



## Measuring self-esteem instability through a single-administration scale: Still a fruitless endeavor?



Matt C. Howard<sup>1</sup>

5811 USA Drive S., Rm. 346, Mitchell College of Business, University of South Alabama, Mobile, AL 36688, United States

### ARTICLE INFO

#### Article history:

Received 25 May 2016

Received in revised form 25 August 2016

Accepted 7 September 2016

Available online xxxx

#### Keywords:

Self-esteem instability

Self-esteem

Scale development

Replication

### ABSTRACT

Studies on self-esteem instability (SEI) must obtain multiple measurements of self-esteem, such as daily for two weeks, to calculate SEI. This measurement method is called the “gold standard.” The intensive nature of the gold standard forces SEI to be studied through multiple-day, controlled, student-sample designs, which restricts generalizability and possible variables studied. We attempt to address these concerns through creating a new single-administration SEI scale, called the SEI Measure (Studies 1 and 2), and analyzing the relationship of the SEI Measure and prior single-administration SEI scales with the gold standard (Studies 3 and 4). The SEI Measure and existing scales only have moderate correlations with the gold standard, suggesting that they do not adequately gauge SEI. As these studies are the most robust investigation into single-administration SEI scales to date, we suggest that the study of such scales should be refocused, and our results also provide direction for new research avenues. Innovative single-administration SEI measures, such as implicit measures, may still adequately gauge SEI, and authors should consider perceived SEI separately from SEI itself. Perceived SEI may be an important aspect of self-perceptions that leads to distinct personal outcomes that are not predicted by SEI or other commonly studied variables.

© 2016 Published by Elsevier Ltd.

Self-esteem is one of the oldest (Nicole, 1696) and widest (Myers, Willse, & Villalba, 2011) studied constructs in psychology, and it influences a wide array of outcomes significant to daily life, such as depression and life satisfaction (Brown, Andrews, Harris, Adler, & Bridge, 1986; Meier, Semmer, & Hupfeld, 2009; Milevsky, Schlechter, Netter, & Keehn, 2007). One's amount of self-esteem, however, is not the only aspect of self-esteem that influences personal outcomes. Self-esteem instability (SEI; also called stability or variability) also has important psychological implications that are largely independent of one's amount of self-esteem (Chabrol, Rousseau, & Callahan, 2006; Kernis & Waschull, 1995; Okada, 2010). Research has shown SEI to incrementally predict depression and life satisfaction, among several other outcomes, beyond self-esteem itself (De Man, Gutiérrez, & Sterk, 2001; Meier et al., 2009; Oosterwegel, Field, Hart, & Anderson, 2001).

Traditionally, SEI has been operationalized as the variations between repeated measurements of self-esteem, often calculated through within-person standard deviations from self-reported scales administered over the course of several days (Kernis, 2005; Kernis, Cornell, Sun, Berry, & Harlow, 1993; Kernis & Waschull, 1995). Some studies have gauged SEI through administering a survey every day for a week

(Kernis, Grannemann, & Barclay, 1989), whereas others have opted for more measurement occasions and administered two surveys every day across two weeks (Zeigler-Hill, 2006). Nevertheless, of all existing studies, most SEI research has been limited to multiple-day studies in largely controlled settings, due to the need for extensive repeated measurements. Further, the vast majority of these studies have solely used student samples, as student samples are much easier to obtain for intensive repeated measurements compared to adult or specialized samples.

This restriction to multiple-day, controlled, student-sample studies limits the validity and generalizability of all SEI research. A history of authors have noted concerns with such studies, particularly in regards to external validity (Cook et al., 1979; Shadish, Cook, & Campbell, 2002). Without more naturalistic studies, which can ensure adequate external validity, any observed results cannot be certain to apply to real-world settings. Also, the methodological restrictions to SEI research limit the possible variables that can be studied. For example, it is very difficult to observe the effects of SEI on workplace outcomes, such as performance and job satisfaction, and the same is true for other variables that are largely limited to adult populations. In the current article, we aim to address the limited nature of SEI research due to measurement constraints.

Our primary objective is to create a single-administration scale of SEI. While the study of SEI requires the repeated measurement of self-esteem, it may be possible to create a single-administration measure of *perceived* SEI that adequately converges with the traditional

E-mail address: [MHoward@SouthAlabama.edu](mailto:MHoward@SouthAlabama.edu).

<sup>1</sup> I thank Ryan Smith for his feedback on previous ideas that contributed to the current manuscript, Mike Hoffman for his feedback on the current manuscript, as well as the Reviewer and Editors for their very helpful comments.

measurement of SEI. Such a scale could allow for the study of SEI using alternative research designs, particularly those with superb external validity. Further, the scale could be applied in a broader range of contexts than currently possible in SEI research, including more naturalistic settings, and it could be used to study a broader range of variables. In fact, the scale may even open completely new areas of research for SEI, such as investigating the influence of SEI on workplace outcomes. Therefore, existing areas of study could be more robustly investigated, and new areas of study could be opened through the creation of a single-administration SEI scale.

Our secondary objective is to reanalyze three existing single-administration SEI scales (Chabrol et al., 2006; Kernis, Grannemann, & Barclay, 1992; Rosenberg, 1965). Each of these measures have certain noteworthy concerns, such as their small relationships with the traditional measurement of SEI, but much is still unknown about their psychometric properties and/or other aspects of validity. Through investigating these aspects, important information may be uncovered about the ideal methods to gauge SEI through a single-administration scale. Likewise, it may be possible that prior concerns with these scales are unfounded, and they are adequate measures of SEI. If so, the current article could aid the study of SEI through encouraging future use of these potentially underutilized measures, achieving all the benefits of creating a new scale altogether.

Most importantly, however, the analyses of the new and existing scales may provide definitive evidence regarding the ability of SEI to be gauged through a single-administration scale. While prior analyses of the individual scales have uncovered concerning features, it is possible the SEI is simply unable to be gauged through a single-administration scale. In these prior analyses, researchers could not differentiate the concerning aspects from the scale and the nature of the construct itself. Only through an omnibus analysis of all the scales, as done in the current article, can clear inferences about this research question be achieved. Therefore, while the current article is framed through the lens of scale development and (re)analysis, it is also an investigation into the nature of SEI.

To achieve these goals, the current article is organized as such: First, we provide a review of SEI and its measurement. Second, we present a series of studies that creates a new single-administration measure of SEI. Third, in two separate studies, we test the convergent validity of the new and existing measures of SEI with the traditional method of gauging the construct. Fourth, we discuss the implications of the results for research and practice.

## 1. Background

### 1.1. Self-esteem instability and the gold standard

Self-esteem is often considered to be a primary indicator of psychological health, and those with high self-esteem are seen as more mentally fit (Sedikides, Rudich, Gregg, Kumashiro, & Rusbult, 2004; Taylor & Brown, 1988; Wilkinson, 2004). The construct has been shown to be a resilient barrier to negative events, preventing stress, depression, and anxiety (De Cremer, 2002; Meier et al., 2009; Zeigler-Hill & Showers, 2007), among other detrimental personal outcomes (Cassidy, O'Connor, Howe, & Warden, 2004; Myers & Rosen, 1999; Stokes & Peterson, 1998). Due to the focus on psychological health in self-esteem research, similar relationships have also been studied in regards to SEI.

SEI is the, “propensity to exhibit variability in self-feelings across time” (Kernis et al., 1993, p. 1190). Although related, SEI is not entirely dependent on self-esteem. That is, both those with high and low self-esteem can express both high and low SEI. Also, as suggested by Kernis et al. (1993), SEI is “a dispositional quality that interacts with contextual factors to produce specific patterns of fluctuations [in self-esteem]” (p. 1190). The construct likely arises from a variety of personal attributes that determine reactions to external events (i.e. salient self-

identities, sensitivity to evaluations, source of self-esteem), and SEI is believed to be a relatively stable personality trait (Chabrol et al., 2006; Kernis et al., 1992; Rosenberg, 1965).

Further, research has likewise shown SEI to be related to many of psychological health outcomes. Like self-esteem, SEI is related to stress, depression, and anxiety (Kernis et al., 1993; Roberts, Kassel, & Gotlib, 1995), and it has also been linked to the ultimate outcome of poor psychological well-being, suicide (de Man & Gutierrez, 2002). Research has also linked SEI to various aspects of interpersonal interaction, particularly experienced social anxiety and emotional reactions to (dis)approval (Crocker, Luhtanen, Cooper, & Bouvrette, 2003; Kernis et al., 1993; Seery, Blascovich, Weisbuch, & Vick, 2004), suggesting that SEI may be a contributor to certain social tendencies or even phobias.

