Objective To extend the health-related quality of life assessment literature by examining the reliability and validity of Sizing Me Up in a community sample of nontreatment-seeking overweight and obese children. Methods Participants included 302 students (M age = 10.34) recruited from 6 elementary schools, who completed self-report measures of health-related quality of life and weight-related health. Results 134 overweight and obese children were included in the analyses to establish validity for the measure. Confirmatory factor analyses supported a 5-factor first-order factor structure with 1 second-order factor representing a total score. Convergent and criterion-related validity was established among overweight and obese children. Estimates of internal consistency for the 5-factor Sizing Me Up factor structure indicated that the subscales are not acceptable outside of a structural equation modeling approach. Conclusion Outside of a structural equation modeling framework, only the total score appears to be appropriate in this population. Key words assessment; confirmatory factor analysis; overweight and obesity; quality of life.

Health-related Quality of Life (HRQOL) is a multidimensional construct that refers to an individual’s functioning as directly affected by an illness or its treatment (Spieth & Harris, 1996). HRQOL is increasingly being recognized and used as a primary health outcome in both randomized clinical trials and phenomenological investigations because of its correlations with biological indicators of disease and its unique ability to capture the subjective experience of an individual, ensuring that the patient's perspective on her/his condition is understood (Spieth & Harris, 1996). HRQOL can be reliably and validly measured both narrowly within particular illness groups, as well broadly across a spectrum of ill and healthy groups (Limbers, Newman, & Varni, 2008; Palermo et al., 2008). General HRQOL measures have clear strengths for gathering epidemiological data; but often lack the specificity to detect small changes in HRQOL that may be unique to the disease experience for a particular illness population (Palermo et al., 2008; Zeller & Modi, 2008). It is commonly reported that disease-specific measures are appropriate for use in patients with a particular condition, but are not appropriate for the collection of normative data from healthy controls (Palermo et al., 2008).

Weight-related HRQOL is a sub-construct of general HRQOL that holds promise for achieving a patient-centered perspective of the overweight and obesity experience, and if measured properly, may provide a proximal indicator of treatment success that is likely to demonstrate clinically meaningful change before weight-related outcomes. A robust body of literature attests to the challenges of children and adolescents with overweight and obesity, including impairments in general HRQOL as well as HRQOL issues that are more specific to weight-related concerns (Zeller & Modi, 2008).

In response to identified gaps in the literature, Zeller and Modi (2009) developed the Sizing Me Up (SMU) questionnaire with the stated purpose of assessing obesity-specific HRQOL in school-aged children (5–13 years old) in a community sample of overweight and obese children.
Between the SMU and teasing/marginalization (a avoidance (a consisting of emotion (a .70), positive social attributes (a .68), and teasing/marginalization (a .71) scales. Correlations between the SMU and the PedsQL® (Varni et al., 2001) were significant (Total HRQOL, r = .52; Maximum estimated correlation, r = .85).

Despite these promising initial results, the literature on assessing weight-related HRQOL in children is scarce and a number of significant gaps remain. First, the initial SMU validation study was conducted exclusively in a treatment-seeking sample of children with obesity, consequently limiting the generalizability of the findings. As indicated by Zeller and Modi (2009), SMU requires application to a nontreatment-seeking sample in order to establish its generalizability to community samples of overweight and obese children. This is important for school- and community-based interventions for pediatric obesity, and to help inform policy decisions designed to improve the HRQOL of children with obesity and overweight generally. In addition, the original validation study was concerned with obesity-specific HRQOL only. It is not known how the measure functions as a general weight-related HRQOL instrument (i.e., including overweight children).

Second, Zeller and Modi (2009) used exploratory factor analysis for the initial validation study. Exploratory factor analysis is a data-driven approach in which all items are allowed to load on all constructs; as a result, constructs are formed by observing high loadings of individual items on a scale and relatively low loadings of the same item on other scales (Tabachnick & Fidell, 2001). Although exploratory factor analysis is often appropriate early in the measure development process, it does not provide a test of a priori hypothesized factor loadings of items on theoretically meaningful latent constructs. Therefore, confirmatory factor analytic techniques are considered superior to exploratory analyses because confirmatory factor analysis (CFA) is a theory-driven technique that tests the degree to which the data fit a specified model (Brown, 2006). This approach is considered a more strenuous test of a measure’s factor structure because individual items are forced to load on theoretically derived latent constructs, and the total model is evaluated based on its fit to observed patterns in the data (Brown, 2006).

