It works both ways. Enhancing explicit self-esteem using the self-reference task

Laurențiu P. Maricuțoiu⁎, B. Keith Payne, Dragoș Iliescu

⁎ Corresponding author at: West University of Timișoara, Department of Psychology, 4 Vasile Pârvan Blvd., Room 303, 300223 Timișoara, Romania.
E-mail address: laurentiu.maricutoiu@e-uvt.ro (L.P. Maricuțoiu).

Handling editor: Sean McCrea
Keywords:
Self-reference task
Evaluating learning
Self-esteem
Well-being

ABSTRACT

The self-reference task (SRT) is an evaluative learning paradigm that uses the positive valence of the self to change the attitude towards new and neutral stimuli. In this contribution, we present evidence regarding the possibility of changing the attitudes towards the self, following a modified SRT. In three independent research studies, we provided participants in experimental groups with an SRT that paired the self with highly positive pictures. After the completion of the modified SRT, participants in the experimental group reported significantly higher explicit self-esteem (in all studies) and more positive self-views (Study 1 and Study 2), as compared with a control group. In the second study, students completed the modified SRT each day for two weeks. The participants in the experimental group reported more positive explicit self-esteem and higher psychological well-being (e.g., higher academic engagement and lower academic burnout), as compared to the control group. However, Study 3 did not provide evidence regarding the effect of the modified SRT on self-reported mental health. Our results suggest that asking individuals to relate the self with another stimulus can be used in both directions to transfer valence from the self to external targets, and from external targets to the self.

1. Introduction

Individuals generally have a positive evaluation of themselves (Bosson, Swann, & Pennebaker, 2000). Moreover, the positivity of the self-evaluations can be transferred to any owned object through the mere ownership effect (Nuttin, 1985). Although the mere ownership effect was initially used for measuring implicit self-esteem (Koole, Dijksterhuis, & van Knippenberg, 2001), more recent developments in the field of evaluative learning investigated whether it can be used for changing attitudes through the self-referencing task (SRT - Perkins, Forehand, & Greenwald, 2005).

The SRT (Perkins et al., 2005) is an evaluative learning paradigm that recently gained the interest of researchers. Unlike other well-established learning paradigms, the particularity of the SRT is that it uses the self as a valenced stimulus to modify the valence of previously neutral stimuli. Although it was first considered a particular form of evaluative conditioning, recent theoretical developments suggested that the SRT might be a new type of evaluative learning effect, based on intersecting regularities (Hughes, De Houwer, & Perugini, 2016). In a recent meta-analysis of 53 studies, Mattavelli, Richetin, Gallucci, and Perugini (2017) reported that the SRT is effective at forming or changing the attitudes towards the new stimuli, at implicit and explicit levels.

All previous studies paired the self with neutral stimuli and aimed to demonstrate that the positive valence of the self can be transferred to these stimuli. Given that the SRT effect proved to be reliable and robust (Mattavelli et al., 2017), in the present research we are interested in investigating whether the request to classify the self and other stimuli in the same category can transfer the valence from the stimuli to the self. More specifically, we are interested to see whether the valence of the self can be enhanced, after an SRT in which the paired stimuli are also positively valenced. If the new stimuli generate intense positive emotional reactions, we expect to find a reversed transfer of valence (i.e., from the stimuli to the self).

1.1. Conditioning the positivity of the self

Theoretical developments in the research field of attitudes (Greenwald & Banaji, 1995) suggested that a positive self-view is an attitude towards the self. Following this paradigm shift, psychologists took into consideration both explicit self-views (or discursive self-esteem, resulted from conscious self-appraisal processes) and implicit self-
views (or implicit self-esteem, resulted from automated associations with the self) (Bosson et al., 2000).

Given the fact that positive self-views can be conceptualized as attitudes towards the self, researchers used psychological theories of attitude formation and attitude change to explain the formation and the temporal dynamics of self-esteem. In this vein, Dijksterhuis (2004) investigated whether implicit (or automated) self-esteem can be changed using a subliminal evaluative conditioning task. In a series of five experimental research studies, Dijksterhuis (2004) exposed participants to repeated associations between a self-related stimulus and positive stimuli and did not use any discursive instructions. Following this procedure, higher levels of implicit self-esteem were found in the case of the experimental groups, as compared with neutral control groups. Covin (2008, Experiment 2) and Grumm, Nestler, and Von Collani (2009, Experiment 1 and Experiment 3) used an evaluative conditioning procedure similar to the one used by Dijksterhuis (2004) and reported similar results regarding implicit self-esteem. In another study, Baccus, Baldwin, and Packer (2004) used an evaluative conditioning procedure that paired self-relevant information with smiling faces within a computer game, and reported higher implicit self-esteem in the experimental group, as compared with the control group. Moreover, Baccus et al. (2004) reported null effects of the evaluative conditioning on explicit self-esteem, which suggested that stimulus pairing tasks are effective for changing implicit, and not explicit self-appraisals.

