Running Head: INDIRECT MEASURES OF SELF-ESTEEM

Through a Glass, Less Darkly?

Reassessing Convergent and Discriminant Validity in Measures of Implicit Self-Esteem

Almut Rudolph[1]

Michela Schröder-Abé[1]

Astrid Schütz[1]

Aiden P. Gregg[2]

Constantine Sedikides[2]

[1]Chemnitz University of Technology

[2]University of Southampton

Key words: implicit self-esteem, explicit self-esteem, Implicit Association Test, Extrinsic Affective Simon Task, Single Category Implicit Association Test, Go/No-Go Association Task

Sponsors: German Research Foundation, grant SCHU 1459/2; Economic and Social Research Council, grant RES-000-23-0331

Address for correspondence:

Almut Rudolph

Department of Psychology,

Chemnitz University of Technology

D-09107 Chemnitz, Germany

email: almut.rudolph@phil.tu-chemnitz.de

Abstract

Self-esteem has been traditionally assessed via self-report (explicit self-esteem: ESE). However, the limitations of self-report have prompted efforts to assess self-esteem indirectly (implicit self-esteem: ISE). It has been theorized that ISE and ESE reflect the operation of largely distinct mental systems. However, although low correlations between measures of ISE and ESE empirically support their discriminant validity, similarly low correlations between different measures of ISE do not support their convergent validity. We explored whether such patterns would re-emerge if more newly developed, specific, and reliable ISE measures were used. They did, although some convergent validity among ISE measures emerged once confounds due to conceptual mismatch, individual differences, and random variability were minimized. Nonetheless, low correlations among ISE measures are not primarily due to the usual psychometric suspects, and may be the result of other factors including subtle differences between structural features of such measures.

Through a Glass, Less Darkly?

Reassessing Convergent and Discriminant Validity in Measures of Implicit Self-Esteem

To investigate people's attitude towards themselves—their *self-esteem*—psychologists have traditionally relied on self-report (*explicit self-esteem* or ESE; Rosenberg, 1965). Fortunately, when reporting their self-esteem, people are reasonably knowledgeable about themselves, honest with themselves, and honest with others. Nonetheless, people sometimes lack self-insight ("How do I feel about myself *really*?"; Wilson, 2002), deceive others ("I really think I'm useless, but I better pretend to be great!"; Schlenker & Leary, 1982), or even deceive themselves ("I'm great—even if everyone hates me!"; Paulhus, Fridhandler, & Hayes, 1997). Hence, self-reports of self-esteem, though tolerably valid, still contain some systematic error.

One possible way to curtail such error is to employ *indirect measures* of self-esteem (Greenwald & Farnham, 2000). Such measures are designed to reveal people's attitude towards themselves from their reactions to self-related stimuli (e.g., first and last names, first-personal pronouns), typically under conditions where people are either unaware of, or lack control over, the measurement process (e.g., Rudolph, Schröder, & Schütz, 2006). Consider unawareness: the *initials preference task* (*IPT*; Koole, Dijksterhuis, & van Knippenberg, 2001) requires respondents to rate all letters of the alphabet for likeability, whereupon people typically exhibit an unknowing preference for their initials.[1] Or consider uncontrollability: the *Implicit Association Test (IAT*; Greenwald, McGhee, & Schwartz, 1998) requires respondents to co-classify self-related and self-unrelated stimuli alongside positive and negative stimuli. Respondents are also required to go as quickly as they can without making errors. However, they typically find responding more difficult—and hence go more slowly—when the four target categories are configured one way (e.g., *Self* with *Bad*, *Non-Self* with *Good*) rather than another (e.g., *Self* with *Good*, *Non-Self* with *Bad*). What indirect measures assess is often termed *implicit self-esteem* (*ISE*). The properties of ISE have been assumed to reflect properties of the indirect measures used to assess it (e.g., ISE is unconscious and automatic; Greenwald & Farnham, 2000). However, it remains controversial whether and to what extent indirect measures operate via wholly implicit processes (for a discussion, see De Houwer & Moors, 2007).

Standard dual-process models (e.g., Strack & Deutsch, 2004) suggest that ISE and ESE, being subserved by modular cognitive systems, should yield measures that are largely independent. In addition, different measures of each construct, in virtue of tapping into the same modular system, should exhibit reasonable intercorrelations (though see Marsh & Craven, 2006; Sakellaropoulo & Baldwin, 2007). Given such expectations, prior empirical research has yielded two patterns, one reassuring, the other troubling. The first suggests *discriminant validity*. In particular, measures of ISE and ESE typically show weak correlations (Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005), although these can be augmented under theoretically specified conditions (e.g., Jordan, Whitfield, & Zeigler-Hill, 2007; Koole et al., 2001; Olson, Fazio, & Hermann, 2007). However, the second pattern suggests a *lack of convergent validity*: different measures of ISE typically fail to exhibit the predicted intercorrelations (Bosson, Swann, & Pennebaker, 2000). If valid, this pattern implies one of two things: either (a) ISE exists and is heterogeneous; or (b) ISE does not even exist. The matter remains unresolved. Nonetheless, most measures of ISE *do* converge insofar as they register a pronounced average self-positivity bias (e.g., Greenwald & Farnham, 2000; Gregg & Sedikides, 2008). In addition, such measures exhibit meaningful antecedents and consequences (see Koole & DeHart, 2007, for an overview). Hence, there are some reasons to believe (a) over (b).