To address these limitations, the primary aim of the current study was to examine the construct validity of SMU in a community sample of 4th and 5th grade overweight and obese children using CFA in a structural equation modeling (SEM) framework. This age of children was chosen because they represent a developmentally homogeneous group that can be reasonably expected to participate in group administration, and are available in all grade school systems (e.g., 6th grade starts Junior High for many children). The confirmation of the published factor structure using CFA will increase confidence in the measure’s underlying structure for use in nonclinical samples. Further, specification of the factor structure in a community sample allows future studies to apply the SMU measure in school-based interventions for overweight and obese children. This aim partially answers the question, “Are the scoring conventions established in a treatment-seeking sample appropriate for nontreatment-seeking samples?”. Thus, the first hypothesis of the current study was that a five-factor structure with a single second-order construct consistent with Zeller and Modi (2009) underlies SMU in a nontreatment-seeking sample.

Second, the criterion-related validity of SMU was assessed by specifying predictable associations between the five factors identified by Zeller and Modi (2009) and body mass index (BMI) percentile, and modeling the association between BMI percentile and each construct measured. The associations were entered into the model as regression paths using BMI percentile to predict each latent construct of the SMU measure. Building on other investigations of weight-related quality of life measures (Kolotkin et al., 2006), it was hypothesized that BMI percentile would be associated with poorer weight-related HRQOL as indicated by significant and positive associations between BMI percentile and the SMU physical scale, emotion scale, social scale, and the teasing and marginalization scale; and a significant negative association between BMI percentile and the positive social attributes scale. Overall, it was hypothesized that BMI percentile would be significantly and positively associated with the SMU single second-order factor (i.e., total score).

To further examine the construct validity of the measure, the current study tested the convergent validity of SMU by examining significant associations between latent SMU factors identified by Zeller and Modi (2009), and among the intercorrelations of the latent factors underlying
SMU and the PedsQL. By using a CFA framework to conduct this test, the current study advances the literature by evaluating the associations of theoretically similar constructs modeled without measurement error, thereby providing a purer estimate of the true intercorrelation between the constructs (Brown, 2006). Based on the correlations reported by Zeller and Modi (2009), it was hypothesized that all of the SMU factors are significantly and moderately correlated except for positive social attributes, which should only be correlated with social avoidance. Based on the widely reported theoretical position that disease-specific measures are only appropriate for children with a particular condition (e.g., Palermo et al., 2008), it was hypothesized that the measure would demonstrate different psychometric properties in normal weight children compared with overweight and obese children. Additionally, it was hypothesized that the SMU scales would demonstrate good convergent validity with the PedsQL scales as demonstrated by small-to-moderate positive associations between the physical SMU scale and the physical PedsQL scale, the emotion SMU scale and PedsQL emotion scale, the social avoidance SMU scale and the social PedsQL scale, the teasing and marginalization SMU scale and the social PedsQL, and the total scores of the PedsQL and SMU. Finally, the initial evaluation of SMU did not test the theoretical assumption that a weight-related HRQOL measure adds additional information beyond what is available from a broad and general measure. It was hypothesized that the SMU would demonstrate the incremental validity relative to the PedsQL in a nested models analysis applying model constraints to correlations between the two instruments.

Methods
Participants
Participants were 4th and 5th grade students enrolled in one of six Lawrence, Kansas public schools. Parental consent and child assent were obtained from 307 participant families. However, height and weight data of five children were not available. The final sample for analysis included 302 participants. For the purposes of the current study, participants were categorized into overweight and obese (n = 134) and healthy weight groups (n = 168). Healthy weight children were retained because the idea that weight-related HRQOL cannot be assessed in this sample is an assumption currently unexamined in empirical studies. Table I provides additional demographic data.

Sampling Procedure and Questionnaire Administration
Following receipt of approval from the University of Kansas Institutional Review Board, Unified School District 497, and respective institute/school principals and classroom teachers, a convenience sample of 4th and 5th grade students were recruited from six elementary schools in Lawrence, Kansas. Of those approached, 88.7% (n = 307) provided parental consent and participant assent to participate. As noted earlier, of the 307 children with consent (and assent), 302 (98.3%) completed all study measures and procedures. Participants were gathered in a cafeteria or classroom during a convenient time determined by school personnel. Survey packets containing study questionnaires and other instruments part of a larger evaluation of self-esteem, physical activity, and body image were distributed. Research assistants were available to read measures to self-identified students requiring accommodation.

