but none showed age DIF. Of the six Self-Concept items, one item (Item 29) showed DIF with respect to gender comparisons, but no items were suspect for comparisons of sample type or age. Of the seven Social Context items, three (Items 9, 12 and 24) had patterns of DIF for gender, sample and age comparisons. When making comparisons of groups with items that have been shown to have DIF, Gregorich (2006) recommends using estimates of the latent variables from confirmatory factor analysis programs from models that allow adjustments for the tendency of one group to endorse a DIF item at a higher or lower level than another group rather than using the typical convention of summing (or averaging) the observed item responses.

Discussion

With some adjustments to the three AQOL-MHS scales, we were able to identify a 21-item version that has either full or partial measurement invariance for each of the scales. Using new data that combined participants from mental health clinics and community sources, we found evidence that factor loadings were invariant, suggesting that the items are consistently related to the underlying latent construct in each of the scales for older and younger adolescents, boys and girls, and those from the community versus the clinic sample.

Where full measurement invariance is achieved, group differences can be directly compared from observed scores. When partial invariance rather than full invariance is attained, there are one or more items that vary in endorsement probability beyond expectations due to differences in latent QOL. In these cases, the observed bias can be remedied by using factor scores to obtain a more precise estimate. These corrected estimates take into account the differences in the relevant item mean responses. Generally speaking, we found that the bias in the uncorrected summed scores was small; all of the correlations were above .90. For clinical purposes, the summed score provides a good approximation to the refined estimates. Nevertheless, for research that compares the level of the construct across groups, we recommend computing factor scores.

Limitations and Future Research Directions

Our study was limited by the small sample size in the clinic group and the fact that these participants were recruited from public clinics. Further research is warranted in a larger clinical sample and to characterize item functioning of responses from patients in private care or for specific types of disorders. Also, our results are based on responses of adolescents in Puerto Rico and they will not necessarily generalize to other samples of youth. Nevertheless, having a valid and reliable research measure will enhance the possibility of conducting research with Latinos in the broader US population. Future research will be of value to test the instrument with both Latino and non-Latino adolescents.

Among the limitations to our study are that we retained and discarded items using the same data that was used to verify our invariance analyses. This typically leads to an overestimation of the fit of the models. We encourage investigators using the AQOL-MHS to verify the fit of these models in new data. Likewise, because our results are based on

secondary data analyses, the specific groups examined were determined by the variables found in these data sets. We hope future testing of the AQOL-MHS will include other variables of interest.

Our results contribute to the tradition of establishing invariance by addressing possible measurement bias in diverse populations and thus advance the QOL literature (Bagheri et al. 2014; Limbers et al. 2008; Mora et al. 2009). The equivalent factor structure of the AQOL-MHS across three different group comparisons has significant implications, with special relevance placed on being able to screen and assess adolescents in both clinical and non-clinical groups. The new version also has improved clinical applicability by being shorter, resulting in a useful and potentially important instrument for measuring QOL. Going forward, we support the use of the 21 item AQOL-MHS by researchers interested in the construct of HRQOL, and by clinicians who study or treat youth with mental health problems.

Acknowledgments

Funding sources: This study was supported by the National Institutes of Health research award #HD060888 and #GM109326, and was partly supported by grants HL079966 and HL117191 from the U.S. National Institutes of Health (NIH), and by The Heinz Endowments.

