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

SCHOLARONE™
Manuscripts

Lifespan Self-Esteem   1

Abstract

This paper 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, internally consistent, and 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.

        Keywords: *global self-esteem, self-esteem development, age differences, self-esteem measurement, scale construction*

Word count: 200/200

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 (Orth & 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 may 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., Alessandri, Vecchione, Eisenberg, & Laguna, 2015), and the wording of specific items may 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 the present 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

William James (1985/1892) first described self-esteem as the "ratio of *our actualities* to *our supposed potentialities*" (italics added, p. 54). 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, & Barnes, 1985; Wigfield, Eccles, Mac Iver, Reuman, & Midgley, 1991). Although domain specific measures have been valuable for informing specific outcomes such as sport performance (Marsh, Gerlach, Trautwein, Lüdtke, & Brettschneider, 2007), physical activity (Crocker, Sabiston, Kowalski, McDonough, & Kowalski, 2006), academic motivation and achievement (see Bong & Skaalvik, 2003), global self-esteem has been more extensively studied in the literature than has domain self-esteem (see Donnellan et al., 2015) and is the focus of the current scale development efforts.

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, 2012, 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 three and five, 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 five (the youngest age self-report global self-esteem has been recorded) provide self-esteem reports that are reliable, unidimensional (Marsh, Craven, & Debus, 1991; van den Bergh & de Rycke, 2003), consistent across one 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 may be formed from an early age (see Crocker & Park, 2012). At the same time, certain contingencies may 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 & Pike, 1984; Harter, 1982, 2012) and

Marsh's Self-Description Questionnaire (SDQ: e.g., Marsh et al., 1991, 1998; Marsh, Parker, &

Barnes, 1985; Marsh & O'Neill, 1984; Marsh, Smith, & Barnes, 1985). These existing measures

have seemingly adequate psychometric properties for assessing self-esteem within particular

developmental periods (reviewed in Donnellan, Trzesniewski, & Robins, 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 on pp. 9-10 in the Supplemental Online Materials (SOM) contain more detailed comparisons of

the changes in the scales across developmental periods, and only the main issues will be

summarized here. For the SDQ: (1) 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, (2) the total number of items

varies across scale versions, (3) the number of positively and negatively worded items varies, (4)

the descriptive language varies, and (5) the ordering of items varies across scale versions. For

Harter's self-esteem inventories: (1) the total number of items in the global self-esteem subscales

varies across versions, (2) the adolescent and college student versions each change the stem of

the items, (3) the qualifier is removed or changed for the emotion (e.g., very happy vs. happy),

(4) the descriptive language changes from, for example, "happy/unhappy" to

"pleased/disappointed," (5) the ordering of items changes, (6) there is no global subscale for

children younger than age 8, and (7) the two-step, forced choice response format occasionally

proves cumbersome leading to a worrisome amount of unusable data (e.g., Donnellan,

Trzesniewski, & Robins, 2015; Wichstrøm, 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 the present study were to (1)

develop a short, unidimensional global self-esteem scale that can quickly be administered to

participants across the lifespan and (2) 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

---

[1] One single-item scale has been developed and tested with children as young as 9: the Single-Item Self-Esteem
Scale (SISE; Robins et al., 2001).

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, SPPC, 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 towards the self), which resulted in an initial pool of 13 items (see Table 3

on pp. 11-12 in the SOM 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 4 items. A 4-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 (i.e., a

3-item scale would just be identified). Finally, a 4-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 on pp. 11-17 in the

SOM.

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 dataset 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,* or 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

meets 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:

18-24, 25-29, 30-39, 40-49, 50-59, 60-69, 70-79, and 80-89. Age was selected in three ways: (1)

the title of the survey described the desired age group, (2) the project page repeated the desired

age group above the link to the survey, and (3) 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" from the two options of agree and 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.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Sample 2. Due to low responses rates in middle adulthood (ages 39 to 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, who 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 39-49 and 100 participants between ages 50-59).

Age was selected in three ways: (1) Clearvoice sent the link to individuals within the

requested age range, (2) 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 (3)

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 p. 1 in the SOM;

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 upon

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 to 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 p. 1 in the SOM) 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 to 17) were recruited in Spring 2014 from Qualtrics for

$6 per response, resulting in 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 Kindergarten through 8th grade 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 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 ($\beta = -.06$, $p = .21$).

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 1st grade classrooms were split into smaller groups of students (range 3 to 6 children per group) for survey administration and data collection. Students

took about one hour to complete the survey, and sessions were split into two, half-hour sessions

for Kindergarten and 1st grade classes.

Children in the school sample completed four and two practice questions (four for LSE

and two for Kerns' Security Scale) to learn the scales before administration of scales described

below. 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 was 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 one year later in

their school classrooms. Children were between 6 and 15 years old at the second time point.

Measures

Lifespan Self-Esteem   13

       Self-esteem measures. *Lifespan Self-Esteem scale (LSE).* The SOM (pp. 11-17) describes

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

towards yourself?" and "How do you feel about doubts towards 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 (1) additional redundancies

and (2) 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 smiley 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 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 smiley faces (see pp. 3-4 in the SOM for display of full scales).

The SOM (pp. 5-8) includes a description of the remaining measures used to evaluate convergent validity (i.e., Self-Perception Profile for Children global subscale; SPPC; Harter, 1982, 2012; Rosenberg Self-Esteem scale; RSE; Rosenberg, 1965; Single-Item Self-Esteem scale; SISE; Robins, Hendin, & Trzesniewski, 2001; Self-Description Questionnaire global subscale; SDQ; 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' 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 through 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, 2,064) = 21.31, MSE = 13.54, p < .01; adjusted means (for age): 3.59

vs. 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)[2], assent administrator (F (11, 401) = 1.26, MSE = .68, p = .25, average SD =

0.77)[2], reader of the survey items (F (10, 402) = 1.67, MSE = .90, p = .09, average SD = 0.74)[2],

---

[2] There are too many groups to report all mean scores, but descriptive statistics are available upon request.

research assistants (F (7, 322) = .42, MSE = .23, p = .89, average SD = 0.75)[2], 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 will not be 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: 5-7, 8-13, 14-17, 18-24, 25-29, 30-39, 40-49, 50-59, 60-69, 70-79, and 80-89[3]. 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 accounting for the nested data (output are available

upon 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 exploratory factor

analysis (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 one 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

---

[3] The final sample included one individual who was 93 years old. However, we restricted the oldest age group to ages 80-89 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.

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 above 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 multi-group CFA using the pre-selected 11

age groups. We found a small (β = -.06, p = .01) effect of continuous age using the MIMIC

model, and multi-group models supported metric invariance across the lifespan (see pp. 19-21 in

the SOM 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 inter-item 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

to 15 years) who completed the LSE at two time points (separated by one year). We calculated a

correlation between students' scores across the two occasions. We expected the LSE to be

somewhat stable across one year, given stability coefficients found in past literature (meta-

analytic r = .31 for 6-8-year-olds, k = 5; Trzesniewski, Donnellan, & Robins, 2003).

Surprisingly, the one-year test-retest 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 5-7-year-olds (r = .48, p < .01)

and 8-13-year-olds ($r = .62$, $p < .01$) showed consistent scores one 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 ($r = .29_{\text{teachers}}$, $p < .01$; $r = .26_{\text{parents}}$, $p = .04$). The correlation

between teachers' and children's ratings was not significantly different from the correlation

between parent and children's ratings across all grades ($z = .17$, $p > .25$). Regressions indicated

that age did not moderate the relation between teachers and children ($\beta = -.29$, $p > .25$), or

between parents and children ($\beta = .43$, $p > .25$). Specifically, the correspondence for children

aged 5 to 7 was .23 ($p = .05$) with teachers and .16 ($p > .25$) with parents, and the

correspondence for children aged 8 to 12 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 (r's = .61 to .73, all p's < .01; disattenuated r's = .70 to .84; see Table 5).

    Correlates of Self-Esteem. The LSE had evidence of criterion-related validity: scores

were consistently correlated 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

---

[4] All effect sizes reported are standardized beta coefficients.

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 interactions predicting LSE were

nonsignificant); however, some correlations were significantly moderated by age. These are

summarized in the SOM (p. 23) 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.

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, Trzesniewski, Potter, & Gosling, 2001), narcissism (Ackerman

& Donnellan, 2013), depression (Ulrich 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 SOM (p. 18) 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

$= .02$, $\beta = -.02$, $p = .41$, $R^2 = .00$; quadratic $B = .00$, SE $= .00$, $\beta = .23$, $p < .01$, $R^2 = .04$; cubic B

$= -.00$, SE $= .00$, $\beta = -.37$, $p < .01$, $R^2 = .06$). 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 through 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 $= .14$, $p > .25$; means: US $= 3.66$, not US $= 3.80$, average $SD = 0.90$), *gender* (F (1,

2,453) = .60, MSE = .41, p > .25; means: male = 3.74; female = 3.76, average SD = 0.86),

ethnicity (F (5, 2,119) = 2.44, MSE = 1.62, p = .03, average SD = 0.85)[5], *education* (for Samples

1 through 3: r = -.04, p = .07, average SD = 0.77)[3], *adolescent mother education* (Sample 4: r =

.10, p = .09, average SD = 0.94)[3], or *adolescent father education* (Sample 4: r = -.03, p > .25,

average SD = 0.89)[3] (see SOM pp. 27-36 for descriptive statistics for all groups).

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.

That is, 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 8-13 on pp. 27-

36 of the SOM 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, Trzesniewski, & Robins, 2011 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

---

[5] There are too many groups to report all mean scores, but descriptive statistics are reported in the SOM (pp. 27-36).

Lifespan Self-Esteem   23

antecedents and consequences of self-esteem across developmental periods. Thus, we created the

Lifespan Self-Esteem (LSE) scale 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,

2012) 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 one year, converged

reasonably with parent and teacher ratings of their self-esteem, and were meaningfully correlated

with relevant variables (e.g., attachment, narcissism). Most importantly, 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 smiley faces of the

LSE seemed to aid children's comprehension of the response options. In addition, the smiley

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

(2012), 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, the current study suggests that self-report scale development

efforts at this age for constructs other than self-esteem is 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 the current 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 somewhere 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., 2015; 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 current

samples 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,

Allemand, Robins, & Fend, 2014b). 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, Robins, & Widaman, 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 the current study in more diverse samples. Cross-cultural validity of the LSE will be an important step 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 for 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.

References

Ackerman, R. A., & Donnellan, M. B. (2013). Evaluating self-report measures of narcissistic entitlement. *Journal of Psychopathology and Behavioral Assessment*, *35*(4), 460–474. Journal Article. http://doi.org/10.1080/00223891.2012.732636 . http://dx.doi.org/10.1007/s10862-013-9352-7

Alessandri, G., Vecchione, M., Eisenberg, N., & Laguna, M. (2015). On the factor structure of

Lifespan Self-Esteem   27

the Rosenberg (1965) General Self-Esteem Scale. *Psychological Assessment*, *27*, 621–635.

Ames, D. R., Rose, P., & Anderson, C. P. (2006). The NPI-16 as a short measure of narcissism.

Journal of Research in Personality, *40*, 440–450. Journal Article.

