The Lifespan Self-Esteem Scale: Initial Validation of a New Measure of Global Self-Esteem

Michelle A. Harris, M. Brent Donnellan & Kali H. Trzesniewski

To cite this article: Michelle A. Harris, M. Brent Donnellan & Kali H. Trzesniewski (2017): The Lifespan Self-Esteem Scale: Initial Validation of a New Measure of Global Self-Esteem, Journal of Personality Assessment, DOI: 10.1080/00223891.2016.1278380

To link to this article: http://dx.doi.org/10.1080/00223891.2016.1278380
The Lifespan Self-Esteem Scale: Initial Validation of a New Measure of Global Self-Esteem

Michelle A. Harris, M. Brent Donnellan, and Kali H. Trzesniewski

Department of Human Ecology, University of California, Davis; Department of Psychology, Texas A&M University

ABSTRACT

This article introduces the Lifespan Self-Esteem Scale (LSE), a short measure of global self-esteem suitable for populations drawn from across the lifespan. Many existing measures of global self-esteem cannot be used across multiple developmental periods due to changes in item content, response formats, and other scale characteristics. This creates a need for a new lifespan scale so that changes in global self-esteem over time can be studied without confounding maturational changes with alterations in the measure. The LSE is a 4-item measure with a 5-point response format using items inspired by established self-esteem scales. The scale is essentially unidimensional and internally consistent, and it converges with existing self-esteem measures across ages 5 to 93 (N = 2,714). Thus, the LSE appears to be a useful measure of global self-esteem suitable for use across the lifespan as well as contexts where a short measure is desirable, such as populations with short attention spans or large projects assessing multiple constructs. Moreover, the LSE is one of the first global self-esteem scales to be validated for children younger than age 8, which provides the opportunity to broaden the field to include research on early formation and development of global self-esteem, an area that has previously been limited.

Global self-esteem reflects the subjective evaluation of the self and is one of the most widely studied constructs in the social sciences (see Donnellan, Trzesniewski, & Robins, 2015). Global self-esteem can be assessed across cultures (Schmitt & Allik, 2005), is correlated with consequential life outcomes (see Steiger, Allemand, Robins, & Fend, 2014; Trzesniewski et al., 2006), and appears to be a risk factor for the development of depression and anxiety (Sowislo & Orth, 2013). In short, global self-esteem is an important psychological attribute that demonstrates both stability and change across the lifespan (Ort & Robins, 2014).

The most commonly used measure of global self-esteem is the Rosenberg Self-Esteem Scale (RSE; Rosenberg, 1965). Despite its popularity, the RSE has potentially limited use for research across the lifespan because researchers conducting large studies that measure many constructs might prefer a scale with fewer than 10 items. In fact, many researchers have adapted the RSE to use only a subset of the original 10 items, with no justification for the chosen items. Moreover, there are ongoing debates about its factor structure (e.g., Alessandrini, Vecchione, Eisenberg, & Laguna, 2015), and the wording of specific items might not be suitable for younger children (e.g., “I feel that I am a person of worth, at least on an equal basis with others”). Researchers sometimes use different scales for individuals of different ages, but this makes it difficult to study the development of self-esteem across the lifespan, given that researchers risk confounding age-related changes with changes in assessment tools. Although advanced psychometric techniques like item response theory (e.g., Hambleton, 1991) could be used to equate scales, a more straightforward solution is to develop a relatively small set of items that are suitable for a wide range of ages. We engaged in this task, and the goal of this study is to provide initial evidence for the validity of a short, new measure of global self-esteem suitable for use across the lifespan—the Lifespan Self-Esteem Scale (LSE).

Defining global self-esteem

James (1892/1985) first described self-esteem as the “ratio of our actualities to our supposed potentialities” (p. 54, italics added). This foundational definition emphasizes the essentially subjective nature of this self-judgment. More recent treatments have followed this Jamesian perspective and emphasized that self-esteem involves subjective feelings of both self-acceptance and self-respect (see Rosenberg, 1989). Thus, self-report methods seem especially well-suited for assessing self-esteem given the phenomenological nature of the construct.

Two major types of self-esteem have been studied with self-report measures: global and domain-specific self-esteem. Global self-esteem is the general, subjective evaluation of the self, whereas domain-specific self-esteem focuses on self-evaluations in developmentally relevant domains such as academic abilities, peer relations, and physical appearance (e.g., Marsh, Parker, &
Assessing global self-esteem across the lifespan

There is debate in the developmental literature as to when children can first provide valid reports of global self-esteem. One influential perspective is that global self-esteem “is not a concept that can be verbalized in children’s repertoire until [middle childhood]” (Harter, 2012b, p. 3). In light of this view, many researchers assume that global self-esteem cannot be assessed until around age 8. Instead, researchers typically inquire about domain-specific self-perceptions, such as those related to academic contexts or peer relationships prior to age 8 because these self-perceptions are more concrete (e.g., I am good at numbers, I have a lot of friends; Harter & Pike, 1984).

However, research from other areas focused on children’s ability to think about the self is inconsistent with the idea that global self-esteem cannot be assessed in young children. For example, between the ages of 3 and 5, children first develop the ability to form autobiographical memories, or a sense of self through time (see Fivush & Nelson, 2004) and are able to start remembering at least short-term past events and integrate them into a reasonably coherent self-view. In addition, emerging research shows that children can provide reliable ratings of their emotions, such as worry and anxiety (Lagattuta, Sayfan, & Bamford, 2012). These findings suggest that children as young as age 5 can potentially report on their global self-esteem. Consistent with this suggestion, research has suggested that young children can provide valid self-reports of global self-esteem. Indeed, these studies have found that children as young as age 5 (the youngest age at which self-report global self-esteem has been recorded) provide self-esteem reports that are reliable, unidimensional (Marsh, Craven, & Debus, 1991; van den Bergh & de Ruyck, 2003), consistent across 1 year (Marsh, Craven, & Debus, 1998), and convergent with informant ratings of child self-esteem.

Furthermore, theorists have suggested there are similar correlates of self-esteem across age groups, such as perceived acceptance by significant others (see Harter, 2012a) and various contingencies of self-esteem that might be formed from an early age (see J. Crocker & Park, 2012). At the same time, certain contingencies might be more prevalent at specific developmental periods; for example, physical appearance and peer approval become salient issues for global self-esteem during early and middle adolescence (Harter, 2012a). However, there is limited empirical research on whether correlations actually vary with age, largely due to the need for a measure of global self-esteem that can be administered across the lifespan. In light of these theoretical discussions and the handful of studies testing global self-esteem at age 5, we sought to develop a self-report measure that could be used with children as young as age 5 and throughout the lifespan.

Is there a need for a new measure?

