

Enhancing implicit self-esteem in children: Can a smile make you feel worthwhile?

These experiments represent a first exploration of implicit and explicit self-esteem in children, and test their malleability using the evaluative conditioning procedure that was successfully employed in adults (Baccus, Baldwin, & Packer, 2004). Experiment 1 showed that children for whom self-relevant information was repeatedly paired with smiling adult faces displayed somewhat higher implicit self-esteem and less aggression than controls. These findings were not replicated, however, in Experiment 2 that used a pre-post test design. In Experiment 3, repetition of the conditioning procedure did not produce significant effects on self-esteem or psychological functioning. Overall, pairing smiling faces with self-relevant information had no effect on children's implicit or explicit self-esteem, and no consistent relations between implicit and explicit self-esteem measures were found. Implications and future research are discussed.

Where: *Netherlands Journal of Psychology*, Volume 66, 11-25

Received 21 April 2009; accepted 11 June 2009

Keywords: Implicit self-esteem; Self-esteem enhancement; Evaluative conditioning; Children; Aggression

Authors: Jorg Huijding*, Arjan E. R. Bos** and Peter Muris*

Self-esteem, the overall evaluation of one's worth or value as a person, is a construct that plays an important role in the mental and social development of children and adolescents (e.g., Harter, 1999). Research has shown that high self-esteem is related to parental approval, peer-support, adjustment, and success in school (Steinberg & Sheffield Morris, 2001). Conversely, children with low self-esteem have been found to be less successful in school and are less accepted by their peers (e.g., Mann, Hosman, Schaalma, & De Vries, 2004; Bos, Muris, Mulken, & Schaalma, 2006). In addition, low self-esteem is related to several forms of child and adolescent psychopathology including anxiety (Beck, Brown, Steer, Kuyken, & Grisham, 2001; Muris, Meesters, & Fijen, 2003a), depression (Harter, 1993; Mann et al., 2004), and eating pathology (e.g. Muris, Meesters, Van de Blom, & Mayer, 2005; Stice, 2002). Self-esteem has also been linked with externalising problems such as aggressive and delinquent behaviours. However, there is still debate concerning the exact nature of this relation. While some researchers have found low self-esteem to be related to externalising problems (Donnellan, Trzesniewski, Robins, Moffitt, & Caspi, 2005), others have argued that externalising problems are

related to high self-esteem as a result of egotism (Baumeister, Smart, & Boden, 1996). The majority of research on self-esteem has focused on self-reported, or explicit, self-esteem (see Butler & Gasson, 2005 for an overview), which refers to conscious and deliberate evaluations of the self, self-relevant objects and situations. More recently, however, it is increasingly acknowledged that there are also more automatic self-evaluative processes that are not necessarily conscious (e.g., Greenwald & Banaji, 1995; Koole, Dijksterhuis, & Van Knippenberg, 2001; Dijksterhuis, 2006). Such self-evaluative processes have been subsumed under the concept of implicit self-esteem. Although the concept of implicit self-esteem is relatively young, there is accumulating evidence that it may be an important determinant of social and mental functioning (see for a review Koole & DeHart, 2007). For instance, Conner and Barrett (2005) demonstrated that implicit but not explicit self-esteem predicted participants' spontaneous daily experiences of a range of negative emotions such as boredom, shame, anger and sadness over an extended period of time. Other studies have shown that individuals with low implicit self-esteem show a lower level of aspiration after failure (Greenwald & Farnham, 2000), and

* Erasmus University
Rotterdam, the Netherlands

** Maastricht University,
the Netherlands
Correspondence to: Jorg
Huijding, Institute of Psy-
chology, Erasmus University
Rotterdam, PO Box 1738,
3000 DR Rotterdam,
the Netherlands,
e-mail: huijding@fsw.eur.nl

experience more anxiety during a very personal interview (Spalding & Hardin, 1999) compared with individuals with high implicit self-esteem. In addition, individuals who score high on a measure of social anxiety show lower implicit self-esteem than individuals who score low on such a measure (De Jong, 2002; Tanner, Stopa, & De Houwer, 2006). Taken together, the existing data suggest that both a low explicit and a low implicit self-esteem are associated with a range of negative consequences. Importantly, the effects of implicit and explicit self-esteem seem to be complementary rather than identical (Dijksterhuis, Albers, & Bongers, 2007). That is, they each appear to explain unique variance on various outcome measures. This underlines the importance of including measures of both implicit and explicit self-esteem when studying the effects of self-esteem on psychological functioning.

Given the risk that low self-esteem may have for a healthy development, several interventions have been developed that aim to enhance children's self-esteem (see Bos et al., 2006). A meta-analysis by Haney and Durlak (1998) showed that such interventions have modest effects on explicit self-esteem (with a mean effect size of 0.27). Recent evidence in the adult literature suggests that implicit self-esteem can also be successfully enhanced (Baccus, Baldwin, & Packer, 2004; Dijksterhuis, 2004). For instance, Baccus et al. (2004) found that implicit self-esteem can be increased using an evaluative conditioning paradigm in which self-relevant stimuli (e.g., first name, date of birth) were consequently paired with pictures of smiling faces, and other-relevant stimuli were always followed by neutral or angry faces. Interestingly, this procedure not only seemed to promote implicit self-esteem relative to other-esteem, but also appeared to have an effect on a measure of aggression. That is, participants who received positive reinforcement during the conditioning task were less aggressive than participants for whom self and other-relevant stimuli were equally often followed by pictures of smiling, neutral and angry faces. This suggests that implicit self-esteem can be enhanced and that this positive effect generalises to behaviours that are assumed to be related to self-esteem, such as aggression. Meanwhile, as self-relevant stimuli were always followed by smiling faces, but other-relevant stimuli always by negative and neutral faces it is difficult to know whether the effects were due to a strengthening of *positive* associations with the self, a strengthening of *negative* associations with others, or a combination of both¹. In addition, there are no published studies that have examined whether implicit self-esteem can be enhanced in children and adolescents. In

fact, to our knowledge, there are no published studies at all that focus on implicit self-esteem in children, and, consequently, there are virtually no data available concerning the relation between implicit and explicit self-esteem in children, and the psychological correlates of implicit self-esteem outside adulthood.

The main purpose of the present series of experiments was to take a first step in filling the gap in our knowledge concerning implicit self-esteem in childhood by assessing implicit and explicit self-esteem in various samples of primary school children. We employed the conditioning paradigm that was successfully used by Baccus and colleagues (2004) and tested whether this experimental procedure can be used to enhance implicit and explicit self-esteem in primary school children. In order to test the net effect of the positive conditioning procedure for self-relevant stimuli, the conditioning procedure of Baccus et al. (2004) was slightly modified. That is, instead of always pairing other-relevant stimuli with negative or neutral faces, we paired these stimuli equally often with positive, negative or neutral faces (see method section of Experiment 1 for details). In all experiments we focused on youths in middle childhood, because children in this age range are assumed to be capable of integrating both positive and negative evaluations about themselves and as such form a more realistic representation of their overall self-worth (Harter, 1999). As such, this age range seems suitable for examining inter-individual differences in implicit and explicit self-esteem, and the relations between both types of self-esteem and other variables of interest, such as aggression and psychological functioning. Younger children typically make very positive self-attributions, due to an inability to distinguish between the real and the ideal self, an overestimation of their own virtuosity, and all-or-nothing thinking (Harter, 1999).

Experiment 1 was designed to test whether we could replicate the basic findings of Baccus et al. (2004) in a group of primary school children. Children were given a conditioning task that closely followed the procedure as described by these researchers. Subsequently, children completed measures of implicit self-esteem, explicit self-esteem, and aggression. Because Experiment 1 only included post-assessments and any lack of effects might be caused by differences between the experimental and the control group prior to the conditioning procedure, Experiments 2 and 3 used a pre-post test design. As a subsidiary issue, Experiment 2 was designed to test whether any effects of the conditioning procedure on implicit or explicit self-esteem and aggression

¹ Note that even though Baccus and colleagues analysed only the 'self' trials in order to avoid this problem, research on the implicit measure that was used suggests that the approach of analysing just trials from one of the two target categories is not a valid one (Nosek, et al., 2007). As a result, one cannot conclude that the results of this earlier study were entirely about associations with the self only.

would be stable over longer time periods. Therefore, this experiment included a follow-up assessment of one month. Finally, Experiment 3 tested whether the effect on children's self-esteem and general psychological functioning can be enhanced if the conditioning task is completed repeatedly. Because previous studies consistently observed differences between boys and girls in explicit self-esteem (e.g., Kling, Hyde, Showers, & Buswell, 1999; Robins, Trzesniewski, Tracy, Gosling, & Potter, 2002), we examined the influence of gender in all experiments.