Measures
Anthropometric Data
In the current study, data including children’s date of birth and height and weight were obtained from school records and were used to compute a BMI percentile score. Height and weight data were collected by a school nurse at a regular annual health screening. Anthropometric assessments occurred within approximately one month of questionnaire administration. Conversion of height, weight, sex, and age values to standardized BMI scores (BMI percentile)

<table>
<thead>
<tr>
<th>Demographics</th>
<th>BMI percentile ≥ 85 (n = 134)</th>
<th>BMI percentile &lt; 85 (n = 168)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>10.22 (SD = .69)</td>
<td>10.34 (SD = .76)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>56.0%</td>
<td>43.5%</td>
</tr>
<tr>
<td>Female</td>
<td>44.0%</td>
<td>56.5%</td>
</tr>
<tr>
<td>BMI percentile</td>
<td>94.4 (SD = 4.3)</td>
<td>50.6 (SD = 23.6)</td>
</tr>
<tr>
<td>Race/ethnicity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White not Hispanic</td>
<td>54.4%</td>
<td>66.6%</td>
</tr>
<tr>
<td>Black not Hispanic</td>
<td>7.5%</td>
<td>3.6%</td>
</tr>
<tr>
<td>Hispanic</td>
<td>6.7%</td>
<td>6.5%</td>
</tr>
<tr>
<td>Asian</td>
<td>8.2%</td>
<td>4.2%</td>
</tr>
<tr>
<td>American Indian</td>
<td>4.5%</td>
<td>6.5%</td>
</tr>
<tr>
<td>Other</td>
<td>14.9%</td>
<td>8.9%</td>
</tr>
<tr>
<td>Did not report</td>
<td>3.8%</td>
<td>3.7%</td>
</tr>
</tbody>
</table>
was performed using a SAS application provided by the Centers for Disease Control and Prevention (CDC, 2007). Higher BMI percentile is indicative of poorer health. Commonly used categorical definitions for normal weight (BMI percentile ≤ 84), overweight (BMI percentile ≥ 85), and obese (BMI percentile ≥ 95) were used (CDC, 2007).

General HRQOL
The PedsQL (Varni et al., 2001) is a 23-item questionnaire that measures self-reported HRQOL using questions designed to assess how much each item has been a problem for the child in the last month. The PedsQL uses an ordinal scale with anchors of never, almost never, sometimes, often, and almost always. Higher scores indicate better HRQOL. Previous studies of the PedsQL have found evidence for a four-factor solution comprising Physical (α = .80); Emotional (α = .73); Social (α = .71); and School (α = .68) HRQOL scales, as well as a unidimensional total score made up of the four subscales (Varni et al., 2001). The measure has demonstrated good internal consistency and appears to discriminate appropriately between well and ill groups (Varni et al., 2001; Varni et al., 2003).

Weight-Related HRQOL
Weight-related HRQOL was assessed using the 22-item SMU self-report questionnaire designed for use with 5- to 13-year-old obese children (Zeller & Modi, 2009). The instrument is made up of items that orient the participant to the health-related component of the assessment by asking how much the item is true during the past month “...because of your size.” Participants respond to the questionnaire using an ordinal scale with anchors of none of the time (1), a little (2), a lot (3), and all the time (4). Higher scores indicate better HRQOL. As noted earlier SMU has acceptable reliability estimates and evidence of convergent validity with the PedsQL in a treatment seeking sample. The SMU has a Flesch–Kincaid readability index score of 2.1. The SMU is a copyrighted instrument. However, the entire measure is available as an online supplement associated with the original validation study (Zeller & Modi, 2009).

Missing Data
To achieve a complete dataset, the Expectation Maximization algorithm was used in the PRELIS program bundled with Lisrel 8.8 (Jöreskog & Sörbom, 2006). Imputation using the Expectation Maximization algorithm is a method of stochastic imputation considered to be consistent with the best statistical practices in applied psychology (Schlomer, Bauman, & Card, 2010). Due to the small amount of missing data (1.18%) and the assumption that data were missing completely at random only a single imputation was necessary (Schafer 1999).

Results
Data Screening
Initial data screening revealed that the variables included in both SMU and the PedsQL were not normally distributed. Specifically, several items were significantly positively skewed. Therefore, the data were modeled using Robust Maximum Likelihood (RML) and evaluated using the Satorra–Bentler χ² scaled test of model fit. Additionally, data screening revealed that the PedsQL physical scale included one item with no variance (i.e., hard to take bath or shower). That is, all participants responded that this was “never” a problem. Therefore, this item was eliminated from all analyses using the PedsQL.