However, an even gloomier possibility exists: *both* patterns could be artifacts of measurement error. Measures of ISE have a reputation for unreliability (Bosson et al., 2000).

Such unreliability could obscure latent correlations, and falsely suggest that ESE and ISE diverge when they do not, or that different indices of ISE fail to converge when they do.

In this article, we reconsider the convergent and discriminant validity of ISE and ESE. In particular, we investigate whether and to what extent two factors—(a) the reliability and sensitivity to self-positivity bias of different measures of ISE, and (b) the conceptual correspondence between what different measures of ISE assess—moderate the relation between measures of ISE and ESE, and between different measures of ISE. We then attempt to formulate concrete and constructive recommendations for future research, and make some empirically informed theoretical interpretations.

Some years ago, a study concluded that the IAT and the IPT were the most reliable and valid measures of ISE available (Bosson et al., 2000). Since then, however, indirect measures have proliferated. In particular, three new measures have emerged that—unlike the IAT—permit associations toward an object to be assessed in isolation: the *Single-Category IAT (SC-IAT*; Karpinski & Steinman, 2006), the *Extrinsic Affective Simon Task* (*EAST*; De Houwer, 2003), and the *Go/No-Go Association Task* (*GNAT*; Nosek & Banaji, 2001). These new measures are of interest, because they assess self-related evaluations independently of other-related evaluations—an advantage, given that variations in the theoretically irrelevant non-self category confound performance on the self-esteem IAT (Karpinski, 2004). In addition, due to acknowledged problems concerning the effect size and reliability of the EAST, an improved variant of the EAST, namely the *Identification EAST* (*ID-EAST*), has been devised (De Houwer & De Bruycker, 2007). Furthermore, the present authors have devised a potentially more reliable version of the IPT, namely the *Duplicate IPT* (*D-IPT*). To update the literature, we conducted three studies to compare and contrast the older IAT and IPT with the newer SC-IAT, EAST, ID-EAST, GNAT, and D-IPT as putative indices of ISE. In addition to using new measures of ISE, we also applied more recently developed algorithms (e.g., *D*-index; Greenwald, Nosek, & Banaji, 2003) to maximize validity, and employed standard indices of internal consistency (based on equivalent split-halves, and incorporating warranted Spearman-Brown adjustments).

STUDY 1

In our first study, we evaluated three different measures of ISE: an IPT, an EAST, and an IAT. We quantified their internal consistency, their test-retest stability one-week apart, their intercorrelations, and their correlations with ESE.

Method

Participants and Procedure

A total of 102 students (80 female; $M_{AGE} = 22.7$) participated.[2] They began by providing basic demographic data and by generating an ID code that contained their first and last initials. Next, they completed three measures of ISE in a fixed order: an IPT, an IAT, and an EAST. Finally, participants completed a measure of ESE. They were then dismissed, but returned exactly one week later to redo the three measures of ISE.

Measures of ISE and ESE

IPT. As in Bosson et al. (2000), participants rated each letter of the alphabet on a scale from 1 (*I dislike this letter very much*) to 7 (*I like this letter very much*). To derive an initials preference index that controlled for general letter popularity and personal rating tendencies, we followed the guidelines provided by Koole et al. (2001). We derived Spearman-Brown corrected split-half estimates of internal consistency from correlating ratings of first and last initials.

IAT. The IAT conformed to the canonical five-block structure and procedure (see Appendix; Greenwald & Farnham, 2000). We kept critical block order constant to reduce method variance. We computed the IAT index using the scoring algorithm (the *D*-index) recommended by Greenwald et al. (2003). Higher scores reflect an automatic preference for

Self over *NonSelf*. The IAT's internal consistency was based on a Spearman-Brown corrected split-half correlation, the split-halves being derived from alternating pairs of trials in both critical blocks. This served to ensure (a) that both halves were maximally comparable and (b) that attribute and target trials were equivalently represented in each half.[3]

EAST. The EAST featured the same general structure and response options as described by De Houwer (2003). As for the IAT, we computed an EAST index using the *D*-algorithm. Higher scores reflect stronger automatic liking for *Self*. We computed internal consistency on the basis of a split-half, as above.

ESE. The total score from the 32-item Multidimensional Self-Esteem Scale (*MSES*; Schütz & Sellin, 2006) served as index of ESE. Each item featured a seven-point scale with one of two types of endpoints (1 = *Not at All* to 7 = *Very Much*; 1 = *Never* to 7 = *Always*.)

