Comparison of Two Measures of Weight Criticism in Youth: Associations With Physical Activity Engagement and Attitudes, Weight Status, and Health-Related Quality of Life

Laurie A. Gayes, MA, and Ric G. Steele, PhD, ABPP
Clinical Child Psychology Program, University of Kansas

All correspondence concerning this article should be addressed to Ric G. Steele, PhD, ABPP, 2010 Dole Human Development Center, 1000 Sunnyside Avenue, Lawrence, KS 66045, USA. E-mail: rsteele@ku.edu

Received February 28, 2014; revisions received and accepted August 26, 2014

The objective of this study was to examine the degree to which 2 measures of weight criticism, the Weight Criticism During Physical Activity (WCA) scale and the Perceptions of Teasing Scale (POTS), represent distinct constructs and in what circumstances each is most appropriately used. A community sample of 307 fourth and fifth graders completed these measures, as well as measures of health-related quality of life, physical activity engagement, and attitudes toward physical activity. Body mass index was also calculated. Results of confirmatory factor analysis indicated that the WCA scale and the POTS represented correlated but distinct constructs and related differently to measures of physical activity and weight status. Findings suggested that the WCA scale may be representing criticism regarding athletic competency, not criticism of weight status as described in the literature. The POTS subscales appear to be accurately described and used in the literature.

Key words obesity; peers; psychosocial functioning; structural equation modeling.

Peer victimization is a common experience for many elementary school youth (Nansel et al., 2001), and for many of these children, victimization is a chronic problem with potentially damaging effects. Numerous studies have shown that child peer victimization is associated with emotional suffering that surpasses simple social distress (Hawker & Boulton, 2000). Instead, the impact of peer victimization manifests as compromised psychological and emotional functioning, including depression (Kaltiala-Heino, Rimpela, Marttunen, Rimpela, & Rantanen, 1999), body dissatisfaction (Jensen & Steele, 2009), loneliness (Hawker & Boulton, 2000), poor coping skills (Faith et al., 2002), and low self-esteem (Hawker & Boulton, 2000).

Overweight and obese children are particularly at risk for experiencing pervasive and damaging stigmatization by peers (Gray, Kahhan, & Janicke, 2009). Children who are overweight or obese are particularly susceptible to weight-related criticism, a type of peer victimization targeting weight or body shape. Rates of peer victimization and weight-related criticism are twice as high as for overweight youth as compared with normal weight peers (Hayden-Wade et al., 2005) and results in perceptions of negative personal attributes (e.g., lazy, stupid; Davison & Birch, 2004), increased dislike from classmates (Kraig & Keel, 2001), and fewer friendship nominations from peers (Strauss & Pollack, 2003). Weight-related criticism has also been linked to behavioral consequences such as greater preference for isolative and sedentary behaviors (Hayden-Wade et al., 2005), greater perceived barriers to engagement in physical activity (Gray et al., 2009), and unhealthy dieting behavior (Thompson et al., 2007), including binge-eating behavior (Neumark-Sztainer et al., 2002).

Despite the recognized importance of identifying and intervening on weight criticism in youth, there appears to be little guidance from the literature about how to best select an appropriate measurement tool. This lack of guidance is particularly noteworthy considering a recent meta-analysis found that type of teasing measure used is a
statistically significant moderating factor on the relationship identified between weight-related teasing (WRT) and the outcome variable (e.g., body dissatisfaction, dietary restraint; Menzel et al., 2010), suggesting that it is essential to carefully select the appropriate measure to suit the empirical question. The Weight Criticism During Physical Activity (WCA) scale (Faith, Leone, Ayers, Heo, & Pietrobelli, 2002) and the Perceptions of Teasing Scale (POTS; Thompson, Cattarin, Fowler, & Fisher, 1995) have both been designed to understand and track weight-related criticism experienced by youth. As described in more detail below, the POTS is a measure with two subscales commonly used for capturing weight-related criticism. However, the WCA scale is a less well-known measure that may capture a unique perspective on weight-related criticism by specifically targeting teasing that occurs during physical activity, a time when weight issues may be particularly salient.

Despite the emergent research on these two measures, the current literature is not clear as to how these measures relate to one another and in what circumstances or with which populations each should be used. Although the POTS weight criticism subscale and the WCA scale are moderately correlated with each other (r = .40; Faith et al., 2002), there has not been a direct comparison of how they differentially relate to physical activity outcomes or how the two subscales of the POTS correlate separately with the WCA scale. The moderate correlation between the two measures suggests that they might represent related but distinct concepts. The present study is designed to build on previous literature to clarify the relationship between weight-related criticism and the WCA scale in a sample of preadolescents.