Overview of Analyses
The null and alternative models were specified using LISREL 8.8 (Jöreskog & Sörbom, 2006). Because the manifest data collected in the current study were ordinal and skewed, the polychoric correlation matrix with an asymptotic covariance matrix was analyzed in all structural models. All models were evaluated by examining the Satorra–Bentler χ² test of significance, comparative fit index (CFI), non-normed fit index (NNFI), and the root mean squared error of approximation (RMSEA). Model fit was considered to be acceptable if the CFI and NNFI were above .90 and the RMSEA was below .1. For nested model comparisons, χ² change tests were considered significant at the p < .05 level (Cheung & Rensvold, 2002). An a priori decision was made to include all SMU items in the final models even if factor loadings were not significant. This decision was made because the current study was focused on questions of generalizability and theoretical significance. Measure revision was not a goal of the current project.

The original five-factor model of the SMU was initially assessed in a subsample of both children with overweight and children with obesity before assessing factorial invariance across these groups. This decision was to capitalize on statistical power and to promote ease of data management. After establishing the factor structure, factorial invariance was assessed across normal weight and a combined group of children with overweight and obesity. The nine-factor model (SMU subscales and PedsQL subscales) was tested only in overweight and obese children.

All reliability statistics were taken from the final model with nine first-order and two second-order factors and are calculated using the formula \[ \rho = \frac{(\Sigma \lambda_i)^2}{[(\Sigma \lambda_i)^2 + (\Sigma \theta_i)^2]} \]
(where $\lambda$ = the unstandardized factor loadings and $\theta$ = the unstandardized error terms) as this method provides an estimate of the true scale reliability in a CFA framework (Raykov, 2004). Because this measure of internal consistency is a representation of true scale reliability, readers familiar with Cronbach’s alpha can easily interpret the values using the same cutoffs commonly applied to Cronbach’s alpha (i.e., acceptable $\geq .70$). In the case of the second-order factors, error terms are replaced with the error variance of the first-order factors and the factor loadings are replaced with the disattenuated loadings of first-order factors on the second-order factor (Ping, 2004). This method of reliability estimation is free from many of the biases that are present in traditional methods of estimating scale reliability (e.g., Cronbach’s alpha; Raykov, 2004).

Finally, a Minimal Clinically Important Difference (MCID) score was calculated to give clinicians a guideline for the magnitude of change in a given scale that would indicate that a change in treatment was clinically relevant (Jaeschke et al., 1989). Consistent with other studies of HRQOL assessment in the literature, the MCID was calculated by taking the square root of the result of one minus the internal consistency (i.e., $\rho$) times the standard deviation (i.e., Standard Error of Measurement; Varni et al., 2003; Wyrwich, Tierney, & Wolinsky, 1999).

**Primary Analyses**

As noted earlier, the current study aimed to evaluate the construct validity of SMU in a community sample of 4th and 5th grade children with overweight and obesity using a CFA in an SEM framework. Specifically, this aim addresses convergent validity both within SMU factors and between SMU factors and similar factors on the PedQL. Additionally, criterion validity was assessed by examining the association between BMI percentile and SMU factors.

**Null Model**

The null model for overweight and obese children was specified as all 22 SMU items with error terms freely estimated and factor loadings fixed to 0.0. BMI percentile and sex were included as exogenous variables specified by a single indicator to allow for regression tests using these variables in the alternative model. This is the mathematical equivalent of the statement, “SMU items do not measure latent constructs and capture only error.” The null model demonstrated poor fit to the data, $\chi^2 (265, n = 134) = 1890.00$, $p < 0.001$, RMSEA = .219. The Aim 1 alternative model was assessed using this null model.

**Alternative Model**

Also in the sample of children with overweight and obesity, the five-factor model of SMU was estimated using the RML estimator with a ridge constraint of 1.0. Results of the five factor solution revealed close fit to the data, $\chi^2 (236, n = 134) = 311.58$, $p < 0.001$, RMSEA = .049, CFI = .96, NNFI = .95. The Satorra–Bentler $\chi^2$ improved significantly compared with the null model, $\Delta \chi^2 = 1578.42$, $p < .05$. All of the lambda loadings were significant, except for item number 8 “Stood up for or helped other kids because of your size.” Overall, results indicated that Zeller and Modi’s (2009) factor structure is adequate for use in community samples of overweight and obese children.

