

Critique of the Piers-Harris Children's Self-Concept Scale II

Chapman University

## Critique of the Piers-Harris Children's Self-Concept Scale II

### **Description of the Test**

The current paper aims to discuss and analyze the Piers-Harris Children's Self-Concept Scale II (Piers-Harris 2), authored by Ellen V. Piers and David S. Herzberg and published by WPS. The first version of the Piers-Harris Children's Self-Concept Scale was developed in 1963 to measure self-concept in youth. The original authors defined self-concept as a consistent set of attitudes about one's own behavior and attributes that are reflected through descriptions and evaluations (Piers & Herzberg, 2002). In 2002, Piers and Herzberg developed the Piers-Harris 2, which included a new normative data sample, computerized administration and scoring tools, and revisions to test items. The Piers-Harris 2 is a self-report questionnaire comprised of 60 items within six domain subscales. It is intended to be administered to children and adolescents aged seven to 18. In 2018, Piers, Shemmassian, and Herzberg released the Piers-Harris Children's Self-Concept Scale III. However, the current critique only evaluates the contents and properties of the Piers-Harris 2.

### **Content and Use**

#### **Practical Information**

The Piers-Harris 2 includes a manual with descriptions of the contents, background information, and instructions on how to administer, score, and interpret the assessment. The computerized version of the measure includes an answer sheet, the necessary software for scoring and interpretation, and an AutoScore form to supplement the software. A paper-and-pencil version of the test includes the answer sheet and the AutoScore form. The Piers-Harris 2 manual can no longer be purchased from the publisher due to the release of the latest version. A pack of 100 answer sheets is available for purchase in English for \$25.00 and in Spanish for

\$29.00. Lastly, the AutoScore Form can be purchased in a pack of 20 for \$63.00. The entire Piers-Harris 2 can be administered to a child in an estimated 10 to 15 minutes. However, there is no time limit on the assessment, and a child may utilize as much time necessary to complete the test with accuracy. The assessment can be administered to children individually or in groups as needed. The manual suggests that students complete the assessment in a room that is free of distractions while seated comfortably at a desk or table that is appropriate for writing.

### **Characteristics of the Test**

The Piers-Harris 2 is recommended for use with youth who have at least 2<sup>nd</sup> grade reading abilities. Further, the authors suggest that the test instructions and items be read aloud by the administrator for children in the 2<sup>nd</sup> to 4<sup>th</sup> grades. This can aid in the accuracy of responses because younger children will not have to focus on reading and sounding out low-frequency words in the text (e.g., “disappointed”, “important”). The Piers-Harris 2 is not recommended for administration with children who are unwilling or unable to participate. The authors state that children who are uncooperative, prone to exaggerating answers, or have unorganized thoughts may not accurately reflect their self-concept (Piers & Herzberg, 2002). However, the children who will benefit most from being tested or screened using the Piers-Harris 2 are likely to be referred for reasons related to maladaptive internalizing and externalizing behaviors. The exclusion statement made by the authors may discourage professionals from using the Piers-Harris 2 to measure self-concept with certain populations of children that can benefit from it.

The Piers-Harris 2 manual gives clear and concise instructions on how to administer the test. The instructions for the questionnaire should be read aloud by the test administrator. The instructions use language that is easy for children to understand, such as referring to test items as “sentences”. Additionally, children are encouraged to answer based on “the way [they] really

feel inside” rather than the way they think they should feel, or the way others think of them (Piers & Herzberg, 2002, pg. 7). The manual offers guidance on how to respond to children who are having a difficult time choosing between the yes and no anchors. The authors suggest that assessors prompt children to choose the answer that best reflects the way they usually feel (Piers & Herzberg, 2002).