Experiment 1

Method

Participants

A total of 133 primary school children (67 boys, 66 girls) aged between 9 and 12 years ($M = 11.2$, $SD = 0.71$) participated after obtaining written consent from the children themselves and their parents. All participants were randomly assigned to either the experimental or the control condition. There were no differences with regard to gender distribution, $\chi^2(1, N = 133) < 1$, and age, $t(131) < 1$, between the experimental and the control condition.

Materials

Explicit self-esteem. To assess global self-esteem the children completed the Dutch version of the Rosenberg Self-Esteem Scale (RSES; Rosenberg, 1965). The RSES consists of ten items that assess global feelings of self-esteem on a four-point Likert scale. The internal consistency of the RSES was acceptable in the present sample (Cronbach's $\alpha = .67$).

Children also completed the Dutch version of the Self-Perception Profile for Children (SPPC; Harter, 1985; Veerman, Straathof, Treffers, Van den Bergh, & Ten Brink, 1997). The 36-item SPPC is the most widely used self-report measure for assessing self-esteem in youths, and the Dutch version of this scale has satisfactory reliability with good internal consistency and test-retest stability (Muris, et al., 2003). In the present sample the internal consistency of the total scale was good ($= .90$).

Implicit self-esteem. Similar to Baccus et al. (2004), the present experiment employed the self-esteem IAT (Greenwald & Farnham, 2000) as a measure of implicit self-esteem. The IAT was adapted especially for the use with children following the procedure described by Field and Lawson (2003). The tasks were run on laptop computers using E-prime software (Version 1.0, Schneider, Eschman, & Zuccolotto, 2002) to ensure high precision accuracy of reaction time recordings. The IAT required children to categorise words into

self and other categories, and *positive* and *negative* categories using two response keys. In accordance with the standard IAT design (Greenwald, McGhee, & Schwarz, 1998), the task consisted of two critical sets of trials, the *compatible block* and the *incompatible block*. In the *compatible block*, children were asked to use one key to categorise self (e.g., self, me, my) and positive words (e.g., nice, good, smart), and the other key to categorise other (e.g., other, they, them) and negative words (e.g., stupid, dumb, bad). In the *incompatible block*, children were instructed to use one key to categorise self and negative words and the other key to categorise other and positive words. The idea behind the task is that participants will perform better when the two concepts that share a response key are somehow associated in memory then when they are not. Following this, implicit self-esteem can be computed as the difference in performance between the compatible and incompatible blocks (see for details Greenwald, Nosek, & Banaji, 2003).

The entire IAT procedure consisted of seven phases: (1) practice categorising *self* and *other* words (10 trials); (2) practice categorising *positive* and *negative* words (10 trials); (3) practice the *compatible* categorisation of all words (20 trials); (4) critical *compatible* categorisation of all words (40 trials); (5) practice of reversed key assignments for the self and other words (30 trials); (6) practice the *incompatible* categorisation of all words (20 trials); (7) critical *incompatible* categorisation of all words (40 trials). Words were categorised by pressing the 'E' or the 'I' on the keyboard. Key assignments were counterbalanced across participants. Before each stage children received instructions on the response keys that represented each of the categories, and were then asked to categorise the words as fast and accurately as possible. To remind children of these instructions during the task, category labels were shown in the upper left and right corners of the screen throughout each stage. Each trial started with the presentation of a fixation dot in the middle of the computer screen. After 500 ms the fixation dot was replaced by the stimulus word. If a word was miscategorised then a red 'X' appeared under the stimulus. Both the stimulus and the cross then remained on the screen until the correct categorisation was made. Note, however, that only the latency until the initial response was recorded. After the correct categorisation was made the stimulus was immediately replaced by the fixation dot for the next trial. In total the task took about ten minutes for the children to complete.

An IAT score was calculated for each child on the basis of the scoring algorithm as proposed by

2 Similar to Dijksterhuis (2004) the order of compatible and incompatible blocks was not counterbalanced, because the focus of the present study was not on absolute IAT effects but rather on the comparison of IAT effects between the two conditions. The major advantage of using only one IAT version was that between-subjects method variance was reduced resulting in more power.

Table 1 Mean scores (standard deviations) on self-esteem and aggression measures for both conditions in Experiment 1

	Condition	
	Experimental (n = 67)	Control (n = 66)
RSES	31.4 (4.1)	30.8 (3.8)
SPPC	98.4 (1.5)	100.0 (1.5)
IAT	0.76 (0.32) ^a	0.68 (0.28) ^b
Aggression	1233.0 (172.7) ^a	1817.2 (172.1) ^b

RSES = Rosenberg Self-Esteem Scale; SPPC = Self-Perception Profile for Children; IAT = Implicit Association Test; Aggression = aggression index based on the Hot Sauce test. For the SPPC and Aggression estimated marginal means (standard errors) are reported. IAT scores are similar to Cohen's *d* for the difference between the compatible (self + positive, others + negative) and incompatible (self + negative, others + positive) blocks, with positive scores indicating more positive self-esteem. For the IAT $n = 61$ for the experimental and $n = 64$ for the control condition. For Aggression $n = 63$ for both the experimental and the control condition. Means in the same row with different superscripts differ with $p < .1$.

Greenwald and colleagues (2003). More specifically, we calculated the recommended D4 measure that includes both combined practice and test phases and replaces errors with a 600 ms penalty. The Spearman-Brown corrected split-half reliability of this measure was reasonable ($r = .59$). Data of five children in the experimental and three children in the control condition were excluded from the analyses because they had more than 25% errors on the IAT. In addition, data from one child in the experimental group was excluded because he was more than 3 standard deviations removed from the mean score of his group.

Aggression. Aggression was measured by means of an adapted version of the 'hot sauce test' (Lieberman, Solomon, Greenberg, & McGregor, 1999). Under a cover story that we wanted their help in a taste experiment, children were presented with a cup that was filled for a quarter with a red, hot (spicy) sauce. Children were told that another participant had tasted and rated the sauce as very spicy, and had then decided for them that they should eat this amount of it. They were then asked to taste the sauce and indicate how much sauce they would like this other participant to eat in return. Children indicated the amount of sauce on a drawing of a cup. Finally, each child rated on a 100 mm Visual Analogue Scale (VAS) how much they liked or disliked the sauce (0 = very tasty, 100 = very nasty). As an index of aggression we took the product of the amount of sauce (distance in mm from the bottom of the cup up to the line marked by the children) and the VAS rating of the sauce.

Data from seven children (3 from the experimental and 4 from the control condition) were missing because these children refused to complete the hot sauce task.

Conditioning procedure. The conditioning procedure was designed after Baccus et al. (2004). Participants started by entering personal information about themselves (i.e., first name, last name, date of birth, place of birth, first letter of first name) into the computer. Then they were told that a word would appear randomly in one of the quadrants on the computer screen and that their task was to click on that word as fast as possible. They were instructed that clicking on the correct quadrant would cause an image to be displayed briefly (for 400 ms.) in that quadrant. This procedure was repeated for 240 trials. The presented words were randomly chosen from the personal information that was entered at the start of the session (self-relevant words) as well as from a pre-programmed list of other-relevant words that fitted the same categories. Half of the trials were self-relevant and the other half were other-relevant. In the experimental condition self-relevant items were always followed by an image of a smiling adult face (120 trials), whereas other-relevant items were followed randomly by an image of a smiling, angry or neutral adult face (120 trials). In the control condition all items (self and other-relevant) were followed randomly by an image of a smiling, angry or neutral adult face. The faces were equally often male and female and derived from the MacBrain Face Stimulus Set (NimStim)³. The conditioning procedure took about seven minutes to complete.

Procedure

All children were tested individually in a separate room at their school. After the initial instruction, participants first completed the RSES. Then, children were seated in front of the laptop computer and completed either the experimental or the control version of the conditioning task. Following the conditioning task they completed the IAT. Upon finishing the IAT, participants were asked to fill in the Self-Perception Profile for Children. Finally, the children completed the 'hot-sauce test'. Children needed about 30 minutes to complete all tests.