Back, M. D., Kufner, A. C. P., Dufner, M., Gerlach, T. M., Rauthmann, J. F., & Denissen, J. J.

A. (2013). Narcissistic admiration and rivalry: Disentangling the bright and dark sides of

narcissism. Journal Article.

Barry, C. T., Frick, P. J., & Killian, A. L. (2003). The relation of narcissism and self-esteem to

conduct problems in children: A preliminary investigation. *Journal of Clinical Child &*

Adolescent Psychology, *32*, 139–152. Journal Article.

Behrend, T. S., Sharek, D. J., Meade, A. W., & Wiebe, E. N. (2011). The viability of

crowdsourcing for survey research. *Behavior Research Methods*, *43*, 800–813.

http://doi.org/10.3758/s13428-011-0081-0

Bleidorn, W., Arslan, R. C., Denissen, J., Rentfrow, P., Gebauer, J., Potter, J., & Gosling, S.

(2015). Age and gender differences in self-esteem — A cross-cultural window. *Journal of*

Personality and Social Psychology, *Advance on*, 1–15.

Bong, M., & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are

they really? *Educational Psychology Review*, *15*, 1–40.

http://doi.org/10.1023/A:1021302408382

Brennan, K. A., Clark, C. L., & Shaver, P. R. (1998). Self-report measurement of adult

attachment. In J. A. Simpson & W. S. Rholes (Eds.), *Attachment Theory and Close*

Relationships (pp. 46–76). New York, NY: Guilford Press.

Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. Guilford Publications.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source

of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, *6*, 3–5.

http://doi.org/10.1177/1745691610393980

Cole, J. C., Rabin, A. S., Smith, T. L., & Kaufman, A. S. (2004). Development and validation of

a Rasch-derived CES-D short form. *Psychological Assessment*, *16*, 360–372.

http://doi.org/10.1037/1040-3590.16.4.360

Crocker, J., & Park, L. E. (2012). Contingencies of self-worth. In M. R. Leary & T. J (Eds.),

Handbook of self and identity (2nd ed., pp. 309–326). New York, NY: The Guilford Press.

Crocker, P. R. E., Sabiston, C. M., Kowalski, K. C., McDonough, M. H., & Kowalski, N. (2006).

Longitudinal Assessment of the Relationship Between Physical Self-Concept and Health-

Related Behavior and Emotion in Adolescent Girls. *Journal of Applied Sport Psychology*,

18(3), 185–200. JOUR. http://doi.org/http://dx.doi.org/10.1080/10413200600830257

Davis-Kean, P. E., & Sandler, H. M. (2001). A meta-analysis of measures of self-esteem for

young children: A framework for future measures. *Child Development*, *72*, 887–906.

Journal Article.

de Leeuw, E. D. (2012). Improving data quality when surveying children and adolescents:

Cognitive and social development and its role in questionnaire construction and pretesting.

In *Meeting of the Advisory Scientific Board of Statistics*. Stengevik, Sweden.

Diener, E., Emmons, R., Larsen, J., & Griffin, S. (1985). The Satisfaction with Life Scale.

Journal of Personality Assessmemt, *49*, 71–75.

http://doi.org/10.1207/s15327752jpa4901_13

Donnellan, M. B., Oswald, F. L., Baird, B. M., & Lucas, R. E. (2006). The mini-IPIP scales:

Tiny-yet-effective measures of the Big Five factors of personality. *Psychological*

Assessment, *18*, 192–203. http://doi.org/10.1037/1040-3590.18.2.192

Donnellan, M. B., Trzesniewski, K. H., & Robins, R. W. (2015). Assessing self-esteem. In G. J.

Boyle, D. H. Saklofske, & G. Matthews (Eds.), *Measures of Personality and Social*

Psychological Constructs (pp. 7–15). Elsevier.

Donnellan, M. B., Trzesniewski, K. H., Robins, R. W., Moffitt, T. E., & Caspi, A. (2005). Low

self-esteem is related to aggression, antisocial behavior, and delinquency. *Psychological*

Science, *16*, 328–335. Journal Article.

Erol, R. Y., & Orth, U. (2011). Self-esteem development from age 14 to 30 years: A longitudinal

study. *Journal of Personality and Social Psychology*, *101*, 607–619.

http://doi.org/10.1037/a0024299

Fan, X., & Thompson, B. (2001). Confidence intervals for effect sizes. *Educational and*

Psychological Measurement, *61*, 517–531.

Fivush, R., & Nelson, K. (2004). Culture and language in the emergence of autobiographical

memory. [References]. *Psychological Science*, *15*(9), 573–577.

http://doi.org/10.1111/j.0956-7976.2004.00722.x

Furr, R. M. (2010). The double-entry intraclass correlation as an index of profile similarity:

Meaning, limitations, and alternatives. *Journal of Personality and Social Assessment*, *92*, 1–

15. http://doi.org/10.1080/00223890903379134

Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data collection in a flat world: The

strengths and weaknesses of Mechanical Turk samples. *Journal of Behavioral Decision*

Making, *26*, 213–224. http://doi.org/10.1002/bdm.1753

Hambleton, R. K. (1991). Fundamentals of item response theory. Sage Publications.

Harter, S. (1982). The Perceived Competence Scale for Children. *Child Development*, *53*(1), 87–

97. Journal Article.

Harter, S. (2012a). Emerging self-processes during childhood and adolescence. In M. R. Leary &

J. Tangney (Eds.), *Handbook of self and identity* (2nd ed., pp. 680–715). New York, NY:

The Guilford Press.

Harter, S. (2012b). *Self-Perception Profile for Children manual and questionnaires*.

Harter, S., & Pike, R. (1984). The Pictorial Scale of Perceived Competence and Social

Acceptance for young children. *Child Development*, *55*, 1969–1982. Journal Article.

http://doi.org/10.2307/1129772

James, W. (1985). *Psychology: The briefer course*. Notre Dame: University of Notre Dame

Press.

Jöreskog, K. G., & Goldberger, A. S. (1975). Estimation of a model with multiple indicators and

multiple causes of a single latent variable. *Journal of the American Statistical Association*,

70, 631–639. http://doi.org/10.2307/2285946

Kerns, K. A., Klepac, L., & Cole, A. (1996). Peer relationships and preadolescents' perceptions

of security in the child-mother relationship. *Developmental Psychology*, *32*, 457–466.

Journal Article.

Kling, K. C., Hyde, J. S., Showers, C. J., Buswell, B. N., Garvey, A., Glassburn, D., … Onorato,

N. (1999). Gender differences in self-esteem: A meta-analysis. *Psychological Bulletin*, *125*,

Lifespan Self-Esteem 31

470–500.

Lagattuta, K. H., Sayfan, L., & Bamford, C. (2012). Do you know how I feel? Parents

underestimate worry and overestimate optimism compared to child self-report. *Journal of*

Experimental Child Psychology, *113*, 211–232. http://doi.org/10.1016/j.jecp.2012.04.001

Leary, M. R. (2006). What are the most pressing issues facing researchers? In M. H. Kernis

(Ed.), *Self-esteem issues and answers: A sourcebook of current perspectives* (pp. 424–429).

New York, NY: Psychology Press.

Little, T. D., & Rhemtulla, M. (2013). Planned missing data designs for developmental

researchers. *Child Development Perspectives*, *7*, 199–204.

http://doi.org/10.1111/cdep.12043

Marsh, H., Craven, R., & Debus, R. (1998). Structure, stability, and development of young

children's self-concepts: A multicohort-multioccasion study. *Child Development*, *69*, 1030–

1053. Journal Article.

Marsh, H. W., Craven, R. G., & Debus, R. (1991). Self-concepts of young children 5 to 8 years

of age: Measurement and multidimensional structure. *Journal of Educational Psychology*,

83, 377–392. Journal Article.

Marsh, H. W., Gerlach, E., Trautwein, U., Lüdtke, O., & Brettschneider, W.-D. (2007).

Longitudinal study of preadolescent sport self-concept and performance: Reciprocal effects

and causal ordering. *Child Development*, *78*(6), 1640–1656. JOUR.

http://doi.org/http://dx.doi.org/10.1111/j.1467-8624.2007.01094.x

Marsh, H. W., & O'Neill, R. (1984). Self description questionnaire III: The construct validity of

multidimensional self-concept ratings by late adolescents. *Journal of Educational*

Measurement, *21*, 153–174. http://doi.org/10.1111/j.1745-3984.1984.tb00227.x

Marsh, H. W., Parker, J., & Barnes, J. (1985). Multidimensional adolescent self-concepts: Their relationship to age, sex, and academic measures. *American Educational Research Journal*, *22*, 422–444. http://doi.org/10.3102/00028312022003422

Marsh, H. W., Smith, I. D., & Barnes, J. (1985). Multidimensional self-concepts: Relations with sex and academic achievement. *Journal of Educational Psychology*, *77*, 581–596. http://doi.org/10.1037/0022-0663.77.5.581

Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods*, *44*, 1–23. http://doi.org/10.3758/s13428-011-0124-6

Muthén, L. K., & Muthén, B. O. (n.d.). Mplus User's Guide. Los Angeles, CA: Muthén & Muthén.

Orth, U., Maes, J., & Schmitt, M. (2015). Self-esteem development across the life span: A longitudinal study with a large sample from Germany. *Developmental Psychology*, *51*, 248–259. http://doi.org/http://dx.doi.org/10.1037/a0038481

Orth, U., & Robins, R. W. (2014). The development of self-esteem. *Current Directions in Psychological Science*, *23*, 381–387. http://doi.org/10.1177/0963721414547414

Orth, U., Robins, R. W., & Widaman, K. F. (2012). Life-span development of self-esteem and its effects on important life outcomes. *Journal of Personality and Social Psychology*, *102*(6), 1271–1288. JOUR. http://doi.org/http://dx.doi.org/10.1037/a0025558

Orth, U., Robins, R. W., Widaman, K. F., & Conger, R. D. (2014). Is low self-esteem a risk factor for depression? Findings from a longitudinal study of Mexican-origin youth.

Developmental Psychology, *50*(2), 622–633. JOUR.

http://doi.org/http://dx.doi.org/10.1037/a0033817

Orth, U., Trzesniewski, K. H., & Robins, R. W. (2010). Self-esteem development from young

adulthood to old age: A cohort-sequential longitudinal study. *Journal of Personality and*

Social Psychology, *98*, 645–658. Journal Article. http://doi.org/10.1037/a0018769

Robins, R. W., Hendin, H. M., & Trzesniewski, K. H. (2001). Measuring global self-esteem:

Construct validation of a single-item measure and the Rosenberg Self-Esteem Scale.

Personality and Social Psychology Bulletin, *27*, 151–161. Journal Article.

http://doi.org/10.1177/0146167201272002

Robins, R. W., Tracy, J. L., Trzesniewski, K. H., Potter, J., & Gosling, S. D. (2001). Personality

correlates of self-esteem. *Journal of Research in Personality*, *35*(4), 463–482. Journal

Article. http://doi.org/10.1006/jrpe.2001.2324

Robins, R. W., Trzesniewski, K. H., Tracy, J. L., Gosling, S. D., & Potter, J. (2002). Global self-

esteem across the life span. *Psychology and Aging*, *17*(3), 423–434. Journal Article.

http://doi.org/10.1037/0882-7974.17.3.423

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton

University Press.