When reviewing these and other studies, however, the scope of SEI research is notably limited. Although extremely important, the majority of studies only investigate the relationship of SEI with psychological health outcomes. While this topic may be the primary focus of research on self-esteem and SEI, the former construct has also been applied in much wider contexts. For example, authors have studied the relationship of self-esteem with several aspects of work, such as employee performance and job satisfaction (Judge & Bono, 2001; Judge, Erez, & Bono, 1998); other individual differences, placing self-esteem in the larger nomological network of related constructs (Baumeister, Smart, & Boden, 1996; Scheier, Carver, & Bridges, 1994); and an array of theories, many created for the specific purpose of understanding self-esteem (Greenberg, Pyszczynski, & Solomon, 1986; Leary & Baumeister, 2000). Alternatively, very few similar studies have been performed on SEI. The limited nature of SEI research may be due to the required measurement method for the construct, which may be preventing SEI from becoming as widespread in research as self-esteem itself.

The traditional method to gauge SEI is to administer several iterations of the Rosenberg Self-Esteem Scale (RSES; Rosenberg, 1965), which is a general measure of global self-esteem. The RSES consists of ten questions, including “All in all, I am inclined to feel that I am a failure” and “I take a positive attitude towards myself.” Further, the scale’s instructions can be modified to instruct participants to report “how they feel in general” or “how they feel in the present moment,” whereas the latter is typically used in SEI research.

To gauge SEI using the RSES, the scale is administered multiple times over the course of several days, and the standard deviation of these scores is considered to be representative of the construct (Kernis, 2005; Kernis & Waschull, 1995; Kernis et al., 1993). In current research, there is no standard duration to administer the multiple iterations of the RSES. Previous studies have used every day for one week (Kernis et al., 1989), twice a day for four days (Kernis et al., 1992), twice a day for two weeks (Zeigler-Hill, 2006), and several others (Kernis, 2005; Kernis & Waschull, 1995; Kernis et al., 1993). In general, however, researchers administer surveys daily or twice daily for one or two weeks. This method of measuring SEI has been called the “gold standard” (Chabrol et al., 2006), and we also use this terminology.

While the gold standard is believed to be effective and accurate, it creates certain limitations in SEI studies. First, the reliance on studies lasting multiple days or weeks slows the progress of SEI research. Whereas studies on other constructs can apply speedier methods, such as cross-sectional designs, no such luxury is available for SEI research. Further, although cross-sectional designs have notable limitations, they are able to provide initial inferences about research questions, which may be replicated using more advanced methodological designs (Cook et al., 1979; Shadish et al., 2002). These initial inferences are unable to be discovered in SEI studies, preventing researchers from inferring which research questions may be most valuable to pursue. This may cause some researchers to be hesitant to study SEI.

Second, SEI studies are largely limited to relatively controlled settings. Currently, when obtaining repeated measurements, some studies require participants to complete daily surveys in a lab or classroom

setting (i.e. administered before class starts; de Man & Gutierrez, 2002; Kernis et al., 1989; Kugle, Clements, & Powell, 1983; Waschull & Kernis, 1996). These studies force participants to remove themselves from their environments, possibly altering their self-esteem. Also, other studies give a packet of daily measures and instruct participants to complete these measures at designed time points, but concerns about compliance have been noted in regards to this method (Greenier et al., 1999; Kernis, Grannemann, & Mathis, 1991; Kernis et al., 1992; Kernis et al., 1993; Kernis, Greenier, Herlocker, Whisenhunt, & Abend, 1997; Knowles et al., 2007; Roberts et al., 1995; Seery et al., 2004; Waschull & Kernis, 1996; Kirkpatrick, Nezek, Smith, & Paddock, 2007; Zeigler-Hill & Abraham, 2006). While certain technologies, such as cellphones, may allow for daily measurements in a completely naturalistic setting, thereby avoiding these concerns, applications of these devices have yet to be seen in the majority of SEI research.

Third, obtaining measurements over multiple days is extremely difficult for most populations, causing the vast majority of SEI studies to use student samples (de Man & Gutierrez, 2002; Greenier et al., 1999; Kernis et al., 1989; Kernis et al., 1991; Kernis et al., 1992; Kernis et al., 1993; Kernis et al., 1997; Knowles et al., 2007; Kugle et al., 1983; Roberts et al., 1995; Seery et al., 2004; Waschull & Kernis, 1996; Webster et al., 2007; Zeigler-Hill & Abraham, 2006). The reliance on the student participant in research is well documented, and many authors suggest that studies using these samples are not generalizable to the general population (Bello, Leung, Radebaugh, Tung, & Van Witteloostuijn, 2009; Greenberg, 1987; Turner & Upshur, 2002). Also, many variables cannot be studied in student samples, such as workplace outcomes. Thus, while SEI is relatively well-studied in students, much less is known about SEI in the general population.

Together, the methodological restrictions of SEI studies poses certain concerns for validity, generalizability, and other aspects of research design. These concerns could be addressed through creating a single-administration SEI scale. A single-administration SEI scale could encourage speedy research designs, using adult populations, in naturalistic settings, while analyzing a wider range of variables. We create a single-administration SEI scale in the current article. Before doing so, we review three previous attempts at creating such a scale.

### 1.2. Single-administration self-esteem instability measurement

The earliest single-administration SEI scale is the five-item Rosenberg Stability Scale (RSS; Rosenberg, 1965), which includes the item, “Does your opinion of yourself tend to change a good deal or does it always continue to remain the same?” Although the RSS was created over fifty years ago, few studies have adapted its use. This may be due to the scale’s very small relationship with the gold standard ( $r = .01$ ), certain wording issues (Kernis et al., 1992), and Rosenberg’s (1965) own criticisms of the scale. Thus, the RSS does not appear to be a suitable single-administration measure of SEI.

The second single-administration SEI scale must be administered in conjunction with the RSES. Kernis et al. (1992) created a scale in which participants rated the extent that their answers to each RSES item would change on a daily basis. Much like the RSS, this scale has also seen limited usage. The sparse usage of Kernis et al.’s (1992) scale is likely due to the creators’ own criticisms in its original publication. The authors noted that their scale was weakly correlated to the gold standard ( $r = .22$ ), and believed that the scale was unsuccessful in measuring SEI. Kernis et al. (1992) also noted certain aspects that may have caused the scale to perform poorly. Participants may be unable to estimate their changes to the specific perceptions described in the RSE items (i.e. “At times I think I am no good at all”), as changes to these specific perceptions may occur subconsciously. Instead, participants may only be able to perceive general changes to their self-esteem (i.e. My self-esteem changes often) – if at all. Therefore, in the creation of a new single-administration measure, it may be beneficial to solely gauge perceptions about general changes to self-esteem.

The final scale is the Instability of Self-Esteem Scale (ISES; Chabrol et al., 2006). This scale gauges SEI through four questions that indirectly gauge changes in self-esteem (i.e. “Sometimes I feel worthless, at other time I feel that I am worthwhile”). Like the other scales, the ISES has also seen limited usage. Although it has an extremely strong relationship with the gold standard ( $r = .81$ ; Chabrol et al., 2006), other concerns limit its usage. First, most aspects of the ISES’s validity are unknown, such as its concurrent validity. Second, little investigation into its psychometric properties has been performed, such as its factor structure. Third, the scale is only four items, which draws questions to its criterion validity. Fourth, the scale has similar wording to the RSS. Fifth, the sole investigation into the relationship between the ISES and the gold standard only included 19 participants, and the results may have been due to the small sample size. Despite these concerns, the ISES shows that a single-administration SEI scale may strongly relate to the gold standard, which is promising for future research. Also, many of these concerns may be addressed through additional research. For instance, if the ISES is shown to strongly relate to the gold standard with a larger sample, then it is almost assuredly a satisfactory single-administration SEI scale, and it could be used to derive valid inferences SEI.

Together, none of these three scales are entirely satisfactory, but the ISES shows potential to measure SEI. In the current article, a new single-administration SEI scale is created to avoid the concerns with prior measures, such as the focus on specific perceptions, and the three prior SEI scales are reanalyzed to ensure their (in)ability to gauge SEI. If the undiscovered aspects of these scales prove to be concerning, then the new scale may be most suitable for gauging SEI. On the other hand, if the undiscovered aspects prove to be satisfactory, then the existing scales may already be suitable. Thus, we create a new single-administration SEI scale, followed by an exploration into the validity of the new and existing scales.

## 2. Scale development

### 2.1. Study 1 - initial scale creation

To create the new scale, an over-representative item list was created and subsequently reduced, which helps ensure that all aspects of the construct are gauged, also known as content validity (Hinkin, 1995, 1998). For this process, 35 items were created based on prior research. These items were meant to gauge general self-perceptions about SEI, rather than specific self-perceptions akin to the ISES. Example items are “My self-worth is unstable” and “Certain events make my self-esteem change.” To reduce the item list, a modified item-sort task was performed based on prior suggestions (Anderson & Gerbing, 1991; Howard & Melloy, 2016).