There are existing self-esteem inventories designed for multiple age groups, such as Harter’s Self-Perception Profiles (SPP; e.g., Harter, 1982, 2012b; Harter & Pike, 1984) and Marsh’s Self-Description Questionnaire (SDQ; e.g., Marsh et al., 1991, 1998; Marsh & O’Neill, 1984; Marsh, Parker, & Barnes, 1985; Marsh, Smith, & Barnes, 1985). These existing measures have seemingly adequate psychometric properties for assessing self-esteem within particular developmental periods (reviewed in Donnellan et al., 2015); however, they are not as useful for studying self-esteem across developmental periods, because of differences across measures within each family of scales (e.g., SPP for Children, Adolescents, and Adults).

Specifically, the existing families of measures use different items for different ages, and in the case of the SDQ, different response options for individuals of different ages. Tables 1 and 2 in the Supplemental Online Materials contain more detailed comparisons of the changes in the scales across developmental periods, and only the main issues are summarized here. For the SDQ, (a) for children younger than 8 years old, the researchers used one-on-one interviews and a two-step, forced-choice response format that is turned into a 4-point scale, whereas for older children and adolescents, the scale is group administered and consists of a 6-point (for older children) or 8-point (for adolescents) scale: (b) the total number of items varies across scale versions: (c) the number of positively and negatively worded items varies: (d) the descriptive language varies: and (e) the ordering of items varies across scale versions. For Harter’s self-esteem inventories, (a) the total number of items in the global self-esteem subscales varies across versions, (b) the adolescent and college student versions each change the stem of the items: (c) the qualifier is removed or changed for the emotion (e.g., very happy vs. happy), (d) the descriptive language changes from, for example, “happy/unhappy” to “pleased/disappointed,” (e) the ordering of items changes, (f) there is no global subscale for children younger than age 8, and (g) the two-step, forced-choice response format occasionally proves cumbersome, leading to a worrisome amount of unusable data (e.g., Donnellan et al., 2015; Wichstrom, 1995; Yeager & Krosnick, 2011). As it stands, there is no existing multi-item inventory that is ideal for studying self-esteem across developmental periods from childhood to old age.1 Thus, there is a need for a new short measure of global self-esteem.

Overview of the development of the LSE

Although efforts have been made to develop self-esteem measures that cover specific phases of the lifespan, there is still no single measure of global self-esteem that can be used across a wide range of ages. Accordingly, the primary goals of this study were to (a) develop a short, unidimensional global self-esteem measure that could be used with children as young as age 5 and throughout the lifespan.

1 One single-item scale has been developed and tested with children as young as age 5: the Single-Item Self-Esteem Scale (SISE; Robins, Tracy, Trzesniewski, Potter, & Gosling, 2001).
scale that can quickly be administered to participants across the lifespan, and (b) evaluate the structure, reliability, and validity of this new scale in participants ranging in age from 5 (when children might theoretically have an understanding and ability to reliably report on their global self-esteem) to 93. The scale will be deemed adequate for use across the lifespan if internal consistency estimates and test–retest coefficients are similar for all ages, and if similar convergent validity and nomological network correlations are found across age groups.

We first identified items from the RSE, SPP for Children, and SDQ–I that could be understood or simplified to be relatable to individuals across the lifespan. We found 11 items from across these scales that could be simplified, and we wrote 2 new items to capture some additional concepts (i.e., feeling of doubts toward the self), which resulted in an initial pool of 13 items (see Table 3 in the Supplemental Online Materials for original items and item adaptions). We opted to use a 5-point response scale given that this approach is relatively simple, and we suspected it would provide enough differentiation to prove useful for assessing global self-esteem across different ages. After pilot testing with the full pool of items, and taking into consideration the arguments in Robins, Hendin, and Trzesniewski (2001) regarding the possibility of capturing global self-esteem with relatively few items, we winnowed the scale down to four items. A four-item scale provides a short, multi-item measure suitable for large projects assessing multiple constructs while still being long enough to test unidimensionality with confirmatory factor analysis (CFA; i.e., a three-item scale would just be identified). Finally, a four-item scale is feasible for populations with diminished attention spans (e.g., young children). We provide a detailed description of scale development procedures and extensive pilot studies with children and adults in the Supplemental Online Materials.

**Method**

**Participants**

Analyses are based on data from five samples. Table 1 shows the final sample sizes and demographic characteristics of each sample. The combined data set included 2,714 individuals across ages 5 to 93 (39% male). Of the total sample size, 268 (9.9%) of individuals did not provide their age.

**Sample 1**

Participants were recruited in Fall 2013 through Mechanical Turk (MTurk; for details see Behrend, Sharek, Meade, & Wiebe, 2011; Buhrmester, Kwang, & Gosling, 2011; Mason & Suri, 2012) for a task called “Survey about experiences and self-views” (key words: survey, study, psychology, people). Participants with at least a 95% approval rate for their previously completed tasks on MTurk (called human intelligence tasks [HITs]) were invited to participate in exchange for a nominal level of compensation (US$1). They were allotted 2 days to complete the survey after accepting the HIT. MTurk samples are often more diverse than samples from university subject pools (especially in terms of age) and appear to provide data that meet standards for reliability and validity (see Behrend et al., 2011; Buhrmester et al., 2011). Participants were asked their country of origin, ethnicity, and language spoken at home.

Eight separate HITs were created to ensure similar sample sizes for age-stratified subsamples. The a priori goal was to recruit 200 participants in each of the following age groups: 18 to 24, 25 to 29, 30 to 39, 40 to 49, 50 to 59, 60 to 69, 70 to 79, and 80 to 89. Age was selected in three ways: (a) the title of the survey described the desired age group, (b) the project page repeated the desired age group above the link to the survey, and (c) an item asked participants to write their exact age in years. This third age filter was used to calculate age groups for all analyses. Two attention filters were included in the middle and at the end of the survey. Filters asked participants to select either “agree” or “disagree.” Participants who did not follow the instructions for both filter items were excluded from analyses. Participants identified their age, gender, country of origin, race, and highest level of education attained. The final sample size was 1,413 participants ranging in age from 18 to 93.

**Sample 2**

Due to low response rates in middle adulthood (ages 39–59) through MTurk, an additional sample was recruited in Winter 2014 through the market research division of the Qualtrics organization. Qualtrics charged $5 per response for a total of $1,000 and contracted with an outside vendor, Clearvoice, which maintains a database of individuals across the United States and other countries. Participants in this study were recruited mainly from the U.S. panel. Participants included in the Clearvoice database come from a broad range of online and offline sources. Participants within our requested age criteria were randomly selected to take part in the survey. Selected participants were e-mailed a link to a Web-based survey. The survey for each age group remained open until the target number of participants exceeded the number contracted (100 participants between ages 39 and 49 and 100 participants between ages 50 and 59).