Rosenberg, M. (1989). *Society and the adolescent self-image* (Revised Ed). Book, Middletown,

CN: Wesleyan University Press.

Rosnow, B. R. L., Rosenthal, R., & Rubin, D. B. (2000). Contrasts and correlations in effect-size

estimation. *Psychological Science*, *11*, 446–453.

Scheier, M. F., Carver, C. S., & Bridges, M. W. (1994). Distinguishing optimism from

neuroticism (and trait anxiety, self-mastery, and self-esteem): A reevaluation of the Life

Orientation Test. *Journal of Personality and Social Psychology*, *67*, 1063–1078.

http://doi.org/10.1037/0022-3514.67.6.1063

Schmitt, D. P., & Allik, J. (2005). Simultaneous administration of the Rosenberg Self-Esteem

Scale in 53 nations: exploring the universal and culture-specific features of global self-

esteem. *Journal of Personality and Social Psychology*, *89*, 623–642.

http://doi.org/10.1037/0022-3514.89.4.623

Sowislo, J. F., & Orth, U. (2013). Does low self-esteem predict depression and anxiety? A meta-

analysis of longitudinal studies. *Psychological Bulletin*, *139*, 213–240.

http://doi.org/10.1037/a0028931

Steiger, A. E., Allemand, M., Robins, R. W., & Fend, H. A. (2014a). Low and decreasing self-

esteem during adolescence predict adult depression two decades later. *Journal of

Personality and Social Psychology*, *106*, 325–38. http://doi.org/10.1037/a0035133

Steiger, A. E., Allemand, M., Robins, R. W., & Fend, H. A. (2014b). Low and decreasing self-

esteem during adolescence predict adult depression two decades later. *Journal of

Personality and Social Psychology*, *106*, 325–338. http://doi.org/10.1037/a0035133

Steiger, A. E., Fend, H. A., & Allemand, M. (2015). Testing the vulnerability and scar models of

self-esteem and depressive symptoms from adolescence to middle adulthood and across

generations. *Developmental Psychology*, *51*(2), 236–247. JOUR.

http://doi.org/http://dx.doi.org/10.1037/a0038478

Trzesniewski, K. H., Donnellan, M. B., Moffitt, T. E., Robins, R. W., Poulton, R., Caspi, A., …

Lifespan Self-Esteem   35

Fend, H. A. (2006). Low self-esteem during adolescence predicts poor health, criminal

behavior, and limited economic prospects during adulthood. *Developmental Psychology*, *42*,

381–390. Journal Article. http://doi.org/10.1037/0012-1649.42.2.381

Trzesniewski, K. H., Donnellan, M. B., & Robins, R. W. (2003). Stability of self-esteem across

the life span. *Journal of Personality and Social Psychology*, *84*, 205–220. Journal Article.

http://doi.org/10.1037/0022-3514.84.1.205

van den Bergh, B. R. H., & de Rycke, L. (2003). Measuring the multidimensional self-concept

and global self-worth of 6- to 8-year-olds. *The Journal of Genetic Psychology: Research

and Theory on Human Development*, *164*, 201–225. Journal Article.

Wagner, J., Lang, F. R., Neyer, F. J., & Wagner, G. G. (2013). Self-esteem across adulthood:

The role of resources. *European Journal of Ageing*, *11*(2), 109–119. JOUR.

http://doi.org/http://dx.doi.org/10.1007/s10433-013-0299-z

Wigfield, A., Eccles, J. S., Mac Iver, D., Reuman, D. A., & Midgley, C. (1991). Transitions

during early adolescence: Changes in children's domain-specific self-perceptions and

general self-esteem across the transition to junior high school. *Developmental Psychology*,

27, 552–565.

Zuckerman, M., Li, C., & Hall, J. A. (2016). When men and women differ in self-esteem and

when they don't: A meta-analysis. *Journal of Research in Personality*, *64*, 34–51.

	Really Sad	Sad	Neutral	Happy	Really Happy
1. How do you feel about yourself?					
2. How do you feel about the kind of person you are?					
3. When you think about yourself, how do you feel?					
4. How do you feel about the way you are?					

Figure 1. Items and response options for the Lifespan Self-Esteem scale (LSE).
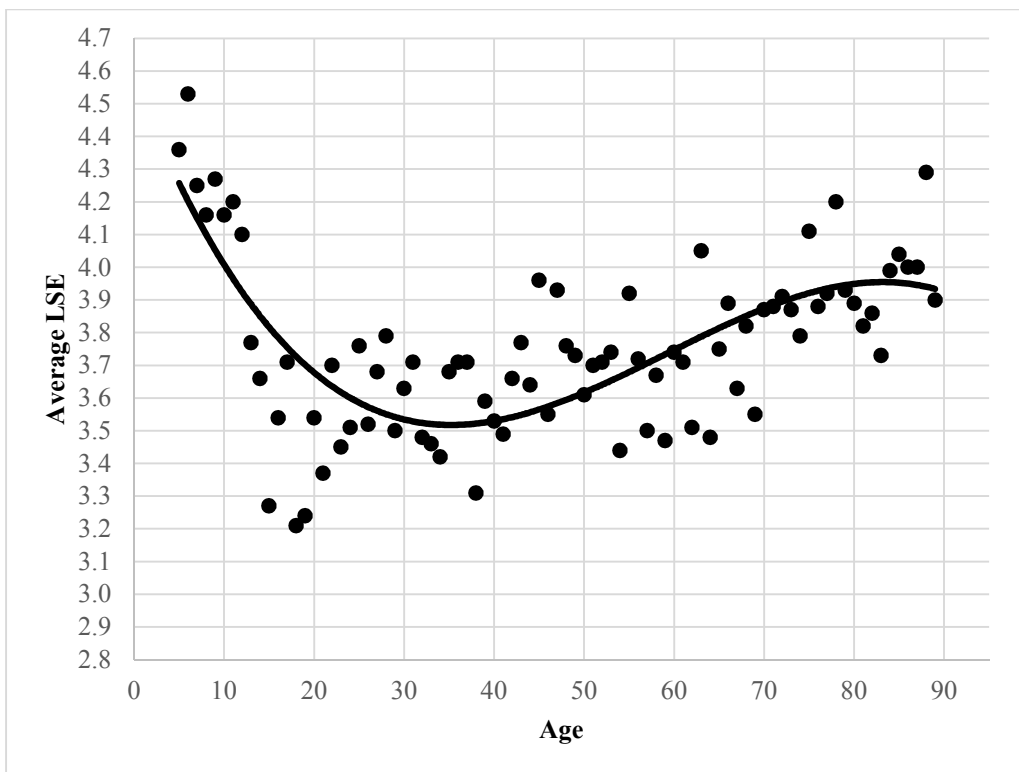
Figure 2. Mean Lifespan Self-Esteem scale (LSE) scores by continuous age.

Lifespan Self-Esteem 1

Table 1

Measures Administered to Samples 1 through 5

Sample	1	2	3	4	5
Age range	18 to 93	39 to 59	60 to 89	14 to 17	5 to 13
N	1,413	201	451	211	438
% Male	48%	38%	41%	23%	45%
% From US	96%	93%	97%	89%	NA
% White or Caucasian	75%	83%	95%	60%	NA
% BA or lower	84%	82%	82%	81% (Mom), 81% (Dad)	NA
LSE	✔	✔	✔	✔	✔ Preceded by four practice questions
SPPC	✔	✔	✔	✔	
RSE	✔	✔	✔	✔	
SISE	✔ 5-point scale	✔ 6-point scale	✔ 6-point scale	✔ 6-point scale	✔ 5-point scale (LSE format)
SDQ	✔				
NARQ	✔	✔	✔	✔	
CES-D	✔	✔	✔	✔	
ECR-R	✔ 3-point scale	✔ 7-point scale	✔ 7-point scale	✔ 7-point scale	
Mini-IPIP	✔	✔	✔	✔	
Parent attachment				✔	✔ Preceded by two practice questions
NPI-16-C				✔	✔
LSE-parent					✔
LSE-teacher					✔
Self-esteem explanation	✔ provided with 6 examples and instructed, "Please check which of the following…"	✔ Did not receive checklist	✔ provided with 6 examples and instructed, "Please check which of the following…"	✔ provided with 6 examples and instructed, "Please check which of the following…"	✔ provided with 6 examples and instructed, "If you read these, read all of them…"; only given to third school
Trait mood					✔
LOT-R	✔	✔	✔		
SWLS	✔	✔	✔		

Note. N = sample size, 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-

Lifespan Self-Esteem   2

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.

Table 2

Descriptive Statistics for Practice Items for the LSE with Child Sample 5

	Ages 5-7			Ages 8-10		
Item	Min-Max	*M*	*SD*	Min-Max	*M*	*SD*
"How do you feel about the chores you do at home?"	1-5	3.62	1.17	1-5	3.17	0.79
"How do you feel about getting shots from the doctor?"	1-5	2.43	1.47	1-5	2.72	1.09
"How do you feel about getting presents for your birthday?"	2-5	4.88	0.47	3-5	4.83	0.41
"How would you feel about being eaten by a T-Rex?"	1-5	2.03	1.51	1-5	1.66	1.12

Note. Min = minimum, max = maximum, *M* = mean, *SD* = standard deviation.

Table 3

Factor Loadings and Eigenvalues from Exploratory Factor Analysis on the LSE by Age Group

Items	Age Group										
	5-7	8-13	14-17	18-24	25-29	30-39	40-49	50-59	60-69	70-79	80-89
"How do you feel about yourself?"	.64	.78	.86	.83	.85	.83	.81	.84	.79	.69	.88
"How do you feel about the kind of person you are?"	.68	.68	.68	.66	.77	.75	.76	.79	.69	.64	.8
"When you think about yourself, how do you feel?"	.55	.80	.86	.83	.88	.86	.92	.92	.84	.82	.87
"How do you feel about the way you are?"	.61	.85	.83	.84	.87	.87	.87	.83	.88	.84	.86
Initial Eigenvalue	2.16	2.34	2.78	2.69	3.05	2.87	2.99	3.12	2.83	2.65	3.13
Second Eigenvalue	0.72	0.76	0.54	0.64	0.43	0.50	0.47	0.41	0.63	0.66	0.42

Table 4

Alpha Coefficients and Inter-Item Reliability for the LSE by Age Group

Age group	N	Cronbach's Alpha	95% CI		Inter-Item r
All	2,542	.89	.88	.90	.67
5-7	111	.71	.61	.79	.39
8-13	246	.86	.83	.89	.60
14-17	356	.88	.85	.90	.65
18-24	274	.87	.84	.89	.62
25-29	245	.91	.88	.92	.70
30-39	253	.90	.87	.92	.68
40-49	241	.90	.88	.92	.70
50-59	227	.91	.89	.93	.72
60-69	199	.88	.85	.90	.64
70-79	196	.84	.79	.87	.56
80-89	197	.91	.89	.93	.72

Note. N = sample size, 95% CI = 95% Confidence Intervals, Inter-Item r = Inter-Item Correlation.