To address this gap in the literature, our first aim for this study was to conduct a confirmatory factor analysis (CFA) using structural equation modeling (SEM) with the POTS and the WCA scale to determine the degree to which the measures represent distinct constructs. To our knowledge, there has been only one published comparison of the POTS and the WCA scale (Faith et al., 2002). This study builds on the finding of a moderate correlation between the two measures by analyzing the relationship between the measures within an SEM framework, providing an estimation of the correlation between measures without the potentially distorting presence of measurement error. Despite their similar descriptions in the literature, we hypothesized that our CFA would reveal only moderate latent correlations between the two measures, with stronger relationships between the two measures of weight criticism than with the measure of general competency.

Our second aim was to examine the degree to which the measures are invariant, or consistently valid, across males and females and across overweight and not-overweight weight status groups. Sex differences have been identified in relation to teasing in the literature (Faith et al., 2002; Jensen & Steele, 2009), and owing to these measures targeting weight issues, it seemed important to test for differences in responses due to weight status. SEM was used to test measurement invariance at the level of factor structure, loadings, and means to determine the degree of equality in response to the measures. We hypothesized that measurement invariance across neither sex nor weight status would be tenable.

Finally, the current study aims to assess the construct validity of the measures by modeling their association with other constructs related to physical activity and weight status, including health-related quality of life, attitudes toward physical activity, engagement in physical activity, and body mass index (BMI). It was hypothesized that both measures of teasing would correlate with these outcomes, but that the WCA scale (due to it being a more targeted measure of teasing around physical activity) would be associated more negatively with physical activity and attitudes toward physical activity than either subscale of the POTS. It was further hypothesized that both the WCA scale and the POTS weight criticism subscales would correlate with BMI, but that the competency subscale of the POTS would not. Finally, it was hypothesized that each measure of teasing would be associated with lower health-related quality of life.

Method

Participants

Participants included 307 fourth- and fifth-grade elementary school children recruited through a Midwestern public school district. Eligibility criteria included (a) the child was enrolled in either fourth or fifth grade, (b) the student spoke and read English, and (c) the child’s parent or custodial caregiver provided informed consent for participation. The recruitment process included sending 558 consent forms home to parents, of which 354 were returned, 330 indicated consent, and 307 children participated in data collection. Demographic characteristics of the sample and means (and standard deviations) of all study variables are presented in Table 1.

Procedures

A description of the study and a consent form were sent home to all parents of fourth and fifth graders in six schools, selected to provide a diverse sample of children.
Gayes and Steele

frequency with which they have experienced the forms of sports or exercise (e.g., ‘’People make fun of you when you play sports or exercise’’; Faith et al., 2002). Children rate the frequency with which they have experienced the forms of teasing on a scale from 1 (never) to 5 (very often) and the degree to which the teasing bothered them on a scale from 1 (not upset) to 5 (very upset) within a timeframe of kindergarten until the present. This measure is scored by summing the intensity and frequency score for each item. The scale was originally tested with fifth- through eighth-grade students. Internal consistency for this measure has been reported to be good (α = .83; Faith et al., 2002), and scores have been shown to (negatively) correlate with mild-intensity leisure activity (Faith et al., 2002). In the present sample, internal consistency was estimated at α = .76.

Perceptions of Teasing Scale
The POTS is an extensively used measure of teasing in children (Janicke et al., 2007; Jensen & Steele, 2010; Stern et al., 2007), and contains both a six-item measure of WRT (e.g., ‘’People made fun of you because you were heavy’’) and a five-item measure of competency-related teasing (CRT; e.g., ‘’People said you acted dumb’’). Children rate the frequency (from kindergarten until the present) with which they experienced the forms of teasing on a scale from 1 (never) to 5 (very often) and the degree to which it bothered them on a scale from 1 (not upset) to 5 (very upset). This scale is scored by summing the intensity and frequency scores for each item. The scale has been demonstrated to have good internal consistency (α = .94) and has been validated against psychological measures related to teasing (Faith et al., 2002; Jensen & Steele, 2010). The measure was originally developed with undergraduate university students, but has been used with children as young as fifth-grade students (Faith et al., 2002). In the present sample, alpha for the WRT and CRT subscales was .87 and .75, respectively.