Using the same null and alternative model, participants were grouped into overweight ($n = 56$) and obese ($n = 78$) groups. Factor loadings, intercepts, and mean level information was constrained across overweight and obese groups in sequential steps. Based on the RMSEA test (i.e., “Does the 90% RMSEA confidence interval overlap with the 90% RMSEA confidence interval of the alternative model?”; Little, 1997) all three constraints were tenable, indicating that the factor structure of the SMU is the same across both overweight and obese groups. Therefore, the same scoring conventions are appropriate across overweight and obese groups.

Next, children with overweight and obesity were once again combined into one grouping. Using the same null and alternative model identified, responses from the 168 healthy weight subjects were entered as an independent group to test the assumption of factorial invariance across healthy and unhealthy weight categories. The loadings and intercepts were constrained across healthy weight and overweight groups in sequential steps. Based on the RMSEA test, both constraints were untenable (90% CI configural model = .00-.034, 90% CI weak invariance model = .429-.889, 90% CI strong invariance model = .0590-.0745). These tests indicate that neither the factor loadings nor intercepts of SMU are meaningfully similar across the two groups. The previous results suggest that scores should not be compared across children with healthy weight and children with unhealthy weight. In many cases, this finding was a result of limited variation at the item level for normal weight children. Although some variance was observed for all items, it was small enough in most cases to cause model misfit. Therefore, normal weight children were excluded from all subsequent analyses, and children with overweight and obesity were grouped together.

Sex and BMI percentile were included in the model as independent exogenous latent variables predicting all five
of the previously specified endogenous latent variables. None of the paths from sex to any of the other latent constructs were significant. Partial support for the hypothesis that BMI is associated with SMU factors was observed as: (a) the regression path from BMI percentile to the physical SMU scale was small but significant \((\gamma = .19)\), (b) the path from BMI percentile to the SMU social scale was small but significant \((\gamma = .13)\), and (c) the path from BMI percentile to the SMU teasing and marginalization scale was small but significant \((\gamma = .11)\). Therefore, criterion validity was established with three of the five first-order SMU scales.

**Nine-Factor Null Model**

To assess the associations between SMU and the PedsQL, a null model was specified with the (combined) 44 items from SMU and the PedsQL. As stated previously, one PedsQL item was excluded due to restriction of range. The null model was specified as all 44 manifest variables with error terms freely estimated and factor loadings fixed to 0.0. The null model evidenced poor fit to the data, \(\chi^2 (1016, n = 134) = 6489.77, p < 0.001, \text{RMSEA} = .201\). All nine-factor alternative models were assessed using this null model.

**Nine-Factor Alternative Model**

The nine-factor model including both SMU and the PedsQL was estimated using the RML estimator with a ridge constraint of 1.0. As in the previous model, BMI percentile and sex were allowed to enter the model freely. Only the three paths identified as significant in the evaluation of SMU alone remained significant. Nonsignificant regression paths were pruned from the final nine-factor model. Results of the final nine-factor solution revealed acceptable fit to the data, \(\chi^2 (974, n = 134) = 1453.50, p < 0.001, \text{RMSEA} = .061, \text{CFI} = .91, \text{NNFI} = .91\), and was used to examine the convergent validity between the HRQOL scales.

**Convergent Validity Between HRQOL Scales**

To assess convergent validity between the two HRQOL scales, latent intercorrelations between SMU factors and PedsQL scales were estimated as well as latent correlations between SMU factors. The hypothesis of convergent validity was partially supported. The correlation between the physical \((\psi = .22)\) scales of the SMU and the PedsQL was positive and significant. Similarly, the association between the PedsQL social scale and SMU teasing and marginalization scale \((\psi = .27)\) was positive and significant. Hypothesized intercorrelations between the SMU emotional and social avoidance scales and the PedsQL social scale were not significant.