Table 5

Convergent Validity Correlations between the LSE and Existing Self-Esteem Measures

	SDQ	SPPC	RSE	SISE
LSE Items and Composite				
"How do you feel about yourself?"	.65[*]	.53[*]	.61[*]	.62[*]
"How do you feel about the kind of person you are?"	.57[*]	.48[*]	.54[*]	.52[*]
"When you think about yourself, how do you feel?"	.67[*]	.56[*]	.62[*]	.64[*]
"How do you feel about the way you are?"	.65[*]	.58[*]	.61[*]	.60[*]
LSE	.73[*]	.61[*]	.68[*]	.69[*]
Existing Self-Esteem Measures				
SDQ	-			
SPPC	.73*	-		
RSE	.88*	.66*	-	
SISE	.80*	.60*	.67*	-
Alpha reliability	.96	.85	.91	.75[a]
N	1,247	2,082	2,127	2,516

Note. *p < .01. SDQ = Self-Description Questionnaire global subscale, SPPC = Self-Perception Profile for Children global subscale, RSE = Rosenberg Self-Esteem scale, SISE = Single-Item Self-Esteem scale, LSE = Lifespan Self-Esteem scale, *N* = sample size. [a]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).

Table 6

Criterion Validity and Nomological Network Correlations with the LSE

Item	Parent attachment	Avoidance	Anxiety	Admiration	Rivalry	NPI-16-C	CES-D	O	C	E	A	N	Mood	LOT-R	SWLS
1	.48*	.23*	.27*	.35*	-.18*	.14*	-.55*	.08*	.29*	.28*	.16*	-.48*	.44*	.58*	.54*
2	.37*	.26*	.21*	.35*	-.23*	.10	-.44*	.12*	.32*	.23*	.27*	-.40*	.35*	.50*	.43*
3	.44*	.22*	.30*	.38*	-.16*	.17*	-.54*	.02	.30*	.29*	.14*	-.48*	.37*	.59*	.54*
4	.48*	.25*	.27*	.37*	-.18*	.14*	-.51*	.08*	.34*	.26*	.17*	-.45*	.41*	.55*	.51*
N	619	2,123	2,122	2,147	2,147	551	2,132	1,873	1,873	1,875	1,873	1,873	482	1,857	1,862
LSE	.53*	.27*	.30*	.42*	-.21*	.16*	-.59*	.09*	.36*	.30*	.21*	-.52*	.48*	.64*	.58*

Note. *p* < .01. NPI-16-C = Narcissistic Personality Inventory 16 Adapted for Children, CES-D = Center for Epidemiologic Studies – Depression scale, O = Openness, C = Conscientiousness, E = Extraversion, A = Agreeableness, N = Neuroticism, LOT-R = Life Orientation Test – Revised, SWLS = Life Orientation Test – Revised.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

Table 7

Criterion Validity and Nomological Network Correlations for the LSE and Established Self-Esteem Measures

	Parent attachment	Avoidance	Anxiety	Admiration	Rivalry	NPI-16-C	CES-D	O	C	E	A	N	Mood	LOT-R	SWLS
LSE	.53*	.27*	.30*	.42*	-.21*	.16*	-.59*	.09*	.36*	.30*	.21*	-.52*	.48*	.64*	.58*
SDQ	-	.40*	.52*	.39*	-.28*	-	-.71*	.21*	.45*	.40*	.25*	-.62*	-	.76*	.63*
SPPC	.41*	.24*	.29*	.31*	-.26*	.31*	-.54*	.11*	.36*	.26*	.25*	-.52*	-	.61*	.54*
RSE	.55*	.33*	.45*	.33*	-.34*	.32*	-.70*	.20*	.44*	.28*	.29*	-.59*	-	.71*	.55*
SISE	.46*	.21*	.26*	.45*	-.14*	.22*	-.54*	.04	.33*	.36*	.18*	-.52*	.59*	.63*	.55*
Mean r[1]	.48	.30	.39	.37	-.26	.28	-.63	.14	.40	.33	.24	-.56	.59	.68	.57

Note. *p < .01. Sample sizes ranged from 211 to 2,125. LSE = Lifespan Self-Esteem scale, SDQ = Self-Description Questionnaire global subscale, SPPC Self-Perception Profile for Children global subscale, RSE = Rosenberg Self-Esteem scale, SISE = Single-Item Self-Esteem scale, NPI-16-C = Narcissistic Personality Inventory 16 Adapted for Children, CES-D = Center for Epidemiologic Studies – Depression scale, , O = Openness, C = Conscientiousness, E = Extraversion, A = Agreeableness, N = Neuroticism, LOT-R = Life Orientation Test – Revised, SWLS = Satisfaction with Life Scale. [1]We transformed criterion validity correlations with all measures (omitting the LSE) to Fisher's z scores, averaged them, and then back transformed to the r metric.

1

Supplementary Online Materials

Table of Contents

2

Instructional Manipulation Check Items

Sample 2

Research in decision making shows that people, when making decisions and answering questions, prefer not to pay attention and minimize their effort as much as possible. Some studies show that over 50% of people don't carefully read questions. If you are reading this question and have read all the other questions, please select the box marked 'other' and type 'Decision Making' in the box below. Do not select "predictions of your own behavior." Thank you for participating and taking the time to read through the questions carefully!

What was this study about?

A Predictions of your own behavior
B Lions
C Tigers
D Other _____

Sample 3

For this survey, we want to make sure our respondents are taking the time to read and understand each question. Please show that you are paying attention by selecting 'Lions' from the list of options below.

A Predictions of your own behavior
B Lions
C Tigers
D Other _____

3

LSE-parent

Please use the rating scale below to answer the following questions about how you think your child feels about him/herself. Circle one face for each question.

	Really Sad	Sad	Neutral	Happy	Really Happy
1. How does your child feel about him/herself?					
2. How does your child feel about the kind of person s/he is?					
3. When your child thinks about him/herself, how does s/he feel?					
4. How does your child feel about the way s/he is?					

4

LSE-teacher

Development of the Self Study: Teacher survey   Student name: _____

Please complete one page for each child on the roster included in this envelope. You can return this to us at the end of the class period or to the school office. Thank you very much!

	Really Sad	Sad	Neutral	Happy	Really Happy
1. How does _____ feel about him/herself?	🙁	🙁	😐	🙂	😃
2. How does _____ feel about the kind of person s/he is?	🙁	🙁	😐	🙂	😃
3. When _____ thinks about him/herself, how does s/he feel?	🙁	🙁	😐	🙂	😃
4. How does _____ feel about the way s/he is?	🙁	🙁	😐	🙂	😃

Remaining Measures

Self-Perception Profile for Children (SPPC) Global Subscale (Ages 14 to 93)

The SPPC (Harter, 1982, 2012) global subscale consisted of 6 items assessing global self-worth (α

= .85). Items were presented in a two-step structured alternative format (e.g., individuals first chose

which of two descriptions were more like them and then indicate whether the descriptions were sort of

true for them or really true for them). Items were adapted by replacing the word 'children' with 'people'

for all participants. A sample item is, "Some people are often unhappy with themselves BUT other

people are pretty pleased with themselves." Responses were scored from 1 to 4.

Rosenberg Self-Esteem Scale (RSE; Ages 14 to 93)

The RSE (Rosenberg, 1965) consisted of 10 items assessing global self-esteem (α = .91). Items

were administered using a 5-point scale ranging from 1 (*Strongly Disagree*) to 5 (*Strongly Agree*). A

sample item is, "I feel that I have a number of good qualities."

Single-Item Self-Esteem Scale (SISE; Ages 5 to 93)

The SISE (Robins, Hendin, & Trzesniewski, 2001) was administered using a 5-point scale ranging

from 1 (*Strongly Disagree*) to 5 (*Strongly Agree*), and some participants saw a 6 (*Not Applicable*)

response option by error (see Table 1 in the manuscript; 1% of participants chose this option). The item

text was, "I see myself as someone who has high self-esteem." Alpha reliability cannot be computed

with one item, but retest reliability and validity information for this scale have been reported elsewhere

(Robins et al., 2001).

6

Self-Description Questionnaire (SDQ) Global Subscale (Ages 18 to 93)

The SDQ (Marsh et al., 1991) consisted of six items and was administered using an 8-point scale

ranging from 1 (*Definitely False*) to 8 (*Definitely True*). A sample item is, "Overall, I have a lot of self-

confidence." Alpha reliability was .96.

Narcissistic Personality Inventory 16 Adapted for Children (NPI-16-C; Ages 5 to 17)

To create a shorter measure of narcissism than the 40-item NPI-C (Barry, Frick, & Killian, 2003),

we adapted the wording from the NPI-16 (Ames, Rose, & Anderson, 2006) for use with children. That is,

we selected the items from the NPI-C that were used in the NPI-16 (i.e., items 4, 7, 9, 12, 13, 14, 21, 20,

23, 24, 30, 32, 24, 35, 39, and 40). The final measure included these 16 items in the original forced-

choice format of both previous narcissism measures (α = .76). A sample item is, "I am able to do more

things than other people, or I can learn a lot from other people."

Narcissistic Admiration and Rivalry Questionnaire (NARQ; Ages 14 to 93)

The NARQ (Back et al., 2013) contains 18 items, but the first item ("I am great") was omitted by

error. Therefore, the NARQ in this study had 17 items and used a 5-point scale ranging from 1 (*Strongly

Disagree*) to 5 (*Strongly Agree*). There were two subscales: admiration (8 items) and rivalry (9 items). A

sample item for admiration (α = .83) is, "I deserve to be seen as a great personality," and for rivalry (α =

.88) is, "I react annoyed if another person steals the show from me."

Center for Epidemiologic Studies – Depression Scale (CES-D) Short Form (Ages 14 to 93)

The short form of the CES-D (Cole, Rabin, Smith, & Kaufman, 2004) consisted of 10 items with a

response scale ranging from 1 (*Rarely or None of the Time*) to 4 (*Most of the Time*) regarding symptoms

of depression over the last two weeks. A sample item is, "I was bothered by things that don't usually

bother me." Alpha reliability was .87.

Experiences in Close Relationships – Revised (ECR-R; Ages 14 to 93)

The ECR-R (Brennan, Clark, & Shaver, 1998) consisted of 36 items with two subscales regarding

feelings of avoidance and anxiety in close relationships. A sample item for avoidance (α = .93) is, "I

prefer not to show a partner how I feel deep down," and a sample item for anxiety (α = .93) is, "I worry

about being abandoned." Some participants used a 3-point scale (1 = *Strongly Disagree*, 2 = *Neutral*, 3 =

Strongly Agree), and other participants used a 7-point scale ranging from 1 (*Strongly Disagree*) to 7

(*Strongly Agree*) due to error (see Table 1 in the manuscript). POMP scores (Cohen, Cohen, Aiken, &

West, 1999) were created separately for these samples, and the POMP-scored variables were used in all

of the analyses.

Kerns' Security Scale (Ages 5 to 17)

Kerns' Security Scale (Kerns, Klepac, & Cole, 1996) consisted of 15 items (α = .74) and was used

to assess participants' perceptions of relationships with their parents. We used a 5-point scale ranging

from 1 (*Not At All True For Me*) to 5 (*Really True For Me*), and response options were represented by

pictorial rectangles increasing in size. A sample item is, "I find it easy to trust my mom and dad." Practice

questions (see Table 1 in the manuscript for samples who received practice questions) were, "I find it

easy to draw," and "I do not like playing outside."