The paper version and the computerized version of the Piers-Harris 2 questionnaire are presented differently. The formatting of the computerized version is more organized and easier to follow because the test items are clearly differentiated from the identifying information section. On the written form, students are required to fill in their identifying information just as they do for the test items. All test forms are titled “The Way I Feel About Myself”, which is an age-appropriate phrase to describe to self-concept (Piers & Herzberg, 2002). Finally, the Piers-Harris 2 asks students to identify their race/ethnicities with only six available options (Asian, Hispanic, White, Black, Native America, and Other). These options are limited in accurately representing children of diverse racial and ethnic backgrounds.

The initial item set in the original Piers-Harris included 80 items and was condensed to 60 items for the Piers-Harris 2. Items were removed if their wording was outdated, they were too specific to one sex, or they contained words that children were unfamiliar with (Piers & Herzberg, 2002). The anchors are dichotomous, with yes responses reflecting items that are true or mostly true for the child, and no responses reflecting items that are false or mostly false for the child. The test includes reverse scoring items (e.g., “I am smart” and “I am dumb about most things”) and similar items that measure the same concepts in different ways (e.g., “I am good in my schoolwork” and “I am a good reader”) (Piers & Herzberg, 2002). Reverse scored items are expected to have opposite responses, and similar items are expected to have the same responses.

These test construction components are imperative to include to reduce response biases. Measuring self-concept requires subjective scoring by the child, meaning the responses rely on personal impressions rather than objective, observable qualities (Salvia, Ysseldyke, & Bolt, 2013). In general, self-concept measures are limited because they require a self-report on an individual's belief about himself, which does not necessarily represent his actual behaviors. There may be a discrepancy between self-concept and behaviors that can be evaluated through further data collection.

Six domain subscales were developed to measure self-concept in the Piers-Harris 2. There are 14 items that measure behavioral adjustment (BEH), 16 items that measure intellectual and school status (INT), 11 items that measure physical appearance and attributes (PHY), 14 items that measure freedom from anxiety (FRE), 12 items that measure popularity (POP), and 9 items that measure happiness and satisfaction (HAP). Some test items belong to multiple domains, and Piers and Herzberg (2002) considered labeling items under one domain that they best fit in, but there was a significant drop in internal consistency when this was tested. While the domains are presented with examples from the test, the terms they represent are not defined or described in this section. Including definitions for each term would be beneficial for users of the Piers-Harris 2 to understand how these domains relate to self-concept.

### **Characteristics of the Test Manual**

The Piers-Harris 2 manual includes information on administering, scoring, and interpreting the assessment. Chapter 1 provides an introduction, general description, and some improvements made from the Piers-Harris 1. Chapter 2 details the instructions on how to administer and score the assessment and provides a sample computerized AutoScore Form. This chapter gives clear information on how to deliver the test instructions and how to score the

computerized form. Chapter 3 includes guidelines for interpretation (e.g., standard scores, scales, and item responses). Chapter 4 informs readers about the assessment's restandardization from the Piers-Harris 1 to the Piers-Harris 2, item development, and scale construction. Lastly, chapter 5 includes technical properties (e.g., reliability, validity) and future directions.

In general, the Piers-Harris 2 test manual is not clearly organized. Whereas the table of contents gives an overview of the chapter contents, some detailed information is located in different chapters. For example, information on reliability can be found in chapter 4 (development and restandardization) and chapter 5 (technical properties). Additionally, the subscale domains are listed in the introduction, but are not defined until chapter 3. The manual includes contents of the Piers-Harris 2 while also giving significant attention to the changes made from the Piers-Harris 1. The organization of the information related to the changes is not clear for readers to follow. Similarly, the self-concept rationale and theory are not specified until chapter 4, which makes it difficult for readers to understand the background of the assessment concepts for the first three chapters.

### **Standardization Sample and Norms**

An entirely new normative data sample was used for the Piers-Harris 2 revision. The original standardization sample was recruited from a rural Pennsylvania public school system in the 1960s, which was very limited for generalization purposes. The Piers-Harris 2 developers improved their normative sample by basing their participant demographic information on the 2000 US Census data. The new sample consisted of 1,387 elementary, middle, and high school students from various backgrounds and geographic regions in the United States.