Physical Activity
Physical activity was assessed using the Fels Physical Activity Questionnaire (PAQ; Treuth, Hou, Young, & Maynard, 2005). The measure was developed for use with preadolescents. This eight-item questionnaire contains three “open” questions in which participants list activities and frequency of participation. The remaining five questions use a Likert scale to evaluate physical activity during leisure time, transportation, and chores at home. The Fels PAQ has been reliably compared with objective accelerometry measures of physical activity (Treuth et al., 2005).

Body Mass Index
BMI z score is the preferred measure for evaluating obesity among children and adolescents (Krebs et al., 2007). BMI z

| Table I. Demographic Characteristics of the Sample and Means and Standard Deviations of All Study Variables |
|-----------------|-----------------|-----------------|
| Variable | Percentage | Mean (SD) |
| Age | 9.85 (0.72) |
| Sex | | |
| Male | 49.35% |
| Female | 50.64% |
| BMI | 20.25 (4.54) |
| Race/ethnicity | | |
| White, not Hispanic | 64.7% |
| Black, not Hispanic | 5.2% |
| Hispanic | 6.5% |
| Asian | 4.9% |
| American Indian | 3.2% |
| Other | 13.3% |
| Did not report | 1.6% |
| POTS | | |
| Weight criticism | 1.61 (1.39) |
| Competency | 2.85 (1.66) |
| WCA | 2.44 (1.48) |
| Attitudes | | |
| Negative | 2.11 (0.61) |
| Positive | 3.60 (0.63) |
| HRQOL | 85.08 (11.80) |
| Physical activity | 9.04 (2.12) |

Note. For this sample, average BMI z score was 0.751 (SD = 1.09). Mean of WCA intensity scores was 1.61 (SD = 0.16), and mean of WCA frequency scores was 2.27 (SD = 0.12). Mean of POTS weight intensity scores was 1.23 (SD = 0.08) and mean of POTS weight frequency was 2.44 (SD = 0.23). Mean of POTS competency intensity scores was 1.82 (SD = 0.20) and frequency was 2.27 (SD = 0.22). Means and standard deviations for the teasing measures were higher than to those reported by Faith et al. (2002), also in a community sample (i.e., POTS girls = 0.33, POTS boys = 0.33; WCA girls = 1.94, WCA boys = 1.87).

To incentivize return of consent forms (no matter whether consent was given), the mascot of the authors’ university visited schools if at least 80% of the forms were returned. Participating students completed research measures in the cafeteria or classroom with help from research assistants, who read questions aloud. The measures pertaining to this study were nested within a larger evaluation of psychosocial variables associated with physical activity and weight. These procedures were approved by the human subjects review board at the authors’ university.

Measures
Weight Criticism During Physical Activity Scale
Criticism during physical activity was assessed using the WCA scale. This six-item measure describes various examples of teasing that children may experience during physical activity (e.g., ‘’People make fun of you when you play sports or exercise’’; Faith et al., 2002). Children rate the frequency with which they have experienced the forms of teasing on a scale from 1 (never) to 5 (very often) and the degree to which the teasing bothered them on a scale from 1 (not upset) to 5 (very upset) within a timeframe of kindergarten until the present. This measure is scored by summing the intensity and frequency score for each item. The scale was originally tested with fifth- through eighth-grade students. Internal consistency for this measure has been reported to be good (α = .83; Faith et al., 2002), and scores have been shown to (negatively) correlate with mild-intensity leisure activity (Faith et al., 2002). In the present sample, internal consistency was estimated at α = .76.

Perceptions of Teasing Scale
The POTS is an extensively used measure of teasing in children (Janicke et al., 2007; Jensen & Steele, 2010; Stern et al., 2007), and contains both a six-item measure of WRT (e.g., ‘’People made fun of you because you were heavy’’) and a five-item measure of competency-related teasing (CRT; e.g., ‘’People said you acted dumb’’). Children rate the frequency (from kindergarten until the present) with which they experienced the forms of teasing on a scale from 1 (never) to 5 (very often) and the degree to which it bothered them on a scale from 1 (not upset) to 5 (very upset). This scale is scored by summing the intensity and frequency scores for each item. The scale has been demonstrated to have good internal consistency (α = .94) and has been validated against psychological measures related to teasing (Faith et al., 2002; Jensen & Steele, 2010). The measure was originally developed with undergraduate university students, but has been used with children as young as fifth-grade students (Faith et al., 2002). In the present sample, alpha for the WRT and CRT subscales was .87 and .75, respectively.