**Second-Order Factor Structure**

To test the overall construct validity of weight-related HRQOL and determine the utility of a total score for SMU, second-order weight-related HRQOL and general HRQOL constructs were specified. Weight-related HRQOL was made up of the five latent constructs derived from SMU. General HRQOL was made up of the four latent constructs derived from the PedsQL (representing the PedsQL total score). The two-factor second-order model demonstrated close fit to the data, \(\chi^2 (892, n = 134) = 1171.66, p < 0.001, \text{RMSEA} = .049, \text{CFI} = .95, \text{NNFI} = .94\) (Figure 1). Similar to the results

---

**Figure 1.** Second-Order Factor Structure, \(\chi^2 (892, n = 134) = 1171.66, p < 0.001, \text{root mean squared error of approximation} = .049, \text{comparative fit index} = .95, \text{non-normed fit index} = .94\).
from the first-order structure, sex was not associated with either second-order factor, and BMI percentile was significantly associated with the weight-related HRQOL factor (i.e., SMU total score; \( \gamma = .11 \)) and the general HRQOL total score (i.e., PedsQL total score; \( \gamma = .09 \)). Due to a relatively low loading of the positive social attributes scale (\( \gamma = -.53 \)), a three-factor higher-order model, with positive social attributes as a unique second-order factor, was tested to ensure that the five SMU scales represent a unitary construct. This model demonstrated significantly worse fit (\( \Delta \chi^2 = 3.86, p < .05 \)), was less parsimonious than the two-factor model, and was rejected. Therefore, the two-factor higher-order model appears to be most appropriate, given the current data.

The HRQOL assessment literature typically assumes that disease-specific and general quality of life are distinct constructs that should be measured independently (Palermo et al., 2008). To address this assumption statistically, similar constructs across SMU and the PedsQL were constrained to 1.0, using nested model comparisons. The observation of model misfit after a constraint was applied, was considered to be an indication that overweight and obesity confers a unique impairment on HRQOL over and above that conferred by broad and general HRQOL. In the current analysis, the nine-factor model with significant BMI percentile regressions was used. To allow for testing of nested models, the measurement model for the current test included nonsignificant estimates for the latent constructs that should be measured independently. The hypothesis that SMU assesses unique weight-related HRQOL constructs was supported. Specifically, when the latent correlations of each corresponding pair of scales for the SMU and PedsQL were constrained to 1.0 the resulting nested model \( \chi^2 \) comparison indicated that the constraint was untenable due to significant change in the \( \chi^2 \) statistic (SMU physical and PedsQL physical, \( \Delta \chi^2 = 129.35, p < .05 \); SMU emotion and PedsQL emotion, \( \Delta \chi^2 = 226.38, p < .05 \); SMU social avoidance PedsQL social, \( \Delta \chi^2 = 143.43, p < .05 \); SMU teasing and marginalization and PedsQL social, \( \Delta \chi^2 = 9.5, p < .05 \)). Finally, to test the hypothesis that weight-related HRQOL and general HRQOL are distinct constructs, the latent second-order measurement model was used, and the latent correlation between the two higher-order HRQOL constructs was fixed to 1.0. The nested model \( \chi^2 \) comparison indicated that the constraint was untenable due to a significant change in the \( \chi^2 \) statistic (\( \Delta \chi^2 = 30.05, p < .05 \)). In sum, these findings indicate that the constructs measured by SMU are distinctly different from the constructs assessed by the PedsQL.

### Transformed Means and Standard Deviations

To aid with interpretation, the PedsQL and SMU were each transformed to a 0 to 100 scale, as recommended by their respective authors. Scaled means and standard deviations are shown in Table II. Mean comparisons between overweight and obese groups were not possible due to low statistical power.

One method of establishing cut-off scores for population level HRQOL measures is to subtract one standard deviation from the total mean score (Varni et al., 2003). Following a similar procedure, the current SMU mean score minus one standard deviation (78.30 – 10.93 = 67.37) was almost identical to the mean reported in Zeller and Modi’s (2009) initial validation study (~68) of treatment-seeking overweight or obese children.

### Reliability of First and Second Order Factors and Minimal Clinically Important Difference

Reliability statistics (calculated as \( \rho \) to provide factor reliability statistics) for the current study were marginally acceptable for all of the first-order factors except for the SMU positive social attributes and teasing and marginalization scales (Table III). However, the total score reliability estimate was much higher for both the SMU and PedsQL total scores than for their subscales. MCIDs are provided for all the factors examined in the nine-factor and