Results and Discussion

Positivity Bias. All indices of ISE, like the index of ESE, yielded significant effects that were both positive in sign and large in magnitude, although the EAST index lagged behind the others (Table 1). Thus, all indices of self-esteem, implicit and explicit, converged at a directional level, revealing a general bias towards positive self-evaluation.

Internal Consistency and Temporal Stability. As Table 1 shows, the IAT displayed satisfactory levels of internal consistency. However, levels were less satisfactory for the IPT, and unsatisfactory for the EAST. In addition, only the IAT (r_{tt} = .54) and the IPT (r_{tt} = .56) but not the EAST (r_{tt} = .18) showed satisfactory levels of temporal stability over one week.

Intercorrelations. Correlations between the three indices of ISE hovered around zero on both measurement occasions (Table 2). In addition, neither the IPT nor the IAT index correlated significantly with the ESE index on either occasion; and although the EAST index did on one occasion, its failure to do so on another, combined with its trifling internal consistency, suggest sampling error as the most likely explanation.

The overall pattern implies that, although indices of ISE and ESE show directional convergence, they do not show convergence at the level of individual scores. It seems that either the underlying "elephant" of ISE (Bosson et al., 2000) is an illusory beast, or that different indices of ISE map on to very different parts of that underlying "elephant." However, given that the IAT and IPT at least showed a degree of internal consistency, the lack of emergent relations does not appear to have been solely an artifact of measurement unreliability.

STUDY 2

We proceeded to test a further indirect measure of ISE designed to assess automatic attitudes specifically toward the *Self*—the SC-IAT—in conjunction with structurally improved versions of previous measures (i.e., the ID-EAST and D-IPT), plus an IAT. We examined their relative psychometric properties, intercorrelations, and correlations with ESE.

Method

Participants and Procedure

A total of 60 students (11 female; $M_{AGE} = 22.4$) participated. In fixed order, participants completed four measures of ISE (an IAT, a D-IPT, an ID-EAST, and a SC-IAT) followed by one measure of ESE. Due to computer problems, data from eleven participants on the ID-EAST were lost. One participant was excluded due to extreme scores.

Measures of ISE and ESE

D-IPT. We administered the IPT as in Study 1, except that all letters were now presented, not just once, but *twice*, in the same fixed random order. The repetition was designed to increase its reliability. We derived an overall initial preference index by averaging the two initial preference scores (calculated as before) computed separately from each of the rated alphabets. Internal consistency was based on a Spearman-Brown corrected split-half correlation using the (now averaged) preference scores for first and last initials.

IAT. The layout of the IAT, and the computation of its results, was the same as in Study 1. The only difference was the addition of a few extra target stimuli (e.g., *I*, *mine*, *their*, *them*) to vary the representation of self (also added to the SC-IAT and ID-EAST).

SC-IAT. Unlike the IAT, the two critical blocks of the SC-IAT required participants to classify stimuli into one of three categories using two keys. We indexed automatic liking for self by an analogue of the *D*-index. Internal consistency was estimated as for the IAT.

ID-EAST. The ID-EAST contains a structural modification designed to ensure that the stimuli it contains are processed based on semantics rather than on features.[4] We adopted a single category ID-EAST, using only the *Me* target category (Appendix).[5] Trial data were aggregated, and internal consistency was estimated, just as in the original EAST.

MSES. As in Study 1, ESE was measured with the MSES (Schütz & Sellin, 2006).

Results and Discussion

Positivity Bias. Again, all indices yielded significant effects that were both positive in sign and large in magnitude, with the ID-EAST index lagging behind the others (Table 1). Thus, a convergent directional bias towards positive self-evaluation re-emerged.

Internal Consistency.[6] Both the IAT and SC-IAT displayed high levels of internal consistency, with the value for the latter numerically exceeding that for the former (Table 1). In addition, both the D-IPT and the ID-EAST showed reasonable levels of internal consistency, higher than with the IPT and original EAST in Study 1.

Intercorrelations. Overall, the pattern replicated Study 1 (Table 2). First, no correlation between any index of ISE and the ESE index approached significance. Second, none of the intercorrelations between indices of ISE attained significance, although one approached it. However, given that, thanks to methodological innovations to the EAST and IPT, the internal consistency was even higher than in Study 1, measurement unreliability is unlikely to be the whole explanation for the absence of significant correlations.

Nonetheless, from an exploratory perspective, one might ask why the sole marginal correlation found was between the ID-EAST and the SC-IAT. We believe the answer is that, despite some structural differences, both these indirect measures *corresponded conceptually*, in that they reflected specific self-evaluations uncomplicated by any salient other-comparisons. By reflecting the same target, they achieved greater convergent validity.

STUDY 3

Our investigations of measures of ISE had yet to include a promising methodology: the GNAT. This measure was originally designed to provide an alternative to the IAT assessing automatic associations towards individual objects. In Study 3, therefore, we duly examined the reliability and sensitivity to self-positivity bias of the GNAT as a potential measure of ISE. We additionally examined the relation of the GNAT to another indirect measure of self-esteem, the IAT, as well as to a traditional direct measure (Rosenberg, 1965).