The sample was divided into age groups ranging from seven to 18 years, which reflects the age range designed for the test. Age ranges were divided into six groups, beginning with 7-8-

year old's, and ending with 17-18-year-old participants. All age groups consisted of at least 100 participants, which is suitable for research purposes. However, the test developers did not specify how many participants represent a single age within the age group. This may be a limitation because students of specific ages may not be represented. Sex was adequately represented in the normative sample when compared to 2000 Consensus data, with 49.7% identifying as male, and 50.3% identifying as female.

The new standardization sample was further characterized by a variety of race and ethnic backgrounds. Whereas some groups were accurately represented by the 2000 Census data (e.g., 18.4% of the normative sample and 14.7% of the Census sample identified as Black), other groups were highly overrepresented (e.g., 68% of the normative sample and 60.9% of the Census sample identified as White) and underrepresented (e.g., 7.4% of the normative sample and 17.1% of the Census sample identified as Hispanic/Latino). Similarly, some United States geographic regions were represented in the standardization sample, but they do not accurately represent the 2000 Census data. The represented regions are categorized as Northeast, Midwest, South, and West, and the procedure for grouping states in each of these regions is not described. In the normative sample, 30.6% of participants are from the Midwest (while the 2000 Census reports 22.9%) and 13.3% are from the West (while the 2000 Census reports 22.5%).

Finally, the Piers-Harris 2 standardization sample included head of household education level as an indicator of socio-economic status. However, only 673 participants reported on education level, which is not representative of the entire sample. Additionally, test users should interpret results related to this sample norm with caution because education level and socio-economic status are not synonymous entities. In the normative sample, the 4-year college degree or more education level is highly overrepresented (47.3%) when compared to the 2000 Census

data (28.7%). This factor can limit the test's generalizability because the sample is geared towards students whose parents or caretakers have a college degree. The remaining standardization sample education levels include: some college (14.9%), high school graduate (27%), and less than high school graduate (10.8%).

### **Scores and Interpretation**

The Piers-Harris 2 manual provides detailed instructions for scoring the assessment. First, the authors describe how to score the computerized version of the Piers-Harris 2 using the AutoScore Form. Once a child completes the assessment, professionals are instructed to scan the form for any incomplete items or items that have both yes and no items circled. These items should be omitted when calculating the raw score. If the assessment contains seven or more invalid responses, the manual suggests that the test should not be scored or interpreted.

A scoring sheet and profile sheet are used for scoring both written and computerized tests. The scoring sheet is used to calculate raw scores by allocating one point for the yes or no response that is consistent with higher self-concept, and zero points for the response that reflects lower self-concept. To calculate the total raw score, professionals are instructed to add and record the number of responses scored as one point. Similarly, the raw scores for the six domains are calculated on the scoring sheet by summing the total points for items within a subscale.

An Inconsistent Responding (INC) index and a Response Bias (RES) index are included on the scoring sheet. The INC index can be utilized to detect random response patterns and to evaluate inter-scale correlations, defined as the relationship between items on the same subscale (Piers & Herzberg, 2002). The INC index presents 15 pairs of items that are expected to correlate (e.g., a child who reports being worried a lot is also expected to report being afraid often). Users are instructed to add the number of pairs that do not correlate to evaluate the INC index raw



score. A raw score of four or more on the INC index may indicate random responding. The RES index assesses for response bias, defined as the tendency to agree or disagree with a test item for a reason unrelated to the item content (Piers & Herzberg, 2002). The RES index is a count of the number of yes responses; a higher RES score indicates a positive response bias tendency and a lower RES score indicates a negative response bias tendency. A RES score of greater than 40 or less than 18 flags users of response bias concerns.

The Piers-Harris 2 profile sheet assists users in converting raw scores to standard scores that can be interpreted in relation to the standardization sample. The current assessment converts the total and six domain raw scores into t-scores and percentile ranks. The Piers-Harris 2 T-score, or standard score, transformation involves converting the raw scores into a normal distribution with a mean of 50 and standard deviation of 10. The manual states that transformed t-scores that fall within the 40 to 60 range are considered average when compared to the normative sample.