---

**Table II. Scaled Means and Standard Deviations by Weight Category**

<table>
<thead>
<tr>
<th>Scale</th>
<th>Overweight (n = 56)</th>
<th>Obese (n = 62)</th>
<th>Very obese (n = 16)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>SMU physical</td>
<td>92.78</td>
<td>12.00</td>
<td>89.30</td>
</tr>
<tr>
<td>SMU emotional</td>
<td>86.94</td>
<td>22.85</td>
<td>84.60</td>
</tr>
<tr>
<td>SMU social avoidance</td>
<td>93.38</td>
<td>14.02</td>
<td>93.19</td>
</tr>
<tr>
<td>SMU positive social</td>
<td>49.03</td>
<td>17.59</td>
<td>49.01</td>
</tr>
<tr>
<td>SMU Teasing/marginalization</td>
<td>94.66</td>
<td>12.32</td>
<td>93.60</td>
</tr>
<tr>
<td>SMU total</td>
<td>79.82</td>
<td>11.25</td>
<td>78.34</td>
</tr>
<tr>
<td>PedsQL physical</td>
<td>90.94</td>
<td>9.92</td>
<td>88.88</td>
</tr>
<tr>
<td>PedsQL emotional</td>
<td>79.29</td>
<td>20.75</td>
<td>78.95</td>
</tr>
<tr>
<td>PedsQL social</td>
<td>87.59</td>
<td>16.33</td>
<td>86.45</td>
</tr>
<tr>
<td>PedsQL school</td>
<td>80.71</td>
<td>17.25</td>
<td>76.37</td>
</tr>
<tr>
<td>PedsQL total</td>
<td>85.20</td>
<td>11.28</td>
<td>83.23</td>
</tr>
</tbody>
</table>

Note: The 0-100 point scales for each measure are derived by linearly transforming the item level data such that 0 indicates poorer HRQOL and 100 indicates higher quality HRQOL. Very obese = BMI ≥ 99th percentile; very obese children were classified as obese for purposes of the CFA and are separated here for descriptive purposes only.
second-order models; however, relatively lower reliability estimates for the first-order factors make the second-order MCIDs the most meaningful estimate of clinically significant changes. Previous data are not available for MCID for SMU; however, the total score MCID identified for the PedsQL is consistent with previous reports (e.g., Varni et al., 2003).

**Discussion**

This study fills a gap in the HRQOL assessment literature by evaluating the construct validity of SMU in a non-treatment-seeking sample of 4th and 5th grade children with overweight and obesity. Findings from the CFA confirmed the SMU five-factor first-order structure with one second-order factor previously proposed by Zeller and Modi (2009) among treatment-seeking children with obesity. Moreover, construct validity was partially established for the SMU scales and total score. Criterion validity was established for the physical, social avoidance, teasing and marginalization scales, and the total score. This lends evidence that these scales assess latent variables that are significantly associated with weight-related health and should fluctuate with changes in BMI.

Generally, results support the use of the SMU in community samples of overweight and obese children. This was an important finding because it speaks to the generalizability of the measure in the context of research studies. Greater generalizability in measurement allows for easier transportation of research findings in treatment-seeking populations to community-based interventions. Adding to the evidence for the construct validity of the SMU, convergent validity was established with the PedsQL for the SMU physical scale, teasing and marginalization scale, and total score. Departing from the associations discovered by Zeller and Modi (2009), the current investigation did not observe significant associations between the SMU and PedsQL emotion scales or the social avoidance and social scales. This may be because the degree of impairment in the current sample was not sufficient to elicit the associations found in a treatment-seeking sample. As noted earlier, the initial validation sample was composed entirely of treatment-seeking children with obesity and had poorer mean HRQOL on the SMU than the current sample (Zeller & Modi, 2009). Alternatively, by removing measurement error, the current analysis produced a more accurate true score estimate (Brown, 2006), and it may be that these scales do not share an association in latent space. Additional studies are needed to determine definitively whether these correlations are limited to samples of obese children or whether they reliably disappear when measurement error is removed from the analysis.

The results of the invariance test comparing overweight and obese children to healthy weight children revealed that the assumption of factorial invariance does not hold across these two distinct groups. The results of the current study do not support the same SMU scoring conventions in children with a healthy weight due to a different underlying measurement structure. This is consistent with leading HRQOL assessment theory suggesting that disease-specific HRQOL measures are not appropriate for population-level assessments or prevention interventions; for these purposes, broad and general HRQOL measures are the appropriate choice (Palermo et al., 2008). By way of explanation, it may be that the items of SMU are not understood in the same way by healthy weight children because they do not have the experience of being overweight or obese. Much the same way that a healthy child would not be expected to report on cancer-specific, cystic fibrosis-specific, or inflammatory bowel disease-specific quality of life, healthy weight children may not have experiences that relate meaningfully to items assessing the clinical conditions of overweight and obesity.