Age was selected in three ways: (a) Clearvoice sent the link to individuals within the requested age range; (b) one age filter was placed at the beginning of the survey, which directed participants to the last page if they indicated they were outside the range of desired ages; and (c) participants were asked to write their exact age in years. This third age filter was used to calculate age groups for all analyses. Two attention filters were included in the middle and at the end of the survey. The first filter asked participants to select “agree” from the two options of agree and disagree. The second filter was in the form of a paragraph (see the Supplemental Online Materials; Goodman, Cryder, & Cheema, 2013). Participants who did not follow the instructions for both filter items were excluded from analyses. Participants identified their age, gender, country of origin, race, and highest level of education attained.

Due to limits on the length of the survey placed by Qualtrics, we used a three-form planned missing design (see Little & Rhemtulla, 2013) to reduce the number of items each participant completed while still gathering data on all constructs measured in Sample 1. We created a base form to include all demographic items and items from our self-esteem scale as well as a random selection of items from each of the remaining scales, which were evenly split across three forms: A, B, and C.
Therefore, participants were randomly assigned to complete one of three combinations of forms: Base and A, Base and B, or Base and C. Each combination resulted in a total of approximately 100 items. Exact items in each form are available on request. The final size of Sample 2 was 201 participants ranging in age from 39 to 59.

**Sample 3**

Due to low response rates from older adults (ages 60–89) through MTurk, a third sample was recruited in Winter 2014 through Qualtrics. Recruitment procedures were the same as Sample 2 (e.g., Qualtrics, Clearvoice, planned missing design), except Qualtrics charged $6 per response for a total of $2,700. Age was selected in the same ways as in Sample 2. Two attention filters similar to those used for Sample 2 were used; however, the paragraph filter was abbreviated for the participants in Sample 3 (see the Supplemental Online Materials) to reduce participant fatigue. Participants identified their age, gender, country of origin, race, and highest level of education attained. The final size of Sample 3 was 451 participants ranging in age from 60 to 89.

**Sample 4**

Adolescents (ages 14–17) were recruited in Spring 2014 from Qualtrics for $6 per response, for a total of $1,254. Recruitment procedures and age filters were similar to those for Sample 3 (e.g., Qualtrics, Clearvoice, planned missing design). Attention filters were the same as those used for Sample 3. To provide overlapping data across age, adolescents completed the versions of the attachment and narcissism measures completed by both children and adults in Samples 1, 2, 3, and 5 (see Table 1). Participants identified their age, gender, country of origin, race, and highest level of education attained by both of their parents. The final size of Sample 4 was 211 participants ranging in age from 14 to 17.

**Sample 5**

We thought supervised collection of data from children in elementary and middle school would produce higher quality data; thus we recruited a large sample of children and administered surveys in their school classrooms. The final size of Sample 5 was 438 participants ranging in age from 5 to 15. Principals of three K–8 private schools in northern California agreed to participate in a study on self-esteem development. For two schools, parents of every student received a consent form explaining the study and asking them to provide consent for their child to participate. For the third school, passive consent procedures were used, so that all parents were informed of the data collection procedures and date and could elect to pull their child out of the study. There was no effect of school on self-esteem scores, controlling for age ($\beta = .06, p = .19$), nor was there an interaction between school and age ($\beta = -.06, p = .21$).

---

**Table 1. Measures administered to Samples 1 through 5.**

<table>
<thead>
<tr>
<th>Measure</th>
<th>Sample 1</th>
<th>Sample 2</th>
<th>Sample 3</th>
<th>Sample 4</th>
<th>Sample 5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age range</strong></td>
<td>18–93</td>
<td>39–59</td>
<td>60–89</td>
<td>14–17</td>
<td>5–13</td>
</tr>
<tr>
<td>N</td>
<td>1,413</td>
<td>201</td>
<td>451</td>
<td>211</td>
<td>438</td>
</tr>
<tr>
<td>% male</td>
<td>48%</td>
<td>38%</td>
<td>41%</td>
<td>23%</td>
<td>45%</td>
</tr>
<tr>
<td>% from U.S.</td>
<td>96%</td>
<td>93%</td>
<td>97%</td>
<td>89%</td>
<td>N/A</td>
</tr>
<tr>
<td>% White or Caucasian</td>
<td>75%</td>
<td>83%</td>
<td>95%</td>
<td>82%</td>
<td>N/A</td>
</tr>
<tr>
<td>% BA or lower</td>
<td>84%</td>
<td>82%</td>
<td>82%</td>
<td>81% (mother), 81% (father)</td>
<td>N/A</td>
</tr>
<tr>
<td>LSE</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>SPPC</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>RSE</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>SISE</td>
<td>✓ 5-point scale</td>
<td>✓ 6-point scale</td>
<td>✓ 6-point scale</td>
<td>✓ 6-point scale</td>
<td>✓ 5-point scale (LSE format)</td>
</tr>
<tr>
<td>SDQ</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>NARQ</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>CES–D</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>ECR–R</td>
<td>✓ 3-point scale</td>
<td>✓ 7-point scale</td>
<td>✓ 7-point scale</td>
<td>✓ 7-point scale</td>
<td>✓ 7-point scale</td>
</tr>
<tr>
<td>Mini-IPIP</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td><strong>Parent attachment</strong></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>NPI–16–C</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>LSE–Parent</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>LSE–Teacher</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Self-esteem explanation</td>
<td>✓ Provided with 6 examples and instructed, &quot;Please check which of the following…”</td>
<td>✓ Did not receive checklist</td>
<td>✓ Provided with 6 examples and instructed, &quot;Please check which of the following…”</td>
<td>✓ Provided with 6 examples and instructed, &quot;Please check which of the following…”</td>
<td>✓ Provided with 6 examples and instructed, &quot;If you read these, read all of them…&quot;) only given to third school</td>
</tr>
<tr>
<td>Trait mood</td>
<td>✓ LOT–R</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>✓ SWLS</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Note. BA = bachelor’s degree; LSE = Lifespan Self-Esteem Scale; SPPC = Self-Perception Profile for Children global subscale; RSE = Rosenberg Self-Esteem Scale, SISE = Single-item Self-Esteem Scale; SDQ = Self-Description Questionnaire global subscale; NARQ = Narcissistic Admiration and Rivalry Questionnaire, CES–D = Center for Epidemiologic Studies–Depression scale; ECR–R = Experiences in Close Relationships–Revised; Mini-IPIP = Mini-International Personality Item Pool; NPI–16–C = Narcissistic Personality Inventory–16 Adapted for Children; LOT–R = Life Orientation Test–Revised; SWLS = Satisfaction with Life Scale.
All classes were awarded an ice cream party if 80% of their parents returned the consent forms (regardless of whether they provided or refused consent). Parents were also provided a questionnaire and given the option of completing it on paper or online through a provided link. All teachers were asked to complete the LSE for each participating child in their class. The research assistants leading the child surveys provided each teacher with an envelope containing one printed survey for each participating child, and they instructed teachers to return the envelope within a week of data collection.