Percentile ranks are also used in the Piers-Harris 2 to demonstrate the percentage of the normative sample that scored at or below the current child's score (Piers & Herzberg, 2002). Generally, percentile ranks are simpler scores for families to understand when interpreting a child's assessment results. However, percentile ranks are often misinterpreted as percent correct or interval data. Therefore, users must ensure that they are presenting and describing percentile ranks accurately to families and other professionals. Appendix D of the manual displays a table for converting raw scores into T-scores and percentile ranks. One major limitation of the Piers-Harris 2 scoring system is the omission of confidence intervals within score transformation procedures. Confidence intervals are essential for interpreting standard scores because they take

measurement error into consideration. Thus, the Piers-Harris 2 scores should be interpreted with caution because the amount of error that exists in the measurement is unknown.

The Piers-Harris 2 manual classifies T-scores into three categories of interpretation. T-scores of 60 and above for the total self-concept scale are considered to be in the high range. T-scores of 40 to 59 are in the average range, and t-scores of 39 or below are classified in the low range. T-scores for the domain subscales are also categorized in the above average range ( $t \geq 56$ ), the average range ( $t = 40-55$ ), and the low range ( $t \leq 39$ ). The authors offer implications for characteristics associated with high, average, and low range scores. For instance, the manual states that children with scores in the high range for the total self-concept scale are typically confident in many domains, are likely to have competent relationships with family and peers, and generally view themselves as happy. Conversely, children with total scores in the low range are described as doubtful of their self-worth, viewing themselves as less competent than their peers, and may be in need of clinical attention for anxiety and depression.

These general characteristics associated with differing scores may be beneficial in viewing results in a realistic manner. However, assigning definitive qualities to children's scores may also be detrimental. Professionals are required to interpret students' scores with fidelity based on the individualized data they gathered, and not based on a general description of qualities they are expected to have. It is unclear where the test developers retrieved the information to make these conclusions about high, average, and low range scores. Therefore, professionals would benefit from using them only as general descriptions to reference, and not as guidelines for interpretation of individual students' self-concept. Additionally, the Piers-Harris 2 manual suggests that assessment results should be interpreted with other sources of information

(e.g., parent interviews, child observations) to make appropriate intervention recommendations for children with varying self-concept scores. (Piers & Herzberg, 2002)

### **Psychometric Properties**

#### **Reliability**

The Piers-Harris 2 manual defines reliability as a psychological test's score stability. The manual also states that a reliable test should result in consistent scores over time and should be free of measurement error. The test developers determined reliability by using measures of internal consistency and test-retest from both versions of the Piers-Harris standardization sample data. (Piers & Herzberg, 2002)

**Internal Consistency.** According to the Piers-Harris 2 manual, internal consistency refers to how well a subdomain is sampled by the test items. Internal consistency can also be interpreted as the consistency in a test-taker's response pattern. The original Piers-Harris did not measure internal consistency but relied on a sample of nine additional studies that reported internal consistency estimates of the test's TOT score. These studies used alpha scores and the split-half method, described as splitting the test into halves and measuring how one half correlates to the other. The Kuder-Richardson 20 (KR 20) method was also used, which is an appropriate measure for the Piers-Harris because it is an estimate for dichotomous data. All nine studies reported alpha scores ranging from 0.88 to 0.93, however, no information was presented on these studies' samples and whether they are similar to the Piers-Harris standardization sample. (Piers & Herzberg, 2002)

The Piers-Harris 2 measured internal consistency estimates on the normative sample using alpha scores for the TOT scale and the six subdomains. The TOT estimated alpha score was .91, which is adequate for clinical and educational purposes. The estimated alpha scores for

the six domains ranged from 0.74 to 0.81. For subscales with a limited number of items, a 0.70 alpha estimate is adequate. However, test users should be cautious when using domain scale scores as resources for educational decision making because they include a substantial amount of measurement error. More specifically, Piers-Harris 2 users should be aware that the POP ( $\alpha = 0.74$ ), PHY ( $\alpha = 0.75$ ) and HAP ( $\alpha = 0.77$ ) subscale internal consistency estimates are the least reliable.