## 3. Method

### 3.1. Participants

Study 1 included 14 participants ( $M_{age} = 21.92$ ,  $SD_{age} = 6.11$ , 12 female, 7 Caucasian) recruited from a student subject pool in return for course credit. An attention check was used, which asked participants to mark a certain response (i.e. Please mark the third option to ensure that you are paying attention). If a participant did not provide the correct response, their data was removed. All reported statistics reflect the sample after these participants were removed.

### 3.2. Procedure

Participants signed-up for the study via a website, and they gave their digital informed consent. Then, participants were provided a thorough definition of SEI, and they were asked to indicate the extent that they believed each of the 35 items from the over-representative item list gauged SEI. Response options ranged from 0 (This item does NOT

measure SEI) to 3 (This item measures SEI perfectly). Afterwards, participants were disclosed about the purpose of the study.

#### 4. Results and Discussion

The average rating of each item was calculated. No decisive a priori cutoff was chosen. As concise measures are often preferred in research and practice, however, it was desired to reduce the survey to approximately 10 items. Six items met the cutoff of 2.9, twelve met the cutoff of 2.8, and twenty-eight met the cutoff of 2.7. For this reason, only the items that met the cutoff of 2.8 were retained, resulting in an initial measure of 12 items. These 12 items are included in Appendix A, and are henceforth entitled the SEI Measure. With the initial measure created, the following studies investigate its psychometric properties and validity.

##### 4.1. Study 2 - exploratory factor analysis

Any measure must have a stable and identifiable factor structure to be reliable and valid. For this reason, we investigate the factor structure of the SEI Measure in Study 2.

#### 5. Method

##### 5.1. Participants

Study 2 included 202 participants ( $M_{age} = 29.90$ ,  $SD_{age} = 9.26$ , 59% female, 82% American) recruited from Amazon's Mechanical Turk (mTurk) in return for a small amount of monetary compensation. This website is an online platform which connects individuals willing to perform small tasks on their computer, such as taking a survey, with those who need these tasks performed. Previous studies have shown results using mTurk as valid (Buhrmester, Kwang, & Gosling, 2011; Paolacci & Chandler, 2014; Shapiro, Chandler, & Mueller, 2013).

##### 5.2. Procedure

Participants signed-up for the study via mTurk. They provided their digital informed consent and completed the survey. Then, they were disclosed about the purpose of the study.

##### 5.3. Measures

###### 5.3.1. SEI measure

The SEI Measure created in Study 1 was administered in Study 2.

#### 6. Results and discussion

Based on the suggestions of prior scale development researchers (Hinkin, 1995, 1998; Howard, 2016), a principal axis factoring method with a direct oblimin rotation was chosen to perform the exploratory factor analysis (EFA). Further, three methods were considered to determine the number of emergent factors: a visual scree plot analysis, a parallel analysis, and the Keiser criterion (O'Connor, 2000; Patil et al., 2008). Visual scree plot analyses involve plotting the scale's eigenvalues, and all factors are retained that occur before the "elbow" or sudden drop in eigenvalues. Parallel analyses compare the scale's eigenvalues against the average eigenvalues of multiple randomized datasets (1000 replications), and factor is retained if its eigenvalue is greater than its respective randomized dataset eigenvalue. The Keiser criterion retains all factor with an eigenvalue greater than 1.0. Lastly, as suggested by prior authors (Hinkin, 1995, 1998; Howard, 2016), only items with factor loadings above .40 were retained.

With these a priori decisions, an EFA was performed. The visual scree plot analysis, parallel analysis, and Keiser criterion all suggested that the SEI Measure is unidimensional. A significant "elbow" occurred after the first factor (*eigenvalues* = 9.011, .681, .414, and .359); only the

eigenvalue of the first factor was greater than the randomized dataset (*parallel analysis eigenvalues* = 1.488, 1.365, 1.275, and 1.196); and only the first factor had an eigenvalue greater than one. Also, Table 1 includes the factor loadings of the SEI measure. All items load very strong on this single factor (>.75), and the Cronbach's alpha was .97. The collective results indicate that the scale has satisfactory psychometric properties and internal consistency.

##### 6.1. Study 3 - convergent and concurrent validity

Studies 1 and 2 provide initial support for the SEI Measure, but several other aspects must be shown before the scale can be considered valid. In Study 3, the scale's concurrent validity is studied through its relationship with state and trait self-esteem. Prior studies have shown SEI to moderately and negatively relate to state and trait self-esteem (Kernis, 2005; Kernis et al., 1993; Okada, 2010; Sowislo, Orth, & Meier, 2014), which is also expected with the SEI Measure. Also, the scale's convergent validity is studied through analyzing the scales' relationship with the gold standard of SEI measurement. As mentioned, the gold standard is obtained through multiple administrations of a single self-esteem measure, such as once a week for two weeks, and it is believed to be the only method to obtain a true measure of SEI. For any single-administration SEI scale to be valid, it must demonstrate an exceptional relationship with the gold standard. In Study 3, we analyze the relationship of the SEI Measure with the gold standard. Lastly, we noted three existing single-administration SEI scales: the RSS, Kernis's SEI scale, and the ISES. The ISES may be an adequate single-administration measure of SEI, but much is still unknown about it. In Study 3, we also analyze the relationship of these three scales with the SEI Measure, self-esteem, and the gold standard.

#### 7. Method

##### 7.1. Participants

Study 3 included 129 participants ( $M_{age} = 19.03$ ,  $SD_{age} = 1.43$ , 85% female, 76% Caucasian) recruited from a student subject pool in return for course credit. As detailed below, some participants were removed due to non-response. All reported statistics reflect the sample after these participants were removed.

##### 7.2. Procedure

Participants signed-up for the study via a website, and they gave their digital informed consent. Then, they completed a survey that included all measures noted below. After the initial survey, participants completed daily measures of self-esteem for two weeks (fourteen total occasions). After the study period ended, they were disclosed about the purpose of the study. If participants failed to complete at least ten of the daily surveys, their data was removed.

**Table 1**  
Exploratory factor analysis item loadings of the SEI measure in study 2.

	Factor loading
Item 1	.80
Item 2	.83
Item 2	.89
Item 4	.78
Item 5	.86
Item 6	.89
Item 7	.89
Item 8	.86
Item 9	.90
Item 10	.87
Item 11	.85
Item 12	.83

### 7.3. Measures

**SEI Measure.** The 12-item SEI Measure was administered ( $\alpha = .95$ ).

**RSS.** The 5-item RSS was administered ( $\alpha = .79$ ).

**Kernis's SEI scale.** Kernis's 10-item SEI scale was administered ( $\alpha = .91$ ).

**ISES.** The 4-item ISES was administered ( $\alpha = .89$ ).

**RSES.** The 10-item RSES was administered during each measurement occasion ( $\alpha = .91$ ), and participants were asked to "answer in regards to how you feel right now." This scale was used to gauge state self-esteem (first measurement occasion), trait self-esteem (average across all measurement occasions), and the gold standard (standard deviation across all measurement occasions). These uses of the RSES have been commonly applied in prior SEI studies (Greenier et al., 1999; Meier, Orth, Denissen, & Kühnel, 2011; Seery et al., 2004).

### 7.4. Single-item self-esteem scales

Due to noted concerns with the RSES, such as its potential ceiling effect or uncertain factor structure (Goldsmith, 1986; Greenberger, Chen, Dmitrieva, & Farruggia, 2003), two additional self-esteem scales were given during each measurement occasion. The first was the single-item self-esteem scale (SISE A), which has been shown to be reliable and valid for measuring self-esteem (Robins, Hendin, & Trzesniewski, 2001). The SISE asks participants' level of agreement to the item, "I have high self-esteem." The second was a self-created single-item self-esteem scale (SISE B), which asked participants to rate the level of their current self-esteem (1 – Extremely Low to 9 – Extremely High). In addition to the RSES, these two scales were used to gauge state self-esteem, trait self-esteem, and the gold standard. Both single-item scales asked participants to "answer in regards to how you feel right now."

## 8. Results and discussion

Table 2 includes the correlations and Cronbach's alphas of Study 3. The SEI Measure had large and negative relationships with state self-esteem (Average  $r = -.56$ ) and trait self-esteem (Average  $r = -.47$ ). The RSS had large and negative relationships with state self-esteem (Average  $r = -.45$ ) as well as moderate and negative relationships with trait self-esteem (Average  $r = -.34$ ). Kernis's SEI scale had moderate and negative relationships with state self-esteem (Average  $r = -.35$ ) and trait self-esteem (Average  $r = -.33$ ). The ISES had moderate-to-large and negative relationships with state self-esteem (Average  $r = -.42$ ) as well as moderate and negative relationships with trait self-esteem (Average  $r = -.30$ ). These results support the construct validity of the scales.