Data were collected in Spring 2013 for the first school, Fall 2013 for the second school, and Winter 2014 for the third school. Trained research assistants visited the classrooms of each grade during school hours and administered the survey to the group via either paper copies or iPads. Students in the kindergarten and first-grade classrooms were split into smaller groups of students (3–6 children per group) for survey administration and data collection. Students took about 1 hr to complete the survey, and sessions were split into two half-hour sessions for kindergarten and first-grade classes.

Children in the school sample completed four and two practice questions (four for LSE and two for Kerns’s Security Scale) to learn the scales before the administration of scales described next. Children were reminded throughout the survey that they could choose whichever response option best described their answer, that there were no right or wrong answers, and that the researcher did not know what the child was going to say. Children were asked to circle the figure or touch the picture on the iPad that corresponded to their answer after the research assistant read each question aloud. Then, they covered their answers with pictorial “game boards” provided by the researchers. Instructions for adolescents and adults simply asked them to choose the face that best described their response.

Children identified their age and gender, and data were recorded for grade, teacher, and school. Children in second grade and older were instructed to write their birthday, and children in kindergarten and first grade were instructed to write their age. Then, the school staff for the third school provided birthdays for all children. For children with missing ages or birthdays (e.g., from the first and second schools), an age was inserted that corresponded with the approximate average age of the child’s class. That is, Kindergarteners with missing age information were assigned the age of 5, first graders were assigned the age of 6, and each grade was assigned the next year of age, ending with eighth graders assigned the age of 13.

**Longitudinal subsample.** A subsample (n = 91) of Sample 5 participated 1 year later in their school classrooms. Children were between 6 and 15 years old at the second time point.

**Measures**

**Self-esteem measures**

**Lifespan Self-Esteem Scale.** The Supplemental Online Materials describe scale development, piloting samples, and results in full. The initial item pool for the LSE consisted of 13 items assessing global self-esteem administered on a 5-point scale (1 = really sad, 2 = sad, 3 = neutral, 4 = happy, 5 = really happy). After examination of items during piloting, we eliminated one item (“How do you feel about yourself?”) due to excessive redundancy with other items, as well as two other items (“How do you feel about confidence toward yourself?” and “How do you feel about doubts toward yourself?”) due to higher confusion among child samples and lower factor loadings than the other 10 items in an exploratory factor analysis (EFA). We then reviewed the items for (a) additional redundancies, and (b) understandability by the youngest participants. This process produced four items that required no additional explanations for the youngest children and had good psychometric properties. These four items make up the final scale (see Figure 1).

The response options were also illustrated with faces depicting the appropriate feeling (really sad = crying face, sad = slight frown, neutral = flat mouth, happy = slight smile, really happy = open-mouthed smile). A sample item is, “How do you feel about yourself?” To evaluate whether children younger than 8 appeared to understand how to use the response scales, we examined the practice questions. Practice questions (see Table 1 for description of samples who received practice questions) were, “How do you feel about chores you do at home?,” “How do you feel about going to the doctor?,” “How do you

![Figure 1. Items and response options for the Lifespan Self-Esteem Scale.](image-url)
feel about getting presents for your birthday?,” and “How would you feel about being eaten by a T-Rex?” Children of all ages used the 5-point scale appropriately; that is, children across both age groups used the full range of the scale, and the means were high for positively worded questions and low for negatively worded questions (see Table 2). In addition, we calculated the average standard deviation for all the practice items for each age group and found these did not significantly differ in an independent samples t test, t(386) = 1.38, p = .17.

**LSE-Parent and –Teacher.** Parents (N = 61) and teachers (N = 282) rated children’s self-esteem with the LSE–Parent and –Teacher forms. They completed the same LSE items that were given to their children, but the wording was changed to focus on the self-esteem of the child. For example, “How do you feel about yourself?” was changed to “How does your child feel about him or herself?” and “How does ______ feel about him or herself?” Parents and teachers were instructed to respond according to their best guess for their child’s self-esteem, and they were assured that they did not have to be correct. They used the same 5-point scale illustrated by faces (see the Supplemental Online Materials for a display of full scales).

The Supplemental Online Materials include a description of the remaining measures used to evaluate convergent validity (i.e., Self-Perception Profile for Children global subscale [SPPC; Harter, 1982, 2012b], RSE [Rosenberg, 1965], SISE [Robins, Hendin, & Trzesniewski, 2001]; SDQ global subscale [Marsh et al., 1991]), criterion-related validity (i.e., Center for Epidemiologic Studies–Depression scale [CES–D; Cole, Rabin, Smith, & Kaufman, 2004]), and the nomological network of self-esteem (i.e., Narcissistic Personality Inventory–16 Adapted for Children [NPI–16–C; see Ames, Rose, & Anderson, 2006; Barry, Frick, & Killian, 2003]; Narcissistic Admiration and Rivalry Questionnaire [NARQ; Back et al., 2013]; Experiences in Close Relationships–Revised, Avoidance and Anxiety scales [Brennan, Clark, & Shaver, 1998]; Kerns’s Security Scale, Parent attachment [Kerns, Klepac, & Cole, 1996]; Mini-International Personality Item Pool [Mini-IPIP, Donnellan, Oswald, Baird, & Lucas, 2006]; Trait mood, Life Orientation Test–Revised [LOT–R; Scheier, Carver, & Bridges, 1994]; Satisfaction with Life Scale [SWLS; Diener, Emmons, Larsen, & Griffin, 1985]).

**Preliminary results**

Before conducting analyses on the full sample, we tested whether the method of data collection was related to average levels of self-esteem (given that sample age is confounded by collection method, we controlled for age in these preliminary analyses). Across adult and adolescent samples (i.e., Samples 1–4), we tested for mean differences by recruitment panel (MTurk or Qualtrics), recruitment group (Qualtrics Sample 1, 2, or 3), and planned missing form (A, B, or C). There was a significant difference between participants recruited through MTurk and Qualtrics, F(1, 2064) = 21.31, MSE = 13.54, p < .01; adjusted means (for age): 3.59 versus 3.77, respectively, average SD = 0.79. We ran relevant psychometric analyses controlling for recruitment panel and found no differences in results, so we report findings without controlling for recruitment panel. Next, we found no differences in mean self-esteem for participants recruited through the different Qualtrics panels, F(2, 856) = .30, MSE = .16, p = .74; adjusted means: Panel 1 = 3.84, Panel 2 = 3.75, Panel 3 = 3.92, average SD = 0.74. Finally, LSE scores did not differ by the planned missing form to which participants were randomly assigned, F(2, 856) = .33, MSE = .17, p = .72; adjusted means: Form A = 3.80, Form B = 3.79, Form C = 3.84, average SD = 0.74.