Using the reflective-impulsive model (Strack & Deutsch, 2004) and the associative-propositional model (Gawronski & Bodenhausen, 2006), Grumm et al. (2009) argued that different experimental manipulations should have different effects on implicit and explicit self-esteem, depending on the characteristics of these experimental manipulations. An experimental manipulation that changes the affective reaction towards the self will affect only implicit self-esteem and will also have to address self-related propositions to successfully change explicit self-esteem (Grumm et al., 2009). In two studies, Grumm et al. (2009) used propositional manipulations to change explicit self-esteem (Experiment 2), and a combination of affective and propositional manipulations to show that a change in implicit self-esteem can lead to a change in explicit self-esteem (Experiment 3).

Starting from similar assumptions, Ebert, Steffens, von Stulpnagel, and Jelenec (2009) investigated whether learning processes are present within an implicit association test (IAT - Greenwald, McGhee, & Schwartz, 1998). Ebert et al. (2009) used a modified version of the IAT for self-esteem (Greenwald & Farnham, 2000), and presented the participants only the blocks that paired the “self” category with positive attributes, and the “others” category with negative attributes. In the other experimental condition, the “self” category was paired with negative attributes, and the “others” category was paired with positive attributes. Ebert et al. (2009) reported enhanced implicit self-views in the “self + positive” condition, as compared with the “self + negative” condition.

In conclusion, previous research studies reported that implicit and explicit self-views could be changed using evaluative learning paradigms (i.e., evaluative conditioning), or using classification tasks similar with the SRT (i.e., the unbalanced IAT task used by Ebert et al., 2009). Taken together, these findings suggest that it is possible to find a reversed SRT effect.

1.2. The self-referencing task

In their initial study on the SRT, Perkins et al. (2005) used four types of stimuli: self-concept stimuli (e.g., I, mine), other-related stimuli (e.g., they, their), pictures of digital clocks, and pictures of analog clocks. Participants were asked to press the same key (the ‘d’ key) when the stimulus displayed on a computer screen represented the “self” or an analog clock, and to press the ‘k’ key when the stimulus represented the “other” or a digital clock. Following this simple classification task, the participants reported preferences towards the analog clocks, if this type of clocks shared the same key with the “self” stimuli.

Initially, Perkins and Forehand (2012) used the balance congruity principle (Greenwald et al., 2002) to explain the transfer of valence behind the SRT. From this perspective, the positive valence and the self-concept share a first order link prior to the SRT, and this link is assessed by implicit or explicit self-esteem measures. As they complete the SRT, individuals create a first-order link between the self-concept and the new set of stimuli, and this generates a second-order link between the positive valence of the self and the new stimuli. Therefore, the establishment of the second-order link is the reason we find positively valenced stimuli after the completion of an SRT.

More recently, Hughes et al. (2016) suggested that the SRT effect is the result of intersecting regularities. In the SRT, the “self” category and one set of stimuli (e.g., analog clocks) share a common response key throughout the task, and another common response key is shared by the “others” category and another set of stimuli (e.g., digital clocks). Hughes et al. (2016) argue that in the SRT there are four operant contingencies (i.e. if I, then press key A; if others, then press key I; if digital clock, then press key A; if analog clock, then press key I), and these operant contingencies intersect in terms of a common response (i.e., key A or key I). The intersecting regularities is an evaluative learning procedure which is different from the simple stimulus pairing involved in evaluative conditioning paradigms and can explain how individuals create associations between the self and the other categories presented in an SRT (Hughes et al., 2016; Mattavelli et al., 2017).

2. The present study

Previous research studies provided converging results that implicit valence of the self can be enhanced through subliminal conditioning (Covin, 2008; Dijksterhuis, 2004) or by conducting a categorization task similar with the SRT (Ebert et al., 2009). Because the underlying learning effects are presumed to be implicit, these research studies did not use any explicit measures for assessing the manipulation effects and relied exclusively on implicit measures. However, Baccus et al. (2004) reported significant effects of a conditioning paradigm on implicit self-esteem, not on explicit self-esteem. Taken together, these findings support the idea that evaluative learning paradigms can be used successfully to change the valence of the self-concept.