Physical Activity
Physical activity was assessed using the Fels Physical Activity Questionnaire (PAQ; Treuth, Hou, Young, & Maynard, 2005). The measure was developed for use with preadolescents. This eight-item questionnaire contains three “open” questions in which participants list activities and frequency of participation. The remaining five questions use a Likert scale to evaluate physical activity during leisure time, transportation, and chores at home. The Fels PAQ has been reliably compared with objective accelerometry measures of physical activity (Treuth et al., 2005).

Body Mass Index
BMI z score is the preferred measure for evaluating obesity among children and adolescents (Krebs et al., 2007). BMI z
score was calculated using the formula defined by the CDC (BMI = kg/m²) through a program provided by the CDC (CDC, 2007a). Student height and weight measurements were collected by public school officials during the first academic term as part of a district-wide mandatory health assessment. Conversion of height and weight values to BMI z score values was performed using an SAS program provided by the CDC (CDC, 2007b).

**Attitudes Toward Physical Activity**
Attitudes toward physical activity were assessed using a self-report measure designed to examine children’s beliefs about the consequences of being physically active (Physical Activity Attitudes Questionnaire [PAAQ]); (Motl et al., 2000; Nelson, Benson, & Jensen, 2009). This measure is composed of an eight-item positive subscale reflecting positive attitudes toward physical activity and an eight-item negative subscale representing negative attitudes toward physical activity. All items follow the stem “If I were to be physically active on most days...” Children rate various positive (e.g., “...it would help me cope with stress”) and negative (e.g., “...it would be boring”) completions of this stem on a 5-point scale with endpoints of 1 (disagree a lot) and 5 (agree a lot). Both subscales have demonstrated acceptable reliability and validity (Motl et al., 2000; Nelson et al., 2009). In the current sample, the alpha coefficient for the positive and negative subscales was .69 and .75, respectively.

**Health-Related Quality of Life: PedsQL 4.0 Generic Core Scales**
The PedsQL 4.0 Generic Core Scales (PedsQL) is a 23-item self-report measure of health-related quality of life (HRQoL). The measure yields scores on four subscales: physical, emotional, social, and school functioning. These scores where then entered into SEM analyses as domain representative parcels. The PedsQL has been demonstrated to have good reliability and validity. Internal consistency statistics are consistently >0.70 (Varni, Seid, & Kurtin, 2001). In this sample, the alpha coefficient was estimated at .88.

**Analytic Procedures**
SEM was used to analyze our research questions. SEM is a well-regarded technique that allows researchers to flexibly and precisely analyze relationships among latent variables (Kline, 2005). Mplus version 7 (Mutheén & Mutheén, 2012) was used to conduct all analyses. Robust maximum likelihood was used as the estimator for analyses. Analyses were conducted in two stages. First, a CFA was conducted to assess for discriminant validity between the WCA weight criticism subscale, the POTS competency subscale, and the WCA scale, as well as to assess for factorial invariance across between high- and low-weight status groups and between sexes (i.e., to determine whether these groups responded to items in the same manner). For each construct, parcels were created. Parceling is recommended because parcels have higher reliability, higher ratio of common-to-unique variance, as well as more, tighter, and more equal intervals than items (Little, Cunningham, Shahar, & Widaman, 2002). First, each factor was examined with all items loading separately. Once the significance of all item loadings was assured, parcels were formed using the serpentine method based on the loading structure of the items so that parcels are constructed with items of balanced weight. To set the scale, the latent variances were set equal to 1.0. Full information maximum likelihood was used to handle missing data.

At each stage of the CFA, model fit was assessed by examining the fit statistics RMSEA, CFI, TLI/NNFI, and change in $\chi^2$. According to interpretation guidelines suggested by Little (2013), the RMSEA was considered a poor fit if the value was >0.10, a mediocre fit if the value fell between 0.10 and 0.08, an acceptable fit if the value fell between 0.08 and 0.05, a good/close fit if it fell between 0.05 and 0.02, and a great fit if the value was <0.01. Similarly, Little (2013) suggests that CFI values and TLI/NNFI <0.85 represent a poor fit, values 0.85–0.90 represent a mediocre fit, values 0.90–0.99 represent an acceptable fit, values 0.95–0.99 represent a very good fit, and values >0.99 represent an outstanding fit. Model fit was also assessed by testing whether the difference in $\chi^2$ values between nested models was statistically significant. For each analysis, sample size exceeded the number of parameters estimated in the model, yielding reliable estimates (Kline, 2005).