Across the child sample (i.e., Sample 5), we found no differences (again controlling for age) by school, F(2, 411) = 1.88, MSE = 1.03, p = .15; adjusted means: School 1 = 4.17, School 2 = 4.07, School 3 = 4.21, average SD = 0.76; teacher, F(27, 386) = .92, MSE = .50, p = .59, average SD = 0.73; assent administrator, F(11, 401) = 1.26, MSE = .68, p = .25, average SD = 0.77; reader of the survey items, F(10, 402) = 1.67, MSE = .90, p = .09, average SD = 0.74; research assistants, F(7, 322) = .42, MSE = .23, p = .89, average SD = 0.75; whether the survey was conducted in the students’ classrooms or in another room, F(1, 412) = 2.60, MSE = 1.50, p = .11; adjusted means: classroom = 4.21, another room = 4.08, average SD = 0.73; or whether students had completed the survey on paper or a tablet, F(1, 410) = 1.73, MSE = .95, p = .19; adjusted means: paper = 4.20, tablet = 4.08, average SD = 0.77. Thus, these potential confounds and qualifiers are not discussed further.

**Analytic approach**

Given the large sample size, we set an alpha level of p < .01 to determine statistical significance in all analyses. In addition, we used age as a continuous variable and present results using the full sample as well as 11 age-stratified subsamples for ease of presentation and discussion: 5 to 7, 8 to 13, 14 to 17, 18 to 24, 25 to 29, 30 to 39, 40 to 49, 50 to 59, 60 to 69, 70 to 79, and 80 to 89. In addition, because Sample 5 had students nested within classes, we ran initial psychometric analyses using a multilevel design in Mplus 7.31 (Muthén & Muthén, 1998–2012). Results were similar to those found using SPSS and not

---

**Table 2. Descriptive statistics for practice items for the Lifespan Self-Esteem Scale with Child Sample 5.**

<table>
<thead>
<tr>
<th>Item</th>
<th>Ages 5–7</th>
<th></th>
<th>Ages 8–10</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Minimum–</td>
<td>Maximum</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>How do you feel about the chores you do at home?</td>
<td>1–5</td>
<td>3.62</td>
<td>1.17</td>
<td>1–5</td>
</tr>
<tr>
<td>How do you feel about getting shots from the doctor?</td>
<td>1–5</td>
<td>2.43</td>
<td>1.47</td>
<td>1–5</td>
</tr>
<tr>
<td>How do you feel about getting presents for your birthday?</td>
<td>2–5</td>
<td>4.88</td>
<td>0.47</td>
<td>3–5</td>
</tr>
<tr>
<td>How would you feel about being eaten by a T-Rex?</td>
<td>1–5</td>
<td>2.03</td>
<td>1.51</td>
<td>1–5</td>
</tr>
</tbody>
</table>

---

*There are too many groups to report all mean scores, but descriptive statistics are available on request.

3The final sample included one individual who was 93 years old. However, we restricted the oldest age group to ages 80 to 89 for simplification and ease of presentation, given that results did not differ when including the 93-year-old. Thus, for all analyses using continuous age, ages ranged from 5 to 93, but for all analyses using the 11 age-stratified groups, ages ranged from 5 to 89.
accounting for the nested data (output are available on request); therefore, we report all findings from SPSS for consistency across samples.

We tested six psychometric properties of the full sample as well as for each age group separately. First, we evaluated dimensionality using techniques related to EFA. Next, we computed Cronbach’s alpha coefficients and their 95% confidence intervals to assess internal consistency. Then for a subsample, we computed test–retest correlations with LSE scores 1 year later as another measure of reliability. Next, we computed self-informant correlations (i.e., from parents and teachers) for a subsample of children with available data. Finally, we correlated LSE scores with four other established self-esteem scales to examine convergent validity and with 15 other measures theorized to be related to self-esteem. We wanted to make sure the LSE showed a similar pattern of association with these variables as do existing measures of self-esteem.

To establish that the LSE did not have measurement-related changes when administered to individuals of different ages, we split the sample into the 11 age-stratified groups (or the groups with available assessments for test–retest reliability and convergent validity, for example) and compared psychometric properties across the age groups. We used confidence intervals, z-score comparisons, and regression models with continuous age interaction terms when appropriate to compare psychometric properties. We also used two structural models to determine measurement invariance across age. We followed these analyses with checks of moderation by different data collection procedures (e.g., paper vs. computerized version, research assistant) and sample characteristics (e.g., gender, ethnicity).

Results

Dimensionality, reliability, and validity

Unidimensionality

First, we tested whether a single-factor model was appropriate (i.e., we evaluated unidimensionality), given that we intended the LSE items to represent a single construct. Therefore, we ran an EFA on the whole sample and found support for unidimensionality based on the eigenvalues (the initial eigenvalue for the whole sample was 3.01, whereas the second eigenvalue was less than 1) and the factor loadings, which ranged from .73 to .86. We then split the sample into the 11 age-stratified groups identified earlier and again found support for unidimensionality in that only the first eigenvalue was above 1 for each group (see Table 3).

We also conducted tests of measurement invariance across age using the multiple indicators, multiple causes (MIMIC) model that treats age as a continuous variable (see Brown, 2015; Jöreskog & Goldberger, 1975) and a traditional multigroup CFA using the preselected 11 age groups. We found a small ($\beta = -.06, p = .01$) effect of continuous age using the MIMIC model, and multigroup models supported metric invariance across the lifespan (see the Supplemental Online Materials for details of invariance tests).

Internal consistency

After evaluating unidimensionality, we assessed reliability by calculating Cronbach’s alpha coefficients and their 95% confidence intervals (see Fan & Thompson, 2001) for the full sample and for each age group. The four items of the LSE generated a strong Cronbach’s alpha coefficient across the full sample (.89) as well as for each age group (range = .84–.91). Confidence intervals for all age groups overlapped with at least one other group, with a majority of groups overlapping almost completely. See Table 4 for exact alpha coefficients and average interitem correlations. These results suggest that the LSE is internally consistent when administered to participants of different ages.