Mini-International Personality Item Pool (Mini-IPIP; Ages 14 to 93)

The Mini-IPIP (Donnellan, Oswald, Baird, & Lucas, 2006) consisted of 20 items intended to

measure the Big 5 personality trait domains. We used a 5-point scale ranging from 1 (*Very Inaccurate*) to

8

5 (*Very Accurate*). A sample item for extraversion (α = .80) is, "I am the life of the party." A sample item

for agreeableness (α = .78) is, "I sympathize with others' feelings." A sample item for conscientiousness

(α = .71) is, "I get chores done right away." A sample item for neuroticism (α = .74) is, "I have frequent

mood swings." Finally, a sample item for openness (α = .74) is, "I have a vivid imagination."

Mood (Ages 5 to 13)

We included one item to assess trait mood: "How do you usually feel in general?" Participants

responded using the same 5-point smiley face Likert scale as the LSE.

Life Orientation Test Revised (LOT-R; Ages 18 to 93)

The LOT-R (Scheier, Carver, & Bridges, 1994) consisted of 10 items designed to measure

optimism (α = .83). We used a 5-point scale ranging from 1 (*Strongly disagree*) to 5 (*Strongly agree*). A

sample item is, "Overall, I expect more good things to happen to me than bad."

Satisfaction with Life Scale (SWLS; Ages 18 to 93)

The SWLS (Diener, Emmons, Larsen, & Griffin, 1985) consisted of five items intended to measure

global life satisfaction (α = .91). We used a 5-point scale ranging from 1 (*Strongly disagree*) to 5 (*Strongly

agree*). A sample item is, "In most ways my life is close to my ideal."

Differences across Measures

In the manuscript, we summarized differences across measures within two major families of

scales that have been used to study self-esteem in multiple age groups (i.e., the SPPC and the SDQ).

Tables 1_{SOM} and 2_{SOM} detail the differences in item content, number of items, ordering, descriptive

language, and response options across the versions.

Table 1_{SOM}: *SPPC Global Subscales by Age Group - Abbreviated*

Children	Adolescents	College Students	Adults
Some **kids** are **not very happy** with **the way they do a lot of things** BUT Other kids think **the way they do things is fine**.	"teenagers;" "happy with themselves;" "most of the time;" positively worded option first	"students;'" "often dissatisfied;" negatively worded option first	"adults;" "dissatisfied" (no qualifier); negatively worded option first
Some **kids** like the kind of person they are BUT Other kids often **wish** they were someone else.	"teenagers"	"students"	"adults;" "would like to be"
Some **kids** are very happy being the way they are BUT Other kids **wish** they were different.	"teenagers"	"students"; "would really rather be"; negatively worded option first	"adults;" "would like to be"
Some **kids** are often **unhappy** with themselves BUT Other kids are **pretty** pleased with themselves.	"teenagers;" "disappointed"	"students;" "disappointed;" "usually quite pleased"	"adults;" "disappointed;" "quite pleased"
Some **kids** don't like the way they are leading their **life** BUT Other kids **do** like the way they are leading their **life**.	"teenagers"	"students;" "really like;" "often don't like;" positively worded option first	"adults;" "like" (no qualifier); "lives;" positively worded option first
Some **kids** are **happy with themselves as a person** BUT Other **kids are often not happy with themselves**.	NA	"students;" "usually like;" "don't like"	"adults;" "sometimes question whether they are a worthwhile person;" "feel that they are a worthwhile person;" negatively worded option first

Note. Full items are only shown for child version. Bolded words identify words that change in later versions; the changes are listed in quotes for each version.

Table 2_{som}: *SDQ Global Subscales by Age Group*

10

SDQ-I	SDQ-II	SDQ-III
5-point scale: "False to True" (group setting), "No always" to "Yes always" (individual setting)	*6-point scale: "False" to "True"*	*8-point scale: "Definitely false" to "Definitely true"*
A lot of things about me are good.	NA	Overall, I have a very good self-concept.
I do lots of important things.	If I really try I can do almost anything I want to.	Overall, nothing that I do is very important.
NA	NA	Overall, I do lots of things that are important.
In general I like being the way I am.	NA	Overall, I am not very accepting of myself.
NA	NA	Overall, I am pretty accepting of myself.
Overall I have a lot to be proud of.	Overall I have a lot to be proud of.	Overall, I have a lot of self-confidence.
I can do things as well as most other people.	I can do things as well as most people.	Overall, I lack self-confidence.
Other people think I am a good person.	Overall, I am a failure.	Overall, I have a lot of respect for myself.
When I do something, I do it well.	Most things I do, I do well.	Overall, I don`t have much respect for myself.
I am as good as most other people.	Overall, most things I do turn out well.	Overall, I have pretty positive feelings about myself.
Overall, I am no good.	Overall, I am no good.	Overall, I have a very poor self-concept.
I can't do anything right.	I can't do anything right.	Overall, I have pretty negative feelings about myself.

Scale Development

Our goal was to create a measure that could be used across the life span with a sample ranging

from young children to older adults. Thus, our first step was to identify items that young children could

understand. This allowed us to address potential difficulties with administering surveys to young

children (e.g., limited verbal comprehension, suggestibility, attention; see de Leeuw, 2012). We assumed

that if young children could understand the items, the items should be understandable by older children

and adults. The table below displays all original item wording, sources of items, and adapted item

wording.

Table 3SOM**:** *Sources and Adaptations of Item Wording for Creation of LSE Final Items*

Original wording from existing measures	Adapted wording for initial item pool	Adapted wording for pilot testing of LSE (original 12 items)	Final wording used in LSE
Other children are pretty *pleased* with themselves (SPPC)	I am pretty pleased with myself	How do you feel about yourself?	How do you feel about yourself?
Other children *do* like the way they are leading their life (SPPC)	I like the way I am leading my life	How do you feel about the way you are leading your life?	
Some children are *happy* with themselves as a person (SPPC)	I am happy with myself as a person	How do you feel about yourself as a person?	
Some children *like* the kind of person they are (SPPC)	I like the kind of person I am	How do you feel about the kind of person you are?	How do you feel about the kind of person you are?
Some children are very *happy* being the way they are (SPPC)	I am very happy being the way I am	How do you feel about *being* the way you are?	
Other children think the way they do things is *fine* (SPPC)	I think the way I do things is fine	How do you feel about the way you do things?	
I take a positive	I feel positive about	When you think about	When you think

attitude towards myself (RSE)	myself	yourself, how do you feel?	about yourself, how do you feel?
A lot of things about me are good (SDQ)	When I think about myself, I feel good	*(removed due to repetition with previous item)*	
On the whole, I am satisfied with myself (RSE)	I like myself	How do you feel about liking yourself?	
I feel that I have a number of good qualities (RSE)	I have a number of good qualities	How do you feel about your qualities?	
In general, I like being the way I am (SDQ)	I like the way I am	How do you feel about the way you are?	How do you feel about the way you are?
(New item)	I am confident about myself	How do you feel about confidence towards yourself?	
(New item)	I have no doubts about myself	How do you feel about doubts towards yourself?	

Note. SPPC = Harter's Self-Perception Profile for Children. RSE = Rosenberg Self-Esteem scale. SDQ = Marsh's Self-Description Questionnaire.

Original version of the LSE

We started with a version of the LSE global that attempted to simplify the two-step, forced-choice response format of the Harter items while still reducing any desirability bias. Children were presented with an item (e.g., "How do you feel about yourself?") and were then asked to choose a happy, neutral, or sad smiley face. Then, children were asked to rate the extent to which they felt that emotion, on a two-point rectangle scale (i.e., a small rectangle representing a little sad/happy and a large rectangle representing a lot sad/happy). Response labels varied based on the question (e.g., happy/sad vs. like/dislike). Practice questions consisted of positive and negative events so that children had the chance to use all options of the scale.

Child pilot sample 1*.* The original version of the LSE global was piloted with a group of 38 children between the ages of 5 and 10 (*M* = 7.32, *SD* = 1.61). All children were recruited through a local Farmer's Market in Northern California, and parents brought children to the lab at a later date for an individual assessment with the researcher. Parents were informed that the goal of the project was to

13

study the development of the self and that the study would take no longer than one hour. There were

23 girls (61%) and 15 boys, and a majority of the participants (71%) were White.

Adult pilot sample 1. We also piloted the original LSE with a sample of 30 adults on MTurk so

that we could have a comparison sample. That is, we could not be sure about the psychometric

characteristics of responses from young children without comparing them to a sample that is known to

be able to respond to questions about self-esteem (i.e., an adult sample).

Procedure. The LSE was administered in an individual interview with children in pilot sample 1.

Adults in pilot sample 1 independently completed the measure online. Before administration of the LSE,

children in sample 1 completed six practice questions to learn the response format. They were reminded

that they could choose whichever face and rectangle combination described their opinion the best, that

there were no right or wrong answers, and that the researcher did not know what the child was going to

say. Then, one trained researcher read each question of the LSE aloud and provided examples or

rephrases of difficult questions.

Children of all ages used the two-step format 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 3_{SOM}). In addition, we calculated the average standard

deviation for all the practice items within each age group and found that these did not significantly differ

in an independent samples *t*-test (t (22) = .22, p = .83).

Results. Children older than 8 years and adults had high reliability (α = .96 and .93, respectively)

and inter-item correlations (r = .67 and .49, respectively). However, reliability and inter-item correlations

were lower for children between 5 and 7 years old (α = .21; r = .06). Despite low reliability among

younger children, their scores still correlated with their parent's rating of their self-esteem (r = .21), as

did scores of older children (*r*'s = .66 & .90). In addition, adult scores on the LSE were strongly related to

scores on the RSE (r = .72). Therefore, we adjusted the scale (see current version of LSE below) and

14

piloted it a second time. We decided to test a continuous scale because this format is easier to

administer to participants and matches many other established scales.

Table 4$_{SOM}$: *Descriptive Statistics for Practice Items for the LSE with Child Pilot Sample 1*

	Ages 5-7			Ages 8-10		
Item	**Min-Max**	***M***	***SD***	**Min-Max**	***M***	***SD***
"How do you feel about the toys you have at home?"	3-5	4.33	0.98	1-5	3.45	1.57
"How do you feel about your chores?"	1-5	3.91	1.64	1-5	2.36	1.43
"What do you think about candy?"	1-5	3.62	1.66	1-5	3.45	1.57
"What do you think about vegetables?"	1-5	3.85	1.82	1-5	3.45	1.57
"How do you feel about getting shots from the doctor?"	1-5	2.00	1.41	1-5	2.27	1.42
"How do you feel about ice cream?"	2-5	4.62	0.96	3-5	4.45	0.93

Current version of LSE

We manipulated item wording (e.g., How do you feel about yourself vs. I like myself), response

options (e.g., happy vs. good), and pictorial representations of response options (i.e., rectangles of

increasing size vs. large rectangles for the end points and the smallest rectangle for neutral; rectangles

vs. smiley faces; different types of smiley faces). We also asked children to explain real responses (e.g.,

why did you circle a '4' and not a '5'?) and hypothetical answers (e.g., what would it mean if I circled a

'2'?). This process resulted in 13 items, but one item was completely redundant with another (i.e., "How

do you feel about yourself?"), so we kept 12 items for further pilot testing. The 12 original items and

inter-correlations are presented in Table 4$_{SOM}$.