The second stage of analyses was to examine the latent regressive relationships between the two measures of teasing and other physical activity and weight-related constructs. A structural model was run incorporating the WCA scale, the two POTS subscales, HRQoL, positive and negative attitudes toward physical activity, BMI, and physical activity. For the PAAQ, parcels were created using a serpentine method based on the loading structure of the items so that parcels are constructed with items of balanced weight. For the HRQoL and physical activity constructs, facet representative parcels were created based on the four preexisting subscales in the measure. BMI consisted of a single-item indicator. To set the scale, the latent variances were set equal to 1.0.

**Results**
The first aim of these analyses was to assess the degree to which the WCA scale and POTS subscales represent
distinct constructs. To address this question, the factor structure of the WCA scale, POTS weight criticism subscale, and POTS competency subscale was examined in a CFA. These three measures formed three latent constructs with no cross-loadings. The initial freely estimated measurement model with item parcels demonstrated good model fit $[\chi^2 (24, N = 308) = 65.974, p \leq .0001$, RMSEA $= 0.075$, CFI $= 0.968$, NNFI/TLI $= 0.952]$. The correlation between the two POTS subscales was .54, the correlation between the WCA scale and the POTS weight criticism subscale was .46, and the correlation between the WCA scale and the POTS competency subscale was .71. Without parcels, the model fit was reduced according to the CFI, but strengthened according to the RMSEA $[\chi^2 (116, N = 308) = 199.570, p \leq .0001$, RMSEA $= 0.048$, CFI $= 0.913$, NNFI/TLI $= 0.899]$. These moderate correlations suggest that these measures are related but distinct and that the WCA scale and the POTS competency scale are highly related.

The second aim was to test establish factorial invariance across several dimensions. First, invariance was tested between a high-weight status group (i.e., BMI > 85th percentile) and a low-weight status group (i.e., BMI ≤ 85th percentile). The WCA scale, POTS weight criticism scale, and POTS competency scale were examined separately owing to differences in their descriptions for how they relate to weight status. The results demonstrated that each scale passed invariance testing (Table II). Results also indicated mean differences between groups only on the POTS weight criticism measure, where the high-weight status group demonstrated significantly higher weight criticism scores than the low-weight status group. Next, invariance across sex was examined on a model with the three scales together, as no differences were expected between the measures on their relationship with sex. The initial freely estimated configural model demonstrated acceptable model fit $[\chi^2 (48, N = 308) = 128.072, p \leq .0001$, RMSEA $= 0.104$, CFI $= 0.941$, NNFI $= 0.911]$. The test of weak invariance also demonstrated acceptable model fit $[\chi^2 (54, N = 308) = 136.831, p \leq .0001$, RMSEA $= 0.100$, CFI $= 0.939$, NNFI $= 0.918]$. The test of strong invariance also demonstrated acceptable model fit $[\chi^2 (60, N = 308) = 146.381, p \leq .0001$, RMSEA $= 0.097$, CFI $= 0.936$, NNFI $= 0.923]$. As no significant change in model fit resulted, strong invariance was established.

Our third aim was to examine the relationships between the measures of teasing and other constructs related to weight and physical activity to examine the construct validity of the teasing measures. A model including the two POTS subscales, the WCA scale, HRQoL, positive and negative attitudes toward physical activity, physical activity engagement, and BMI (a single-item indicator) was tested to examine these associations. First, a CFA was run between the WCA scale, POTS weight criticism subscale, POTS competency subscale, HRQoL, positive attitudes toward physical activity, negative attitudes toward physical activity, physical activity, and BMI without any regressive paths specified. This model demonstrated acceptable model fit $[\chi^2 (248, N = 308) = 473.086, p \leq .0001$, RMSEA $= 0.054$, CFI $= 0.911$, NNFI $= 0.893]$. Next, nonsignificant regressive paths were then pruned, and model fit did not change significantly $[\chi^2 (262, N = 308) = 486.329, p \leq .0001$, RMSEA $= 0.053$, CFI $= 0.911$, NNFI $= 0.899]}$.

In this final model (see Figure 1), the WCA scale and the POTS subscales were significantly correlated with one another. The WCA scale was significantly positively correlated with negative attitudes toward physical activity and significantly negatively correlated with HRQoL. The POTS weight criticism subscale was significantly positively correlated with physical activity and BMI and significantly negatively correlated with HRQoL. The POTS competency subscale was significantly positively correlated with positive attitudes toward physical activity and significantly negatively related to HRQoL. These findings further suggest that the POTS and the WCA scale do not assess the same underlying construct.