Test–retest reliability

As a second indicator of reliability, we used data from a subsample of students from one school (n = 91, 36% male, mean age = 9.69 years, age range = 6–15 years) who completed the LSE at two time points (separated by 1 year). We calculated a correlation between students’ scores across the two occasions. We expected the LSE to be somewhat stable across 1 year, given stability coefficients found in past literature (meta-analytic r = .31 for 6- to 8-year-olds, k = 5; Trzesniewski, Donnellan, & Robins, 2003). Surprisingly, the 1-year test–retest correlation

Table 3. Factor loadings and eigenvalues from exploratory factor analysis on the Lifespan Self-Esteem Scale by age group.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>How do you feel about yourself?</td>
<td>.64</td>
<td>.78</td>
<td>.86</td>
<td>.83</td>
<td>.85</td>
<td>.83</td>
<td>.81</td>
<td>.84</td>
<td>.79</td>
<td>.69</td>
<td>.88</td>
</tr>
<tr>
<td>How do you feel about the kind of person you are?</td>
<td>.68</td>
<td>.68</td>
<td>.68</td>
<td>.66</td>
<td>.77</td>
<td>.75</td>
<td>.76</td>
<td>.79</td>
<td>.69</td>
<td>.64</td>
<td>.80</td>
</tr>
<tr>
<td>When you think about yourself, how do you feel?</td>
<td>.55</td>
<td>.80</td>
<td>.86</td>
<td>.83</td>
<td>.88</td>
<td>.86</td>
<td>.92</td>
<td>.92</td>
<td>.84</td>
<td>.82</td>
<td>.87</td>
</tr>
<tr>
<td>How do you feel about the way you are?</td>
<td>.61</td>
<td>.85</td>
<td>.83</td>
<td>.84</td>
<td>.87</td>
<td>.87</td>
<td>.87</td>
<td>.83</td>
<td>.88</td>
<td>.84</td>
<td>.86</td>
</tr>
<tr>
<td>Initial eigenvalue</td>
<td>2.16</td>
<td>2.34</td>
<td>2.78</td>
<td>2.69</td>
<td>3.05</td>
<td>2.87</td>
<td>2.99</td>
<td>3.12</td>
<td>2.83</td>
<td>2.65</td>
<td>3.13</td>
</tr>
<tr>
<td>Second eigenvalue</td>
<td>.72</td>
<td>.76</td>
<td>.54</td>
<td>.64</td>
<td>.43</td>
<td>.50</td>
<td>.47</td>
<td>.41</td>
<td>.63</td>
<td>.66</td>
<td>.42</td>
</tr>
</tbody>
</table>

Table 4. Alpha coefficients and interitem reliability for the Lifespan Self-Esteem Scale by age group.

<table>
<thead>
<tr>
<th>Age group</th>
<th>N</th>
<th>Cronbach’s alpha</th>
<th>95% CI</th>
<th>Interitem r</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>2,542</td>
<td>.89</td>
<td>.88–.90</td>
<td>.67</td>
</tr>
<tr>
<td>5–7</td>
<td>111</td>
<td>.71</td>
<td>.61–.79</td>
<td>.39</td>
</tr>
<tr>
<td>8–13</td>
<td>246</td>
<td>.86</td>
<td>.83–.89</td>
<td>.60</td>
</tr>
<tr>
<td>14–17</td>
<td>356</td>
<td>.88</td>
<td>.85–.90</td>
<td>.65</td>
</tr>
<tr>
<td>18–24</td>
<td>274</td>
<td>.87</td>
<td>.84–.89</td>
<td>.62</td>
</tr>
<tr>
<td>25–29</td>
<td>245</td>
<td>.91</td>
<td>.88–.92</td>
<td>.70</td>
</tr>
<tr>
<td>30–39</td>
<td>253</td>
<td>.90</td>
<td>.87–.92</td>
<td>.68</td>
</tr>
<tr>
<td>40–49</td>
<td>241</td>
<td>.90</td>
<td>.88–.92</td>
<td>.70</td>
</tr>
<tr>
<td>50–59</td>
<td>227</td>
<td>.91</td>
<td>.89–.93</td>
<td>.72</td>
</tr>
<tr>
<td>60–69</td>
<td>199</td>
<td>.88</td>
<td>.85–.90</td>
<td>.64</td>
</tr>
<tr>
<td>70–79</td>
<td>196</td>
<td>.84</td>
<td>.79–.87</td>
<td>.56</td>
</tr>
<tr>
<td>80–89</td>
<td>197</td>
<td>.91</td>
<td>.89–.93</td>
<td>.72</td>
</tr>
</tbody>
</table>

Note. 95% CI = 95% confidence intervals; Interitem r = interitem correlation.
was considerably higher than those found in previous literature using other self-esteem scales \((r = .58, p < .01)\). Importantly, splitting the sample into two age groups based on their age at Wave 1, both 5- to 7-year-olds \((r = .48, p < .01)\) and 8- to 13-year-olds \((r = .62, p < .01)\) showed consistent scores 1 year later and similar test–retest reliability \((z = -0.87, p = .19)\). This provides further evidence that children younger than 8 years old can provide consistent responses to self-report measures of global self-esteem.

**Correspondence with informant ratings**

Teachers’ \((N = 282)\) and parents’ \((N = 61)\) ratings of children’s self-esteem were related to self-reports of self-esteem among children in the school sample across all grades \((r_{\text{teachers}} = .29, p < .01); r_{\text{parents}} = .26, p = .04)\). The correlation between teachers’ and children’s ratings was not significantly different from the correlation between parents’ and children’s ratings across all grades \((z = -17, p = .43)\). Regressions indicated that age did not moderate the relation between teachers and children \((\beta = -.29, p = .66)\), or between parents and children \((\beta = .43, p = .79)\). Specifically, the correspondence for children aged 5 to 7 was .23 \((p = .05)\) with teachers and .16 \((p = .48)\) with parents, and the correspondence for children aged 8 to 13 was .29 \((p < .01)\) with teachers and .26 \((p = .13)\) with parents.\(^4\)

**Convergent validity**

We tested convergent validity by examining correlations with established measures of global self-esteem: the SDQ, SPPC, RSE, and SISE. There was good convergent validity: Scores on the LSE were associated with scores on existing measures of self-esteem \((rs = .61–.73, all p < .01);\) disattenuated \(rs = .70–.84\); see Table 5).

**Correlates of self-esteem**

The LSE had evidence of criterion-related validity: Scores were consistently related with depression. In addition, self-esteem was correlated with all hypothesized variables, including attachment security (to parents and romantic partners), narcissism, depression, openness, conscientiousness, extraversion, agreeableness, and neuroticism in ways consistent with the existing literature on global self-esteem (see Table 6). Furthermore, we included three measures to determine the distinctiveness of the LSE from established scales with similar emotional wording (i.e., trait mood, optimism, and life satisfaction). Table 6 shows that correlations between the LSE and these three measures were moderate, indicating the LSE is not isomorphic with these measures.