**Table 2**  
Correlations of all measures in study 3.

	1	2	3	4	5	6	7	8	9	10	11	12	13
1.) SEI measure	.95												
2.) RSS	.72**	.79											
3.) Kernis SEI scale	.43**	.39**	.91										
4.) ISES	.70**	.72**	.39**	.89									
5.) RSES	-.54**	-.47**	-.43**	-.50**	.91								
6.) SISE A	-.57**	-.42**	-.28**	-.38**	.65**	N/A							
7.) SISE B	-.56**	-.46**	-.33**	-.38**	.67**	.82**	N/A						
8.) Gold standard SEI two week – RSES	.31**	.25**	.24**	.29**	-.32**	-.20*	-.25**	N/A					
9.) Gold standard SEI two week – SISE A	.43**	.36**	.23*	.44**	-.38**	-.40**	-.39**	.60**	N/A				
10.) Gold standard SEI two week – SISE B	.43**	.38**	.31**	.37**	-.39**	-.36**	-.47**	.64**	.76**	N/A			
11.) Average of SE two week – RSES	-.47**	-.39**	-.42**	-.40**	.91**	.65**	.63**	-.34**	-.34**	-.37**	N/A		
12.) Average of SE two week – SISE A	-.47**	-.30**	-.28**	-.23*	.71**	.83**	.79**	-.23**	-.32**	-.34**	.78**	N/A	
13.) Average of SE two week – SISE B	-.47**	-.33**	-.30**	-.27**	.72**	.79**	.82**	-.23**	-.29**	-.31**	.77**	.94**	N/A

Note: Variables above the dotted line were administered only on Day 1.

Variables below the dotted line are the standard deviations of measures administered every day for two weeks.

\*  $p < .05$ .

\*\*  $p < .01$ .

The intercorrelations of the SEI Measure, the RSS, and the ISES were very strong (Average  $r = .71$ ); however, the relationships of Kernis's SEI scale with the other SEI scales were only moderate-to-large in size (Average  $r = .40$ ). These results support the convergent validity of the SEI Measure, RSS, and the ISES, but they fail to support Kernis's SEI scale.

Lastly, the SEI Measure resulted in only moderate correlations with the gold standard (Average  $r = .39$ ). The same was true for the RSS (Average  $r = .33$ ), Kernis's SEI scale (Average  $r = .26$ ), and the ISES (Average  $r = .37$ ). Also, some authors have suggested that the observed relationships between SEI and various outcomes may be due to their common association with trait self-esteem (Greenier et al., 1999; Kernis, Lakey, & Heppner, 2008; Seery et al., 2004). We test this possibility through regressing the gold standard onto trait self-esteem and the single-administration SEI measures, separately (Table 3). Trait self-esteem was entered in the first step and the SEI measures were entered in the second step of the regressions, which produces a metric of change in variance accounted for by the predictors. No measure was significantly related to the gold standard, as calculated from the RSES, when accounting for trait self-esteem (all  $p > .05$ ). When the gold standard was calculated through the SISE A or B, the single-administration SEI scales were significant predictors while accounting for trait self-esteem (all  $p < .01$ ), except for the Kernis SEI Scale (SISE A,  $p > .05$ ); however, these effect sizes were relatively small. The measures only accounted for five to fourteen percent of the variance in the gold standard, as denoted in the  $R^2$  change, which is well below the standard to be considered representative of a construct (Hinkin, 1995, 1998; House & Rizzo, 1972). Together, these results fail to support the convergent validity of all single-administration SEI measures.

### 8.1. Study 4 – replication

The results of Study 3 are concerning for the single-administration measurement of SEI. Any single-administration measure of SEI should have an extremely strong correlation ( $>.70$ ) with the gold standard of SEI measurement for it to be valid. As this was not seen in Study 3, none of the single-administration SEI scales can be considered valid measures. These results, however, should be considered with past research. As mentioned, in the original creation of the SISE, Chabrol et al. (2006) showed that the scale had a .81 correlation with the gold standard. When reviewing their methods, they calculated the gold standard differently than most prior research. While most prior studies use the standard deviation of measures given daily or twice daily for a week or two weeks, Chabrol et al. (2006) used the standard deviation of measurements given once a week for five weeks. This difference in timespan may cause the differences between their study and Study 3. Therefore, in Study 4, we analyze the relationship of the single-administration

**Table 3**  
Regression results of study 3.

	Gold standard SEI two week - RSES				Gold standard SEI two week - SISE A				Gold standard SEI two week - SISE B			
	β	t	β	t	β	t	β	t	β	t	β	t
Constant		8.694**		3.583**		8.170**		1.298		8.212**		1.597
1.) Average self-esteem	-.358	-4.269**	-.272	-2.897**	-.325	-3.825**	-.160	-1.758	-.328	-3.867**	-.162	-1.786
2.) SEI measure			.185	1.973			.352	3.860**			.356	3.924**
ΔR <sup>2</sup>		.13**		.03		.11**		.10**		.11**		.10**
Constant		8.694**		4.582**		8.170**		2.723**		8.212**		2.885**
1.) Average self-esteem	-.358	-4.269**	-.307	-3.387**	-.325	-3.825**	-.237	-2.777**	-.328	-3.867**	-.228	-2.652**
2.) RSS			.131	1.449			.292	3.422**			.301	3.500**
ΔR <sup>2</sup>		.13**		.02		.11**		.08**		.11**		.08**
Constant		8.654**		5.397**		8.199**		5.011**		8.270**		4.630**
1.) Average self-esteem	-.358	-4.253**	-.313	-3.381**	-.324	-3.799**	-.284	-3.221**	-.329	-3.859**	-.259	-2.971**
2.) Kernis SEI scale			.108	1.165			.147	1.665			.234	2.688**
ΔR <sup>2</sup>		.13**		.01		.11**		.02		.11**		.05**
Constant		8.694**		4.314**		8.170**		2.478*		8.212**		3.294**
1.) Average self-esteem	-.358	-4.269**	-.288	-3.174**	-.325	-3.825**	-.237	-2.946**	-.328	-3.867**	-.245	-2.924**
2.) ISES			.173	1.910			.386	4.799**			.308	3.674**
ΔR <sup>2</sup>		.13**		.03		.11**		.14**		.11**		.09**

\* p < .05.  
\*\* p < .01.

SEI scales with the gold standard as calculated through daily measurements over two weeks as well as weekly measurement over five weeks.

for course credit. As detailed below, some participants were removed due to non-response. All reported statistics reflect the sample after these participants were removed.

**9. Method**

**9.1. Participants**

Study 4 included 44 participants ( $M_{age} = 19.39, SD_{age} = 1.83, 70\%$  female, 59% Caucasian) recruited from a student subject pool in return

**9.2. Procedure**

All procedures for Study 4 were identical to Study 3; however, participants completed the daily surveys for two weeks, followed by a weekly survey for three weeks (five total weeks). This allowed for the

**Table 4**  
Correlations of all measures in study 4.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1.) SEI Measure	.96																		
2.) RSS	.69**	.81																	
3.) Kernis SEI Scale	.28	.45**	.92																
4.) ISES	.60**	.57**	.52**	.90															
5.) RSES	-.59**	-.55**	-.31*	-.60**	.86														
6.) SISE A	-.51**	-.27	-.04	-.24	.60**	N/A													
7.) SISE B	-.50**	-.46**	-.17	-.27	.62**	.76**	N/A												
8.) Gold Standard SEI Two Week - RSES	.31*	.31*	.36*	.42**	-.35*	-.07	-.20	N/A											
9.) Gold Standard SEI Two Week - SISE A	.42**	.33*	.25	.46**	-.56**	-.41**	-.46**	.76**	N/A										
10.) Gold Standard SEI Two Week - SISE B	.43**	.22	.36*	.47**	-.37*	-.12	-.25	.76**	.77**										
11.) Average of SE Two Week - RSES	-.48**	-.39**	-.12	-.48**	.84**	.54**	.54**	-.33*	-.44**	-.34*	N/A								
12.) Average of SE Two Week - SISE A	-.23	-.11	.17	-.15	.55**	.68**	.61**	-.11	-.40**	-.10	.62**	N/A							
13.) Average of SE Two Week - SISE B	-.11	-.10	.21	-.01	.45**	.55**	.60**	-.11	-.40**	-.05	.55**	.93**	N/A						
14.) Gold Standard SEI Five Week - RSES	.25	.18	.27	.40**	-.28	-.15	-.17	.85**	.71**	.65**	-.24	-.17	-.13	N/A					
15.) Gold Standard SEI Five Week - SISE A	.32*	.34*	.33*	.46**	-.57**	-.45**	-.50**	.62**	.79**	.55**	-.36*	-.26	-.24	.77**	N/A				
16.) Gold Standard SEI Five Week - SISE B	.24	.34*	.49**	.45**	-.42**	-.07	-.27	.71**	.63**	.68**	-.21	.05	.04	.75**	.77**	N/A			
17.) Average of SE Five Week - RSES	-.42**	-.39**	-.09	-.47**	.83**	.51**	.53**	.31*	-.43**	-.31*	.96**	.63**	.55**	-.29	-.44**	-.23	N/A		
18.) Average of SE Five Week - SISE A	-.30*	-.14	.18	-.20	.55**	.59**	.62**	-.17	-.39**	-.13	.62**	.92**	.54**	-.31*	-.38*	-.02	.69**	N/A	
19.) Average of SE Five Week - SISE B	-.18	-.13	.22	-.09	.45**	.53**	.64**	-.17	-.39**	-.13	.57**	.88**	.91**	-.27	-.34*	-.04	.64**	.92**	N/A