In the present paper, we develop the ideas suggested by Ebert et al. (2009), and we use the SRT as a reliable paradigm for evaluative learning. We anticipate that the introduction of highly positive stimuli in the SRT will result in an enhancement of the self-concept valence. Using the balance congruity principle (Greenwald et al., 2002), we expect to enhance the positivity of the self-concept by creating a second-order link between the self-concept and the positive valence of the stimuli, in addition to the existing first-order link between the self-concept and positive characteristics. Therefore, the general hypothesis of the present research study is:

H1. The participants that pair the self-concept with positive stimuli will report more positive self-views, as compared with participants in the control group.

3. Study 1

3.1. Method

3.1.1. Participants

127 undergraduate students (72.44% female, mean age = 23.3 years, SD age = 5.89 years) participated in this study, in exchange for partial course credit. Following the analysis of the SRT tasks, 22 participants were excluded from the statistical analyses because they had < 80% correct classifications. An a priori, two-tailed, power analysis indicated that a sample size of 64 participants in each of the two groups should provide 80% power to detect a moderate effect size (d = 0.50).
3.1.2. Measures

Implicit self-esteem was assessed using the Initials Preference Task (IPT - Koole et al., 2001). The IPT is a paper-and-pencil measure for implicit self-esteem, in which participants assessed how much they like each letter of the alphabet, on a 7 points Likert scale (1 – not at all; 7 – very much). The individual’s preference for own initials (as compared with the other letters in the alphabet) is an indirect effect of high self-esteem (Greenwald & Banaji, 1995). We used the suggestions of Baccus et al. (2004) to compute the IPT score (high scores indicate high self-esteem), using the ipsatized double-correction algorithm. The internal consistency (split-half) of the IPT on the present sample was 0.65.

Explicit self-esteem was assessed using the popular Rosenberg Self-esteem Scale (Rosenberg, 1965). This questionnaire has 10 items, the response range scales from 1 (strongly disagree) to 4 (strongly agree), and high scores indicate high self-esteem. For the present sample, the internal consistency (Cronbach’s alpha) of the scale was 0.86.

We used the Positivity Scale (Caprara et al., 2012) to assess explicit positivity. The Positivity Scale has 8 items that assess the individual tendency towards a positive outlook regarding one’s self, one’s life, and one’s future. The items are cumulated in a single score (large values indicate positive views) and are evaluated using a 5 steps Likert scale (1 - strongly disagree to 5 - strongly agree). The internal consistency (Cronbach’s alpha) of the Positivity Scale was 0.84.

3.1.3. Procedure

Participants were randomly assigned to an experimental or to a control group, using the Inquisit randomization algorithm (Inquisit, 2014). Participants in both groups had to complete an adaptation of the SRT (Perkins et al., 2005), followed by the implicit and explicit measures described above. To the best of our knowledge, participants were not aware of the assignment to a particular experimental group and were not aware of the experimental manipulation.

The SRT was presented as a memorization test. Participants were instructed to learn from which category belong the pictures presented on the screen, and to classify these stimuli as fast as they can. In the experimental group, we used four types of stimuli (me vs. others, pleasant vs. neutral), which had to be classified into two categories: me or pleasant, others or neutral. The control group had the same classification task but worked with only two types of stimuli (pleasant vs. neutral). The ‘me’ vs. ‘others’ stimuli were adjectives representative for each category. The pleasant vs. neutral stimuli were pictures selected from the International Affective Picture Set (IAPS; Lang, Bradley, & Cuthbert, 2008), and all stimuli used in this study are presented in Table 1.

The SRT was administered using Inquisit software (Inquisit, 2014), and consisted of four blocks. First, participants completed two practice blocks of 20 trials each, which required them to classify only the positive and negative stimuli. The keys used for the classification were switched from one block to the other. The practice blocks were presented as “memorization blocks”, in which they have to learn which stimulus belongs to which category. Second, participants completed two test blocks (40 trials each) that required them to classify all four types of stimuli into two categories: self or positive image, and others or neutral image. Participants were instructed that the response times to the self vs. others category will be used as a personal benchmark, to evaluate whether they successfully learned the classification of the pictures. Similar to the practice trials, the keys used for the classification were switched from one block to another, but the two categories (me or positive image, and others or neutral image) remained unchanged.

Immediately after the completion of the SRT, the participants completed the explicit measures (self-esteem, positivity), and the implicit measure of self-esteem.

All measures, exclusions, and manipulations in the study are reported in this manuscript.

3.2. Results

The descriptive statistics for each dependent variable are presented in Table 2. We investigated potential distribution problems using the KS test and found that all variables had a normal distribution.