Internal consistency estimates were also provided for the six age groups in the standardization sample. Most age groups, with the exception of the 7- to 8-year-old and 17- to 18-year-old groups, received an alpha TOT score estimate greater than 0.90. As previously stated, an alpha score of 0.90 is adequate for clinical purposes, and the group sample norms falling below this score estimate are less reliable. In addition to the TOT scale, the six subdomains were tested for internal consistency estimates by age group. Most age group subdomains received an alpha estimate of 0.70, with the exception of the PHY 17- to 18-year-old group ( $\alpha = 0.65$ ), the POP 7- to 8-year-old group ( $\alpha = 0.60$ ), and the POP 17- to 18-year-old group ( $\alpha = 0.62$ ). These scores contained the most measurement error and may not be sufficient resources for the interpretation of individual Piers-Harris 2 scores.

The Standard Error of Measurement (SEM), defined by Piers and Herzberg (2002), is an index of how closely a test score is likely to be the true score that would be obtained without measurement error restrictions. Essentially, the SEM is the average amount of error that exists in a measure. A SEM score of 3 or less is considered reliable for a measure. The Piers-Harris 2 TOT scale had a SEM of 3.07, which meets the minimum requirement to be considered reliable.

**Test-retest Reliability.** Test-retest is a measurement of temporal error designed to assess the stability of scores over time. The Piers-Harris 2 test developers stated that self-concept is

relatively stable over time, but less stable in children, and a retest is recommended beyond two weeks following an initial test. Perhaps the greatest limitation of the Piers-Harris 2 psychometric properties is the lack of test-retest reliability evidence. The current authors relied on test-retest coefficients from the original Piers-Harris, although items were changed from the original test to the second edition, and the standardization sample was entirely new. The Piers-Harris 1 test-retest measures will be discussed, but any generalizations in interpretation to Piers-Harris 2 results may not be reliable.

The original Piers-Harris reported on multiple test-retest studies conducted in the 1970s-1990s. One study evaluated the standardization sample's temporal error and stability of over time. The retest was administered four months after the initial test, and had a coefficient of  $r=0.72$ , which meets the minimum requirement of  $r=0.60$  for the retest's time interval. However, it is still important to note for interpretation purposes that only half (48%) of the measurement's error is accounted for.

The nine remaining studies met the minimum requirements to be considered reliable for their test-retest interventions, and their coefficients ranged from  $r=0.65$ - $0.96$ . Interestingly, the study that reported one of the lowest test-retest scores ( $r=0.69$ ) was the study with the shortest test-retest interval (2 weeks), and one of the largest sample sizes of 182 11- to 20-year-olds. A retest administered within two weeks of the initial test is expected to have a coefficient of at least  $r=0.80$ , which means this is not a reliable reference study. Additionally, the study with the largest test-retest score ( $r=0.96$ ) had time interval of three to four weeks with a sample of 10 third to fourth graders. This study's sample size and age group were limited and may not be generalizable to differing populations.

Supplemental articles were examined to determine more accurate test-retest reliability scores for the Piers-Harris 2. DeBettignies and Goldstein (2019) used the Piers-Harris 2 to evaluate the effectiveness of an improv theater class intervention on 52 elementary-aged students' self-concepts. All participants were administered the Piers-Harris 2 at three points in time, approximately 12 weeks apart.