Note: Variables above the dotted lines were administered only on Day 1. Variables between the dotted lines are the standard deviations of measures administered every day for two weeks. Variables below the dotted lines are the standard deviations of measures administered every week for five weeks.  
\* p < .05.  
\*\* p < .01.

calculation of the gold standard through daily administrations for two weeks as well as through weekly administrations for five weeks (seventeen total occasions). If participants completed less than ten of the daily surveys or four or more of the weekly surveys, their data was removed.

### 9.3. Measures

All measures administered in Study 4 were identical to Study 3.

## 10. Results and discussion

Correlations and Cronbach's alphas for the measures administered in Study 4 are included in Table 4. The SEI Measure demonstrated large and negative relationships with state self-esteem (Average  $r = -.53$ ) as well as moderate and negative relationships trait self-esteem (two weeks, Average  $r = -.27$ ; five weeks, Average  $r = -.30$ ). The RSS had moderate-to-large and negative relationships with state self-esteem (Average  $r = -.43$ ) as well as small-to-moderate and negative relationships with trait self-esteem (two weeks, Average  $r = -.20$ ; five weeks, Average  $r = -.22$ ). Kernis's SEI scale had small and negative relationships with state self-esteem (Average  $r = -.17$ ) but small and positive correlations with trait self-esteem (two weeks, Average  $r = .09$ ; five weeks, Average  $r = .10$ ). The ISES had moderate-to-large and negative relationships with state self-esteem (Average  $r = -.42$ ) as well as small-to-moderate and negative relationships with trait self-esteem (two weeks, Average  $r = -.21$ ; five weeks, Average  $r = -.25$ ). These results support the construct validity of the scales, except Kernis's SEI scale.

Also, the intercorrelations of the SEI Measure, the RSS, and the ISES were very strong (Average  $r = .62$ ); however, the relationships of Kernis's SEI scale with the other SEI scales were only moderate-to-large in size (Average  $r = .42$ ). These results support the convergent validity of the SEI Measure, RSS, and the ISES, but they fail to support Kernis's SEI scale.

When calculated by daily measurements for two weeks, the SEI Measure resulted in moderate correlations with the gold standard (Average  $r = .39$ ). The same was true for the RSS (Average  $r = .29$ ) and Kernis's SEI scale (Average  $r = .32$ ), whereas the ISES (Average  $r = .45$ ) had moderate-to-large correlations with the gold standard. Regressions were again performed to analyze the relationship of the SEI measures with the gold standard while accounting for trait self-esteem (Table 5). When regressing the gold standard, as calculated through the RSES, on trait self-esteem and the single-administration SEI measures, the SEI measures only accounted for three to eleven percent of

the variance in the gold standard. When the gold standard was calculated through the SISE A or B, the SEI measures accounted for five to 22 percent of the variance in the gold standard, which is still well below many suggestions for a scale to be considered representative of a construct (Hinkin, 1995, 1998; House & Rizzo, 1972). These results largely replicate those of Study 3, and fail to support the convergent validity of all single-administration SEI measures.

When calculated by weekly measures for five weeks, the SEI Measure resulted in moderate correlations with the gold standard (Average  $r = .27$ ). The same was true for the RSS (Average  $r = .29$ ) and Kernis's SEI scale (Average  $r = .36$ ), whereas the ISES (Average  $r = .44$ ) had moderate-to-large correlations with the gold standard. When regressing the gold standard, as calculated through the RSES, on trait self-esteem and the single-administration SEI measures, the SEI measures only accounted for one to nine percent of the variance in the gold standard. When the gold standard was calculated through the SISE A or B, the SEI measures accounted for five to 26 percent of the variance in the gold standard, still below many standard cutoffs (Hinkin, 1995, 1998; House & Rizzo, 1972). These results are presented in Table 6. Once again, these results fail to support the convergent validity of all single-administration SEI measures. Overall, no single-administration measure of SEI appears to be a satisfactory measure, as suggested by their correlations with the gold standard calculated by daily or weekly measurement occasions.

## 11. Overall discussion

The goal of the current article was to create a new single-administration SEI scale that adequately correlates to the gold standard of SEI measurement as well as to reanalyze existing single-administration SEI scales. To do so, Studies 1 and 2 created the new scale, entitled the SEI Measure, and demonstrated that it had a superb factor structure. Study 3 analyzed the relationship of the SEI Measure and existing single-administration SEI scales with the gold standard as gauged through daily administrations of the RSE for two weeks. Study 4 likewise analyzed the relationship of the SEI Measure and existing single-administration SEI scales with the gold standard, but the gold standard was calculated through daily administrations of the RSE for two weeks as well as weekly administrations for five weeks. Through these four studies, the current article provides the most robust analysis of single-administration SEI scales to date.

The results showed that no single-administration SEI scale, whether new or existing, adequately converged with the gold standard, whether calculated through daily administrations for two weeks or weekly

**Table 5**  
Regression results of study 4 (Gold standard measured daily for two weeks).

	Gold standard SEI two week - RSES				Gold standard SEI two week - SISE A				Gold standard SEI two week - SISE B			
	$\beta$	t	$\beta$	t	$\beta$	t	$\beta$	t	$\beta$	t	$\beta$	t
Constant		3.676**		1.754		4.969**		2.808**		2.081*		.619
1.) Average self-esteem	-.333	-2.288*	-.238	-1.444	-.404	-2.860**	-.325	-2.370*	-.051	-.331	-.006	-.042
2.) SEI Measure			.198	1.200			.341	2.488*			.425	2.988**
$\Delta R^2$		.11*		.03		.16**		.11*		.00		.18**
Constant		3.676**		1.998		4.969**		3.486**		2.081*		1.264
1.) Average self-esteem	-.333	-2.288*	-.252	-1.608	-.404	-2.860**	-.371	-2.715**	-.051	-.331	-.030	-.193
2.) RSS			.209	1.332			.285	2.087*			.218	1.425
$\Delta R^2$		.11*		.04		.16**		.08*		.00		.05
Constant		3.676**		2.596*		4.969**		4.628**		2.081*		1.897
1.) Average self-esteem	-.333	-2.288*	-.295	-2.122*	-.404	-2.860**	-.457	-3.364**	-.051	-.331	-.135	-.913
2.) Kernis SEI scale			.330	2.374*			.322	2.367*			.391	2.649*
$\Delta R^2$		.11*		.11*		.16**		.10*		.00		.15*
Constant		3.676**		1.143		4.969**		2.892**		2.081*		.759
1.) Average self-esteem	-.333	-2.288*	-.169	-1.065	-.404	-2.860**	-.344	-2.654*	-.051	-.331	-.048	-.351
2.) ISES			.343	2.160*			.410	3.170**			.472	3.433**
$\Delta R^2$		.11*		.09*		.16**		.17**		.00		.22**

\*  $p < .05$ .

\*\*  $p < .01$ .