Initial analyses showed that many correlations with nomological network variables were similar across age (i.e., age by criterion variable interaction predicting LSE were nonsignificant); however, some correlations were significantly moderated by age. These are summarized in the Supplemental Online Materials for interested readers. Given that effect sizes were small and patterns were not generally inconsistent, we conclude the LSE has roughly similar levels of criterion validity and a similar nomological network across age.

\(^4\) All effect sizes reported are standardized beta coefficients.

### Table 5. Convergent validity correlations between the LSE and existing self-esteem measures.

<table>
<thead>
<tr>
<th>LSE items and composite</th>
<th>SDQ</th>
<th>SPPC</th>
<th>RSE</th>
<th>SISE</th>
</tr>
</thead>
<tbody>
<tr>
<td>How do you feel about yourself?</td>
<td>.65</td>
<td>.53</td>
<td>.61</td>
<td>.62</td>
</tr>
<tr>
<td>How do you feel about the kind of person you are?</td>
<td>.57</td>
<td>.48</td>
<td>.54</td>
<td>.52</td>
</tr>
<tr>
<td>When you think about yourself, how do you feel?</td>
<td>.67</td>
<td>.56</td>
<td>.62</td>
<td>.64</td>
</tr>
<tr>
<td>How do you feel about the way you are?</td>
<td>.65</td>
<td>.58</td>
<td>.61</td>
<td>.60</td>
</tr>
<tr>
<td>LSE</td>
<td>.73</td>
<td>.61</td>
<td>.68</td>
<td>.69</td>
</tr>
<tr>
<td>Existing self-esteem measures</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>SDQ</td>
<td>—</td>
<td>.73</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>SPPC</td>
<td>.88</td>
<td>.66</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>RSE</td>
<td>.80</td>
<td>.60</td>
<td>.67</td>
<td>—</td>
</tr>
<tr>
<td>SISE</td>
<td>.96</td>
<td>.85</td>
<td>.91</td>
<td>.75</td>
</tr>
<tr>
<td>Alpha reliability</td>
<td>.96</td>
<td>.85</td>
<td>.91</td>
<td>.75</td>
</tr>
</tbody>
</table>

\(^{4}\) Because alpha reliability cannot be computed for a single-item measure, we report reliability for the SISE as calculated by its creators (Robins, Hendin, & Trzesniewski, 2001, p. 154) using the Heise procedure (bases the reliability estimate on the scale’s autocorrelations over multiple time points).

*p < .01.

We also tested whether the LSE showed comparable associations with these other variables as existing self-esteem scales. Table 7 shows correlations for the four other self-esteem measures. Correlations are similar across all of the self-esteem measures, and high self-esteem is positively related to attachment security to parents and romantic partners, narcissistic admiration, narcissism, openness, conscientiousness, agreeableness, extraversion, trait mood, optimism, and life satisfaction; whereas low self-esteem is positively related to narcissistic rivalry, depression, and neuroticism. These patterns are consistent with past research on self-esteem and the Big Five personality traits (Robins, Tracy, et al., 2001), narcissism (Ackerman & Donnellan, 2013), depression (Orth, Robins, Widaman, & Conger, 2014; Steiger, Fend, & Allemand, 2015), and optimism (Scheier et al., 1994). Interestingly, the magnitudes of the LSE correlations tend to fall in between those of the existing measures. For example, narcissistic rivalry correlates with the RSE at \(r = -.34\), with the SISE at \(r = -.14\), and the correlation with the LSE falls in between these two effect sizes \((r = -.21)\). We computed intraclass correlations using the double-entry method (see Furr, 2010) between the vectors of correlations with the 16 other variables for all self-esteem measures. Overall, the pattern of correlations was similar such that the set of correlations for the LSE was strongly associated with the same vectors for the other self-esteem measures we considered (SDQ, \(r = .97\); SPPC, \(r = .98\); RSE, \(r = .97\), and SISE, \(r = .99\)).

### Mean level differences in self-esteem by age

Across all ages, the LSE had a mean of 3.74 \((SD = 0.83)\). Skewness was \(-0.66\). Standard deviations ranged between 0.61 and 0.88, and skewness ranged between \(1.24\) and \(-0.38\). See Table 6 in the Supplemental Online Materials for descriptive statistics for all age groups. We ran a multiple regression analysis to predict mean self-esteem scores for each age (rather than individual scores; see Robins, Trzesniewski, Tracy, Gosling, & Potter, 2002, p. 427; Rosnow, Rosenthal, & Rubin, 2000, pp. 449–450),
with age modeled as linear, quadratic, and cubic functions as the predictor(s). There were significant quadratic and cubic terms for self-esteem (linear $B = −.01$, $SE = .02$, $β = −.02$, $p = .41$, $R^2 = .00$; quadratic $B = .00$, $SE = .00$, $β = .23$, $p < .01$, $R^2 = .04$; cubic $B = −.00$, $SE = .00$, $β = −.37$, $p < .01$, $R^2 = .06$).

Figure 2 shows that average levels of self-esteem are generally lower in older participants than younger participants. However, the pattern was subtle and shows that young children reported the highest levels, participants around age 30 reported the lowest levels, and participants above age 70 scored in between the highest and lowest age groups.

### Mean level differences in self-esteem by demographic groups

Using the alpha of $p < .01$, we tested for mean differences in LSE scores across demographic groups for the adult and adolescent samples (i.e., Samples 1–4) as well as Sample 5 for gender. We found that self-esteem did not differ by country of origin, $F(1, 2,465) = .20$, $MSE = .14$, $p = .65$; M U.S. = 3.66, M not U.S. = 3.80, average $SD = .90$; gender, $F(1, 2,453) = .60$, $MSE = .41$, $p = .44$; M male = 3.74, M female = 3.76, average $SD = .86$; ethnicity $F(5, 2,119) = 2.44$, $MSE = 1.62$, $p = .03$, average $SD = .85$; education (for Samples 1–3, $r = −.04$, $p = .07$, average $SD = .77$); adolescent mother education (Sample 4, $r = .01$, $p = .94$, average $SD = .94$); or adolescent father education (Sample 4, $r = −.03$, $p = .72$, average $SD = .89$).

Overall, psychometric properties of the LSE (i.e., dimensionality and internal consistency) did not seem to appreciably vary by country of origin, gender, ethnicity, education, adolescent mother education, and adolescent father education; however, one difference emerged: Adolescent mother education significantly moderated the association between the SISE and LSE ($p < .01$) and explained 1.8% additional variance in the LSE. Therefore, we split the file by adolescent mother education levels (excluding PhD, JD, MD, or other advanced degrees) and estimated correlations between the LSE and the SISE. They were all significantly related to the LSE, and the maximum difference in magnitude was .37 ($r$s: some high school = .56, high school GED or diploma = .53, some college = .61, associate’s degree = .87, bachelor’s degree = .71, some graduate or professional school = .90, master’s degree = .86). Tables 8 to 13 of the Supplemental Online Materials display psychometric properties of the LSE for all demographic groups.