References

- Ang CS, Tan JP, Fam SY. 2016; Psychometric properties of the Training Parenting Style Scale in a Malaysian sample of adolescents: Factor analysis, internal consistency, and measurement invariance. *Journal of Child and Family Studies*. 25:1505–1514.
- Bagheri Z, Jafari P, Tashakor E, Kouhpayeh A, Riazi H. 2014; Assessing whether measurement invariance of the KIDSCREEN-27 across child-parent dyad depends on the child gender: a multiple group confirmatory factor analysis. *Global Journal of Health Sciences*. 6(5):142–53.
- Chavez LM, Ramirez R, Garcia P, Canino G, Mir K, Ortiz N, Morales LS. 2014; Measurement properties of the Adolescent Quality of Life Mental Health Scale (AQOL-MHS). *Quality of Life Research*. 23(4):1327–1335. [PubMed: 24241819]
- Chavez L, Garcia P, Ortiz N, Shrout PE. 2016; Applying generalizability theory methods to assess continuity and change on the Adolescent Quality of Life-Mental Health Scale (AQOL-MHA). *Quality of Life Research*. 25:3191–3196. [PubMed: 27344319]
- Chavez L, Mir K, Canino G. 2012; Starting from scratch: The development of the Adolescent Quality of Life-Mental Health Scale (AQOL-MHS). *Culture, Medicine & Psychiatry*. 36(3):465–479.
- Chung H, Kim J, Park R, Bamer AM, Bocell FD, Amtmann D. 2016; Testing the measurement invariance of the University of Washington Self-Efficacy Scale short form across four diagnostic subgroups. *Quality of Life Research*. 25(10):2559–2564. [PubMed: 27116230]
- Gregorich SE. 2006; Do self-report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor analysis framework. *Medical Care*. 44(11 Suppl 3):S78–94. [PubMed: 17060839]
- Kline, R. Principles and practice of structural equation modeling, Fourth edition. New York: Guilford; 2016.
- Limbers CA, Newman DA, Varni JW. 2008; Factorial invariance of child self-report across age subgroups: a confirmatory factor analysis of ages 5 to 16 years utilizing the PedsQL 4.0 Generic Core Scales. *Value Health*. 11(4):659–668. [PubMed: 18179671]
- Maes M, Vanhalst J, Spithoven AW, Van den Noortgate W, Goossens L. 2016; Loneliness and Attitudes Toward Aloneness in Adolescence: A Person-Centered Approach. *Journal of Youth and Adolescence*. 45(3):547–567. [PubMed: 26369350]

- Meredith W, Teresi JA. 2006; An essay on measurement and factorial invariance. *Medical Care*. 44(11 Suppl 3):S69–77. [PubMed: 17060838]
- Mora PA, Contrada RJ, Berkowitz A, Musumeci-Szabo T, Wisnivesky J, Halm EA. 2009; Measurement invariance of the Mini Asthma Quality of Life Questionnaire across African-American and Latino adult asthma patients. *Quality of Life Research*. 18(3):371–380. [PubMed: 19221893]
- Muthén, LK, Muthén, BO. *Mplus User's Guide*. Eighth Edition. Los Angeles, CA: Muthén & Muthén; 1998–2017.
- Satorra A, Bentler PM. 2001; A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*. 66(4):507–514.

Table 1

Demographics characteristics of Participants by Sample Type

	Community Sample (n=295)		Clinic Sample (n=59)	
<i>Child</i>				
Gender				
Male	161	54.6%	28	47.5%
Female	134	45.4%	31	52.5%
Age				
12–14	161	54.6%	26	44.1%
15–18	134	45.4%	33	55.9%
<i>Caretaker</i>				
Gender				
Male	6	2.2%	5	8.5%
Female	271	97.8%	54	91.5%
Marital Status				
Married/Living with couple	107	36.3%	18	30.5%
Separated/Divorced/Widowed	65	22.0%	16	27.1%
Never have been married	123	41.7%	25	42.4%
Education				
Less than high school	59	20.1%	11	18.6%
High School	88	29.9%	15	25.4%
Some College or College degree	147	50.0%	33	55.9%
Perception of Poverty				
Lives well	161	54.58%	23	38.98%
Lives check to check	88	29.83%	22	37.29%
Lives poorly	46	15.59%	14	23.73%

Table 2

AQOL-MHS Means and Standard Deviations per Item (26) and Factor Loadings for the Combined Sample
One Factor Solution for the 21 Item AQOL-MHS