**Table 6**  
Regression results of study 4 (Gold standard measured weekly for five weeks).

	Gold standard SEI two week - RSES				Gold standard SEI two week - SISE A				Gold standard SEI two week - SISE B			
	$\beta$	t	$\beta$	t	$\beta$	t	$\beta$	t	$\beta$	t	$\beta$	t
Constant		3.148**		1.626		4.397**		2.524*		1.582		.497
1.) Average self-esteem	-.292	-1.978	-.225	-1.380	-.375	-2.623*	-.308	-2.085*	-.038	-.245	.005	.030
2.) SEI measure			.157	.965			.226	1.536			.242	1.575
$\Delta R^2$		.09		.02		.14*		.05		.00		.06
Constant		3.148**		2.049*		4.397**		2.870**		1.582		.313
1.) Average self-esteem	-.292	-1.978	-.261	-1.617	-.375	-2.623*	-.333	-2.402*	-.038	-.245	.007	.050
2.) RSS			.081	.500			.296	2.135*			.338	2.276*
$\Delta R^2$		.09		.01		.14*		.09*		.00		.11*
Constant		3.148**		2.271*		4.397**		4.152**		1.582		1.213
1.) Average self-esteem	-.292	-1.978	-.269	-1.858	-.375	-2.623*	-.451	-3.411**	-.038	-.245	-.149	-1.082
2.) Kernis SEI scale			.249	1.718			.415	3.140**			.518	3.754**
$\Delta R^2$		.09		.06		.14*		.17**		.00		.26**
Constant		3.148**		.779		4.397**		2.210*		1.582		-.056
1.) Average self-esteem	-.292	-1.978	-.135	-.835	-.375	-2.623*	-.295	-2.201*	-.038	-.245	.002	.013
2.) ISES			.332	2.054*			.400	2.990**			.446	3.177**
$\Delta R^2$		.09		.09*		.14*		.15**		.00		.20**

\*  $p < .05$ .

\*\*  $p < .01$ .

administrations for five weeks. The correlations of these scales with the gold standard ranged from .18 to .47, whereas the cutoff for acceptable convergent validity is often considered to be .70. Likewise, when controlling for trait self-esteem, these scales only accounted for one to 26 percent of the variance in the gold standard, which is well below many preferred cutoffs (Hinkin, 1995, 1998; House & Rizzo, 1972). Aside from Kernis's measure (et al., 1992), the single-administration SEI scales had strong relationships with each other, however, suggesting that they likely gauge the same or theoretically similar constructs – although none likely gauge SEI. Lastly, while not the focus of the current article, Studies 3 and 4 provide further evidence that SEI is moderately and negatively related to state and trait self-esteem, as suggested and found by prior authors (Kernis, 2005; Kernis et al., 2008; Okada, 2010). Although contrary to expectations, these results still have several implications for future research.

## 12. Theoretical implications

First and foremost, the quest for a satisfactory single-administration SEI scale should likely be refocused. Authors have long noted the benefits of such a measure (Chabrol et al., 2006; Kernis et al., 1992; Rosenberg, 1965), and it could indeed provide great advancements to research and practice. People may be unable, however, to accurately perceive their changes in self-esteem, preventing any self-report scale from accurately representing actual changes to self-esteem. The quest for a satisfactory SEI scale that is not self-reported, though, may still be worthwhile to pursue. Great advancements have been seen in the creation of implicit measures and conditional reasoning tests (Bosson, Swann, & Pennebaker, 2000; Gebauer, Riketta, Broemer, & Maio, 2008; Schimmack & Diener, 2003). These methods do not require participants to report conscious self-perceptions, and they may serve as satisfactory methods to gauge SEI in a single administration.

Also, Study 4 provides direct evidence for the need of empirical replication in research. As mentioned, Chabrol et al. (2006) showed that their single-administration SEI scale had an extremely strong correlation with the gold standard, almost double the average correlation observed in the current article when using the same method; however, these authors used a very small sample size for their investigation. Without the attempt at replicating their results in the current article, which showed that their scale may not actually gauge SEI, other researchers may have applied their measure and created inappropriate theoretical inferences about SEI. While the current article is not enough to fully claim that their scale is not valid for gauging SEI, more research is certainly needed before any inferences can be obtained from its application.

Additionally, the current article draws attention to the gold standard of SEI measurement. Almost every SEI study gauges the construct through multiple administrations of the RSES. Studies 3 and 4 also applied this method, but both studies also used two different single-item self-esteem scales to calculate the gold standard. The three gold standards, as calculated by the different scales, had very strong inter-relationships (*Daily for Two Week Average*  $r = .72$ ; *Weekly for Five Week Average*  $r = .76$ ), but they were certainly not identical. Prior research has shown that multiple-item scales are more reliable than single-item scales (Robins et al., 2001; Wanous, Reichers, & Hudy, 1997). In Studies 3 and 4, the variability of participants' self-esteem scores was greater when using the single-item scales compared to the RSE, likewise causing their SEI to be greater. While the current studies cannot determine whether single-item scales are a more accurate method to gauge SEI, it should be recognized that the dominant method to gauge SEI may not be the most ideal and other scales should be considered when calculating the gold standard.

Relatedly, the current article also draws attention to gauging the gold standard through differing amounts and frequencies of measurement occasions. When comparing the two methods in Study 4, daily for two weeks and weekly for five weeks, the inter-correlations ranged from .55 to .85 (*Average*  $r = .69$ ). This result suggests that the different measurement methods do not provide identical results. Once again, the current studies cannot determine which method is more accurate in gauging SEI, but it should be considered that gauging the gold standard through differing amounts and frequencies of measurement occasions may provide differing results.

Lastly, Studies 3 and 4 demonstrated that the multiple single-administration SEI scales had satisfactory convergent validity, except for Kernis's et al., 1992 SEI scale, suggesting that the scales gauge the same or very theoretically similar constructs. Studies 3 and 4 also showed that perceptions of SEI are conceptually distinct from actual SEI. It is possible that perceived SEI has notable relationships with important personal outcomes, causing it to be a worthwhile construct to study – as discussed below.

## 13. Future directions

Despite the unexpected results, the current article opens several avenues for future research. While it is not expected that any of the existing single-administration SEI scales are valid for gauging SEI, authors may want to consider investigating the differences between the study of Chabrol et al. (2006) and Study 4. These authors found that the ISES had a .81 correlation with the gold standard as gauged through weekly measurement occasions for five weeks, but only a .44 correlation

was found in Study 4. The difference in these correlations is statistically significant ( $z = 2.22$ ,  $p < .05$ ), suggesting that a particular cause may have driven these difference. It may be that Chabrol et al. (2006) used high school student participants, whereas Study 4 used college student participants; it may be because their participants were French, whereas the Study 4 participants were American; or it may be because their participants completed paper surveys in the presence of a researcher, whereas Study 4 participants completed their surveys entirely online. Through finding the cause for these disparate results, future researchers could discover methods to create new measures that more accurately gauge SEI or populations that may be able to more accurately report on their SEI.

Further, other approaches may still sufficiently gauge SEI in a single-administration or a very short timeframe. For instance, many implicit measures exist to gauge self-esteem (Bosson et al., 2000; Gebauer et al., 2008; Schimmack & Diener, 2003), and such implicit measures may be altered to gauge SEI. Likewise, participants' self-esteem may be altered in a single lab visit, such as through writing a self-critical story (Dijksterhuis, 2004; Greenberg et al., 1993; Harmon-Jones, Simon, Greenberg, Pyszczynski, & Solomon, 1997), and the extent of a participant's change in self-esteem could be representative of their SEI. Future research should think beyond self-report scales for gauging SEI, such as through the creation of implicit measures and experimental manipulations, and these avenues may still provide all of the suggested benefits of a single-administration scale.

Additionally, while single-administration SEI scales may not adequately gauge SEI, they may still gauge a construct important for future study. Perceived SEI is likely an aspect of self-perceptions, and future research should position perceived SEI in the larger nomological net of related constructs. It is possible that perceived SEI facet of emotional stability, as emotionally stable people are described as *not* moody, temperamental, or fretful (Saucier, 1994; Thompson, 2008). If shown to be the case, theoretical basis for future investigation on perceived SEI could be obtained. Further, future research should also study the outcomes of perceived SEI. While SEI is related to important personal outcomes, such as depression and life satisfaction (de Man et al., 2001; Meier et al., 2009), it is not guaranteed that perceived SEI is also related to such outcomes. Likewise, perceived SEI may more strongly predict certain outcomes than SEI, or it may even be related to outcomes that are not predicted by SEI at all.

Because the current study further reinforced that researchers should obtain the gold standard to gauge SEI, future research should investigate novel methods to obtain multiple measurements of SEI. As mentioned, the difficulty of obtaining multiple measurements of self-esteem prevents most SEI studies from using non-student samples. Through applying novel methods, such as ecological momentary assessments using cellular devices (Beal & Weiss, 2003; Smyth & Stone, 2003), researchers may be able to more easily obtain repeated measurements of self-esteem, allowing for more diverse studies on SEI.

Lastly, it is still possible, albeit unlikely, that perceived SEI is actually the most ideal method to gauge SEI, even better than the gold standard. Just because the perceived SEI measures failed to converge with the gold standard does not necessarily mean that these measures are inferior to the gold standard. Instead, the gold standard itself may not adequately gauge SEI, and the differences between perceived SEI and the gold standard may be due to measurement concerns of the gold standard. To investigate such a possibility, future research should study the nomological net of perceived SEI and compare the results to prior findings on the gold standard. If the relationships of perceived SEI are more theoretically sound compared to the gold standard, researchers should reconsider that the gold standard is the most ideal method to gauge SEI. Likewise, it is also possible that perceived SEI is adequate to gauge certain types or aspects of SEI. The current article only measured daily and weekly changes in self-esteem, but perceived SEI may be apt at gauging other timeframes, such as hourly or yearly changes.