Table 5_{SOM}: *LSE Original Items and Inter-Item Correlations*

		1	2	3	4	5	6	7	8	9	10	11	12
1	**How do you feel about yourself?**	1											
2	How do you feel about the way you are leading your life?	.71*	1										
3	How do you feel about yourself as a person?	.72*	.66*	1									
4	**How do you feel about the kind of person you are?**	.60*	.58*	.69*	1								
5	How do you feel about *being* the way you are?	.66*	.65*	.69*	.70*	1							
6	How do you feel about the way you do things?	.55*	.57*	.53*	.57*	.60*	1						
7	**When you think about yourself, how do you feel?**	.73*	.66*	.70*	.60*	.68*	.62*	1					
8	How do you feel about liking yourself?	.68*	.59*	.68*	.60*	.65*	.55*	.73*	1				
9	How do you feel about your qualities?	.56*	.51*	.61*	.62*	.59*	.57*	.58*	.65*	1			
10	**How do you feel about the way you are?**	.68*	.64*	.69*	.66*	.73*	.61*	.72*	.71*	.67*	1		
11	How do you feel about confidence towards yourself?	.62*	.54*	.59*	.52*	.57*	.54*	.64*	.65*	.59*	.64*	1	
12	How do you feel about doubts towards yourself?	.50*	.47*	.44*	.39*	.46*	.44*	.54*	.52*	.41*	.51*	.57*	1

Note. *p < .01. Final LSE items are in bold. Participants responded on a scale ranging from 1 (*Really Sad*) to 5 (*Really Happy*).

   Adult pilot sample 2*.* Each manipulation was tested through random assignment of adults on

MTurk to the eight different versions (*n* = 50 per group).

16

Child pilot sample 2. Three groups of 4-5-year-old children (*N* = 18) in a daycare center

associated with the university completed selected items from the current version to test for

comprehension of items and different response options. Piloting sessions were conducted by the first

author and trained undergraduate research assistants. The children's instructor was also in the room.

Children were tested in small groups to evaluate whether the survey could be group administered, as

opposed to one-on-one interviews. Children received one printed survey and a marker or crayon and

were seated at a small table in a small room adjacent to a teacher's office. Before entering the room,

the children's teacher informed them that they were going to help a researcher with a special activity.

Then, researchers gave the children the same instructions as those used in sample 5 (e.g., there were no

right or wrong answers).

Items were selected from the LSE for inclusion in this piloting session based on how well they

represented the majority of the LSE items or how difficult they were perceived to be (the most difficult

item was included in piloting). Each of the three groups responded to item 2 of the LSE (i.e., "How do

you feel about the kind of person you are?") using a scale illustrated by increasing rectangles (as in van

den Bergh & de Rycke, 2003). However, the item wording was changed to match the scale format for

this first version (i.e., "I like the kind of person I am"). In addition, Group 1 responded to the same item

using a second scale version for comparison purposes: a scale of rectangles that were largest for the

extreme ends of the scale (i.e., Strongly disagree and Strongly agree) and decreased into the Neutral

response. Next, Group 2 responded to item 3 of the LSE ("When you think about yourself, how do you

feel?") using the smiley faces from the final version of the LSE. Finally, Group 3 responded using smiley

faces without labels, and they answered a second item for which the wording was changed slightly (i.e.,

"I like myself as a person"). In sum, all scale versions tested in the adult pilot sample were presented to

children across the three pilot groups.

Results. Based on young children's explanations of real and hypothetical answers, we concluded

that they understood and preferred the 5-point smiley face scale better than the other versions of the

scale.

For the adult scales, all alpha reliabilities were high and were similar across scale anchors (i.e.,

sad/happy = .93; bad/good = .97; negative/positive = .95, no anchor = .93). In addition, mean scores

were similar across scale anchors for adults (i.e., sad/happy = 3.71, bad/good = 3.65, negative/positive =

3.48, no anchor = 3.74), and all versions correlated well with Marsh's SDQ-I (*r*'s = .67, .85, .87, .72) and

Harter's SPP for Children (*r*'s = .78, .82, .78, .70).

In sum, because the children preferred and understood the smiley face version better, and

because there were no major differences across response options for adults, we retained the smiley face

illustrations and corresponding anchors from "Really Sad" to "Really Happy" for further testing in larger

samples and across the lifespan.

18

Descriptive Statistics by Age Groups

We reported the overall descriptive statistics of the LSE across the whole sample in the

manuscript. Below is a table with descriptive statistics by each of the 11 age-stratified groups.

Table 6SOM**: *LSE Descriptive Statistics by Age Group*

Age Group	*N*	Minimum	Maximum	% 5's	Mean	Standard Deviation	Skewness
All	2,602[a]	1	5	10%	3.74	0.83	-0.66
5-7	125	2	5	33%	4.36	0.74	-1.24
8-13	261	1	5	19%	4.14	0.73	-0.84
14-17	238	1	5	7%	3.56	0.88	-0.38
18-24	278	1	5	4%	3.49	0.85	-0.56
25-29	252	1	5	7%	3.66	0.85	-0.86
30-39	257	1	5	6%	3.57	0.80	-0.48
40-49	242	1	5	8%	3.68	0.80	-0.68
50-59	227	1	5	6%	3.65	0.81	-0.75
60-69	203	1	5	7%	3.70	0.74	-0.43
70-79	200	2	5	4%	3.91	0.61	-0.76
80-89	198	1	5	13%	3.89	0.79	-1.21

Note. [a]Data was missing from 159 participants for self-esteem and from 268 participants for age. % 5's = valid percent of sample with an LSE composite score of 5.

Measurement Invariance

We tested for factorial invariance by age in two ways – using the MIMIC approach (Multiple Indicators Multiple Causes; Brown, 2015, pp. 273-283; Jöreskog & Goldberger, 1975) and traditional multi-group analyses. The MIMIC approach tests for intercept differences and has the advantage that it preserves age as a continuous variable. The multi-group analyses are somewhat limited because age groupings are inherently arbitrary. The advantage, however, is that researchers can conduct separate tests for the equivalence of factor loadings across age groups (metric invariance) and intercept invariance (scalar invariance).

For the MIMIC analyses, we started by identifying a baseline model using the four items as indicators of a latent self-esteem factor. Model fit statistics were as follows: Chi-Square = 68.62, *df* = 2, *p* < .01; RMSEA = .11, CFI = .99; TLI = .97; SRMR = .02 (*N* = 2,602). Standardized factor loadings ranged from .73 to .86 (Range in R^2 = .54 to .73). Given the relatively large Chi-square value, we consulted modification indices with the expectation that correlated residual variances would be present (as is the case for self-esteem measures like the RSE; see Donnellan, Ackerman, & Brecheen, 2016). The largest was for the first and third items (MI = 58.79). Adding that correlated residual improved model fit: Chi-Square = 11.49, *df* = 1, *p* < .01; RMSEA = .06, CFI = 1.0; TLI = .99; SRMR = .01. Standardized factor loadings ranged from .75 to .88 (Range in R^2 = .56 to .78). The modification index for the first and second items was 11.41. We added this correlated residual, and this resulted in a just identified model with no degrees of freedom: Chi-Square = .00, *df* = 0, *p* < .01; RMSEA = .00, CFI = 1.0, TLI = 1.0; SRMR = .00. Standardized factor loadings ranged from .74 to .90 (Range in R^2 = .54 to .80). Thus, we used this as the baseline and added age as a continuous variable covariate.

The model regressing latent global self-esteem on age had the following fit statistics: Chi-Square = 55.99, *df* = 3, *p* < .01; RMSEA = .08, CFI = .99; TLI = .97; SRMR = .02. The regression weight for age was not statistically significant (β = -.00, SE = .00, *p* = .28). Given the relatively large Chi-square value, we

consulted modification indices and found the following values for the items (Item 1: 28.54, Item 2 =

11.86, Item 3 = 33.26, Item 4 = 11.31). We then regressed Item 3 on the age factor to see if there was

evidence of improved fit. This was the case: Chi-Square = 22.12, *df* = 2, *p* < .01; RMSEA = .06, CFI = 1.0;

TLI = .98; SRMR = .01. The standardized estimate was .00 (SE = .00), indicating a very small effect.

Modification indices were also consulted, and item 2 was the only value above 10.00 for age (18.71). We

then made this addition and the model fit better: Chi-Square = 3.25, *df* = 1, *p* < .01; RMSEA = .03, CFI =

1.0; TLI = 1.0; SRMR = .01. The standardized estimate was .06 (SE = .02), indicating a very small effect. No

modification indices for age were above 10.00. The conclusion we drew from these analyses was that

there was some small degree of differential item functioning (DIF); however, the effect was trivial. The

standardized effect for global self-esteem regressed on age was -.06 (SE = .02, *p* = .01) in this model,

suggesting a slight decline in self-esteem with age at the latent level when adjusting for DIF. The

conventional interpretation would be for a very small age effect.

The second approach was to use the age categories outlined in the Introduction. We started

with a baseline model that included a correlation between the residuals for Item 1 and 3 given the

results reported above. We used the automated invariance testing functions in Mplus. The Chi-Square

comparisons indicated that contrast between the metric equivalence model (i.e., factor loadings) and

the configural model was not rejectable (Chi-Square Difference = 34.37, Change in *df* = 30, *p* = .27);

whereas the comparison of the scalar model to the metric model was rejectable (Chi-Square Difference

= 76.66, Change in *df* = 30, *p* = .00). This suggested that factor loadings were consistent across age

groups (metric invariance) but item intercepts were not (scalar invariance). We also tested whether CFI

changed by more than .01 according to criteria set by Cheung and Rensvold (2002) and Little (2013). A

change in CFI of .00 between the metric and configural models supported the conclusion from the Chi-

Square comparisons (i.e., that the metric model was not rejectable), and a change in CFI of .01 between

the scalar and metric models supported the conclusion that the scalar model was rejectable.

21

A visual inspection of the results from the configural models (output available upon request) suggested that item intercepts were higher for the youngest age groups when compared to older groups. We re-ran analyses dropping the two youngest age groups, and we could not reject the hypothesis of metric invariance when comparing metric and configural models (Chi-Square Difference = 25.72, Change in *df* = 24, *p* = .37; Change in CFI = .00), nor could we reject scalar invariance when comparing scalar and metric models (Chi-Square Difference = 41.23, Change in *df* = 24, *p* = .02; Change in CFI = .00).

In sum, we concluded that the LSE exhibited metric invariance across the lifespan according to Chi-Square Difference tests, but the evidence for scalar invariance was more circumspect. Regardless, the effect sizes associated with the difference for the intercepts were likely to be quite small, especially when considering participants ages 14 and up. Accordingly, we suggest that analyses using age and summary LSE composites are more or less justifiable on psychometric grounds across the lifespan. There might be some differences in how younger participants use the scale points, and some caution is warranted when considering how mean levels from participants in childhood compare with mean levels from older ages.

22

Full Correlation Table

In the manuscript, we focused on reporting correlations among each of the criterion and convergent validity scales and the LSE. Below,

we present correlations among all measures included in the study.