One notable feature of the GNAT is that, when used to assess automatic evaluations towards X and Y individually, its results can be combined to create a relative index that conceptually corresponds to the standard IAT index. In particular, if two GNAT blocks respectively assess positive and negative evaluations of *self*, and two more GNAT blocks respectively assess positive and negative evaluations of *non-self*, then all four blocks can assess positive and negative evaluations of self *relative* to non-self—precisely what the IAT assesses. Building on the suggestive results of Study 2 with regard to the lone marginal correlational observed, we tested in Study 3 whether the IAT index would correlate better with a relative GNAT index than with individual GNAT indices, given that the conceptual correspondence would be exact in the former case, but inexact in the latter cases.

We additionally explored the impact on levels of convergent validity of attempting to reduce, first, systematic error (i.e., variance due to individual differences in classification ability), and second, random error (i.e., variance due to measurement unreliability).

Specifically, we (a) compared correlations obtained using the original IAT index (Greenwald & Farnham, 2000) to those based on the newer algorithm (Greenwald et al., 2003), thereby controlling for some systematic error, and (b), compared the second set of correlations to corresponding coefficients estimated in a structural model, thereby controlling for random error. We predicted that the combined use of the improved algorithm and structural modeling would increase convergent validity.

Method

Participants and Procedure

The sample comprised 195 students[7], predominantly young (M_{AGE} = 20.5) and female (85%). Due to participant dropout, technical failures, task non-compliance, or extreme scores, listwise *N*s ranged from 182 to 195 across various analyses. The dataset comprised scores from two self-esteem IATs and GNATs, both run twice with a time interval of one week, along with measures of ESE. The IAT and GNAT featured identical categories and stimuli.

Measures of ISE and ESE

IAT. The IAT in Study 3 resembled those in preceding studies except that it comprised only two critical blocks presented in random order.

GNAT. The GNAT comprised four blocks (*Self-Positive*, *Self-Negative*, *NonSelf-Positive*, and *NonSelf-Negative*) presented in random order, each of which featured two target categories out of a possible four (i.e., two of *Self, NonSelf, Positive, Negative*). Participants attempted to press a key within 750 ms when a word presented matched those categories, and not to press it when a word did not (see Appendix). Accordingly, a response on each trial could be classed as *hit*, *false alarm*, *correct rejection,* or *miss*. Overall accuracy within each block at distinguishing target from non-target items was duly quantified by *d′*, the normalized hit rate minus the normalized false alarm rate (Nosek & Banaji, 2001).

We derived *five* GNAT indices. First, we computed all four *individual d'* indices: *Self P>N* (*d'* in the *Self-Positive* block minus *d'* in the *Self-Negative* block); *NonSelf N>P* (*d'* in the *NonSelf-Negative* block minus *d'* in the *NonSelf-Positive* block); *Positive S>NS* (*d'* in the *Self-Positive* block minus *d'* in the *NonSelf-Positive* block); and *Negative NS>S* (*d'* in the *NonSelf-Negative* block minus *d'* in the *Self-Negative* block). Second, we derived a *relative* index (*Overall*), by computing the average of *Self P>N* and *NonSelf N>P*.

ESE. We used a 10-item questionnaire (*RSES*; Rosenberg, 1965) to assess participants' overall liking for themselves (1 = *strongly agree*; 4 = *strongly disagree*).

Results

Positivity Bias and Reliability. Averaging across sessions, the relative GNAT index showed high sensitivity to self-positivity bias ($d = 1.56$), as well as reasonable internal consistency ($r = .65$) and modest test-retest reliability ($r = .51$). Individual GNAT indices also showed high sensitivity to self-positivity bias (*d*s = .80 to 1.60), but more modest internal consistency (*r*s = .52 to .59) and low test-retest reliability (*r*s = .23 to .38). Corresponding IAT indices were comparable (original index $d = 1.66$, $r_{ic} = .85$, $r_{tt} = .39$; new index $d = 1.54$, $r_{ic} = .67$, $r_{tt} = .29$).[8]

Convergent Validity. Table 3 displays the correlation between each of the five GNAT indices and (a) the original IAT index, (b) the new IAT index, and (c) the new IAT index, estimated as part of a different structural model. In each model, we estimated (assuming unequal loadings, unequal error variances, and uncorrelated errors) correlations among three common factors: each GNAT index, the IAT, and the RSES. To derive manifest variables on the basis of which to estimate latent correlations, we created four parcels for each measure, consisting of equivalent split-halves derived from each of the two measurement occasions.[9]

As predicted, the correlations involving GNAT indices were always numerically (albeit not significantly) larger when the new IAT index was used as opposed to the original

IAT index. Moreover, these correlations were numerically larger again when estimated from a structural model. Finally, the numerically highest correlation was always obtained between the IAT index and the *Overall* GNAT index. Although the increment in the magnitude and reliability of correlations at each step was admittedly small and nonsignificant, the *combined* increment after taking all three steps was nonetheless theoretically meaningful. For example, if a researcher using our dataset had attempted *neither* to minimize systematic and random error *nor* to examine a pair of conceptually convergent indices, then he or she might well have falsely concluded that the IAT and GNAT (e.g., using the *Positive S>NS* index) did *not* converge ($r = .11$, $p = .14$), whereas had they taken all these steps, they would have correctly concluded that the IAT and GNAT (using the *Overall* index) *did* converge ($r = .27$, $p < .05$). Indeed, when underlying relations are weak, it is critical to maximize all available statistical power and conceptual correspondence.