A multivariate analysis of variance (presented in Table 3) indicated significant differences between the control and the experimental groups regarding the explicit measures, but not the implicit measure of self-esteem. Compared with the control group, the experimental group reported significantly higher levels of explicit self-esteem \((F(1,103)=4.295, p = .041, d = 0.406)\) and higher levels of explicit positivity \((F(1,103)=3.950, p = .050, d = 0.386)\). The statistical power (one-tailed test) of these significant results is 0.66 for an alpha level of 0.05. The two groups were not statistically different regarding the level of implicit self-esteem \((F(1,103) = .343, p = .559, d = .114)\), although the differences were in the same direction as hypothesized.

### Table 2

Descriptive statistics of the Study 1 dependent variables.

<table>
<thead>
<tr>
<th>Source</th>
<th>Dependent variable</th>
<th>Control group ((N = 51))</th>
<th>Experimental group ((N = 54))</th>
<th>All participants ((N = 105))</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>m</td>
<td>SD</td>
<td>m</td>
<td>SD</td>
</tr>
<tr>
<td>Implicit self-esteem</td>
<td>1.182</td>
<td>1.331</td>
<td>1.332</td>
<td>1.296</td>
</tr>
<tr>
<td>Explicit self-esteem</td>
<td>3.980</td>
<td>0.451</td>
<td>3.263</td>
<td>0.451</td>
</tr>
<tr>
<td>Positivity</td>
<td>3.664</td>
<td>0.712</td>
<td>3.898</td>
<td>0.477</td>
</tr>
</tbody>
</table>

Note: Numbers represent the IAPS codes of the stimuli used in the SRT.

### Table 3

Results of the MANOVA analyses (Study 1).

<table>
<thead>
<tr>
<th>Source</th>
<th>Dependent variable</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental manipulation</td>
<td>Implicit self-esteem</td>
<td>0.592</td>
<td>1</td>
<td>0.592</td>
<td>0.343</td>
<td>0.559</td>
</tr>
<tr>
<td></td>
<td>Explicit self-esteem</td>
<td>0.874</td>
<td>1</td>
<td>0.874</td>
<td>4.295</td>
<td>0.041</td>
</tr>
<tr>
<td>Error</td>
<td>Positivity</td>
<td>1.435</td>
<td>1</td>
<td>1.435</td>
<td>3.950</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td>Implicit self-esteem</td>
<td>177.802</td>
<td>103</td>
<td>1.726</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Explicit self-esteem</td>
<td>20.966</td>
<td>103</td>
<td>0.204</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>Positivity</td>
<td>37.424</td>
<td>103</td>
<td>0.363</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Implicit self-esteem</td>
<td>344.941</td>
<td>105</td>
<td>0.874</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Explicit self-esteem</td>
<td>1079.830</td>
<td>105</td>
<td>0.592</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Positivity</td>
<td>1542.734</td>
<td>105</td>
<td>0.592</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3.3. Discussion

The aim of the first study was to investigate whether the intersecting regularities effects can be observed when we use an SRT that links the self with highly positive stimuli. We found small-to-medium effects (d values around 0.40) in the case of explicit measures of self-esteem and positivity, and weak and non-significant effects in the case of implicit self-esteem. The SRT effects on the explicit measures are similar to the averaged effect size reported by Mattavelli et al. (2017) in their meta-analysis. Taken together, results suggest that pairing the self with highly positive stimuli can improve the explicit self-evaluations. An interesting result is that the effect was observed only on the explicit measures, not on the implicit self-esteem measure.

Although the results seem to support the idea that the positive valence was transferred from the stimuli to the self, we cannot exclude the demand characteristics effect (Orne, 1962) as an alternative explanation for these results. From this perspective, it is possible that our participants in the experimental group might have changed their explicit self-evaluations because the assessment moment was immediately after the SRT, not because of an SRT effect. Therefore, one of the objectives of the second study was to surpass this possible limitation.

4. Study 2

The primary objective of the second research study was to replicate the effect observed in Study 1 and to provide more statistical power to these findings. In addition, we also had three secondary objectives. First, we were interested in eliminating the alternative explanation of the demand characteristics effect. To achieve this goal, we introduced a time lag between the completion of the SRT and the self-esteem measurement.

Second, we were interested to see whether the SRT influences how participants evaluate some other types of experiences. In the first study, we have found significant effects on the positivity scale, which suggested that the SRT enhanced participants’ general positive orientation. This result suggests that the SRT effects can extend beyond the evaluations of the self. Therefore, we included two measures of well-being: study engagement and student burnout.

Finally, we were interested to see whether the repetition of the SRT is linked with the magnitude of the learning effect. Therefore, we asked participants to complete daily SRTs, and we evaluated the effect twice (after one week, and at the end of the two weeks).