Table 7_{SOM}: *Pearson's Zero-Order Correlations among all LSE Correlates*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. Parent attachment	1														
2. Avoidance	.10	1													
3. Anxiety	.34*	.17*	1												
4. Admiration	.35*	.13*	-.06*	1											
5. Rivalry	-.20*	-.19*	-.29*	.23*	1										
6. NPI-16-C	.05	.01	.01	.43*	.22*	1									
7. CES-D	-.53*	-.30*	-.42*	-.15*	.32*	-.23*	1								
8. O	-	.23*	.13*	.12*	-.15*	-	-.09*	1							
9. C	-	.24*	.22*	.12*	-.29*	-	-.36*	.12*	1						
10. E	-	.21*	.13*	.44*	-.02	-	-.19*	.15*	.13*	1					
11. A	-	.29*	.08*	.09*	-.49*	-	-.22*	.31*	.28*	.19*	1				
12. N	-	-.19*	-.42*	-.19*	.30*	-	.60*	-.04	-.34*	-.17*	-.18*	1			
13. Mood	.32*	-	-	-	-	.09	-	-	-	-	-	-	1		
14. LOT-R	-	.31*	.35*	.35*	-.34*	-	-.61*	.14*	.39*	.32*	.30*	-.64*	-	1	
15. SWLS	.	.25*	.31*	.31*	-.16*	-	-.52*	.01	.31*	.28*	.15*	-.45*	-	.59*	1

Note. *p < .01

Criterion Validity Correlations by Age

In the manuscript, we stated that some criterion validity correlations with the LSE were

moderated by age. For the significant interactions, we compared the correlations within our 11 age

groups and observed the following differences. Depression interacted with age (β = .06, p < .01) and age

cubed (β = -1.69, p = .000), such that the correlation between depression and self-esteem ranged from -

.61 to -.64 across ages 18 to 39, whereas the correlation ranged from -.52 to -.60 across ages 40 to 89.

Next, extraversion interacted with age (β = -.08, p < .01), such that the correlation between extraversion

and self-esteem ranged from .33 to .40 across ages 18 to 49, whereas the correlation ranged from .17 to

.27 across ages 50 to 89. Optimism interacted with age (β = -.06, p < .01), such that the correlation

between optimism and self-esteem ranged from .60 to .66 across ages 18 to 70, whereas the correlation

was .53 for ages 80 to 89. Finally, life satisfaction interacted with age (β = -.08, p < .01), such that the

correlation between life satisfaction and self-esteem ranged from .58 to .65 across ages 18 to 59,

whereas the correlation ranged from .44 to .54 across ages 60 to 89.

Demographic Differences in LSE Psychometric Properties

Factor Structure

We first tested for differences in factor loadings by splitting the data file by the demographic

variable of interest and running an exploratory factor analysis for each dimension (for a total of 6

dimensions). We found that factor loadings for the four items loaded strongly onto one factor for both

individuals from the US (range: .73 to .86) and for individuals not from the US (range: .76 to .87). In

addition, both groups had the first eigenvalue above 1 (from US: 3.01, not from US: 2.96), and the

second eigenvalue did not rise above 1. That is, the LSE was unidimensional across country of origin.

Next, both males (range: .77 to .86) and females (range: .71 to .87) had strong factor loadings for the LSE

items as well as initial eigenvalues above 1 (males = 3.03, females = 3.01) and second eigenvalues below

1. For ethnicity (excluding the Native Hawaiian or Pacific Islander group due to a low *n* of 7), factor

loadings ranged from .56 to .92 and had an average across groups of .81. Initial eigenvalues ranged from

2.73 to 3.34, and second eigenvalues were all less than 1. Across education levels (Samples 1 through 3),

factor loadings ranged from .61 to .94, initial eigenvalues ranged from 2.81 to 3.11, and all second

eigenvalues were below 1. Factor loadings across levels of adolescent mother education (Sample 4,

excluding the PhD, JD, MD, or other advanced degree group due to a low *n* of 4) ranged from .55 to .98,

initial eigenvalues ranged from 2.57 to 3.26, and all second eigenvalues were below 1. Finally, factor

loadings across levels of adolescent father education (Sample 4, excluding the MA degree group due to

an error in estimating factor loadings: "communality of the variable exceeded 1.0") ranged from .55 to

.98, initial eigenvalues ranged from 2.57 to 3.26, and all second eigenvalues were below 1.

Reliability

Next, we tested for differences in reliability of the scale using Cronbach's alpha coefficients and

their 95% confidence intervals (see Fan & Thompson, 2001). Confidence intervals overlapped for all

groups within each demographic dimension: country, gender, ethnicity, education, and adolescent

mother and father education.

Validity

We tested for differences in convergent and criterion validity by conducting moderated

regressions predicting the LSE from each of the convergent and criterion validity measures, each

demographic variable (dummy coded country such that 0 = not from the US and 1 = from the US, gender

such that 0 = female and 1 = male, and ethnicity such that 0 = not White or Caucasian and 1 = White or

Caucasian), and the interaction between the two. None of the interaction terms with *country of origin*

were statistically significant, and none of the interaction terms accounted for more than 1% of the

variance, with the exception of NPI-16-C*Country explaining 1.2% additional variance (p = .09). *Gender*

significantly interacted with SISE, avoidance, conscientiousness, neuroticism, and optimism; however,

none of these interaction terms explained more than 1% additional variance. Further, NPI-16-C*Gender

explained 2.1% additional variance in LSE scores, yet this interaction term was not significant using our

alpha level of .01 (p = .03). Dummy coded *ethnicities* did not significantly interact with any measures to

account for more than 1% of the variance. Two interaction terms with *education* were significant

(anxiety and conscientiousness), but these terms did not explain more than 1% additional variance. The

interaction between anxiety and *adolescent mother education* explained 2.6% additional variance, but

the interaction term was not significant using our alpha level of .01 (p = .02). Next, the interaction

between the SISE and adolescent mother education explained 1.8% additional variance and was

significant (p = .01). 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). Finally, the

interaction between the SISE and *adolescent father education* explained 1.1% additional variance, but

this interaction term was not significant using our alpha level of .01 (p = .04). Tables 8-13$_{SOM}$ below

display results for the comparisons of psychometric properties of the LSE by demographic groups.

Table 8_{SOM}: *Psychometric Properties of the LSE by Country of Origin*

		United States (US)	Not US
Descriptive Statistics			
Mean		3.75	3.80
SD		0.83	0.75
N		2,368	107
Unidimensionality			
Factor Loadings	LSE 1	.83	.83
	LSE 2	.73	.76
	LSE 3	.85	.87
	LSE 4	.86	.78
Initial Eigenvalue		3.01	2.96
Second Eigenvalue		0.25	0.42
Reliability			
Cronbach's alphas		.89	.88
95% Confidence Intervals		.88 - .90	.84 - .92
Mean inter-item correlations		.67	.65

Validity						
	Model 1 Predicting LSE			**Model 2 Predicting LSE**		
	$\beta_{measure}$	$^{1}\beta_{country}$	R^2	$\beta_{interaction}$	R^2	ΔR^2
SDQ	.73*	-.01	.53	-.00	.53	.00
SPP	.61*	-.03	.37	.00	.37	.00
RSE	.68*	-.04	.47	.00	.47	.00
SISE	.68*	-.03	.46	.00	.46	.00
Parent attachment	.53*	-.06	.28	-.06	.29	.00
Avoidance	.33*	-.04	.11	.00	.11	.00
Anxiety	.40*	-.04	.16	.01	.16	.00
Admiration	.41*	-.03	.17	.02	.17	.00
Rivalry	-.21*	-.04	.05	.01	.05	.00
NPI-16-C	.29*	-.06	.09	-.12	.10	**.01**
CES-D	-.59*	-.05	.35	-.02	.35	.00
O	.09*	-.04	.01	-.01	.01	.00
C	.35*	-.03	.13	-.00	.13	.00
E	.31*	-.03	.10	.01	.10	.00
A	.21*	-.05	.05	.02	.05	.00
N	-.52*	-.04	.27	-.01	.27	.00
Mood²	na	na	na	na	na	na
LOT-R	.64*	-.03	.41	.01	.41	.00
SWLS	.58*	-.01	.34	-.01	.34	.00

Note. *$p < .01$. [1]Country was dummy coded so that 0 = not US, and 1 = US. β = standardized beta coefficient. R^2 = percentage of variance in LSE explained. ΔR^2 = change in R^2 from Model 1 to Model 2. [2]Adult samples did not receive the mood item, and children in Sample 5 were all in the United States.

28

Table 9SOM**: *Psychometric Properties of the LSE by Gender***

		Male	Female
Descriptive Statistics			
Mean		3.73	3.76
SD		0.84	0.81
N		1,067	1,398
Unidimensionality			
Factor Loadings	LSE 1	.82	.84
	LSE 2	.77	.71
	LSE 3	.84	.87
	LSE 4	.86	.85
Initial Eigenvalue		3.03	3.01
Second Eigenvalue		0.39	0.48
Reliability			
Cronbach's alphas		.89	.89
95% Confidence Intervals		.88 - .90	.88 - .90
Mean inter-item correlations		.68	.67

	Model 1 Predicting LSE			Model 2 Predicting LSE		
	$\beta_{measure}$	$^1\beta_{gender}$	R^2	$\beta_{interaction}$	R^2	ΔR^2
SDQ	.73*	-.01	.53	.04	.53	.00
SPP	.60*	.01	.36	.01	.36	.00
RSE	.68*	.01	.47	.03	.47	.00
SISE	.68*	-.05*	.46	**.05***	.47	.00
Parent attachment	.53*	.15	.31	-.09	.32	.01
Avoidance	.33*	.01	.11	**.06***	.11	.00
Anxiety	.40*	0	.16	-.01	.16	.00
Admiration	.42*	-.05	.17	.04	.17	.00
Rivalry	-.23*	.05	.05	.01	.05	.00
NPI-16-C	.26*	.15	.10	-.16	.12	**.02**
CES-D	-.58*	-.01	.34	-.02	.34	.00
O	.09*	-.03	.01	.00	.01	.00
C	.35*	-.01	.13	**.06***	.13	.00
E	.31*	-.05	.10	.02	.10	.00
A	.22*	.03	.04	-.01	.04	.00
N	-.52*	-.08*	.28	**-.07***	.28	.01
Mood	na	na	na	na	na	na
LOT-R	.63*	-.01	.40	**.08***	.41	.01
SWLS	.58*	-.01	.34	.04	.34	.00

Note. *p < .01. [1]Gender was dummy coded so that 0 = female, and 1 = male. β = standardized beta coefficient. R^2 = percentage of variance in LSE explained. ΔR^2 = change in R^2 from Model 1 to Model 2.