Scales	Community Sample (n=294)		Clinic Sample (n=59)		Combined Loadings	Std. Error
	Mean	Std. Dev.	Mean	Std. Dev.		
Emotional regulation						
AQ1	8.00	2.55	7.75	3.18	0.51	0.06
AQ4	6.74	3.67	6.73	3.55	0.60	0.05
AQ7	5.59	4.13	5.37	4.26	0.48	0.05
AQ10	7.67	3.20	6.66	3.63	0.72	0.05
AQ16	5.74	3.80	5.31	3.93	0.50	0.06
AQ19	4.97	3.86	3.61	3.40	0.55	0.05
AQ25	6.00	3.76	4.86	4.18	0.39	0.06
AQ28	6.16	3.84	4.51	4.07		
AQ31	8.38	2.93	7.07	3.51	0.48	0.06
Self-concept						
AQ2	8.78	2.15	7.75	3.04	0.68	0.06
AQ8	9.18	1.78	8.49	2.64	0.67	0.07
AQ14	9.37	1.84	9.19	1.77		
AQ17	8.31	2.50	7.73	2.51		
AQ20	9.32	1.73	8.44	2.34	0.71	0.06
AQ23	6.85	3.02	6.58	3.05	0.38	0.06
AQ26	9.01	1.92	7.73	3.00	0.65	0.07
AQ29	9.37	1.45	9.19	1.87	0.69	0.07
Social context						
AQ3	9.30	1.59	9.41	1.68	0.60	0.08
AQ6	8.20	2.70	7.47	3.34	0.63	0.05
AQ9	9.26	1.62	8.27	2.71	0.55	0.08
AQ12	9.30	1.57	9.53	1.15	0.68	0.07
AQ15	9.39	1.71	9.21	1.76	0.52	0.09
AQ18	9.43	1.51	9.63	0.79		
AQ21	9.65	1.20	8.98	2.37	0.59	0.10
AQ24	8.61	2.43	8.02	3.19	0.55	0.07
AQ27	9.70	1.09	9.00	2.26		

Table 3

Emotional Regulation (8 Items) Model Fit Summary

		Overall Fit Indexes					Comparative Fit Indexes			
Model		χ^2	df	RMSEA	CFI	TLI	Model Comparison	χ^2	df	P
Sample Type										
1	Configural	64.07	40	0.06 (0.03–0.08)	0.94	0.91				
2	Metric	69.22	47	0.05 (0.02–0.08)	0.94	0.93	2 vs 1	5.08	7	0.65
3	Scalar	82.00	54	0.05 (0.03–0.08)	0.93	0.93	3 vs 2	12.91	7	0.07
4	Partial Scalar ^a	76.64	53	0.05 (0.02–0.07)	0.94	0.94	4 vs 2	7.23	6	0.30
Gender										
1	Configural	73.79	40	0.07 (0.04–0.09)	0.92	0.88				
2	Metric	79.51	47	0.06 (0.04–0.9)	0.92	0.90	2 vs 1	5.68	7	0.58
3	Scalar	102.68	54	0.07 (0.05–0.09)	0.88	0.88	3 vs 2	24.07	7	0.00
4	Partial Scalar ^b	88.59	52	0.06 (0.04–0.09)	0.91	0.90	4 vs 2	9.39	5	0.09
Age										
1	Configural	64.32	40	0.06 (0.03–0.09)	0.94	0.92				
2	Metric	74.47	47	0.06 (0.03–0.08)	0.93	0.92	2 vs 1	10.15	7	0.18
3	Scalar	85.70	54	0.06 (0.03–0.08)	0.92	0.92	3 vs 2	11.23	7	0.13

^aOne item intercept was freely estimated across groups (AQ4).

^bTwo item intercepts were freely estimated across groups (AQ1, AQ4).

Note: Models were estimated using Robust Maximum Likelihood (MLR) estimation (Mplus).