Ultimately, the literature on Piers-Harris 2 test-retest reliability is limited. Aside from the DeBettignies and Goldstein (2019) study, additional studies evaluating self-concept using the Piers-Harris 2 continue to rely on the Piers-Harris 1 test-retest scores. Robertson (2013) evaluated self-concept in 280 students attending schools for the gifted using the Piers-Harris 2. A test-retest was not conducted, and the author reported on the TOT score test-retest coefficients presented in the original Piers-Harris. Similarly, two reviews on self-concept measures evaluated the Piers-Harris 2 using the Piers-Harris 1 scores (Butler & Gasson, 2005; Holmbeck et al., 2007). Both reviews presented the Piers-Harris 2 test-retest reliability score ranges as  $r=0.65-0.96$  without stating that the referenced studies used were products of the original Piers-Harris. Essentially, current evidence that the Piers-Harris 2 holds adequate test-retest reliability is lacking, and the interpretation of results regarding temporal error is inconclusive.

### **Validity**

**Content Validity.** According to the Piers-Harris 2 manual, content validity addresses whether a test item's content adequately measures the behavior of interest (Piers and Herzberg, 2002). The authors consider content validity to be the least important validity measure because self-concept is theoretical, which may be best measured by construct validity. As previously stated, the original Piers-Harris consisted of 80 items that were determined to be adequate measures of self-concept. A factor analysis of these items was conducted to identify the six

subscale domains. An expert panel was not utilized to evaluate content validity of the original Piers-Harris items, which limits the overall theoretical significance of the measure.

The Piers-Harris 2 developers condensed the number of test items to 60. A clinical judge was appointed to determine if the removed items contained content overlap with the preserved items. No information about the clinical judge was given (e.g., training, qualifications), other than his or her decision that content validity would be sufficient without the 20 removed items.

**Construct Validity.** Piers and Herzberg (2002) define construct validity as how well a test measures a “theoretical psychological characteristic” (pg. 53). The Piers-Harris 2 construct validity is examined using correlational measures and a factor analysis to evaluate the test’s structural characteristics and convergent validity.

A test’s structural characteristics are the intercorrelations of items within the domain scales. Domain scales are expected to measure separate entities of the overall concept being assessed. The Piers-Harris 2 test developers conducted an inter-scale correlation analysis to determine how closely the subscales are interrelated based on the number of items they share. A larger correlation coefficient reflects a greater number of shared items between two subscales. For example, the popularity domain correlates with the physical appearance and attributes domain at  $r=0.66$ , and with the behavioral adjustment domain at  $r=0.30$ . The differences in the correlations are expected because the POP and PHY domains share more items. The authors concluded that each domain scale correlated most with the TOT self-concept score, suggesting that each domain better represents overall self-concept rather than particular components of the concept. (Piers & Herzberg, 2002)

A factor analysis was included to determine how well the domain scales measure different aspects of self-concept. Six common factors that represent self-concept were discovered

through similarities between the items. For instance, one factor reflected items that are related to making new friends, maintaining friendships, and perception of popularity. The majority of these items were already identified as part of the POP domain, inferring that the domains adequately measure the factor of self-concept that they intended to measure. Similarly, an item-total correlation was used to evaluate the correlations between the items and their assigned domains. This measure confirmed that 52 of the 60 items significantly correlated to their assigned domains rather than other domains. Although statistically significant, the correlations ranged from  $r=.17$  to  $r=.59$ , which are only considered weak to moderate (Piers & Herzberg, 2002).

Convergent validity measures the correlation between the Piers-Harris 2 and measures of other constructs related to self-concept, including anger/aggressive attitudes and psychological symptoms. To examine the correlation between self-concept and anger/aggressive attitudes, the test developers compared the Piers-Harris 2 to the “Attitudes Towards Guns and Violence Questionnaire” (AGVQ) and the “Aggression Questionnaire” (AQ). Self-concept was expected to have an inverse relationship with these two measures. Significant negative correlations were found between the TOT self-concept score and AGVQ total score ( $r=-0.29$ ), and the TOT self-concept score and the AQ total score ( $r=-0.16$ ). These correlations should be interpreted with caution because they are classified as significant, but weak. (Piers & Herzberg, 2002)