Results

Pre-experimental group differences

Descriptive statistics are shown in Table 1. A 2 (Condition: experimental vs. control) \times 2 (Gender: girls vs. boys) analysis of variance (ANOVA) performed on the RSES total score only yielded a significant main effect of Gender, $F(1, 129) = 4.7$, $p < .05$, partial $\eta^2 = .04$: boys displayed significantly

³ Stimuli can be obtained from the first author. The MacBrain Face Stimulus Set was developed by Nim Tottenham and supported by the John D. and Catherine T. MacArthur Foundation Research Network on Early Experience and Brain Development. Please contact Nim Tottenham at tott0006@tc.umn.edu for more information concerning the MacBrain Face Stimulus Set.

higher RSES scores ($M = 30.8$, $SD = 3.1$) than did girls ($M = 29.6$, $SD = 3.1$). Importantly, there were no group differences on self-esteem before the conditioning procedure. That is, the main effect of Condition and the interaction of Condition with Gender were both non-significant, both $F(1, 129) < 1$.

Post-experimental group differences

SPPC. In order to examine children's explicit self-esteem after the experimental manipulation, SPPC total scores were subjected to a 2 (Condition) \times 2 (Gender) ANCOVA with the pretest RSES scores as the covariate. As can be seen in Table 1, children in the experimental and the control condition had comparable SPPC scores, $F(1, 126) < 1$. As indicated by a significant main effect of Gender, $F(1, 126) = 4.0$, $p < .05$, partial $\eta^2 = .03$, boys displayed generally higher SPPC scores (estimated marginal $M = 101.3$, $SE = 1.7$) than did girls (estimated marginal $M = 97.1$, $SE = 1.3$). The covariate was significant, $F(1, 126) = 29.6$, $p < .01$, $\eta^2 = .19$.

IAT. A 2 (Condition) \times 2 (Gender) ANOVA performed on Children's IAT scores revealed that the main effect of Condition showed a trend towards significance, $F(1, 125) = 2.8$, $p = .097$, partial $\eta^2 = .02$. Table 1 shows that children in the experimental condition exhibited somewhat higher IAT scores than children in the control condition. The main effect of Gender and the Condition \times Gender interaction were non-significant, both $F(1, 123) < 1$.

Aggression. A 2 (Condition) \times 2 (Gender) ANCOVA of the aggression data, with the RSES scores as the covariate, yielded a significant main effect of Condition, $F(1, 121) = 5.7$, $p < .05$, partial $\eta^2 = .05$. As expected, children in the experimental condition showed less aggression than children in the control condition (see also Table 1). The covariate was also significant, $F(1, 121) = 5.4$, $p < .05$, partial $\eta^2 = .04$. The main effect of Gender and the interaction of Condition \times Gender were both non-significant, both $F(1, 121) < 1$. It is important to note that the two groups did not differ in their evaluation of the sauce, $t(124) < 1$.

Relation between Implicit and Explicit Self-esteem. The IAT was not significantly correlated with either the pretest RSES, $r = -.03$, or the posttest SPPC, $r = .06$. The RSES and SPPC did show a significant correlation, $r = .46$, $p < .01$, indicating that higher self-reported self-esteem on the pretest was associated with higher self-reported self-esteem on the posttest.

Relation between self-esteem and aggression. The RSES showed a significant negative correlation with the

measure of aggression ($r = -.22$, $p < .05$), indicating that lower levels of self-reported self-esteem at pretest were associated with higher levels of aggression at posttest. Neither the SPPC nor the IAT correlated significantly with the measure of aggression, $r = -.08$ and $r = .04$, respectively.

Discussion

Consistent with Baccus et al. (2004), participants who completed the experimental conditioning procedure in which self-relevant stimuli were consistently paired with smiling faces, tended to show enhanced implicit self-esteem and less aggression as compared with participants in the control condition. This is the first tentative demonstration that implicit self-esteem may be experimentally enhanced in children, and that a conditioning procedure may have an effect on aggression. It should be noted, however, that both effects were rather small and certainly require replication. In addition, no meaningful relations were found between aggression and measures of implicit and explicit self-esteem, which casts doubt on the idea that self-esteem is related to aggression in this age group. Meanwhile, an alternative explanation is that the currently used measure of aggression, the hot sauce test, did not provide a valid index of aggression in the present context. That is, although the hot sauce test has proven to be a reliable and valid measure of behavioural aggression in adults (Lieberman et al., 1999), it may have been too abstract and complex for children of this age. Related to this, the aggression index was based on a single item, which may have undermined the reliability of the assessment in the present experiments. For these reasons Experiment 2 included an aggression questionnaire that was especially designed to measure aggressive tendencies in youths on a continuous scale.

Although the first experiment yielded potentially interesting results, one important limitation of this experiment is that, despite the random allocation of participants to the experimental and the control condition, it cannot be precluded that the trend towards group differences following the conditioning task was due to pre-existing differences between both conditions, rather than to the experimental manipulation. In order to take potential pre-existing group differences into account a pre-post test design would be required. In addition, given the small size of the effects in Experiment 1, the use of a repeated measures design would have the additional advantage of offering more statistical power.

Following this, Experiments 2 and 3 were designed to further explore whether self-esteem can be

enhanced with the conditioning procedure employed in Experiment 1 using a pre-post test design. In addition, to provide more convincing evidence of changes in implicit self-esteem an additional measure of implicit self-esteem was added, the Name Letter Preference Task (Nuttin, 1985). As a subsidiary issue, Experiments 2 and 3 were designed to test whether any effects of the conditioning procedure would be stable over longer time periods. Therefore, both experiments included a one-month follow-up. Experiment 2 further explored the effect of the conditioning procedure on self-esteem and aggression, and included an alternative aggression measure that assessed both reactive and proactive aggression. Experiment 3 investigated the effect of the conditioning procedure on self-esteem and general psychological functioning, and therefore included the Strengths and Difficulties Questionnaire (Goodman, 2001). Finally, given the small size of the effects as obtained in Experiment 1, an additional goal of Experiment 3 was to examine whether repetition of the conditioning procedure would bolster the effects on children's self-esteem.

Experiment 2

Method

Participants

A total of 106 primary school children (47 boys, 59 girls) were randomly assigned to either the experimental or the control condition after obtaining written consent from the children and their parents. Children were aged between 10 and 13 years ($M = 11.2$, $SD = 0.77$). There were no differences with respect to gender, $\chi^2(1, N = 106) < 1$, and age, $t(104) < 1$, between the experimental and the control condition.

Materials

Explicit self-esteem. To reduce experiment burden only the global self-esteem scale of the SPPC was used in the present experiment. The internal consistency of this scale was good with alphas of .85, .88, and .91, at the pretest, posttest and follow-up assessments respectively. Test-retest stability in the control group was high from pretest to posttest ($r = .90$, $p < .01$) and from posttest to follow-up ($r = .80$, $p < .01$).

Implicit self-esteem. To assess implicit self-esteem we used an IAT that was identical to the one employed in Experiment 1. Spearman-Brown corrected split-half reliability varied over the different assessments (r 's being .64, .39, and .56, respectively). Test-retest correlations in the control group were significant from pretest to posttest, $r = .32$, $p < .05$, and from posttest to follow-up, $r = .39$, $p < .05$. Data were treated similarly as in Experiment

1. Data of children who made more than 25% errors and from children that were more than 3 standard deviations removed from the group mean were excluded from the analyses. All in all, the final N for the analysis was 86 (45 in the experimental and 41 in the control condition).

The Name Letter Preference Task (NLPT; Nuttin, 1985) was used as an additional index of implicit self-esteem. The NLPT assesses how well individuals like their own initials relative to the other letters of the alphabet (see Koole & De Hart, 2007; Krizan & Suls, 2008). In the present experiment children were presented with all 26 letters of the Dutch alphabet, which appeared one by one and in a random order on a computer screen. Beneath each letter five square boxes appeared which were labelled 'not at all beautiful', 'somewhat beautiful', 'beautiful', 'very beautiful', and 'extremely beautiful'. For each letter, the children were asked to click on the box that corresponded best with their evaluation of the letter on screen. Thus, the attractiveness of each letter was rated on a five-point scale. Calculation of the Name Letter Preference Task (NLPT) effect was based on previous studies by Koole and colleagues (2001) and Franck, De Raedt, and De Houwer (2007). As a first step, the rating of each individual letter for each participant was corrected for inter-individual differences in rating tendencies by Z-transforming it using the mean and standard deviation of all letter ratings of that participant. Next, a baseline score was calculated for each letter of the alphabet by taking the mean of each letter for all participants that did not have that letter in their first or last name. For the pretest assessment the baseline scores were calculated on the basis of the entire sample, for the posttest and follow-up assessments, the baseline scores were calculated for each group (experimental, control) separately. Following this, separate name letter effects were calculated for the first letter of the first and last name of each participant by subtracting the baseline score from the participants' Z-transformed rating for that letter. Finally, a total name letter preference score was obtained by averaging both NLPT effects. The final N for the data analysis was 104 (55 in the experimental and 49 in the control condition). Test-retest correlations in the control group were significant both from pretest to posttest, $r = .60$, $p < .01$, and from posttest to follow-up, $r = .63$, $p < .01$. However, correlations between the name letter effects for the first and last name were very low (r 's .01, .00, and .15, respectively), indicating poor internal consistency.