4.1. Method

4.1.1. Participants

114 Psychology students (88.60% female) responded to our recruitment announcements and received informed consent forms. Although most enrolled students (96 persons) completed the first wave, only 76 participants (88.20% female, average age = 22.59 years, SD age = 5.51) completed all outcome measures in all waves. All students that logged in at least once on our Inquisit Web platform received course credit for their participation in the research. Similar to Study 1, participants were randomly assigned to the experimental or control group; this initial assignment was unchanged throughout the study. There were two reminders regarding the participation in this study, one day before each data collection moments (i.e., on day 6 and day 13).

Similar to Study 1, the SRT was presented as a memorization test. However, because the participants had to complete daily SRTs, we changed the positive and the neutral pictures from one day to another. We selected all pictures from the IAPS set, using the valence and intensity ratings available from Lang et al. (2008).

All measures, exclusions, and manipulations in the study are reported.

4.1.4. Data analysis

We had two complementary approaches to analyzing the results. First, we conducted post hoc comparisons of all outcome variables in the T1 (day 7), and in the T2 (day 14), to investigate whether there are any differences between the two experimental conditions.

In the second approach, we used stepwise regression analyses to investigate the incremental effect of the SRT in T2, while controlling the measurement in T1. The purpose of this approach was to investigate the cumulative effect of the SRT from T1 to T2. We can conclude that the SRT has a cumulative effect if the SRT has a predictive effect at T2 which is incremental to the T1 measurement.

4.2. Results

4.2.1. Post-hoc comparisons

The descriptive statistics for all dependent variables are presented in Table 3, for each experimental group in each of the two measurement moments. Similar to Study 1, we used the KS test to check that normality assumptions were respected by all outcome variables. The results presented in Table 2 indicated significant differences at T1 (i.e., day 7) between the experimental and control groups. Compared with the control group, the SRT group reported lower levels of burnout (t(74) = 2.532, p < .013, d = 0.581).

At the end of the 14 weeks (T2), the experimental group was significantly different from the control group on all study outcomes, except for implicit self-esteem (t(74) = 0.732, p = .467, d = 0.167). Taken together, we can conclude that the participants that completed the SRTs reported significantly higher levels of explicit self-esteem and
well-being at the end of the 14 days, as compared with the participants in the control group.

### 4.2.2. Regression analyses

The results presented above suggested that the SRT had a significant effect on how participants evaluated themselves and their level of well-being. However, another objective of this research study was to investigate whether the effect size increases as a result of the SRT practice, from T1 to T2. To achieve this goal, we used stepwise multivariate regression to analyze the data at T2. For each dependent variable measured at T2, we first introduced in the regression analysis the autocorrelation (i.e., the same variable, measured at T1). In the second step, we added the group membership (0 - control group, 1 - experimental group) in the regression analysis. The correlation matrix is presented in Table 4, and the regression estimates are presented in Table 5.

The results of the regression analyses presented in Table 6 suggested that, at T2, the SRT had a significant incremental effect on explicit self-esteem (B = 0.109, SE = 0.054, p = .035) and engagement (B = 0.296, SE = 0.119, p = .014). In the case of these variables the SRT effect was not statistically significant at T1, therefore we can conclude that a cumulative effect occurred from T1 to T2.

Together with these significant incremental effects, our regression analyses also yielded insignificant results. In the case of burnout, the SRTs did not have a significant incremental effect at T2 when its effect was statistically significant at T1. This suggests that the SRT effect did not increase as a result of repetitions from T1 to T2 if it was significant in T1.

### 4.3. Discussion

In the second study, participants completed daily SRTs for two weeks, and we found significant results at about 24 h after the completion of the previous SRT. The results of the second study confirmed the effect of the SRT on explicit self-esteem, and not on implicit self-esteem. In addition, the second study also provided evidence regarding the SRT effect on other variables related to self-esteem, such as well-being (i.e., engagement and burnout). Finally, the results of Study 2 suggested that the SRT effects can be observed at > 18 h after the completion of the task, after a continuous practice of 7 to 14 days.

### 5. Study 3

The third research study aimed at strengthening the evidence regarding the effect of the SRT on self-evaluations. To achieve this goal, we replicated our first study on a larger sample of participants from a different population (i.e., Amazon Mechanical Turk workers). Furthermore, we also included the short-form version of the Depression Anxiety Stress Scales (DASS-21, Henry & Crawford, 2005) to assess self-reported mental health. We added the self-reported mental health to conduct additional investigations regarding the occurrence of an SRT effect on the variables associated with self-esteem. We registered the effect size increases as a result of the SRT practice, from T1 to T2. To achieve this goal, we used stepwise multivariate regression to analyze the data at T2. For each dependent variable measured at T2, we first introduced in the regression analysis the autocorrelation (i.e., the same variable, measured at T1). In the second step, we added the group membership (0 - control group, 1 - experimental group) in the regression analysis. The correlation matrix is presented in Table 4, and the regression estimates are presented in Table 5.