29

Table 10_SOM: *Psychometric Properties of the LSE by Ethnicity*

		American Indian	Asian	Black	Latino/a	White	Other
		Descriptive Statistics					
Mean		3.47	3.53	3.63	3.85	3.71	3.86
SD		0.84	0.79	0.87	0.87	0.81	0.85
N		34	110	136	140	1,704	77
		Unidimensionality					
Factor Loadings	LSE 1	.70	.84	.82	.84	.84	.83
	LSE 2	.56	.73	.60	.73	.75	.92
	LSE 3	.90	.80	.84	.92	.87	.89
	LSE 4	.88	.77	.78	.86	.87	.89
Average loadings		.76	.79	.76	.84	.83	.88
Initial Eigenvalue		2.74	2.85	2.73	3.10	3.08	3.34
Second Eigenvalue		0.67	0.47	0.60	0.43	0.42	0.29
		Reliability					
Alphas		.83	.86	.84	.90	.90	.93
95% CIs		.71 - .91	.82 - .90	.80 - .88	.87 - .93	.89 - .91	.90 - .96
Inter-item *r*		.57	.62	.57	.70	.69	.78

	Validity					
	Model 1 Predicting LSE			**Model 2 Predicting LSE**		
	$\beta_{measure}$	$^1\beta_{ethnicity}$	R^2	$\beta_{interaction}$	R^2	ΔR^2
SDQ	.73*	.02	.53	-.02	.53	.00
SPP	.61*	-.02	.37	-.03	.37	.00
RSE	.69*	-.03	.47	.00	.47	.00
SISE	.68*	.02	.46	-.02	.47	.00
Parent attachment	.55*	-.05	.30	.08	.30	.01
Avoidance	.33*	-.00	.11	-.02	.11	.00
Anxiety	.40*	.01	.16	.00	.16	.00
Admiration	.42*	.07*	.18	-.02	.18	.00
Rivalry	-.22*	.00	.05	-.02	.05	.00

	Model 1 Predicting LSE cont'd			Model 2 Predicting LSE cont'd		
	$\beta_{measure}$	$^{1}\beta_{ethnicity}$	R^2	$\beta_{interaction}$	R^2	ΔR^2
NPI-16-C	.25*	-.00	.06	.08	.07	.01
CES-D	-.59*	-.04	.34	-.01	.34	.00
O	.09*	.03	.01	-.03	.01	.00
C	.35*	.02	.13	-.03	.13	.00
E	.30*	.04	.09	-.04	.10	.00
A	.21*	.01	.04	-.02	.04	.00
N	-.52*	.00	.27	.03	.27	.00
Mood	.62*	.09	.30	-.10	.40	.01
LOT-R	.64*	-.01	.40	-.04	.41	.00
SWLS	.58*	.01	.34	-.02	.34	.00

Validity cont'd

Note. *$p < .01$. [1]Ethnicity was dummy coded so that 0 = White, and 1 = not White. β = standardized beta coefficient. R^2 = percentage of variance in LSE explained. ΔR^2 = change in R^2 from Model 1 to Model 2.

Table 11$_{SOM}$: *Psychometric Properties of the LSE by Target Education Level (Samples 1 through 3)*

		Some HS	HS GED	College	AA	BA	Grad/prof school	MA	Advanced degree
		Descriptive Statistics							
Mean		3.45	3.79	3.66	3.73	3.69	3.78	3.83	3.81
SD		0.86	0.78	0.87	0.75	0.80	0.85	0.66	0.57
N		23	275	565	228	499	125	179	42
		Unidimensionality							
Factor Loadings	LSE 1	.61	.81	.84	.84	.81	.87	.81	.94
	LSE 2	.84	.75	.77	.74	.73	.76	.61	.79
	LSE 3	.86	.86	.88	.87	.88	.84	.91	.86
	LSE 4	.87	.88	.87	.86	.86	.84	.78	.80
Initial Eigenvalue		2.74	2.9	3.04	3.11	3.04	3	3.05	2.81
Second Eigenvalue		0.67	0.65	0.43	0.42	0.42	0.43	0.43	0.63
		Reliability							
Alphas		.86	.90	.91	.89	.89	.90	.90	.91
95% CIs		.74 - .94	.87 - .91	.89 - .92	.87 - .92	.87 - .90	.86 - .93	.82 - .89	.85 - .95
Inter-item *r*		.63	.68	.70	.68	.67	.68	.60	.72
		Validity							
		Model 1 Predicting LSE			Model 2 Predicting LSE			
		$\beta_{measure}$	$^1\beta_{education}$	R^2	$\beta_{interaction}$	R^2	ΔR^2		
SDQ		.73*	.00	.53	-.04	.54	.01		
SPP		.61*	.01	.37	-.03	.37	.00		
RSE		.70*	-.03	.48	-.01	.48	.00		
SISE		.68*	-.01	.46	-.02	.46	.00		
Parent attachment		na	na	na	na	na	na		
Avoidance		.33*	.03	.11	-.03	.11	.00		
Anxiety		.41*	.04	.17	-.06*	.17	.00		
Admiration		.42*	.02	.17	.01	.17	.00		
Rivalry		-.24*	.05	.06	.01	.06	.00		

32

	Model 1 Predicting LSE cont'd			Model 2 Predicting LSE cont'd		
Validity cont'd						
	$\beta_{measure}$	$^1\beta_{education}$	R^2	$\beta_{interaction}$	R^2	ΔR^2
NPI-16-C	na	na	na	na	na	na
CES-D	-.58*	.00	.34	.01	.34	.00
O	.09*	.04	.01	.02	.01	.00
C	.35*	.04	.13	-.07*	.13	.00
E	.30*	.03	.09	-.01	.09	.00
A	.21*	.05	.05	-.04	.05	.00
N	-.52*	.04	.27	.03	.27	.00
Mood	na	na	na	na	na	na
LOT-R	.64*	.01	.41	-.01	.41	.00
SWLS	.58*	.01	.34	-.01	.34	.00

Note. β = standardized beta coefficient. R^2 = percentage of variance in LSE explained. ΔR^2 = change in R^2 from Model 1 to Model 2.

33

Table 12_{SOM}: *Psychometric Properties of the LSE by Mother's Education Level (Samples 4 and 5)*

		Some HS	HS GED	College	AA	BA	Grad/prof school	MA
				Descriptive Statistics				
Mean		3.33	3.66	3.73	3.96	3.84	4.05	3.38
SD		0.90	0.85	0.84	0.87	0.83	0.92	1.17
N		39	37	45	21	53	41	28
				Unidimensionality				
Factor Loadings	LSE 1	.98	.91	.79	.89	.79	.74	.85
	LSE 2	.67	.66	.60	.64	.55	.76	.90
	LSE 3	.87	.87	.86	.92	.75	.93	.91
	LSE 4	.79	.91	.75	.94	.79	.96	.80
Initial Eigenvalue		3.04	3.10	2.69	3.16	2.57	3.14	3.26
Second Eigenvalue		0.51	0.52	0.65	0.54	0.69	0.65	0.36
				Reliability				
Alphas		.88	.90	.84	.91	.81	.90	.92
95% CIs		.74 - .94	.87 - .91	.89 - .92	.87 - .92	.87 - .90	.86 - .93	.82 - .89
Inter-item *r*		.68	.70	.56	.71	.52	.71	.75
				Validity				
		Model 1 Predicting LSE			Model 2 Predicting LSE			
		$\beta_{measure}$	$^1\beta_{education}$	R^2	$\beta_{interaction}$	R^2	ΔR^2	
SDQ		na	na	na	na	na	na	
SPP		.57*	-.00	.32	.09	.33	.01	
RSE		.63*	.01	.40	.09	.40	.00	
SISE		.70*	-.01	.48	.14*	.50	.02	
Parent attachment		.53*	-.00	.28	.08	.29	.01	
Avoidance		.23*	.03	.05	.07	.06	.01	
Anxiety		.25*	-.02	.06	-.18	.09	.03	
Admiration		.47*	-.02	.22	.08	.23	.01	
Rivalry		-.07	.00	.01	.08	.01	0	
NPI-16-C		.30*	-.05	.09	.09	.10	.01	

Validity cont'd						
	Model 1 Predicting LSE cont'd			Model 2 Predicting LSE cont'd		
	$\beta_{measure}$	$^1\beta_{education}$	R^2	$\beta_{interaction}$	R^2	ΔR^2
CES-D	-.59*	-.04	.35	-.09	.36	.01
O	na	na	na	na	na	na
C	na	na	na	na	na	na
E	na	na	na	na	na	na
A	na	na	na	na	na	na
N	na	na	na	na	na	na
Mood	na	na	na	na	na	na
LOT-R	na	na	na	na	na	na
SWLS	na	na	na	na	na	na

Note. There were too few participants with an advanced degree, so this group was dropped from analyses. R^2 = percentage of variance in LSE explained. ΔR^2 = change in R^2 from Model 1 to Model 2.

35

Table 13$_{SOM}$: *Psychometric Properties of the LSE by Father's Education Level (Samples 4 and 5)*

		Some HS	HS GED	College	AA	BA	Grad/prof school	MA	Advanced degree
Mean		3.30	3.70	3.69	3.44	3.78	3.71	3.28	3.27
SD		0.88	0.92	0.80	0.84	0.88	0.96	0.91	0.91
N		30	41	39	16	46	13	17	12
Unidimensionality									
Factor Loadings	LSE 1	.99	.90	.84	.92	.70	.86		.90
	LSE 2	.72	.74	.59	.79	.70	.78	-	.87
	LSE 3	.80	.87	.88	.83	.88	.94		.86
	LSE 4	.77	.89	.82	.82	.93	.99		.99
Initial Eigenvalue		3.04	3.01	3.17	2.83	3.11	2.94	2.51	3.45
Second Eigenvalue		0.51	0.53	0.43	0.70	0.44	0.57	0.73	0.30
Reliability									
Alphas		.88	.91	.86	.90	.88	.94	.80	.94
95% CIs		.80 - .94	.86 - .95	.75 - .91	.79 - .96	.79 - .92	.82 - .97	.57 - .92	.85 - .98
Inter-item *r*		.67	.72	.61	.70	.64	.80	.49	.82

Validity							
	Model 1 Predicting LSE			**Model 2 Predicting LSE**			
	$\beta_{measure}$	$^{1}\beta_{education}$	R^2	$\beta_{interaction}$	R^2	ΔR^2	
SDQ	na	na	na	na	na	na	
SPP	.56*	.01	.32	.04	.32	.00	
RSE	.63*	-.05	.39	.06	.39	.00	
SISE	.69*	.01	.48	.11	.49	.01	
Parent attachment	.53*	-.07	.28	.06	.29	.01	
Avoidance	.22*	-.01	.05	.09	.06	.01	
Anxiety	.25*	-.04	.06	.02	.06	.00	
Admiration	.48*	-.02	.23	.00	.23	.00	
Rivalry	-.07	-.03	.01	-.07	.01	.00	

36

Validity cont'd					
	Model 1 Predicting LSE cont'd			**Model 2 Predicting LSE cont'd**	
	$\beta_{measure}$	$^1\beta_{education}$	R^2	$\beta_{interaction}$	R^2	ΔR^2
NPI-16-C	.29*	-.06	.09	.03	.09	.00
CES-D	-.59*	-.06	.35	-.05	.35	.00
O	na	na	na	na	na	na
C	na	na	na	na	na	na
E	na	na	na	na	na	na
A	na	na	na	na	na	na
N	na	na	na	na	na	na
Mood	na	na	na	na	na	na
LOT-R	na	na	na	na	na	na
SWLS	na	na	na	na	na	na

Note. R^2 = percentage of variance in LSE explained. ΔR^2 = change in R^2 from Model 1 to Model 2. - Communality of var exceeded 1.0.