Aggression. As a measure of aggression we included a ten-item aggression questionnaire

that was especially designed for measuring aggressive tendencies in youths, the Aggression Questionnaire for Youths (AQY). Children were presented with ten situations and asked to rate on a six-point scale, varying from 0 = 'nothing' to 5 = 'very much', how much of a specific agent (e.g., water, itching powder, chewing gum) they would use against another child if they were in that particular situation. Factor analysis demonstrated that the questionnaire consists of two factors, pertaining to *reactive* and *proactive* aggressive actions against others. Supporting the validity of the AQY, pilot work showed both subscales are significantly related the Child Rating Scale for Aggression (CRA: Meesters, Muris, & Van Rooijen, 2007). Both subscales displayed acceptable to good internal consistency. For the reactive subscale alphas ranged from .87 to .92, while for the proactive subscale the alphas ranged from .65 to .79. Further, in spite of the experimental manipulation, test-retest correlations ranged from .71 (pretest to follow-up) to .86 (posttest to follow-up), for the reactive subscale and from .60 (pretest to follow-up) to .70 (posttest to follow-up) for the proactive subscale, suggesting adequate test-retest stability over time.

Conditioning procedure. The conditioning task was identical to the one used in Experiment 1.

Table 2 Mean scores (standard deviations) on various self-esteem and aggression measures in both conditions obtained at the pretest, posttest, and follow-up assessments in Experiment 2

		Condition	
		Experimental (n = 55)	Control (n = 51)
SPPC	Pre	19.6 (3.1)	20.0 (3.0)
	Post	19.8 (3.2)	20.0 (3.2)
	Fu	20.0 (3.8)	20.1 (3.1)
IAT	Pre	0.39 (0.33)	0.44 (0.40)
	Post	0.58 (0.26)	0.60 (0.34)
	Fu	0.59 (0.31)	0.63 (0.28)
NLPT	Pre	0.62 (0.62)	0.61 (0.74)
	Post	0.82 (0.68)	0.76 (0.74)
	Fu	0.80 (0.70)	0.72 (0.74)
AQY	Pre	11.4 (7.6)	12.1 (8.0)
	Post	12.3 (8.8)	14.0 (8.6)
	Fu	12.6 (9.0)	13.0 (8.0)

SPPC = Self Perception Profile for Children; IAT = Implicit Association Test (positive scores indicate positive self-esteem); NLPT = Name Letter Preference Task (positive scores indicate positive self-esteem); AQY = Aggression Questionnaire for Youths; FU = follow-up. N's vary across variables and assessments due to missing and/or excluded data; reported means are based on the data included in the analyses.

Procedure

Children were tested in small groups with a maximum of four children in a separate room at school. The experimenter was present at all times and the children were seated facing away from each other at different ends of the room to ensure that they would not distract each other during the assessments. In order to control for group differences before the conditioning task, a pre-post test design was used. To reduce the potential influence of test-retest effects participants completed the pretest assessment one day before they completed the conditioning task. The posttest assessment was completed immediately after the conditioning task. After one month the experimenter returned for the follow-up assessment. During all assessments children first received a brief introduction and then consecutively completed the IAT, NLPT, SPPC, and the AQY.

Results

Descriptive statistics for children in both conditions at various assessments are shown in Table 2.

Explicit self-esteem

A 2 (Condition: experimental vs. control) x 2 (Gender: girls vs. boys) x 3 (Assessment: pretest, posttest, follow-up) ANOVA on children's SPPC scores showed that none of the main or interaction effects were significant (all F s < 1.2, $p > .2$). Thus, the SPPC scores remained stable over the three assessments, and were independent of Condition and Gender.

Implicit self-esteem

IAT. Children's IAT scores were subjected to a 2 (Condition) x 2 (Gender) x 3 (Assessment) ANOVA, with the last factor being a repeated measure. Only the main effect of Assessment reached significance, $F(2, 81) = 11.1$, $p < .01$, partial $\eta^2 = .22$. Post-hoc comparisons showed that, irrespective of Condition or Gender, children showed increasing IAT scores from the pretest ($M = .45$, $SD = .32$) to the posttest ($M = .62$, $SD = .27$) assessment, $F(1, 71) = 19.6$, $p < .01$, partial $\eta^2 = .22$. The IAT scores remained stable from the posttest to the follow-up assessment, $F(1, 71) < 1$.

NLPT. Children's NLPT effects were subjected to a 2 (Condition) x 2 (Gender) x 3 (Assessment) ANOVA, with the last factor being a repeated measure. Only the main effect of Assessment was significant, $F(2, 99) = 3.4$, $p < .05$, partial $\eta^2 = .06$. Within-subject contrasts indicated that this effect was due to the difference between the pretest ($M = .61$, $SD = .68$) and the posttest ($M = .79$, $SD = .71$) assessment, $F(1, 100) = 6.6$, $p < .05$, partial $\eta^2 = .06$. The difference between the posttest and the follow-up ($M = .76$, $SD = .721$) assessment was not significant, $F(1, 100) < 1$.

Aggression

The AQY scores were analysed by means of a four-way 2 (Condition: experimental vs. control) x 2 (Gender: girls vs. boys) x 2 (Subscale: reactive vs. proactive) x 3 (Assessment: pretest, posttest, follow-up) ANOVA, with the last two factors being repeated measures. As indicated by a significant main effect of Subscale, participants scored higher on the reactive ($M = 10.6$, $SD = 6.4$) than on the proactive subscale ($M = 1.9$, $SD = 2.5$), $F(2, 100) = 3.2$, $p < .05$, partial $\eta^2 = .06$. The main effect of Assessment was also significant, $F(2, 100) = 3.2$, $p < .05$, partial $\eta^2 = .06$. Within-subject contrasts showed that children reported an increase in aggression from the pretest to the posttest ($M = 11.7$, $SD = 7.8$, and $M = 13.1$, $SD = 8.8$, respectively), $F(1, 101) = 6.4$, $p < .05$, partial $\eta^2 = .06$. The difference between the posttest and follow-up assessment ($M = 12.8$, $SD = 8.4$) was not statistically significant, $F(1, 101) < 1$. In addition, a significant main effect of Gender emerged, $F(1, 101) = 10.4$, $p < .01$, partial $\eta^2 = .09$. This main effect was qualified by a significant Subscale x Gender interaction, $F(1, 101) = 7.5$, $p < .01$, $\eta^2 = .07$, indicating that on the reactive subscale boys ($M = 12.3$, $SD = 6.5$) scored higher than girls ($M = 2.2$, $SD = 2.8$), while no such difference emerged on the proactive subscale. None of the other main or interaction effects reached significance.

Relation between implicit and explicit self-esteem

Unexpectedly, the IAT and the NLPT showed no significant correlation on either the pretest ($r = .03$), posttest ($r = .07$), or follow-up assessments ($r = .15$). The IAT was also not significantly related to the SPPC at pretest ($r = -.14$) or follow-up ($r = .05$), but showed an unexpected significant negative correlation with the SPPC at the posttest assessment ($r = -.24$, $p < .05$). The NLT was neither significantly related to the SPPC at pretest ($r = -.02$) or follow-up ($r = .05$), but showed a weak negative correlation with the SPPC at posttest ($r = -.17$, $p = .09$).