The results of the regression analyses presented in Table 6 suggested that, at T2, the SRT had a significant incremental effect on explicit self-esteem (B = 0.109, SE = 0.054, p = .035) and engagement (B = 0.296, SE = 0.119, p = .014). In the case of these variables the SRT effect was not statistically significant at T1, therefore we can conclude that a cumulative effect occurred from T1 to T2.

Together with these significant incremental effects, our regression analyses also yielded insignificant results. In the case of burnout, the SRTs did not have a significant incremental effect at T2 when its effect was statistically significant at T1. This suggests that the SRT effect did not increase as a result of repetitions from T1 to T2 if it was significant in T1.

### 5. Study 3

The third research study aimed at strengthening the evidence regarding the effect of the SRT on self-evaluations. To achieve this goal, we replicated our first study on a larger sample of participants from a different population (i.e., Amazon Mechanical Turk workers). Furthermore, we also included the short-form version of the Depression Anxiety Stress Scales (DASS-21, Henry & Crawford, 2005) to assess self-reported mental health. We added the self-reported mental health to conduct additional investigations regarding the occurrence of an SRT effect on the variables associated with self-esteem. We registered the replication study using the template suggested by van der Veer and Giner-Sorolla (2016), and the anonymized protocol can be accessed at https://osf.io/8rsx2/register/5730e99a9ad5a102c5745a8a.
5.1. Method

5.1.1. Participants
We recruited U.S. participants using the online facilities of the Amazon’s Mechanical Turk. An a-priori power analysis suggested that 105 participants in each experimental group (i.e., 210 participants in total) should provide 0.95 statistical power of finding a two-tailed alpha = 0.05 in the case of a moderate effect size ($d = 0.50$). Initially, 211 persons completed our measures, but we excluded 33 participants because they had ≤ 80% correct classifications in the SRT task. The final sample included 178 participants (55.10% male, 72% with ages between 25 and 45 years old, 50% holding a Bachelor’s degree). Participants received 1.5 USD for their completion of the study.

5.1.2. Measures
Because we intended to replicate Study 1, we used the IPT (Koole et al., 2001) to assess implicit self-esteem, the Rosenberg self-esteem scale (Rosenberg, 1965) for the explicit self-esteem, and the Positivity scale (Caprara et al., 2012) to assess explicit positivity. All these measures had acceptable internal consistency indices (0.672 for implicit self-esteem, 0.923 for explicit self-esteem, and 0.919 for Positivity).

Furthermore, we also included the short-form version of the Depression Anxiety Stress Scales (DASS-21, Henry & Crawford, 2005) to assess self-reported mental health. The DASS-21 (Henry & Crawford, 2005) is a popular questionnaire for self-reported depression, self-reported anxiety, and self-reported stress. Respondents had to rate their answers using a 4 point scale (0–“Did not apply to me at all”; 3–“Applied to me very much or most of the time”), and higher scores indicate poorer mental health.

5.1.3. Procedure
We recruited participants on the Amazon’s Mechanical Turk. On the recruitment page, participants were informed that we were conducting a research study on individual differences regarding learning curves. Participants accessed an online Inquisit script (Inquisit, 2014) that assigned them randomly to an experimental or a control group. Similar to previous studies, the SRT was presented as a memorization test, and we used the scripts and the stimuli from Study 1.

All measures, exclusions, and manipulations in the study are reported.

5.2. Results
We present descriptive statistics on all dependent variables in Table 7. All variables had symmetrical distributions except for DASS-21, which was positively skewed.

Similar to Study 1, we tested our hypotheses using the multivariate analysis of variance. The results presented in Table 8 suggested significant differences between the experimental and control groups only in the case of explicit self-esteem ($F(1, 176) = 10.915, p = .001$, $d = 0.493$). Similar to the previous studies, the explicit self-esteem was higher in the experimental group, as compared with the control group.

Table 7
Descriptive statistics of the Study 3 dependent variables.