Statistically demonstrating that SMU adds incrementally to the understanding of a theoretical construct is an important strength of this study, and has several implications. First, the current study provides empirical support for the widely espoused belief that disease-specific instruments add additional information to a HRQOL assessment over and above broad and general measures (Palermo et al., 2008). Second, when considering markers of clinical progress, children should be assessed using the
The descriptive statistics reported in the current study combined with Zeller and Modi’s (2009) initial evaluation suggest that 68 is a reasonable total score that could prompt the psychologist or pediatrician to have a conversation with the child and their family about the way weight is impairing the child’s HRQOL. In an initial discussion with the child and his/her family, the practitioner may encounter resistance to their suggestions for lifestyle change. In this scenario, the health care provider may be able to use the SMU items to point out areas where the child experiences impairment. The professional may then be able to use these items to call attention to the impairment and effectively increase motivation for behavior change. At follow-up, the baseline total score yielded by the SMU instrument can be assessed for changes of approximately three or more points (i.e., MCID) to determine the impact of the lifestyle changes for that particular child.

Limitations and Future Directions

The findings of the current study are limited by several factors. First, the SMU is designed for use among 5- to 13-year-old children (Zeller & Modi, 2009). However, the current study sample was limited to children aged 8 to 12 years. Therefore, the findings may not generalize to all participants who could respond to the measure. Although one of the strengths of the current study was the use of a nontreatment-seeking sample, the current results cannot address treatment-seeking samples. It is possible that subtle, as of yet, untested differences exist between how the instrument behaves in community-based and treatment-seeking samples. Therefore, additional confirmatory work with the SMU is necessary in treatment-seeking samples including CFA. In addition, although we believe that nontreatment-seeking is an appropriate description of our sample, we did not screen for participation in weight-management treatment. Therefore, it is possible that a small number of children were receiving treatment for their overweight or obesity. Finally, while the racial distribution of the current sample is representative of the geographic region in which the study was conducted, it does not reflect the frequencies of ethnic/racial group memberships in unhealthy weight categories observed in the obesity literature (Ogden, Carroll, Curtin, Lamb, & Flegal, 2010).

As noted earlier, additional confirmatory work is needed to enhance the understanding of the factor structure of SMU. First, future validation studies of SMU should attempt to determine whether a wider range of children can participate in group administrations of the instrument. This would provide information regarding whether the instrument can be used in studies of entire grade schools,
potentially informing community or school-based interventions. Second, a CFA of SMU is still needed in a treatment-seeking sample that closely resembles Zeller and Modi’s (2009) original sample. Additionally, there is still cause for concern that the positive social attributes scale does not constitute a HRQOL factor, and may measure a unique construct. A CFA in a treatment-seeking sample that tested a two-factor versus a one-factor second-order model would add confidence to the assertion that the measure assesses a unitary weight-related HRQOL construct. The current study did not assess test-retest reliability. Similarly, longitudinal invariance studies are needed to demonstrate that the instrument is stable across measurement occasions in the same sample. This information will assure researchers interested in using the measure in the context of longitudinal work or an intervention study that the instrument will not be subject to measurement fluctuations as a function of repeated administrations or the passage of time.

Conclusion

In conclusion, the SMU yields a reliable and valid total score for use in community samples of overweight and obese children. However, it is not appropriate for children with a healthy weight. Consistent with the larger theoretical literature on HRQOL assessment (Palermo et al., 2008), the SMU total score appears to be a precise measure of HRQOL in children with overweight or obesity. SMU offers researchers and clinicians more specific information about a child’s experience of overweight or obesity and a total score that should demonstrate meaningful change at smaller intervals than the PedsQL. It is recommended that clinicians interested in HRQOL among children with overweight or obesity use the SMU total score as part of their assessment battery. Researchers interested in assessing HRQOL in school-aged children should continue to follow the evidence-based assessment recommendations and include both a broad and general HRQOL instrument as well as an obesity-specific measure such as the SMU scale (Palermo et al., 2008).

Acknowledgment

The authors acknowledge Dr. Meg H. Zeller for her helpful comments on an earlier draft of this manuscript.

Funding

This project was made possible by funding from the University of Kansas general Clinical Child Psychology Program research fund, Pioneer Class Dissertation Award, and University of Kansas Graduate College Dissertation Fund.

Conflicts of interest: None declared.

References