Relation between self-esteem and aggression

There were no significant relations at the pretest, posttest or follow-up assessments between the AQY reactive subscale and the IAT (r 's being .06, .04, and .04, respectively), the NLPT (r 's being -.09, -.05, and .06, respectively), or the SPPC (r 's being -.05, -.02, and -.04, respectively). There were also no significant relations at the pretest, posttest or follow-up assessments between the AQY proactive subscale and the IAT (r 's being .08, .04, and -.06, respectively), and the SPPC (r 's being -.01, -.09, and -.01, respectively). The relation between the NLPT and the AQY proactive subscale was inconsistent: At pretest the relation was significantly negative ($r = -.24$, $p < .05$), at posttest it was not significant

($r = .01$), while at follow-up it was marginally significantly positive ($r = .17$, $p = .09$).

Experiment 3

Method

Participants

A total of 151 primary school children (66 boys, 85 girls) and their parents consented to participate. Children were aged between 9 and 13 years ($M = 11.1$, $SD = 0.77$). There were no differences regarding gender, $\chi^2(1, N = 151) < 1$, and age, $t(149) < 1$, between the experimental and the control condition.

Materials

Explicit self-esteem. Similar to Experiment 2 the global self-esteem subscale of the SPPC was used as an index of explicit self-esteem. The internal consistency of this scale was good with alphas of .78, .84, and .90 at the pretest, posttest and follow-up assessments respectively. Test-retest stability in the control group was high from pretest to posttest ($r = .79$, $p < .01$) and from posttest to follow-up ($r = .82$, $p < .01$).

Implicit self-esteem. To assess implicit self-esteem we employed an IAT and an NLPT for which procedures were identical to Experiment 2. Spearman-Brown corrected split-half reliability of the IAT varied across the pretest, posttest and follow-up assessments (r 's being .68, .48, and .35, respectively). Test-retest correlations in the control group were significant both from pretest to posttest, $r = .26$, $p < .05$, and from posttest to follow-up, $r = .44$, $p < .01$. For various reasons (e.g., technical problems, illness, school obligations) data of one or more IAT assessments were missing for 17 children (eight in the experimental and nine in the control group). In addition, data of children who made more than 25% errors or who were more than 3 standard deviations removed from the group mean were excluded from the analysis. All in all, the final N for the analysis was 106 (53 in each group).

The NLPT effect was calculated in the same way as in Experiment 2. Data on one or more NLPT assessments were missing for seven children (four in the experimental and three in the control group). In addition, data that were more than 3 standard deviations removed from the group mean were excluded (at follow-up one boy in the control and one boy in the experimental group). The final N for the data analysis of this test was 136 (71 in the experimental and 65 in the control condition). Test-retest correlations in the control group were significant both from pretest to posttest, $r = .43$, $p < .01$, and from posttest to follow-up, $r = .40$, $p < .01$. Correlations between the name letter effects for

the first and last name were low but significant for the pretest ($r = .18, p < .05$) and the follow-up ($r = .27, p < .01$), but not significant for the posttest ($r = .13, p > .16$).

Psychological functioning. To assess general psychological functioning we employed the Strengths and Difficulties Questionnaire (SDQ; Goodman, 2001). To optimise reliability of the assessments we asked children as well as their parents to fill in this questionnaire, employing the child as well as the parent version of the SDQ. The SDQ is a 25-item self-report measure covering the most important domains of child psychopathology as well as personal strengths. The items describe positive or negative attributes of children and adolescents that have to be scored on a three-point scale (0 = 'not true', 1 = 'somewhat true', 2 = 'certainly true'). The items can be allocated to five subscales of five items each pertaining to emotional symptoms, conduct problems, hyperactivity-inattention, peer problems, and prosocial behaviour. In the present study, however, we only used the total difficulties score, which is obtained by summing the scores on all scales with the exception of the prosocial behavior subscale (range 0-40). The psychometric properties of

Dutch version of the SDQ are satisfactory (Muris, Meesters, & Van den Berg, 2003b). Missing SDQ data were imputed according to the corrected item mean substitution procedure as proposed by Huisman (1999). The parent reports of two children in the control group (one at the pretest assessment and one at the follow-up assessment) were excluded because there were more than two missing answers. The internal consistency of the SDQ total difficulties score in the present sample was good with alphas of .77 and .79 at the pretest assessment and .79 and .78 at the follow-up assessment, for the child and parent version respectively. Test-retest stability in the control group was high, and this appeared true for the self-report ($r = .82, p < .01$) and the parent-report versions of the scale ($r = .73, p < .01$).

Conditioning procedure. The conditioning task was identical to the one used in Experiments 1 and 2.

Procedure

Similar to Experiment 2 all children were tested in small groups with a maximum of four children in a separate room at school. During all assessments children first received a brief introduction and then consecutively completed the IAT, NLPT, and SPPC. Children completed the first SDQ during the pretest assessment. Immediately after completing the pretest assessment participants completed the conditioning procedure for the first time. After one week children completed the conditioning procedure for the second time. After another week children completed the conditioning procedure for the third and final time, immediately followed by the posttest assessment. One month later the experimenter returned for the follow-up assessment, during which children also completed the SDQ for the second time. Parents had received the first SDQ in combination with the informed consent form prior to the start of the experiment and returned it in a closed envelope to the experimenter. They received the second SDQ via their children and were asked to return the completed SDQ when their children received the follow-up assessment.

Results

Descriptive statistics for children in both conditions at various assessments are shown in Table 3.

Explicit self-esteem

A 2 (Condition) x 2 (Gender) x 3 (Assessment) ANOVA on children's SPPC scores showed that none of the main or interaction effects was significant (all F s $< 2.0, p > .13$). Thus, SPPC scores remained stable over the three assessments, and were independent of Condition and Gender.

Table 3 Mean scores (standard deviations) on various measures of self-esteem and psychological functioning in both conditions obtained at the pretest, posttest, and follow-up assessments in Experiment 3

		Condition	
		Experimental (n = 76)	Control (n = 75)
SPCC	Pre	19.6 (2.6)	19.3 (3.1)
	Post	19.3 (2.9)	19.3 (3.0)
	Fu	19.7 (3.1)	19.5 (3.7)
IAT	Pre	0.38 (0.42)	0.57 (0.37)
	Post	0.55 (0.32)	0.60 (0.36)
	Fu	0.55 (0.35)	0.66 (0.27)
NLPT	Pre	0.84 (0.75)	0.83 (0.69)
	Post	0.78 (0.76)	0.58 (0.76)
	Fu	0.71 (0.79)	0.74 (0.74)
SDQ-C	Pre	7.7 (5.2)	9.0 (5.4)
	Fu	6.8 (4.9)	7.5 (5.0)
SDQ-P	Pre	5.1 (3.7)	5.7 (4.8)
	Fu	4.2 (3.9)	5.1 (4.2)

SPPC = Self Perception Profile for Children; IAT = Implicit Association Test (positive scores indicate positive self-esteem); NLPT = Name Letter Preference Task (positive scores indicate positive self-esteem); SDQ-C/P = Strengths and Difficulties Questionnaire, Child and Parent version, respectively; FU = follow-up. N's vary across variables and assessments due to missing and/or excluded data; reported means are based on the data included in the analyses.

Implicit self-esteem

IAT. Children's IAT scores were subjected to a 2 (Condition: experimental vs. control) x 2 (Gender: girls vs. boys) x 3 (Assessment: pretest, posttest, follow-up) ANOVA, with the last factor being a repeated measure. The analysis showed a significant main effect of Assessment, $F(2, 101) = 3.8, p < .05$, partial $\eta^2 = .07$. Within-subject contrasts showed that the difference between the pretest ($M = .47, SD = .40$) and posttest ($M = .58, SD = .34$) assessment was significant, $F(1, 102) = 5.6, p < .05$, partial $\eta^2 = .04$, whereas the difference between the posttest and the follow-up ($M = .61, SD = .32$) assessment was not, $F(1, 100) < 1$. In addition, a significant main effect of Condition emerged, $F(1, 102) = 4.6, p < .05$, partial $\eta^2 = .04$: children in the control condition generally showed higher IAT scores than children in the experimental condition. Finally, a marginally significant main effect of Gender emerged, $F(1, 102) = 3.7, p = .059$, partial $\eta^2 = .04$, indicating that boys tended to have higher IAT scores than girls. None of the other main or interaction effects reached the conventional level of significance (all $F_s < 2.0, p > .15$).

NLPT. A 2 (Condition) x 2 (Gender) x 3 (Assessment) ANOVA on the NLPT effects revealed that none of the main or interaction effects were significant (all $F_s < 2.3, p > .1$), indicating that the NLPT effect did not change significantly over the three assessments, and was independent of Condition and Gender.