Table 4

Self-Concept (6 Items) Model Fit Summary

		Overall Fit Indexes					Comparative Fit Indexes			
Model		χ^2	df	RMSEA (CI)	CFI	TLI	Model Comparison	χ^2	df	P
Sample Type										
1a	Configural	41.08	28	0.05 (0.00–0.08)	0.96	0.94				
2a	Metric	54.82	34	0.06 (0.03–0.09)	0.94	0.92	2a vs 1a	14.38	6	0.02
1	Configural	29.98	18	0.06 (0.02–0.10)	0.96	0.93				
2	Metric	35.66	23	0.06 (0.01–0.09)	0.96	0.94	2 vs 1	5.88	5	0.32
3	Scalar	45.23	28	0.06 (0.02–0.09)	0.94	0.93	3 vs 2	9.98	5	0.08
Gender										
1	Configural	34.23	18	0.07 (0.02–0.10)	0.95	0.91				
2	Metric	40.57	23	0.07 (0.03–0.10)	0.94	0.92	2 vs 1	6.74	5	0.24
3	Scalar	58.92	28	0.08 (0.05–0.11)	0.90	0.89	3 vs 2	25.52	5	0.00
4	Partial Scalar ^a	46.92	27	0.06 (0.03–0.10)	0.93	0.93	4 vs 2	6.57	4	0.16
Age										
1	Configural	31.03	18	0.06 (0.02–0.10)	0.96	0.93				
2	Metric	39.29	23	0.06 (0.03–0.10)	0.95	0.93	2 vs 1	8.34	5	0.14
3	Scalar	48.30	28	0.06 (0.03–0.09)	0.93	0.93	3 vs 2	9.26	5	0.10

^aOne item intercept was freely estimated across groups (item 29).

Note: Models were estimated using Robust Maximum Likelihood (MLR) estimation (Mplus).

Table 5

Social Context (7 Items) Model Fit Summary

		Overall Fit Indexes					Comparative Fit Indexes			
		χ^2	df	RMSEA	CFI	TLI	Model Comparison	χ^2	df	P
Sample Type										
1	Configural	39.51	28	0.05 (0.00–0.08)	0.96	0.94				
2	Metric	38.94	34	0.03 (0.00–0.06)	0.98	0.98	2 vs 1	3.19	6	0.78
3	Scalar	57.51	40	0.05 (0.01–0.08)	0.94	0.93	3 vs 2	35.58	6	0.00
4	Partial Scalar ^a	48.70	39	0.04 (0.00–0.07)	0.96	0.96	4 vs 2	8.16	5	0.15
Gender										
1	Configural	38.90	28	0.05 (0.00–0.08)	0.96	0.94				
2	Metric	45.46	34	0.04 (0.00–0.07)	0.96	0.94	2 vs 1	6.98	6	0.32
3	Scalar	55.81	40	0.05 (0.00–0.08)	0.94	0.94	3 vs 2	12.58	6	0.05
4	Partial Scalar ^b	52.92	39	0.05 (0.00–0.07)	0.95	0.94	4 vs 2	7.14	5	0.21
Age										
1	Configural	34.31	28	0.04 (0.00–0.07)	0.97	0.96				
2	Metric	36.12	34	0.02 (0.00–0.06)	0.99	0.99	2 vs 1	3.33	6	0.77
3	Scalar	46.33	40	0.03 (0.00–0.06)	0.97	0.97	3 vs 2	14.14	6	0.03
4	Partial Scalar ^c	44.92	39	0.03 (0.00–0.06)	0.97	0.97	4 vs 2	7.40	5	0.19

^aOne item intercept was freely estimated across groups (AQ12).

^bOne item intercept was freely estimated across groups (AQ9).

^cOne item intercept was freely estimated across groups (AQ24).

Note: Models were estimated using Robust Maximum Likelihood (MLR) estimation (Mplus).