Measures of psychological symptoms were also evaluated in relation to the TOT self-concept scale. The “My Worst Experience Scale” (MWES) assess for posttraumatic stress disorder symptoms and emotional distress in children. As expected, the total MWES score negatively correlated with the TOT Piers-Harris 2 score ( $r=-0.30$ ). Lastly, the “Overeating Questionnaire” (OQ), a self-report measure of thoughts related to obesity, was anticipated to negatively correlate with self-concept because obesity may be associated with deficits in self-



esteem. The TOT self-concept score significantly correlated with domains of the OQ, including body image ( $r=.34$ ), social isolation ( $r=-0.51$ ), and affective disturbance ( $r=-.50$ ). Generally, convergent validity correlation measures should fall in the  $r \leq 0.6$  range to be considered valid for educational research purposes. The Piers-Harris 2 did not achieve this standard with the related constructs it was compared to. Thus, the overall validity of the Piers-Harris 2 is limited due to strength of these correlations. (Piers & Herzberg, 2002)

**Criterion Validity.** The Piers-Harris 2 manual defines criterion validity as how well a test predicts an individual's performance in other activities (Piers and Herzberg, 2002). Like test-retest reliability, criterion validity data were only reported from the original Piers-Harris. The criterion validity studies were implemented to compare groups that were expected to differ in their self-concept. Examples of comparison groups included: (a) females with mental retardation and females in the normative sample, (b) sixth grade boys with few close friendships and sixth grade boys in a "chum" relationship, (c) children who were physically abused and children who were not abused, and (d) clinically and non-clinically referred children. The authors did not indicate how or why these groups were expected to differ in self-concept, and no data were presented as evidence of criterion validity. Rather, the Piers-Harris 2 manual presents conclusive statements regarding the comparison groups, such as "the groups did not differ on mean Piers-Harris TOT scores" and "as predicted, [adolescent females with mental retardation] scored significantly lower on the Piers-Harris than non-clinical children of the same chronological age..." (Piers & Herzberg, 2002, pg. 69). These conclusions are not valid sources because they are lacking data as evidence. Therefore, Piers-Harris 2 scores used as indicators for an individual's performance in other activities should be interpreted with caution.

## Conclusions

The current paper evaluated the Piers-Harris 2 as an insufficient measure of self-concept based on the test's contents, procedures, and psychometric properties. First, the Piers-Harris 2 can benefit from improvements to its reliability and validity measurements, specifically data evidence for test-retest reliability and criterion validity. The test may be limited in these psychometric properties because self-concept is a theoretical concept that cannot be precisely measured or observed. Additionally, the Piers-Harris 2 developers can improve the test's standardization sample to include a more accurate representation of the United States population. As previously stated, the Piers-Harris Self Concept Scale, 3<sup>rd</sup> Edition was released in 2018, which may have addressed these limitations.

## References

- Butler, R. J., & Gasson, S. L. (2005). Self-esteem/self-concept scales for children and adolescents: A review. *Child and Adolescent Mental Health, 10*(4), 190-201. doi: 10.1111/j.1475-3588.2005.00368.x
- DeBettignies, B. H., & Goldstein, T. R. (2019). Improvisational theater classes improve self-concept. *Psychology of Aesthetics, Creativity, and the Arts*.  
<http://dx.doi.org/10.1037/aca0000260>
- Holmbeck, G. N., Thill, A. W., Bachanas, P., Garber, J., Miller, K. B., Abad, M., ... Zukerman, J. (2007). Evidence-based assessment in pediatric psychology: Measures of psychosocial adjustment and psychopathology. *Journal of Pediatric Psychology, 33*(9), 958-980. doi: 10.1093/jpepsy/jsm059
- Piers, E. V., & Herzberg, D. S. (2002). Piers-Harris children's self-concept scale (2<sup>nd</sup> ed.) (Piers-Harris 2) [Manual]. Torrance, CA: Western Psychological Services.
- Salvia, J., Ysseldyke, J. E., & Bolt, S. (2013). *Assessment in special and inclusive education* (12<sup>th</sup> ed.). Belmont, CA: Cengage Learning.