Lastly, several exploratory analyses were conducted to examine the strength of regressive relationships and indirect effects. When the relationship between the POTS weight criticism scale and the WCA scale was constrained to be equal to the relationship between the POTS competency scale and the WCA scale, model fit was significantly reduced $[\chi^2 (263, N = 308) = 500.502, p \leq .0001$, RMSEA $= 0.054$, CFI $= 0.906$, NNFI $= 0.893]$, confirming that the WCA scale is more similar to the competency scale than the weight criticism scale. Finally, two indirect pathways were tested for statistical significance. The indirect effect of the POTS competency subscale on physical activity through negative attitudes was statistically significant (estimate $= 0.099$, $p = .014$), and the indirect effect of the WCA scale on physical activity through negative attitudes approached significance (estimate $= -0.048$, $p = .057$).

1 A model was assessed without BMI and model fit did not change significantly. Regressive paths became stronger, and the relationship between WCA and physical activity became statistically significant.
Discussion

The present study was designed to contribute to the measurement of WRT by examining the associations between the POTS (Thompson et al., 1995) and the WCA scale (Faith et al., 2002) in a large community sample, and to examine associations among POTS and WCA subscales and other variables related to physical activity. Contrary to hypotheses and descriptions in the literature, a CFA revealed that the WCA scale and the POTS weight criticism subscale represented distinct constructs. To our knowledge, this is the first analysis of the POTS and the WCA scale in an SEM framework, which removes measurement error and produces a more precise true score estimate than traditional statistical techniques (Kline, 2005). This finding is similar to that of Faith and colleagues (2002), who found that the POTS weight criticism subscale and the WCA scale also had a moderate correlation ($r = .4$). Furthermore, the WCA scale was statistically more strongly related to the POTS competency subscale than the weight criticism subscale.

### Table II. Model Fit Statistics for the Tests of Invariance for the Teasing Measures Over High and Low Weight Status Groups

<table>
<thead>
<tr>
<th>Model tested</th>
<th>$X^2$</th>
<th>$df$</th>
<th>$p$</th>
<th>RMSEA</th>
<th>RMSEA 90% CI</th>
<th>CFI</th>
<th>TLI/NNFI</th>
<th>Pass?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>WCA</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Configural invariance</td>
<td>1.270</td>
<td>5</td>
<td>0.938</td>
<td>0.000</td>
<td>0.000–0.026</td>
<td>1.000</td>
<td>1.013</td>
<td>Yes</td>
</tr>
<tr>
<td>Weak invariance</td>
<td>0.551</td>
<td>2</td>
<td>0.759</td>
<td>0.000</td>
<td>0–0.109</td>
<td>1.000</td>
<td>1.012</td>
<td>Yes</td>
</tr>
<tr>
<td>Strong invariance</td>
<td>0.832</td>
<td>4</td>
<td>0.934</td>
<td>0.000</td>
<td>0–0.032</td>
<td>1.000</td>
<td>1.014</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>POTS weight criticism</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Configural invariance</td>
<td>4.612</td>
<td>5</td>
<td>0.737</td>
<td>0.000</td>
<td>0.000–0.113</td>
<td>1.000</td>
<td>1.009</td>
<td>Yes</td>
</tr>
<tr>
<td>Weak invariance</td>
<td>4.080</td>
<td>2</td>
<td>0.130</td>
<td>0.083</td>
<td>0–0.2</td>
<td>0.996</td>
<td>0.987</td>
<td>Yes</td>
</tr>
<tr>
<td>Strong invariance</td>
<td>4.166</td>
<td>4</td>
<td>0.384</td>
<td>0.017</td>
<td>0–0.125</td>
<td>1.000</td>
<td>0.999</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>POTS competency</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Configural invariance</td>
<td>4.599</td>
<td>5</td>
<td>0.467</td>
<td>0.000</td>
<td>0.000–0.108</td>
<td>1.000</td>
<td>1.002</td>
<td>Yes</td>
</tr>
<tr>
<td>Weak invariance</td>
<td>3.213</td>
<td>2</td>
<td>0.200</td>
<td>0.063</td>
<td>0.000–0.186</td>
<td>0.994</td>
<td>0.982</td>
<td>Yes</td>
</tr>
<tr>
<td>Strong invariance</td>
<td>3.565</td>
<td>4</td>
<td>0.468</td>
<td>0.000</td>
<td>0.000–0.117</td>
<td>1.000</td>
<td>1.003</td>
<td>Yes</td>
</tr>
</tbody>
</table>