Psychological functioning

To assess changes in general psychological functioning, SDQ total difficulties scores of the children and their parents obtained during the pretest and follow-up assessment were subjected to two separate 2 (Condition: experimental vs. control) x 2 (Gender: girls vs. boys) x 2 (Assessment: pretest vs. follow-up) ANOVAs, with repeated measures on the last factor. Analysis of the child data showed a significant main effect of Assessment, $F(1, 140) = 24.4, p < .01, \eta^2 = .15$, indicating a decline in SDQ total difficulties scores from the pretest to the follow-up assessment. In addition, a significant main effect of Gender emerged, $F(1, 140) = 5.6, p < .05, \eta^2 = .04$: the SDQ total difficulties score of girls was lower than that of boys. Analysis of the parent data only yielded a significant main effect of Assessment, $F(1, 116) = 5.1, p < .05, \eta^2 = .04$, indicating a decline in the SDQ total difficulties scores from the pretest to the follow-up assessment.

Relation between implicit and explicit self-esteem

In line with the previous experiment, the IAT and the NLPT showed no significant correlation

on either the pretest ($r = .11$), posttest ($r = .02$), or follow-up assessments ($r = .12$). The IAT was also not significantly related to the SPPC at pretest ($r = .09$), posttest ($r = -.01$), or follow-up ($r = .01$). In a similar vein, the NLPT was not significantly related to the SPPC at any of the assessments (r 's being $-.03, -.11$, and $-.04$, respectively).

Relation between self-esteem and psychological functioning

Higher scores on the SDQ, reflecting higher levels of difficulties, were associated with lower levels of self-reported self-esteem on the SPPC. This was true for parent as well as self-reported difficulties at the pretest (r 's being $-.29, p < .01$, and $-.44, p < .01$, respectively) and the follow-up assessment (r 's $-.20, p < .01$, and $-.41, p < .01$, respectively). No significant correlations emerged on any of the assessments between either parent or self-reported difficulties and the implicit self-esteem measures, with the exception of a weak relation between self-reported difficulties and the IAT score at pretest ($r = .18, p < .05$) suggesting that higher IAT scores were related to more self-reported difficulties.

Discussion

The main results of Experiments 2 and 3 can be summarised as follows. First, no specific enhancement effect of the experimental conditioning procedure emerged. That is, in Experiments 2 and 3 children in the experimental and control condition showed a similar increase in implicit self-esteem from the pretest to the posttest as indicated by the IAT. In addition, in Experiment 2 but not Experiment 3, a similar pattern of results emerged for implicit self-esteem as indicated by the NLPT. Second, in both experiments explicit self-esteem as measured by the SPPC did not change as a result of the experimental manipulation. Third, in contrast with predictions, self-reported aggression as indicated by the AQY increased from the pretest to the posttest assessment. Finally, in Experiment 2 various measures of self-esteem and aggression showed no meaningful correlations with each other. In Experiment 3, explicit but not implicit self-esteem was significantly related to psychological functioning.

Although implicit self-esteem did show an increase from the pretest to the posttest assessment on both the IAT and the NLPT in Experiment 2 and on the IAT in Experiment 3, this increase was independent of the experimental condition. One possible explanation might be that the additional attention that participants received during the experiments was enough to increase implicit self-esteem for all children. The picture is not entirely consistent though, as in Experiment

3 the NLPT remained stable from the pretest to the posttest assessment. Nevertheless, it seems safe to conclude that any changes in implicit self-esteem were not specifically due to the pairing of self-relevant stimuli with smiling faces. Moreover, completing the conditioning procedure repeatedly did not seem to have any additional enhancement effects as compared with completing the procedure only once. The findings of Experiments 2 and 3 are also informative with respect to the potential correlates of (implicit) self-esteem in children. In Experiment 2 no meaningful relations were found between a self-report measure of aggression and measures of implicit and explicit self-esteem. This finding casts serious doubt on the idea that self-esteem is related to aggression in this age group. In addition, Experiment 3 showed that explicit self-esteem but not implicit self-esteem was related to both child and parent reports of children's psychosocial difficulties. Possible implications of these findings will be discussed in the general discussion.

General discussion

The major purpose of the present experiments was to take a first step in filling the gap in our knowledge concerning implicit self-esteem in children. In a series of three experiments we explored the malleability of self-esteem using the conditioning paradigm of Baccus and colleagues (2004), the relation between measures of implicit and explicit self-esteem, and their relation with

measures of aggression and general psychological functioning in various samples of primary school children. A summary of the major findings of these experiments is presented in Table 4, and will be discussed below.

Effect of conditioning on implicit and explicit self-esteem

Although the results of Experiment 1 tentatively indicated that the conditioning procedure might be able to enhance children's implicit self-esteem, and additionally seemed to have an effect on aggression, these findings could not be replicated in Experiments 2 and 3. Note in passing that in two other unreported experiments, which were identical to Experiment 1 with the exception that in one experiment child faces and in the other smiley's were used instead of adult faces in the condition task, no effects of the conditioning procedure were found either. This strongly suggests that the pairing of self-relevant information with smiling faces has no specific effect on children's implicit or explicit self-esteem, and that the trend towards group differences found in Experiment 1 either reflects a spurious finding or possibly was the result of pre-existing differences between the experimental and the control group. This raises the question why the currently used experimental conditioning procedure was not able to enhance implicit self-esteem in children, while it was successfully employed for this purpose in a sample of adults (Baccus et al., 2004). There are a number

Table 4 Summary of main results of Experiments 1-3

Exp	Δ ISE	Δ ESE	r I-E SE		SE-Aggression / SDQ				
			RSES	SPPC	SPPC	IAT			
1	.005 (n = 131).	.023 [†] (n = 125)							
				IAT	NLPT		SPPC	IAT	NLPT
2	.002 (n = 86) ^b	.012 (n = 104) ^b	Pre	-.14	-.01	Pre	-.04	.08	-.15
			Post	-.24*	.17 [†]	Post	-.05	.05	-.03
			FU	.05	.05	FU	-.04	.02	-.11
3	.009 (n = 106) ^b	.009 (n = 140) ^{a, b}	Pre	.09	-.03	Pre _{ch}	-.44**	.18*	.07
			Post	-.01	-.11	Pre _{pa}	-.29*	-.16 [†]	-.03
			FU	.01	-.04	FU _{ch}	-.41**	.12	.06
						FU _{pa}	-.20*	-.16	.05

Δ = effect sizes (partial η^2) for posttest differences (Exp. 1)/changes in implicit self-esteem (ISE) or explicit self-esteem (ESE)(Exp. 2, 3); r I-E SE = correlations between implicit and explicit self-esteem measures; r SE-Aggression/SDQ = correlations between implicit and explicit self-esteem measures and the Aggression measure (Hot Sauce test in Exp. 1, and the total score of the Aggression Questionnaire for Youths in Exp. 2) or the Strength and Difficulties Questionnaire (Exp. 3); RSES = Rosenberg Self-Esteem Scale; SPPC = Self-Perception Profile for Children; IAT = Implicit Association Test; NLPT = Name-Letter Preference Task; _{ch} = child report; FU = follow-up; _{pa} = parent report; [†] $p \leq .1$; * $p < .05$; ** $p < .01$; ^a In unexpected direction; ^b Based on Time x Group contrast for pretest to posttest.

of possible explanations for these differential results. First, the lack of effects may be the result of the adaptations we made to the procedure. As mentioned earlier, in the Baccus et al. study self-relevant stimuli were always followed by smiling faces, while other-relevant stimuli were always followed by negative or neutral faces. It is unclear whether the reported effects by Baccus et al. were due to a strengthening of *positive* associations with the self, a strengthening of *negative* associations with others, or a combination of both. To make sure that any effects would be caused by a strengthening of *positive* associations with the self we adjusted the procedure so that other-relevant stimuli were equally often followed by positive, negative or neutral faces. However, as a result of this adaptation the contrast between self and other-relevant stimuli is less sharp which may have reduced the effects of the procedure. If true, the results suggest that the effects found by Baccus et al. were mainly caused by a strengthening of negative associations with other, as the data seem to indicate that this procedure is not effective in strengthening *positive* associations with the self (i.e., increase self-esteem).