Discriminant Validity. The RSES failed to covary even marginally with either index of ISE, both at the level of raw correlations (GNAT: $-.07 < r < -.01$; IAT: $.01 < r < .10$) and estimated structural coefficients (GNAT: $-.11 < r < -.01$; IAT: $r \approx .01$). Thus, both the GNAT and IAT indices were still independent of the RSES index.

General Discussion

We close by making some practical recommendations for researchers wishing to explore ISE empirically and drawing some theoretical conclusions from our findings.

Practical Recommendations

Across three studies, the IAT—and two of its methodological offshoots designed to capture ISE more specifically, namely the SC-IAT and the relative GNAT index—exhibited satisfactory to good levels of reliability. The IPT and EAST exhibited comparatively lower levels, although their improved methodological variants, the D-IPT and ID-EAST, fared somewhat better. Hence, IAT, SC-IAT and GNAT are, for psychometric reasons, to be recommended over IPT and EAST (at least in their original form) for use in research on individual differences in ISE. In addition, the smaller aggregate self-positivity biases obtained for the EAST and ID-EAST (although not for the IPT and D-IPT) suggest less sensitivity to self-positivity bias, and may counterindicate their use.

Theoretical Conclusions

If ISE is a single construct distinct from ESE, then one would expect, all else equal, different indices of ISE to correspond more strongly with one another than with an index of ESE. Having employed several newly developed, reliable, and specific measures of ISE (i.e., the SC-IAT, EAST, and GNAT), what did we find?

First, we found that, despite a pronounced positivity bias for ESE and ISE indices at a directional level, individual ISE scores remained independent of individual ESE scores. Moreover, this independence could not be attributed to measurement unreliability: most indices exhibited satisfactory internal consistency, and estimated structural coefficients were little higher than observed raw correlations. Nor could this independence be attributed to a lack of correspondence between direct and indirect measures of self-esteem: the more specific ISE indices did not correlate any better with ESE than the relative IAT indices. Thus, in the absence of other explanations, our results are in keeping with dual-process models of cognition applied to self-esteem. Indeed, our results are *starkly* in keeping with such models: even though a moderate degree of explicit-implicit correlation is typically observed across a range of topics (Nosek & Smyth, 2007), we found almost *none*. Perhaps self-reported self-esteem is particularly prone to reflect the impact of carefully pondered propositions and of self-presentational concerns (Upshaw & Yates, 1968), and is only more rarely a reflection of introspective insights into spontaneous self-feelings (Jordan et al., 2007; Koole et al., 2001; Olson et al., 2007), given that (a) the self is habitually the focus of so much cognition (e.g., Rogers, Kuiper, & Kirker, 1977), and (b) maintaining a positive self-view is such an urgent motivational priority (Sedikides & Gregg, 2003, 2008; Sedikides & Strube, 1997).

Second, we found, in Studies 1 and 2, that nearly all correlations between indices of ISE—even involving several newer, more specific, and more reliable measures—fell well short of significance. These results echoed previous research casting doubt on the convergent validity of measures of ISE (Bosson et al., 2000). The sole exception here was the marginal correlation obtained between two measures of ISE that were conceptually correspondent (SC-IAT and ID-EAST). Taking our cue from this suggestive result, we took pains in Study 3 *simultaneously* to minimize confounding sources of variance (i.e., conceptual mismatch between indices, individual differences in reaction time, random error of measurement). When we did so, some evidence of convergent validity finally *did* emerge. Yet the level of convergence remained curiously low. The question is why.

We suspect these answer may lie, not merely in the heterogeneity of ISE itself, but also in the contrasting *structural features* that characterize even similar-seeming indirect measures (De Houwer & Moors, 2007; Deutsch & Gawronski, 2008; Payne, Burkley, & Stokes, 2008). For example, due to the "bipolar" layout of the IAT (i.e., A&X *versus* B&Y), asymmetry in the salience of its categories alone is sufficient to engender effects (Rothermund & Wentura, 2004). However, this dynamic is less likely to confound its cousin, the more "unipolar" GNAT, to a similar degree. Moreover, whereas in the IAT accuracy is held constant and reaction time let vary, in the GNAT reaction time is held constant and accuracy let vary. Such disparities in structure may cause corresponding disparities in performance, not only because they recruit different types of classification skill, but also because they elicit different types of classification strategy. We suspect that understanding the structural features of ISE measures will shed light on the reasons for their low convergent validity, and that the devil may be in the procedural details not the underlying construct.