### Figure 1. Pruned model depicting latent regressive paths. $\chi^2 (262, N = 308) = 486.329, p \leq .0001$ (RMSEA = 0.053, CFI = 0.911, NNFI = 0.899).
subscales, further suggesting that the construct the WCA scale is representing is somewhat different than its description in the literature. Further testing of validity confirmed that the WCA scale and the POTS weight criticism scale do not relate to other variables in the same manner, reaffirming that, though related, these measures do not represent the same construct.

The examination of construct validity, or how these measures correlate in latent space with other established measures of related constructs, sheds light on the true constructs these measures assess. When examining the relationship between these measures and BMI, it is first important to discuss that factorial invariance was established with regard to high- and low-weight status groups, suggesting that both measures represented the same constructs across weight status. However, the POTS weight criticism scale was the only scale of the three to have the high-weight status group demonstrate significantly higher scores than the low-weight status group. This finding suggests that both weight status groups conceptualize weight criticism in the same manner but experience different levels of it. Weight status groups experienced nonsignificant differences in criticism according to the WCA scale and the POTS competency scale, suggesting a lack of relation with BMI, which was confirmed through analysis of structural paths.

Although the lack of correlation between the POTS competency scale and BMI may not be surprising, the WCA scale is named and described as a measure of weight criticism—suggesting that it should be associated with weight status. However, our findings suggest that the WCA scale may represent a type of competency criticism instead of weight criticism. The item content and measure structure support this concept, as the items include criticism of physical skills during sports or exercise, not being chosen for a sports team, and being made fun of while playing sports. It is conceivable that factors unrelated to weight status could lead to the perception of teasing during physical activity, such as clumsiness or lack of knowledge of the sport. Furthermore, the scale is constructed so that frequency of criticism contributes to the total score. Potentially, youth with high-weight status engage in physical activity infrequently, and as a result, the criticism that they experience is not being fully captured. With these findings in mind, the WCA scale is not recommended for assessing weight criticism and should be used instead to measure teasing regarding athletic competence.

In contrast, the POTS weight criticism subscale seems to be sensitive to weight status, as it was positively correlated with BMI in latent space, suggesting that the POTS is accurately named and described in the literature. Our findings in this regard is consistent with recent work that has reported significant positive relationships between the POTS weight criticism subscale and weight status (Faith et al., 2002; Hayden-Wade et al., 2005; Nelson et al., 2011) but inconsistent with other studies that have found no relation between BMI and the POTS weight criticism subscale (Jensen & Steele, 2011; Quinlan, Hoy, & Costanzo, 2009). The reasoning behind this discrepancy remains unclear, but out of these studies, the study by Quinlan et al. was the only study that assessed only overweight and obese youth, perhaps suggesting that the degree of overweight does not significantly impact the relationship between BMI and weight criticism within an overweight population. Future studies should further examine this question.

Due to differences between sexes on reports of teasing in the literature (Tiggemann & Anesbury, 2000), examination of measurement equivalence across sexes was critical. These series of tests assessed for factorial invariance, which examines whether males and females respond to the measures in the same manner. Factorial invariance has been established for the POTS previously (i.e., Jensen & Steele, 2010), but has not been previously established for the WCA scale. Factorial invariance using parcels was tentatively established here between sexes, suggesting that the measures were interpreted similarly between males and females. Furthermore, means were not significantly different between sexes, indicating that perceptions of teasing were similar between males and females, inconsistent with some previous studies (Faith et al., 2002; Jensen & Steele, 2009). Likely because this study used a community sample of youth who were mostly of normal weight, low levels of weight criticism and competency teasing experienced in this population may have obscured sex differences that have been detected elsewhere. It is also possible that sex differences emerge later in development.