Another explanation is that any effects of the experimental conditioning procedure were overshadowed by a more general positive effect of an aspecific factor such as increased attention during the experiment. It may well be that such a factor has quite some influence on children, whereas it may have little impact on adults. In line with such an explanation, the results of Experiments 2 and 3 showed that implicit self-esteem was enhanced for all children, independent of the type of conditioning procedure they completed (i.e., experimental or control). Importantly, the pattern of IAT results suggests that this enhancement of implicit self-esteem cannot simply be attributed to, for instance, test-retest effects. That is, with repeated practice participants learn to overcome the crucial interference during the various phases of the IAT. As a result, retest effects on the IAT generally lead to smaller effects. In the present experiments, however, the IAT effects became larger over time. This result was consistent over Experiments 2 and 3, suggesting that it did not reflect a spurious finding. If the effect of the experimental procedure was indeed overshadowed by other factors, conditioning, at least in children, does not seem to be a very efficient way of increasing implicit self-esteem. Related to this, the implicit self-esteem of children in this age group seemed to be fairly positive so that it may be difficult to achieve strong enhancement effects due to a ceiling effect. Adolescents and adults, in contrast, may have been more exposed to (negative) experiences which

results in greater variation of implicit self-esteem. Such an explanation would be in line with the idea that middle childhood is a relatively stable period just before the turbulent stage of adolescence, during which self-esteem typically decreases (e.g., Robins et al., 2002). If this line of reasoning is correct, then it would be interesting to explore at what age the implicit self-esteem of children/adolescents starts to change. In addition, it might be interesting to see whether the implicit self-esteem of a preselected sample of 'at risk' children or children with current psychopathology that are likely to display relatively low levels of implicit self-esteem can be increased using the evaluative conditioning procedure.

While implicit self-esteem seemed to be affected similarly by the experimental and control conditioning procedures in Experiments 2 and 3, there was no effect of the conditioning procedure on explicit self-esteem in any of the experiments, which is in line with previous findings in adults (Baccus et al., 2004). One explanation for the differential effects on implicit and explicit self-esteem might be that measures of implicit self-esteem are more sensitive to temporal changes in self-evaluative associations than a self-report measure like the SPPC. That is, the SPPC subscale used in all experiments asks children how they generally evaluate themselves, rather than how they view themselves at that very moment. For implicit measures, on the other hand, it is known that they are at least partly context dependent (see for instance Blair, 2002 for a review). Following this, the implicit measures may have been more influenced by any additional positive reinforcement and attention that the children received while participating in the experiment than the SPPC.

For exploratory reasons the factor gender was included in all experiments. No gender differences were found with respect to implicit self-esteem in any of the experiments. Thus, although gender role and stereotypes develop early in childhood such differences apparently do not (yet) differentially influence implicit self-esteem in primary school children. Although previous studies indicate that there appears to be a small but consistent gender difference in explicit self-esteem, with males showing higher self-esteem than females (e.g., Kling, et al., 1999; Veerman et al., 1997), such a pattern did not emerge consistently in the present experiments. Meanwhile, mean scores of boys were higher than those of girls.

Relation between measures of implicit and explicit self-esteem

No consistent meaningful relations emerged between the implicit and explicit self-esteem measures. In the adult self-esteem literature meta-analyses found weak but positive mean population estimates for the relation between self-report measures and the IAT (e.g., Hoffman et al., 2005) and the NLPT (Krizan & Suls, 2008). One factor that may have suppressed the correlations in the present experiments is that the IAT always preceded the self-reports (cf. Hoffman et al., 2005; Krizan & Suls, 2008). It seems a consistent finding, however, that the relation between implicit and explicit measures of self-esteem is not strong, neither in children nor in adults. It is not clear what to make of this finding. Some authors interpret this finding as support for the idea that both types of measures reflect different underlying constructs (e.g., Baccus et al., 2004; Franck et al., 2007). Other authors, however, argue that implicit and explicit self-esteem measures tap the same underlying construct and contend that the observed low correlations are due to a lack of reliability of implicit self-esteem measures (e.g., Dijksterhuis et al., 2007). Although the present series of experiments were not designed to address this fundamental issue, it is clear that the reliability of the currently used measures of implicit self-esteem deserves some discussion. For the IAT split-half reliabilities were on average $r = .63$ at the first assessment but lower at subsequent assessments at one week and one month follow-up (around $r = .44$ for the second and third assessment over two experiments). Although these numbers are somewhat lower than what has been reported for adult samples they are still a lot higher than those reported for other latency-based measures in adult samples (e.g., Nosek, Greenwald, & Banaji, 2007). The same is true for the test-retest reliabilities (Bosson, Swann, & Pennebaker, 2000; Nosek et al., 2007). One explanation for the lower reliabilities in these child populations may be that children find the task more difficult than adults, for instance due to a shorter attention span. However, over the first three experiments approximately 4% of the children made more than 25% errors, suggesting that the task was not too difficult. All in all, compared with other latency based measures in adults this version of the IAT performed pretty well. The same cannot be said about the NLPT for which correlations between effects based on the initials of the first and last name ranged between .00 and .27, with a mean correlation of .12. This extremely poor internal consistency limits the maximally possible correlation with other measures to such an extent that significant correlations can hardly be expected. In contrast, the NLPT seems to have a

modest degree of internal consistency in adults (see Bosson et al., 2000). This suggests that the NLPT is not an appropriate test for children in this age range, possibly because children have had much less exposure to their name letters than adults.

Relation between self-esteem and aggression / psychological functioning

The current findings seem to suggest that we can exclude aggression and general psychological functioning from the list of potentially relevant correlates of implicit self-esteem in children. However, similar to what has been mentioned above, the observed low correlations with the NLPT are certainly due to the very poor reliability of this measure, and also correlations with the IAT may have been suppressed due to limited reliability. In addition, the present experiments included unselected samples that predominantly consisted of healthy functioning children. In line with this, implicit self-esteem was generally rather positive, while (self-reported) aggression and difficulties were fairly low. Therefore, it remains possible that significant relations between implicit self-esteem and aggression and other psychological difficulties can be found for instance in samples of children with conduct or psychological disorders. In addition, it also remains possible that implicit self-esteem plays a more prominent role in psychological functioning in children at an older age.

Explicit self-esteem on the other hand was clearly related to general psychological functioning, such that lower levels of self-reported self-esteem were associated with higher levels of difficulties as reported by the children themselves as well as their parents. This finding is in line with other studies that have observed a relation between low explicit self-esteem and a range of negative consequences (e.g., Bos et al., 2006; Harter, 1993; Muris et al., 2003a; Muris et al., 2005). The relation between explicit self-esteem and the aggression measures was low but consistently negative. The direction of this relation is in line with adult studies suggesting that lower levels of self-esteem are associated with higher levels of aggression (Donnelan et al., 2005).

Issues for future research

So what can we learn from these studies? With respect to the question whether pairing of self-relevant information with smiling faces can lead to increases in self-esteem, the present experiments suggest that this is not the case. Given that the effect sizes were very small at best, the lack of effects does not seem to reflect a simple power problem. Furthermore, even though the reliability

of the implicit measures was certainly not high, the reliability of the IAT was much better than has been reported for other implicit measures in adult samples, and hence should be high enough to detect meaningful group differences. Thus, the conditioning procedure did not seem to work as expected. Meanwhile, the possibility cannot be ruled out that this result was due to the use of this specific experimental procedure. Perhaps alternative procedures, such as the subliminal conditioning procedure that was successfully used by Dijksterhuis (2004) to enhance implicit self-esteem in students, may be more successful. Although we cannot think of any *a priori* reason why another evaluative conditioning procedure would be more effective, this remains a potential lead for further research.

With respect to children's implicit self-esteem it is clear that a lot of work remains to be done. As a first step to explore children's implicit self-esteem it seemed logical to employ measures that have been successfully employed in adults, and this seemed particularly true for the IAT as this procedure has been successfully employed in earlier studies of children (e.g., Baron & Banaji, 2006; Field & Lawson, 2003). From this first exploration of implicit self-esteem in children it has become clear, however, that a major challenge for researchers will be to improve the measures in terms of reliability and validity. We hope that the

present findings will be a useful starting point for finding ways to improve measures of implicit self-esteem in children.

Another area where a lot of exciting questions await future research concerns the origins and development of implicit self-esteem. Although the current findings tentatively suggest that implicit self-esteem is generally positive in middle childhood, there are data that suggest that the origins of adult implicit self-esteem lie in youth. DeHart, Pelham, & Tennen (2006) found that college students' as well as their mother's retrospective reports of the parenting style, were related to current levels of implicit self-esteem. A challenge for future research would be to further explore this issue adopting a more developmental approach, for instance by longitudinal studies in which the development of children's implicit and explicit self-esteem is monitored and the relation with symptoms of current and later psychopathology can be explored.