<table>
<thead>
<tr>
<th></th>
<th>Control group (N = 90)</th>
<th>Experimental group (N = 88)</th>
<th>All participants (N = 178)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Implicit self-esteem</td>
<td>0.854 (0.058)</td>
<td>0.860 (0.056)</td>
<td>0.857 (0.064)</td>
</tr>
<tr>
<td>Explicit self-esteem</td>
<td>3.170 (0.431)</td>
<td>3.392 (0.464)</td>
<td>3.279 (0.416)</td>
</tr>
<tr>
<td>Positivity</td>
<td>3.743 (0.642)</td>
<td>3.925 (0.648)</td>
<td>3.833 (0.649)</td>
</tr>
<tr>
<td>DASS score</td>
<td>0.387 (0.501)</td>
<td>0.349 (0.432)</td>
<td>0.368 (0.467)</td>
</tr>
</tbody>
</table>
The results of three independent experiments provided converging evidence for the concept of evaluative conditioning and its role in changing self-esteem. In the first study, a significant effect was observed on the Positivity scale, with a medium effect size ($r = 0.517$) for the relation between explicit self-esteem and positivity. In the second study, a smaller effect size was found ($r = 0.353$), suggesting a moderate correlation. The third study, which involved a larger sample, replicated the effect with a marginally significant result ($r = 0.282$). These findings support the idea that evaluative conditioning can alter self-esteem in a meaningful way.

We obtained marginally significant differences in the case of positivity ($F(1,176) = 3.527, p = .062, d = 0.282$), and insignificant results in the case of self-reported mental health ($F(1,176) = 0.289, p = .592, d = 0.081$) and in the case of implicit self-esteem ($F(1,176) = 0.002, p = .967, d = 0.006$).

Because our explicit outcomes were strongly correlated (i.e., $r(176) = 0.676$ for the relation between explicit self-esteem and positivity), we can infer that the effect holds across different measures. The significant effect on the Positivity scale suggests that the modified SRT has a direct impact on self-esteem, supporting the idea that positive self-referencing can improve general well-being and self-evaluations.

In the third research study, we replicated Study 1 on a U.S. sample. Our results replicated the significant effects of the SRT on explicit self-esteem and the significant effect on implicit self-esteem. Furthermore, we found a weaker effect on the Positivity scale, as compared with the effect found in Study 1. Finally, we did not find an effect of the SRT on self-reported mental health.

6. General discussion

The purpose of this research was to investigate whether the self-referencing task (Perkins et al., 2005) can be used to enhance the positive valence of the self. Previous research studies successfully changed the valence of the self through evaluative conditioning tasks (Baccus et al., 2004; Dijkstra, 2004; Covin, 2008; Grumm et al., 2009). Therefore we expected that we would find more positive self-evaluations, after the completion of an SRT which paired the self with highly positive stimuli.

The results of three independent experiments provided converging evidence regarding the enhancement of explicit positive self-views, following the modified version of the SRT. In our studies, explicit self-esteem was significantly higher in the experimental group (that paired the self with highly positive stimuli), as compared with the control group (that completed the same classification task, but without the self/others category). Furthermore, the results of Study 2 suggested that the effect can be observed at >18 h after the completion of the previous SRT. The effect sizes found in our studies are similar to the ones reported in the meta-analysis conducted by Mattavelli et al. (2017), which suggests that the relations established between the self and other stimuli can lead to changes in linking in both directions. This finding is important because it suggests that pairing the self with neutral stimuli (as it is usually done in classical SRTs) could lead to a decrease of the positive self-related valence. Except for Baccus et al. (2004) and Grumm et al. (2009), previous research studies on evaluative conditioning did not assess the explicit self-esteem. Therefore a general conclusion regarding the learnability of explicit self-valence is premature. However, we believe that our significant findings should encourage future research studies to investigate the evaluative learning effects on both forms of self-esteem (i.e., explicit and implicit). Furthermore, future research should also examine whether the transfer of valence can be observed with less valenced stimuli, or whether the positive valence of the stimuli changes following their pairing with the self.

Interestingly, we did not find a significant result in the case of implicit self-esteem, although the effect size was in the expected direction (i.e., slightly higher in the experimental group, as compared with the control group). Although we used a classification task similar to the one used by Ebert et al. (2009) we obtained smaller effect sizes, and we believe there are four possible reasons for this result. Unlike Ebert et al. (2009, Experiment 1), we paired the “others” category with a neutral set of stimuli, not with a negatively valenced set of stimuli. Therefore, one possible explanation for the smaller effect size is that our positive and neutral stimuli had a more similar valence, as compared with the positive and negative stimuli used by Ebert et al. (2009). The second possible explanation for our smaller effect sizes lies within the design of the research studies. Ebert et al. (2009, Experiment 1) contrasted a “self+/others−” condition with a “self−/others+” condition, while we contrasted a “self+/others0” condition with a controlled group that did not involve the self/others categories. The third possible explanation is related to the measure of implicit self-esteem we used in our studies. Ebert et al. (2009) used the implicit association test, and we used the initials preference task. We did not use the implicit association test because the modified SRT can be considered a task that trains the participant for the “self+positive” part of the implicit association test. Moreover, it is generally acknowledged that implicit measures have poor convergent validity (Falk & Heine, 2015). Therefore the insignificant effect on initials preference task should not be treated as an indicator that the modified SRT has insignificant effects on all measures of implicit self-esteem. Finally, a possible explanation for not finding a significant effect on implicit self-esteem lies in findings reported by Grumm et al. (2009). In their studies, Grumm et al. (2009) suggest that implicit self-esteem is influenced by affective manipulations, while explicit self-esteem is affected by propositional manipulations. In our study, it is possible that the modified SRT is a rather propositional, not an affective manipulation.