## 14. Limitations

As with any article, certain limitations should be noted. Studies 3 and 4 used student samples, which have been previously criticized (Bello et al., 2009; Greenberg, 1987; Turner & Upshur, 2002). While certainly a weakness, the use of student samples is typical of almost all research on SEI, and is difficult to avoid due to the gold standard of SEI measurement.

The use of single-item scales has also been criticized (Robins et al., 2001; Wanous et al., 1997), and some may disagree with the use of single-item scales to calculate the gold standard. These scales were used to keep survey length at a minimum, as prior research has shown that data quality decreases with survey length (Heerwegh & Loosveldt, 2006; Hoerger, 2010). Also, any results obtained from the single-item scales were supported by the results obtained from the RSES. Therefore, the application of these scales is not believed to be a concern.

## 15. Conclusion

The goal of the current article was to create a new single-administration SEI scale that adequately correlates to the gold standard of SEI measurement as well as to reanalyze existing single-administration SEI scales. The results showed that neither the new or existing scales likely gauge SEI, due to their small correlations with the gold standard of SEI measurement. Despite these surprising results, the findings shed light onto new areas for future research, such as the study of perceived SEI apart from SEI.

## Appendix A. SEI measure

In the following, indicate how much you agree or disagree that the statement describes yourself. For these questions, **answer in regards to how you feel right now.**

- 1 – Strongly disagree
- 2 – Disagree
- 3 – Slightly disagree
- 4 – Neither disagree or agree
- 5 – Slightly agree
- 6 – Agree
- 7 – Strongly agree

- 1.) My opinion of myself often changes.
- 2.) My self-esteem is instable.
- 3.) I often have differing opinions about myself.
- 4.) Certain events make my self-esteem change.
- 5.) I feel differently about myself on a day-to-day basis.
- 6.) I have an unstable opinion about myself.
- 7.) My self-esteem goes up and down.
- 8.) My feelings about myself change.
- 9.) My self-esteem level is constantly changing.
- 10.) My self-esteem is NOT constant.
- 11.) I am often unsure how I view myself.
- 12.) My self-worth is unstable.

## References

- Anderson, J. C., & Gerbing, D. W. (1991). Predicting the performance of measures in a confirmatory factor analysis with a pretest assessment of their substantive validities. *J. Appl. Psychol.*, 76(5), 732.
- Baumeister, R. F., Smart, L., & Boden, J. M. (1996). Relation of threatened egotism to violence and aggression: The dark side of high self-esteem. *Psychol. Rev.*, 103(1), 5.
- Beal, D. J., & Weiss, H. M. (2003). Methods of ecological momentary assessment in organizational research. *Organ. Res. Methods*, 6(4), 440–464.
- Bello, D., Leung, K., Radebaugh, L., Tung, R. L., & Van Wittenloostuijn, A. (2009). From the editors: Student samples in international business research. *J. Int. Bus. Stud.*, 40(3), 361–364.