References

Bosson, J. K., Swann, W. B., Jr., & Pennebaker, J. (2000). Stalking the perfect measure of implicit self-esteem: The blind men and the elephant revisited? *Journal of Personality and Social Psychology*, *79*, 631-643.

Byrne, B. M. (2001). *Structural Equation Modeling with AMOS, Basisc Concepts, Applications, and Programming*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

De Houwer, J. (2003). The Extrinsic Affective Simon Task. *Experimental Psychology, 50*, 77-85.

De Houwer, J., & Moors, A. (2007). How to define and examine the implicitness of implicit measures. In B. Wittenbrink, & N. Schwarz (Eds.), *Implicit measures of attitudes: Procedures and controversies*. New York: Guilford Press.

De Houwer, J., & De Bruycker, E. (2007). The identification-EAST as a valid measure of implicit attitudes toward alcohol-related stimuli. *Journal of Behavior Therapy and Experimental Psychiatry, 38*, 95-104.

Deutsch, R., & Gawronski, B. (2008). *When the method makes a difference: Antagonistic effects on "automatic evaluations" as a function of task characteristics of the measure*. Manuscript submitted for publication.

Greenwald, A. G., & Farnham, S. D. (2000). Using the Implicit Association Test to measure self-esteem and self-concept. *Journal of Personality and Social Psychology*, *79*, 1022-1038.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. K. L. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, *74*, 1464-1480.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, *85*, 197-216.

Gregg, A. P., & Sedikides, C. (2008). *Implicit self-esteem and narcissism: Rethinking the link*. Manuscript submitted for publication.

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the Implicit Association Test and explicit self-report measures. *Personality and Social Psychology Bulletin, 31*, 1369-1385.

Jordan, C. H., Whitfield, M., & Zeigler-Hill, V. (2007). Intuition and the correspondence between implicit and explicit self-esteem. *Journal of Personality and Social Psychology*, *93*, 1067-1079.

Karpinski, A. (2004). Measuring self-esteem using the Implicit Association Test: The role of the other. *Personality and Social Psychology Bulletin, 30*, 22-34.

Karpinski, A., & Steinman, R. B. (2006). The Single Category Implicit Association Test as a measure of implicit social cognition. *Journal of Personality and Social Psychology, 91*, 16-32.

Koole, S. L., & DeHart, T. (2007). Self-affection without self-reflection: Origins, models, and consequences of implicit self-esteem. In C. Sedikides, & S. J. Spencer (Eds.), *The Self.* New York: Psychology Press.

Koole, S. L., Dijksterhuis, A., & van Knippenberg, A. (2001). What's in a name: Implicit self-esteem and the automatic self. *Journal of Personality and Social Psychology, 80*, 669-685.

Marsh, H. W., & Craven, R. G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective: Beyond seductive pleasure and unidimensional perspectives. *Perspectives on Psychological Science*, *1*, 133-163.

Nosek, B. A., & Banaji, M. R. (2001). The Go/No-Go Association Task. *Social Cognition*, *19*, 625-666.

Nosek, B. A., & Smyth, F. L. (2007). A multitrait-multimethod validation of the Implicit Association Test: Implicit and explicit attitudes are related but distinct constructs. *Experimental Psychology*, *54*, 14-29.

Olson, M. A., Fazio, R. H., & Hermann, A. D. (2007). Reporting tendencies underlie discrepancies between implicit and explicit measures of self-esteem. *Psychological Science, 18*, 287-291.

Paulhus, D. L., Fridhandler, B., & Hayes, S. (1997). Psychological defense: Contemporary theory and research. In J. Johnson, R. Hogan, & S. R. Briggs (Eds.), *Handbook of Personality Psychology*. New York: Academic Press.

Payne, B. K., Burkley, M. A., & Stokes, M. B. (2008). Why do implicit and explicit attitude tests diverge? The role of structural fit. *Journal of Personality and Social Psychology, 94*, 16-31.

Rogers T. B., Kuiper, N. A., & Kirker W. S. (1977). Self-reference and the encoding of personal information. *Journal of Personality Social Psychology*, *35*, 677-688.

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

Rothermund, K., & Wentura, D. (2004). Underlying processes in the Implicit Association Test (IAT): Dissociating salience from associations. *Journal of Experimental Psychology: General*, *133*, 139-165.

Rudolph, A., Schröder, M., & Schütz, A. (2006). Ein Impliziter Assoziationstest zur Erfassung von Selbstwertschätzung [An Implicit Association Test measuring self-esteem]. In T. M. Ortner, R. T. Proyer, & K. D. Kubinger (Hrsg.), *Theorie und Praxis Objektiver Persönlichkeitstests*. Bern: Verlag Hans Huber.

Sakellaropoulo, M., & Baldwin, M. W. (2007). The hidden sides of self-esteem: Two dimensions of implicit self-esteem and their relation to narcissistic reactions. *Journal of Experimental Social Psychology*, *43*, 995-1001.