Examination of latent regressive paths provides novel information about teasing and related physical activity variables. Contrary to hypotheses, only the POTS weight criticism subscale was directly related to physical activity engagement. Interestingly, the mediation pathway of the POTS competency subscale through positive attitudes to physical activity was statistically significant and the pathway of the WCA scale through negative attitudes to physical activity approached significance. This pattern of results regarding attitudes toward physical activity is consistent with current theoretical models of behavior and decision making (e.g., Social Cognitive Theory, Theory of Planned Behavior). However, the POTS weight criticism subscale was not related to attitudes toward physical activity.
suggesting that weight criticism does not follow this same pathway of influence on behavior. It is conceivable that weight criticism influences behavior through a less cognitive and more emotional pathway, as weight-based victimization has been shown to yield strong emotional responses in youth (Puhl & Luedicke, 2012). These findings are also consistent with recent research finding that positive and negative attitudes toward physical activity are related, but distinct, constructs (Nelson et al., 2009). Overall, our results and the extant literature suggest that both positive and negative attitudes toward physical activity are important variables to include in future studies analyzing physical activity in youth.

Our findings suggest that, for a community sample, athletic competency criticism may be the form of teasing most damaging to physical activity, as the WCA scale was significantly related to negative attitudes toward physical activity, and both POTS subscales were related to positive attitudes toward physical activity or related directly to physical activity in a positive direction. Contrary to predictions, the POTS weight criticism scale was positively correlated in latent space to physical activity engagement. This finding is reminiscent of other work that has suggested that moderate levels of dissatisfaction are associated with improved outcomes (Johnston, Moreno, Regas, Tyler, & Foreyt, 2012), perhaps by inspiring the desire to change the current status, but not creating so much discomfort that change feels impossible or overwhelming. Future longitudinal studies should more thoroughly examine the associations among these measures and physical activity.

Our findings suggest that teasing during childhood is associated with poorer health-related quality of life and has implications for a child’s attitudes toward and engagement in physical activity. This consistently strong negative relationship between teasing and HRQoL is important to highlight in relation to the finding that in some cases higher teasing scores were associated with better findings (i.e., more physical activity, higher positive attitudes toward physical activity), demonstrating that teasing is not an appropriate means by which to achieve these outcomes. Given the poor outcomes associated with teasing found by this study and many others, it seems appropriate to identify levels of teasing experienced by an individual and target teasing in interventions. Such assessment tools could be used in interventions such as universal school-based interventions that have shown an increasing evidence base for decreasing bullying behaviors and increasing social and emotional learning (Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011). Our findings also suggest that there are multiple distinct types of teasing present within different populations and contexts. Interventions should be sensitive to this range of forms of teasing and seek to target the issue in many contexts.

Our findings suggest that the POTS and the WCA scales may be appropriate for differing populations and clinical questions. The POTS weight criticism subscale seems to be most appropriate for identifying perceived criticism due to overweight and obesity, while the POTS competency subscale seems most appropriate for identifying perceived teasing regarding abilities generally, and the WCA scale seems best suited for assessing teasing regarding athletic abilities. A review of the literature did not reveal other validated measures to assess overweight criticism in youth, highlighting the importance of clearly describing the POTS and the WCA scale in the literature.

Several limitations to this study should be considered. First, as mentioned previously, examination of more complex relations among variables requires longitudinal assessment. The potential pathway of attitudes toward physical activity mediating the relationship between teasing and physical activity is worthy of continued exploration in a longitudinal design. Second, this study used largely self-report measures. Objective indicators of physical activity may yield more precise results. Furthermore, this study was conducted with a community sample of youth with a range of weight statuses. Future research should examine the degree to which these same relationships among variables are present and whether sex relationships emerge when tested with an entirely overweight or obese sample. Further research is needed to better understand how teasing interacts with physical activity and other related variables to impact weight status over time. Another limitation lies in the lack of detail and consistency in the literature as to scoring the WCA scale and the POTS (e.g., using only the frequency items, scoring all items). It is possible that differences in scores between studies represent not differences between samples, but differences in measurement. Lastly, the lack of significant diversity within the racial/ethnic composition of the sample limits generalizability.

Teasing appears to be a multidimensional concept encompassing criticism in a range of situations and over a number of topics. Overall, teasing in youth is associated with a range of negative outcomes, including poor health-related quality of life, increased negative attitudes toward physical activity and decreased positive attitudes toward physical activity, and increased weight status. These findings suggest, however, that measurement tools of teasing must be carefully matched to the research and clinical question to capture the presence of teasing appropriately. Precise and accurate measurement of teasing is the first
step toward creating effective interventions for this traumatic childhood event.

Conflicts of interest: None declared.

References


Nansel, T. R., Overpeck, M. D., Pilla, R. S., Ruan, W. J., Simons-Morton, B., & Scheidt, P. (2001). Bullying...
behaviors among US youth: Prevalence and association with psychosocial adjustment.