### Discussion

Global self-esteem is one of the most widely studied variables in the social sciences and one that has been shown to be relevant for mental health (e.g., Sowislo & Orth, 2013) and other outcomes (see Donnellan et al., 2015, for a review). It is also a developmental construct that shows stability and change across the lifespan (Orth & Robins, 2014). However, existing measures are not maximally suitable for studying self-esteem across multiple age groups, limiting the studies that can be conducted and the conclusions that can be made about the antecedents and consequences of self-esteem across developmental periods. Thus, we created the LSE to assess global self-esteem across a wide range of ages with relatively few items.

We found evidence that scores on the LSE were unidimensional, internally consistent, and relatively stable across a 1-year period (at least in childhood and early adolescence). Moreover, scores on the LSE converged with four other established measures of self-esteem and informant ratings. The LSE demonstrated expected patterns of associations with 15 measures of theoretically relevant constructs. Moreover, we found little indication that age moderated the psychometric properties of the LSE, except for lower reliability in the youngest age group. Nonetheless, we found evidence that scores on the LSE were valid in these young age groups. This is inconsistent with some prevailing assumptions about global self-esteem (e.g., Harter, 2012a) but generally consistent with existing empirical evidence suggesting that children as young as 5 years old can provide reliable and valid self-reports of global self-esteem (e.g., Marsh et al., 1998).

Our tentative conclusion is that children as young as 5 can provide reports of global self-esteem with acceptable levels of reliability and validity. This conclusion is based on the three lines of reasoning. Children’s scores were remarkably stable across 1 year, converged reasonably with parent and teacher ratings of their self-esteem, and were meaningfully correlated with relevant variables (e.g., attachment, narcissism). Most important, children’s responses on the LSE had similar properties as those of older children and adults. These findings suggest that self-report assessments are a reasonable way to assess self-esteem in young children. Following the cautions of previous scholars, we took several steps to address the concerns raised over the use of self-report measures with young children. First, although pictorial formats can sometimes hinder consistent
responses to a scale (see Davis-Kean & Sandler, 2001), the faces of the LSE seemed to aid children’s comprehension of the response options. In addition, the faces helped them understand the content of the items, possibly because the images corresponded well with the item wording (e.g., “How do you feel …”). Indeed, other researchers using pictorial scales to assess emotions have collected reliable responses from children as young as 4 years old (Lagattuta et al., 2012). Second, although the reliability of a scale tends to increase with more items, we found evidence that a small number of items can produce scores that demonstrate internal consistency in young children. Based on these efforts, we encourage future researchers to consider these and other issues proposed by Davis-Kean and Sandler (2001), de Leeuw (2011), and Lagattuta et al. (2012) when developing further self-report measures with young children. For example, although it is unclear whether children’s self-reports of narcissism are meaningful at these young ages, this study suggests that self-report scale development efforts at this age for constructs other than self-esteem represent a fruitful area for future research.

Some discussion is warranted regarding our inability to find evidence for some of the age and gender differences in global self-esteem reported in past studies. On the one hand, these results generalized across all the global measure of self-esteem in this study and thus do not appear to be unique to the LSE in these data. For age, the LSE patterns replicated past findings such that younger children reported the highest levels of self-esteem, adolescents reported relatively lower levels than young children, and adults (after age 50) were somewhat in between. However, two discrepancies from some past research were that LSE scores were lowest in middle adulthood (around ages 30–40) and were not relatively lower in the oldest age groups. Previous researchers have found normative increases in global self-esteem across young and middle adulthood (Bleidorn et al., 2016; Orth, Maes, & Schmitt, 2015; Orth, Robins, & Widaman, 2012) and a substantial drop in older adulthood (e.g., Orth, Trzesniewski, & Robins, 2010; Robins et al., 2002). However, other researchers have found both age patterns to be moderated by various factors. For example, Erol and Orth (2011) found a slow increase during adulthood that was moderated by ethnicity such that Blacks and Hispanics gradually increased, whereas Whites decreased from ages 26 to 30. In addition, others have found levels in both middle adulthood and old age to be moderated by health and wealth (see Orth & Robins, 2014; Orth et al., 2010; Wagner, Lang, Neyer, & Wagner, 2013). For instance, Orth et al. (2010) found that after controlling for the time-varying covariates of income, employment, functional health, and chronic health conditions, self-esteem was lower in middle adulthood and only declined slightly from ages 80 to 100. Therefore, it might be the case that the middle adults in the samples used here had relatively low levels of health and wealth, whereas the older adults had relatively high levels of health and wealth, both situations potentially due to our sampling strategy (e.g., using Internet panelists). Future studies should use representative sampling strategies and continue to test for moderators of the age trajectory of self-esteem. The advantage of the LSE is that it is short and thus well suited for such studies, which are typically expensive to conduct.

Besides age differences, we found that LSE scores did not vary by gender, a finding that is inconsistent with some past studies showing that males tend to report higher self-esteem than females in samples spanning in age from 7 to 90 (Kling et al., 1999; Robins et al., 2002; Steiger et al., 2014). However, the effect sizes for these gender differences are generally small (around 0.10–0.20), and one study did not find a gender difference using the RSE with individuals between 16 and 97 years old (Orth et al., 2012). Therefore, the lack of gender difference in LSE scores might not be that surprising (see also Zuckerman, Li, & Hall, 2016).

Beyond these issues, there are other limitations of our work. First, our samples were limited by the relative lack of diversity.
The majority of children in our study were White, middle-class children attending private schools in northern California, and the majority of adolescents and adults spoke English and were from the United States. We had some representation of lower socioeconomic status and different countries of origin, and although the psychometric properties of the LSE generally did not vary by ethnicity or education level, we acknowledge the need to replicate this study in more diverse samples. Cross-cultural validity of the LSE will be an important aspect of future research.

In sum, the LSE is a useful scale for an array of research projects, including large-scale studies assessing multiple constructs (given its short length of only four items), use with populations with shorter attention spans and limited vocabulary skills (e.g., young children), and administration across developmental age groups. We hope the availability of this tool will increase the kinds of studies designed to evaluate the development and correlates of self-esteem across the lifespan and lead to greater insights into the nature of self-esteem. We hope the LSE proves to be a useful tool for studying important questions about the origins and developmental trajectories of self-esteem. Indeed, the short LSE might prove especially valuable in large national and cross-national studies where survey space is at a premium.

**Funding**

This material is based on work supported by the National Science Foundation Graduate Research Fellowship under Grant No. 1650042. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

**References**