- Bosson, J. K., Swann, W. B., Jr., & Pennebaker, J. W. (2000). Stalking the perfect measure of implicit self-esteem: The blind men and the elephant revisited? *J. Pers. Soc. Psychol.*, 79(4), 631.
- Brown, G. W., Andrews, B., Harris, T., Adler, Z., & Bridge, L. (1986). Social support, self-esteem and depression. *Psychol. Med.*, 16(04), 813–831.
- Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk a new source of inexpensive, yet high-quality, data? *Perspect. Psychol. Sci.*, 6(1), 3–5.
- Cassidy, C., O'Connor, R. C., Howe, C., & Warden, D. (2004). Perceived discrimination and psychological distress: The role of personal and ethnic self-esteem. *J. Couns. Psychol.*, 51(3), 329.
- Chabrol, H., Rousseau, A., & Callahan, S. (2006). Preliminary results of a scale assessing instability of self-esteem. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, 38(2), 136.
- Cook, T. D., Campbell, D. T., & Day, A. (1979). *Quasi-experimentation: Design & analysis issues for field settings*. 351. Boston: Houghton Mifflin.
- Crocker, J., Luhtanen, R. K., Cooper, M. L., & Bouvrette, A. (2003). Contingencies of self-worth in college students: Theory and measurement. *J. Pers. Soc. Psychol.*, 85(5), 894.
- De Cremer, D. (2002). The self-relevant implications of distribution rules: When self-esteem and acceptance are influenced by violations of the equity rule. *Soc. Justice Res.*, 15(4), 327–339.
- De Man, A. F., & Gutierrez, B. I. (2002). The relationship between level of self-esteem and suicidal ideation with stability of self-esteem as moderator. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, 34(4), 235.
- De Man, A. F., Gutierrez, B. I., & Sterk, N. (2001). Stability of self-esteem as moderator of the relationship between level of self-esteem and depression. *N. Am. J. Psychol.*, 3(2).
- Dijksterhuis, A. (2004). I like myself but I don't know why: Enhancing implicit self-esteem by subliminal evaluative conditioning. *J. Pers. Soc. Psychol.*, 86(2), 345.
- Gebauer, J. E., Riketta, M., Broemer, P., & Maio, G. R. (2008). "How much do you like your name?" An implicit measure of global self-esteem. *J. Exp. Soc. Psychol.*, 44(5), 1346–1354.
- Goldsmith, R. E. (1986). Dimensionality of the Rosenberg self-esteem scale. *Journal of Social Behavior and Personality*, 1(2), 253.
- Greenberg, J. (1987). The college sophomore as guinea pig: Setting the record straight. *Acad. Manag. Rev.*, 12(1), 157–159.
- Greenberg, J., Pyszczynski, T., & Solomon, S. (1986). The causes and consequences of a need for self-esteem: A terror management theory. *Public self and private self* (pp. 189–212). Springer New York.
- Greenberg, J., Pyszczynski, T., Solomon, S., Pincel, E., Simon, L., & Jordan, K. (1993). Effects of self-esteem on vulnerability-denying defensive distortions: Further evidence of an anxiety-buffering function of self-esteem. *J. Exp. Soc. Psychol.*, 29(3), 229–251.
- Greenberger, E., Chen, C., Dmitrieva, J., & Farruggia, S. P. (2003). Item-wording and the dimensionality of the Rosenberg Self-Esteem Scale: Do they matter? *Personal Individ. Differ.*, 35(6), 1241–1254.
- Greenier, K. D., Kernis, M. H., McNamara, C. W., Waschull, S. B., Berry, A. J., Herlocker, C. E., & Abend, T. A. (1999). Individual differences in reactivity to daily events: Examining the roles of stability and level of self-esteem. *J. Pers.*, 67(1), 187–208.
- Harmon-Jones, E., Simon, L., Greenberg, J., Pyszczynski, T., Solomon, S., & McGregor (1997). Terror management theory and self-esteem: Evidence that increased self-esteem reduced mortality salience effects. *J. Pers. Soc. Psychol.*, 72(1), 24.
- Heerwegh, D., & Loosveldt, G. (2006). An experimental study on the effects of personalization, survey length statements, progress indicators, and survey sponsor logos in Web Surveys. *J. Off. Stat.*, 22(2), 191.
- Hinkin, T. R. (1995). A review of scale development practices in the study of organizations. *J. Manag.*, 21(5), 967–988.
- Hinkin, T. R. (1998). A brief tutorial on the development of measures for use in survey questionnaires. *Organ. Res. Methods*, 1(1), 104–121.
- Hoerger, M. (2010). Participant dropout as a function of survey length in Internet-mediated university studies: Implications for study design and voluntary participation in psychological research. *Cyberpsychology, Behavior, and Social Networking*, 13(6), 697–700.
- House, R. J., & Rizzo, J. R. (1972). Toward the measurement of organizational practices: Scale development and validation. *J. Appl. Psychol.*, 56(5), 388.
- Howard, M. (2016). A review of Exploratory Factor Analysis (EFA) decisions and overview of current practices: What we are doing and how can we improve? *International Journal of Human-Computer Interaction*, 32(1), 51–62.
- Howard, M. C., & Melloy, R. C. (2016). Evaluating item-sort task methods: The presentation of a new statistical significance formula and methodological best practices. *J. Bus. Psychol.*, 31(1), 51–62.
- Judge, T. A., & Bono, J. E. (2001). Relationship of core self-evaluations traits—self-esteem, generalized self-efficacy, locus of control, and emotional stability—with job satisfaction and job performance: A meta-analysis. *J. Appl. Psychol.*, 86(1), 80.
- Judge, T. A., Erez, A., & Bono, J. E. (1998). The power of being positive: The relation between positive self-concept and job performance. *Hum. Perform.*, 11(2–3), 167–187.
- Kernis, M. H. (2005). Measuring self-esteem in context: The importance of stability of self-esteem in psychological functioning. *J. Pers.*, 73(6), 1569–1605.
- Kernis, M. H., & Waschull, S. B. (1995). *The interactive roles of stability and level of self-esteem: Research and theory*.
- Kernis, M. H., Grannemann, B. D., & Barclay, L. C. (1989). Stability and level of self-esteem as predictors of anger arousal and hostility. *J. Pers. Soc. Psychol.*, 56(6), 1013.
- Kernis, M. H., Grannemann, B. D., & Mathis, L. C. (1991). Stability of self-esteem as a moderator of the relation between level of self-esteem and depression. *J. Pers. Soc. Psychol.*, 61(1), 80.
- Kernis, M. H., Grannemann, B. D., & Barclay, L. C. (1992). Stability of self-esteem: Assessment, correlates, and excuse making. *J. Pers.*, 60(3), 621–644.
- Kernis, M. H., Cornell, D. P., Sun, C. R., Berry, A., & Harlow, T. (1993). There's more to self-esteem than whether it is high or low: The importance of stability of self-esteem. *J. Pers. Soc. Psychol.*, 65(6), 1190.
- Kernis, M. H., Greenier, K. D., Herlocker, C. E., Whisenhunt, C. R., & Abend, T. A. (1997). Self-perceptions of reactions to doing well or poorly: The roles of stability and level of self-esteem. *Personal Individ. Differ.*, 22(6), 845–854.
- Kernis, M. H., Lakey, C. E., & Heppner, W. L. (2008). Secure versus fragile high self-esteem as a predictor of verbal defensiveness: Converging findings across three different markers. *J. Pers.*, 76(3), 477–512.
- Knowles, R., Tai, S., Jones, S. H., Highfield, J., Morris, R., & Bentall, R. P. (2007). Stability of self-esteem in bipolar disorder: Comparisons among remitted bipolar patients, remitted unipolar patients and healthy controls. *Bipolar Disord.*, 9(5), 490–495.
- Kugle, C. L., Clements, R. O., & Powell, P. M. (1983). Level and stability of self-esteem in relation to academic behavior of second graders. *J. Pers. Soc. Psychol.*, 44(1), 201.
- Leary, M. R., & Baumeister, R. F. (2000). *The nature and function of self-esteem: Sociometer theory*.
- Meier, L. L., Semmer, N. K., & Hupfeld, J. (2009). The impact of unfair treatment on depressive mood: The moderating role of self-esteem level and self-esteem instability. *Personal Soc. Psychol. Bull.*, 35(5), 643–655.
- Meier, L. L., Orth, U., Denissen, J. J., & Kühnel, A. (2011). Age differences in instability, contingency, and level of self-esteem across the life span. *J. Res. Pers.*, 45(6), 604–612.
- Milevsky, A., Schlechter, M., Netter, S., & Keehn, D. (2007). Maternal and paternal parenting styles in adolescents: Associations with self-esteem, depression and life-satisfaction. *J. Child Fam. Stud.*, 16(1), 39–47.
- Myers, A., & Rosen, J. C. (1999). Obesity stigmatization and coping: Relation to mental health symptoms, body image, and self-esteem. *International journal of obesity and related metabolic disorders: Journal of the International Association for the Study of Obesity*, 23(3), 221–230.
- Myers, J. E., Willse, J. T., & Villalba, J. A. (2011). Promoting self-esteem in adolescents: The influence of wellness factors. *J. Couns. Dev.*, 89(1), 28–36.
- Nicole, P. (1696). *Moral essays, contain'd in several treatises on many important duties*, 4 vols. London: printed for Samuel Manship.
- O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behav. Res. Methods Instrum. Comput.*, 32(3), 396–402.
- Okada, R. (2010). A meta-analytic review of the relation between self-esteem level and self-esteem instability. *Personal Individ. Differ.*, 48(2), 243–246.
- Oosterwegel, A., Field, N., Hart, D., & Anderson, K. (2001). The relation of self-esteem variability to emotion variability, mood, personality traits, and depressive tendencies. *J. Pers.*, 69(5), 689–708.
- Paolacci, G., & Chandler, J. (2014). Inside the turk understanding mechanical turk as a participant pool. *Curr. Dir. Psychol. Sci.*, 23(3), 184–188.
- Patil, V. H., Singh, S. N., Mishra, S., & Donovan, D. T. (2008). Efficient theory development and factor retention criteria: Abandon the 'eigenvalue greater than one' criterion. *J. Bus. Res.*, 61(2), 162–170.
- Roberts, J. E., Kassel, J. D., & Gotlib, I. H. (1995). Level and stability of self-esteem as predictors of depressive symptoms. *Personal Individ. Differ.*, 19(2), 217–224.
- 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. *Personal Soc. Psychol. Bull.*, 27(2), 151–161.
- Rosenberg, M. (1965). *Society and the adolescent self-image*.
- Saucier, G. (1994). Mini-markers: A brief version of Goldberg's unipolar Big-Five markers. *J. Pers. Assess.*, 63(3), 506–516.
- 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. *J. Pers. Soc. Psychol.*, 67(6), 1063.
- Schimmack, U., & Diener, E. (2003). Predictive validity of explicit and implicit self-esteem for subjective well-being. *J. Res. Pers.*, 37(2), 100–106.
- Sedikides, C., Rudich, E. A., Gregg, A. P., Kumashiro, M., & Rusbul, C. (2004). Are normal narcissists psychologically healthy?: Self-esteem matters. *J. Pers. Soc. Psychol.*, 87(3), 400.
- Seery, M. D., Blascovich, J., Weisbuch, M., & Vick, S. B. (2004). The relationship between self-esteem level, self-esteem stability, and cardiovascular reactions to performance feedback. *J. Pers. Soc. Psychol.*, 87(1), 133.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Wadsworth Cengage learning.
- Shapiro, D. N., Chandler, J., & Mueller, P. A. (2013). *Using Mechanical Turk to study clinical populations*. Clinical Psychological Science.
- Smyth, J. M., & Stone, A. A. (2003). Ecological momentary assessment research in behavioral medicine. *J. Happiness Stud.*, 4(1), 35–52.
- Sowislo, J. F., Orth, U., & Meier, L. L. (2014). What constitutes vulnerable self-esteem? Comparing the prospective effects of low, unstable, and contingent self-esteem on depressive symptoms. *J. Abnorm. Psychol.*, 123(4), 737.
- Stokes, J. P., & Peterson, J. L. (1998). *Homophobia, self-esteem, and risk for HIV among African American men who have sex with men*. AIDS Education and Prevention.
- Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychol. Bull.*, 103(2), 193.
- Thompson, E. R. (2008). Development and validation of an international English big-five mini-markers. *Personal Individ. Differ.*, 45(6), 542–548.
- Turner, C. E., & Upshur, J. A. (2002). Rating scales derived from student samples: Effects of the scale maker and the student sample on scale content and student scores. *TESOL Q.*, 49–70.
- Wanous, J. P., Reichers, A. E., & Hudy, M. J. (1997). Overall job satisfaction: How good are single-item measures? *J. Appl. Psychol.*, 82(2), 247.

- Waschull, S. B., & Kernis, M. H. (1996). Level and stability of self-esteem as predictors of children's intrinsic motivation and reasons for anger. *Personal. Soc. Psychol. Bull.*, 22(1), 4–13.
- Webster, G. D., Kirkpatrick, L. A., Nezek, J. B., Smith, C. V., & Paddock, E. L. (2007). Different slopes for different folks: Self-esteem instability and gender as moderators of the relationship between self-esteem and attitudinal aggression. *Self Identity*, 6(1), 74–94.
- Wilkinson, R. B. (2004). The role of parental and peer attachment in the psychological health and self-esteem of adolescents. *Journal of Youth and Adolescence*, 33(6), 479–493.
- Zeigler-Hill, V. (2006). Discrepancies between implicit and explicit self-esteem: Implications for narcissism and self-esteem instability. *J. Pers.*, 74(1), 119–144.
- Zeigler-Hill, V., & Abraham, J. (2006). Borderline personality features: Instability of self-esteem and affect. *J. Soc. Clin. Psychol.*, 25(6), 668.
- Zeigler-Hill, V., & Showers, C. J. (2007). Self-structure and self-esteem stability: The hidden vulnerability of compartmentalization. *Personal. Soc. Psychol. Bull.*, 33(2), 143–159.