Taken together, this first exploration of implicit self-esteem in children reveals a lot of questions and challenges for researchers in this field. Nevertheless we feel safe to conclude that, apparently, most children between the ages of 10 and 13, do not need an additional smile to make them feel worthwhile.

References

- Baccus, J. R., Baldwin, M. W., & Packer, D. J. (2004). Increasing implicit self-esteem through classical conditioning. *Psychological Science, 15*, 498-502.
- Baron, A. S., & Banaji, M. A. (2006). The development of implicit attitudes: evidence of race evaluations from ages 6 to 10 and adulthood. *Psychological Science, 17*, 53-58.
- Baumeister, R. F., Smart, L., & Boden, J. M. (1996). Relation of threatened egotism to violence and aggression: The dark side of high self-esteem. *Psychological Review, 103*, 5-33.
- Beck, A. T., Brown, G. K., Steer, R. A., Kuyken, W., & Grisham, J. (2001). Psychometric properties of the Beck Self-Esteem Scales. *Behavior Research and Therapy, 39*, 115-124.
- Blair, I. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review, 6*, 242-261.
- Bos, A. E. R., Muris, P., Mulkens, S., & Schaalma, H. P. (2006). Changing self-esteem in children and adolescents: A roadmap for future interventions. *Netherlands Journal of Psychology, 62*, 26-33.
- Bosson, J. K., Swann, W. B., & Pennebaker, J. W. (2000). Stalking the perfect measure of implicit self-esteem: The blind man and the elephant revisited. *Journal of Personality and Social Psychology, 79*, 631-643.
- Butler, R. J., & Gasson, S. L. (2005). Self-esteem/self-concept scales for children and adolescents: A review. *Child and Adolescent Mental Health, 10*, 190-201.
- Conner, T., & Barret, L. F. (2005). Implicit self-attitudes predict spontaneous affect in daily life. *Emotion, 5*, 476-488.
- de Jong, P. J. (2002). Implicit self-esteem and social anxiety: Differential self-favouring effects in high and low anxious individuals. *Behaviour Research and Therapy, 40*, 501-508.
- DeHart, T., Pelham, B. W., & Tennen, H. (2006). What lies beneath: Parenting style and implicit self-esteem. *Journal of Experimental Social Psychology, 42*, 1-17.
- Dijksterhuis, A. (2004). I like myself but I don't know why: Enhancing implicit self-esteem by evaluative conditioning. *Journal of Personality and Social Psychology, 86*, 345-355.
- Dijksterhuis, A. (2006). The emergence of implicit self-esteem. *Netherlands Journal of Psychology, 62*, 19-25.
- Dijksterhuis, A., Albers, L. W., & Bongers, K. C. A. (2007). Digging for the real attitude: Lessons from research on implicit and explicit self-esteem. In R. E. Petty, R. H. Fazio & P. Brinol (Eds.), *Attitudes: Insights From the New Wave of Implicit Measures*. Erlbaum.
- Donnellan, M. B., Trzesniewski, K. H., Robins, R. W., Moffitt, T. E., & Caspi, A. (2005). Low self-esteem is related to aggression, antisocial behaviour, and delinquency. *Psychological Science, 16*, 328-335.

- Field, A. P., & Lawson, J. (2003). Fear information and the development of fears during childhood: Effects on implicit fear responses and behavioural avoidance. *Behaviour Research and Therapy*, 41, 1277-1293.
- Franck, E., De Raedt, R., & De Houwer, J. (2007). Implicit but not explicit self-esteem predicts future depressive symptomatology. *Behaviour Research and Therapy*, 45, 2448-2455.
- Goodman, R. (2001). Psychometric properties of the Strengths and Difficulties Questionnaire. *Journal of the Academy of Child and Adolescent Psychiatry*, 40, 1337-1345.
- Greenwald, A. G., & Banaji, M. R. (1995). Implicit Social Cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, 102, 4-27.
- Greenwald, A. G., McGhee, D. E., & Schwarz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74, 1464-1480.
- 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., 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.
- Haney, P., & Durlak, J. A. (1998). Changing self-esteem in children and adolescents: A meta-analytic review. *Journal of Clinical Child Psychology*, 27, 423-433.
- Harter, S. (1985). *Manual for the Self-Perception Profile for Children*. Denver: University of Denver.
- Harter, S. (1993). Causes and consequences of low self-esteem in children and adolescents. In R. F. Baumeister (Ed.), *Self-Esteem: The puzzle of low self-regard* (pp. 87-116). New York: Plenum.
- Harter, S. (1999). *The Construction of the Self: A Developmental Perspective*. The Guilford Press: New York.
- Hoffman, W., Gawronski, B., Geschwendner, 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.
- Huisman, M. (1999). *Item Nonresponse: Occurrence, Causes, and Imputation of Missing Answers to Test Items*. (M&T series 32). Leiden: DSWO Press.
- Kling, K. C., Hyde, J. S., Showers, C. J., & Buswell, B. N. (1999). Gender differences in self-esteem: A meta-analysis. *Psychological Bulletin*, 125, 470-500.
- Koole, S. L., & DeHart, T. (2007). Self-affection without self-reflection: Origins, models, and consequences of implicit self-esteem. In C. Sedikides & S. Spencer (Eds.), *The Self in Social Psychology* (pp. 36-86). New York: Psychology Press.
- Koole, S., 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.
- Krizan, Z., & Suls, J. (2008). Are implicit and explicit measures of self-esteem related: A meta-analysis for the Name-Letter Test. *Personality and Individual Differences*, 44, 521-531.
- Lieberman, J. D., Solomon, S., Greenberg, J., & McGregor, H. A. (1999). A hot new way to measure aggression: hot sauce allocation. *Aggressive Behavior*, 25, 331-348.
- Mann, M., Hosman, C. M. H., Schaalma, H. P., & De Vries, N. K. (2004). Self-esteem in a broad-spectrum approach for mental health promotion. *Health Education Research*, 19, 357-372.
- Meesters, C., Muris, P., Van Rooijen, B. (2007). Relations of neuroticism and attentional control with symptoms of anxiety and aggression in non-clinical children. *Journal of Psychopathological Behavior Assessment*, 29, 149-158.
- Muris, P., Meesters, C., Van de Blom, W., & Mayer, B. (2005). Biological, psychological, and cultural correlates of body change strategies and eating problems in adolescent boys and girls. *Eating Behaviours*, 6, 11-22.
- Muris, P., Meesters, C., & Fijen, P. (2003a). The Self Perception Profile for Children: Further evidence for its factor structure, reliability and validity. *Personality and Individual Differences*, 35, 1791-1802.
- Muris, P., Meesters, C., & Van den Berg, F. (2003b). The Strengths and Difficulties Questionnaire (SDQ): Further evidence for its reliability and validity in a community sample of Dutch children and adolescents. *European Child and Adolescent Psychiatry*, 12, 1-8.
- Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2007). The Implicit Association Test at age 7: A methodological and conceptual review. In: J. A. Bargh (Ed.), *Automatic Processes in Social Thinking and Behavior*. New York: Psychology Press.
- Nuttin, J. M. (1985). Narcissism beyond Gestalt and awareness: The name letter effect. *European Journal of Social Psychology*, 15, 353-361.
- Robins, R. W., Trzesniewski, K. H., Tracy, J. L., Gosling, S. D., & Potter, J. (2002). Global self-esteem across the lifespan. *Psychology and Aging*, 17, 423-434.
- Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.
- Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-prime User's Guide*. Pittsburg: Psychology Software Tools Inc.
- Spalding, L. R., & Hardin, L. R. (1999). Unconscious unease and self-handicapping: Behavioral consequences of individual differences in implicit and explicit self-esteem. *Psychological Science*, 10, 535-539.
- Steinberg, L., & Sheffield Morris, A. (2001). Adolescent development. *Annual Review of Psychology*, 52, 83-110.
- Stice, E. (2002). Risk and maintenance factors for eating pathology: A meta-analytic review. *Psychological Bulletin*, 128, 825-848.
- Tanner, R., Stopa, L., & De Houwer, J. (2006). Implicit views of the self in social anxiety. *Behaviour Research and Therapy*, 44, 1379-1409.
- Veerman, J. W., Straathof, M., Treffers, P., Van den Bergh, B., & Ten Brink, L. (1997). *Competentie belevingsschaal voor kinderen*. Lisse: Swets Test Services.