Schlenker, B. R., & Leary, M. R. (1982). Social anxiety and self-presentation: A conceptualization and model. *Psychological Bulletin*, *92*, 641-669.

Schmukle, S. C., & Egloff, B. (2006). Assessing anxiety with Extrinsic Simon Tasks. *Experimental Psychology*, *53*, 149-160.

Schütz, A., & Sellin, I. (2006). *Die Multidimensionale Selbstwertskala (MSWS).* [The multidimensional self-esteem scale]. Göttingen: Hogrefe.

Sedikides, C., & Gregg, A. P. (2003). Portraits of the self. In M. A. Hogg & J. Cooper (Eds.), *Sage handbook of social psychology* (pp. 110-138). London: Sage Publications.

Sedikides, C., & Gregg, A. P. (2008). Self-enhancement: Food for thought. *Perspectives on Psychological Science, 3*, 102-116.

Sedikides, C., & Strube, M. J. (1997). Self-evaluation: To thine own self be good, to thine own self be sure, to thine own self be true, and to thine own self be better. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology, 29*, 209-269. New York, NY: Academic Press.

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychological Review, 8*, 220-247.

Upshaw, H. S., & Yates, L. A. (1968). Self-persuasion, social approval, and task success as determinants of self-esteem following impression management. *Journal of Experimental Social Psychology, 4*, 143-152.

Wilson, T. D. (2002). *Strangers to ourselves: Discovering the adaptive unconscious.* Cambridge, MA: Harvard University Press.

Footnotes

[1] Although the bias extends to all letters in one's name, it is most pronounced for one's initials (Koole et al., 2001), a pattern that measures of ISE tend to exploit.

[2] We reanalyzed a relevant subset of variables of the data used in Schröder-Abé et al. (2007).

[3] As Schmukle and Egloff (2006) note, this alternating method of deriving split-halves is liable to estimate internal consistency better than one in which the data is simply split into earlier and later trials, because temporal order effects are avoided. A further advantage is that they can be computed across every index of ISE and ESE, thereby increasing comparability.

[4] We adopted the convention to classify target stimuli in the ID-EAST on the basis of their being shown in uppercase or lowercase (De Houwer & De Bruycker, 2007).

[5] In addition, one *Self* stimulus was idiosyncratic: each participant's first name.

[6] For a subset of the sample ($N = 39$), we also computed the six-month temporal stability of IAT, D-IPT, and SC-IAT scores. The IAT and D-IPT showed satisfactory levels ($rs_{tt} = .60$ and .68 respectively), the SC-IAT more modest levels ($r_{tt} = .44$).

[7] We reanalyzed a relevant subset of variables reported by Gregg and Sedikides (2008).

[8] One oddity that emerged was that the newer IAT index exhibited lower internal consistency than the original IAT index, despite the fact that the former was specifically developed to yield higher internal consistency than the latter (Greenwald et al., 2003). In our case, this could have been a procedural artifact, given that our IAT was somewhat abbreviated.

[9] Due to generally low intercorrelations, the SEM fit indices were, according to Byrne's (2001) criteria, located at the boundary values of a mediocre fit (e.g., value for the relative GNAT: $RMSEA = .088$, $CFI = .910$, $TLI = .863$).

Table 1

Study 1 and 2: Descriptive Statistics, Internal Consistency, Directional Significance, and Effect Size of Explicit and Implicit Self-Esteem Indices

Self-Esteem Index	Mean (SD)	Split-Half Reliability	One-Sample *t*	Cohen's *d*
		Study 1		
Explicit				
MSES (T1)	4.69 (.90)	.94	13.38	1.33
Implicit				
IPT_1	.87 (.66)	.51	13.37	1.33
IPT_2	.88 (.72)	.50	12.29	1.22
IAT_1	.62 (.33)	.85	19.18	1.91
IAT_2	.58 (.32)	.83	18.21	1.81
$EAST_1$	.34 (.55)	.16	6.22	.62
$EAST_2$	.45 (.61)	.24	7.39	.74
		Study 2		
Explicit				
MSES	4.75 (.87)	.93	11.15	1.45
Implicit				
D-IPT	.72 (.64)	.69	8.41	1.09
IAT	.64 (.30)	.80	16.67	2.17
SC-IAT	.46 (.29)	.88	12.42	1.62
ID-EAST	.67 (.76)	.64	6.19	.81

Note. $N_{\text{Study 1}} = 102$, $N_{\text{Study 2}} = 60$. MSES = Multidimensional Self-Esteem Scale; (D-) IPT = (Duplicate) Initials Preference Task; (SC-) IAT = (Single Category) Implicit Association Test; (ID-) EAST = (Identification) Extrinsic Affective Simon Task. SD = Standard deviation; Subscripts (1, 2) indicate measurement occasion (one week apart). One-sample *t*-tests to compare the mean of each index with the theoretical midpoint of its scale. All *t*-values reported are significant at $p < .0001$. Split-Half Reliability based on split-half correlations incorporating Spearman-Brown adjustments. Cohen's *d* refers to the standardized difference between theoretical midpoint of scale and the observed mean for each index.