Together with the results on the valence of self-evaluations, we also investigated whether we can observe an SRT effect on concepts closely related to self-esteem. Recent developments suggested that positive self-evaluations are a particular manifestation of positivity, which is a more general individual characteristic that includes evaluations of the “positive view of one’s self, one’s life, and one’s future, as well as one’s confidence in others” (Caprara et al., 2012 p. 702). Therefore, we also included measures of positivity (Caprara et al., 2012), measures of academic well-being (engagement and burnout), or measures of self-reported mental health (DASS-21, Henry & Crawford, 2005). Concerning these variables, our results were mixed. On the one hand, we found significant effects on positivity (statistically significant in Study 1, and marginally insignificant in Study 3) and on well-being variables (academic engagement and academic burnout, Study 2). On the other hand, we found almost null effects in the case of self-reported mental health (Study 3). In Study 3 we found that the modified SRT effect is significant on a latent factor that accounts for the common variance of our explicit variables (i.e., self-esteem, positivity, and self-reported mental health). Therefore, these findings suggest that the SRT effect is
linked with the individual's propensity to make generally positive evaluations regarding his own situational well-being, not only his own person.

Finally, we investigated whether the repetition of the SRT has an impact. We found that when SRT effect was significant on the 7th day of the study (the case of burnout, in Study 2), the effect size at the end of the study was not significantly predicted by the experimental condition. This suggested that, once it is significant, an SRT effect was not significantly improved. However, if the effect was not significant after the first half of Study 2, the repetition of the SRT had a significant effect at the end of the study. At this moment, it is unclear whether this differentiation can be attributed to the dependent variable. Therefore, in order to have a better understanding of the differential effects of this repetition, future research should investigate a more extensive array of positivity-related variables.

Our findings suggest that explicit positivity can be changed using an evaluative learning procedure based on intersecting regularities. This is important because it suggests that one can induce temporal changes in the evaluations of the self and explicit well-being. Because some changes in well-being occurred after the first half of Study 2, it is possible that explicit well-being is more sensitive to change using the modified version of the SRT, as compared with implicit self-esteem.

6.1. Limitations

The findings presented in this paper have some limitations as follows. Firstly, we did not find significant effects on implicit self-esteem although the effect sizes indicated slightly higher levels of implicit self-esteem in the SRT conditions, as compared with the control groups. We believe these results are limited by the fact that we used only one measure for implicit self-esteem (the IPT - Koole et al., 2001). Because the implicit self-esteem measures are generally uncorrelated (Baccus et al., 2004; Bosson et al., 2000), future studies could find larger effect sizes if they use different measures for implicit self-esteem.

Secondly, an alternative explanation of the SRT effect on the explicit measures is the demand characteristics effect (Orne, 1962). In our opinion, this alternative explanation is not supported by the results of Study 2 because i) most dependent variables did not yield significant effects after 7 days and because ii) there was an 18 h time lag between the finalization of the SRT and the completion of the explicit measures. However, future studies should include explicit manipulation checks for the demand characteristic effect.

The third limitation of our studies lies in the nature of our dependent variables. Although we have found significant effects on self-reported measures, we have little information regarding the effects on actual human behavior. Therefore, future studies should also investigate whether the modified SRT effect can also be observed on recordings of actual behaviors.

Finally, another limitation is the absence of a follow-up measurement moment. Although the effects were assessed at >18 h after the completion of the last SRT (Study 2), the studies presented in this paper do not provide any information regarding the duration of the SRT effect. Therefore, future studies should consider the use of follow-up measures, to overcome this limitation.

6.2. Conclusions

The present research provides evidence regarding the evaluative learning processes behind the evaluations of the self and behind the assessments of own well-being. We have shown that evaluative learning processes can explain changes in these evaluations, both in experimental and in naturalistic settings. On a different vein, our results provided support for the general positivity perspective (Caprara et al., 2012), who stated that evaluations of the self are a specific form of a general positive orientation.

Acknowledgments

This work was supported by a grant of the Romanian Ministry of National Education, CNCS – UEFISCUD, project number PNII-RU-TE-2014-4-0398. This organization had no role in the design and implementation of the study.

References


Inquisit 4.0.8.0 (Computer software). (2014). Seattle, WA: Millisecond Software LLC.