Table 2

Study 1 and 2: Correlations between Indices of Explicit and Implicit Self-Esteem

Study 1

	Self-Esteem Index	1	2	3	4
	Explicit				
1	MSES	-	.03	.14	.28**
	Implicit				
2	IPT	-.07	-	-.07	.07
3	IAT	-.06	.06	-	-.08
4	EAST	.09	.07	-.09	-

Study 2

	Self-Esteem Index	1	2	3	4	5
	Explicit					
1	MSES	-				
	Implicit					
2	D-IPT	.11	-			
3	IAT	-.04	.07	-		
4	SC-IAT	-.05	-.07	.09	-	
5	ID-EAST	-.04	-.06	-.03	.25†	-

Note. $N_{\text{Study 1}} = 102$, $N_{\text{Study 2}} = 60$. MSES = Multidimensional Self-Esteem Scale. (D-) IPT = (Duplicate) Initials Preference Task; (SC-) IAT = (Single Category) Implicit Association Test; (ID-) EAST = (Identification) Extrinsic Affective Simon Task.

In Study 1, Time 1 correlations appear below the diagonal, Time 2 correlations above the diagonal (with the MSES administered only once).

† $p < .10$, ** $p < .01$.

Table 3

Study 3: Intercorrelations between Indices of Implicit Self-Esteem (zero-order coefficients and coefficients estimated in a structural model)

GNAT Index	IAT original algorithm	IAT new algorithm	IAT new algorithm (SEM)
Differential Indices			
Self P>N	.14*	.14*	.21†
NonSelf N>P	.11	.15*	.21†
Positive S>NS	.11	.13†	.25*
Negative NS>S	.15*	.17*	.24†
Relative Index			
Overall	.15*	.18*	.27*

Note. N = 195. GNAT = Go/No-Go Association Task; IAT = Implicit Association Test; Self P>N = Self-Positive block minus Self-Negative block; NonSelf N>P = NonSelf-Negative block minus NonSelf-Positive block; Positive S>NS = Self-Positive block minus NonSelf-Positive block; Negative NS>S = NonSelf-Negative block minus Self-Negative block.

† $p < .10$, * $p < .05$.

Appendix

Studies 1, 2, and 3: Structural and Categorical Features of All Indirect Measures of Self-Esteem

Block	*Trial N*	*Task*	*Press Left Key*	*Press Right Key*
		Extrinsic Affective Simon Task (Study 1)		
1	20	P: Attribute's semantic discrimination	Unpleasant	Pleasant
2	20	P: Target's color discrimination	Green[a]	Blue[a]
3-6	30	T: Combined task	Unpleasant + Green[a]	Pleasant + Blue[a]
		Single Target Implicit Association Test (Study 2)		
1	40	P: Attribute discrimination	Pleasant	Unpleasant
2-3	40+80	T: Initial combined task	Pleasant + Me	Unpleasant
4-5	40+80	T: Reversed combined task	Pleasant	Unpleasant + Me
		Identification Extrinsic Affective Simon Task (Study 2)		
1	30	P: Attribute discrimination	Unpleasant	Pleasant
2	30	P: Target's letter case discrimination	Lower case[a]	Upper case[a]
3-5	50 each	T: Combined task	Unpleasant + lower case[a]	Pleasant + upper case[a]
		Implicit Association Test (Study 1 & 2)		
1	24	P: Attribute discrimination	Pleasant	Unpleasant
2	24	P: Target discrimination	Me	Not-Me
3	96	T: Initial combined task	Pleasant + Me	Unpleasant + Not-Me
4	24	P: Reversed target discrimination	Not me	Me
5	96	T: Reversed combined task	Pleasant + Not-Me	Unpleasant + Me
		Implicit Association Test (Study 3)		
1	48	T: Initial combined task	Nice + Me	Nasty + Not-Me
2	48	T: Reversed combined task	Nice + Not-Me	Nasty + Me
		Go/No-Go Association Task (Study 3)		
			Press Space Bar	*Don't Press Space Bar*
1	16+48	P+T: Target / Non-target discrimination	Nice + Me	Nasty + Not-Me
2	16+48	P+T: Target / Non-target discrimination	Nasty + Not-Me	Nice + Me
3	16+48	P+T: Target / Non-target discrimination	Nice + Not-Me	Nasty + Me
4	16+48	P+T: Target / Non-target discrimination	Nasty + Me	Nice + Not-Me
		Sample stimuli (Study 1 & 2): Pleasant (*smile, joy*); Unpleasant (*pain, war*); Me (*self, my*); Not-Me (*other, yours*)		
		Sample stimuli (Study 3): Nice (*excellent, love*); Nasty (*bomb, hatred*); Me (*myself, my*); Not-Me (*they, them*)		

Note. P = practice blocks; T = test blocks.

[a] Stimuli of the target category (*Self* and *Non-Self*) are presented in the defined color or letter case.

A complete list of the stimuli can be obtained from the authors.