Journal of Personality and Social Psychology
2002, Vol. 83, No. 6, 1441–1455

0022-3514/02/$5.00 DOI: 10.1037//0022-3514.83.6.1441

Predictive Validity of an Implicit Association Test for Assessing Anxiety

Boris Egloff and Stefan C. Schmukle
Johannes Gutenberg-University Mainz

The Implicit Association Test (IAT) was adapted to measure anxiety by assessing associations of self (vs. other) with anxiety-related (vs. calmness-related) words. Study 1 showed that the IAT–Anxiety exhibited good internal consistency and adequate stability. Study 2 revealed that the IAT–Anxiety was unaffected by a faking instruction. Study 3 examined the predictive validity of implicit and explicit measures and showed that the IAT–Anxiety was related to changes in experimenter-rated anxiety and performance decrements after failure. Study 4 found that several behavioral indicators of anxiety during a stressful speech were predicted by the IAT. Taken together, these studies show that the IAT–Anxiety is a reliable measure that is able to predict criterion variables above questionnaire measures of anxiety and social desirability.

The easiest way to obtain information about a person's attitudes and traits is certainly to ask him or her to fill out a questionnaire. This assessment method relies on explicit self-reports of the respondent. Questionnaire measures can be administered economically and objectively, and they generally show good reliabilities. Furthermore, the validity of explicit measures has been demonstrated in several studies (Funder, 2001). Thus, they are usually the method of choice when personality dispositions are measured.

Unfortunately, two classes of factors are often confounded with participants' responses to questionnaire items and thus might diminish the validity of the test: *introspective limits* and *response factors* (Greenwald et al., 2002). Simply put, introspective limits refer to the ability of participants to report on the intended content domain: Because they lack awareness, participants might be unable to accurately indicate their true score—independent of their motivation or willingness to comply with the instruction. Response factors, in contrast, refer to the willingness to report on oneself. In the case of response factors, it is assumed that participants are aware of (and thus principally able to indicate) their standing on the dimension of interest but that factors such as demand characteristics (Orne, 1962), evaluation apprehension (Rosenberg, 1969), and faking (Cronbach, 1990) mask the self-report. Introspective limits and response factors can both work independently or in combination and always constitute a challenge for the validity of a test.[1]

Faking, for example, is a major threat to the validity of tests. It refers to conscious distortion to appear overly positive (faking good) or overly negative (faking bad). When testing is done as part of a selection process (e.g., for jobs or promotions) or for decision-making purposes (e.g., culpability in a criminal matter or examination of recruits), the probability of faking is increased. The extent to which conscious distortion is also present in standard psychological research is difficult to estimate (Viswesvaran & Ones, 1999). Unfortunately, efforts to cope with the problem of faking on questionnaire data have met with rather moderate success (Holden, Wood, & Tomashewski, 2001).

Introspective limits, on the other hand, refer to the impressive amount of evidence showing that individuals process information about themselves and their environment not only in an *explicit* (i.e., controlled or conscious) mode but also in an *implicit* (i.e., automatic or nonconscious) mode (Epstein, 1994; Fazio, 1990; Greenwald et al., 2002; Wilson, Lindsley, & Schooler, 2000). Thus, the respondent is not able to report on these cognitive and affective processes because they operate outside of subjective awareness (Nisbett & Wilson, 1977). Nevertheless, implicit pro-

This research was supported by Deutsche Forschungsgemeinschaft Grant EG 143/1-1. We thank Mitja Back, Sebastian Beigel, Sabine Otte, Eva Ringel, Sandra Siekmann, and Christine Volz for their help with data collection and behavior coding. We also thank Mitja Back, Jennifer Dickes, Michael Hock, Heinz Walter Krohne, Andreas Schwerdtfeger, and Florian Weck for helpful comments on a version of this article.

Correspondence concerning this article should be addressed to either Boris Egloff or Stefan C. Schmukle, Department of Psychology, Johannes Gutenberg-University Mainz, D-55099 Mainz, Germany. E-mail: egloff@mail.uni-mainz.de or schmukle@psych.uni-mainz.de

[1] Social desirability, defined as the tendency to portray oneself in a favorable light, constitutes another problematic error tendency in self-reports (Edwards, 1957). Consequently, this has led to the construction of social desirability scales to separate the tendency to give socially desirable responses from responsiveness to the intended content domain (e.g., Crowne & Marlowe, 1960; Paulhus, 1998; Stöber, 2001). Thus, one possible way to cope with the problem of socially desirable responses is to assess this tendency and then to statistically adjust the content test scores for social desirability. On the other hand, some researchers suggest that social desirability scales should be given substantive rather than artifactual interpretations because they claim that validity coefficients decrease rather than increase when self-reports are corrected for social desirability (Borkenau & Ostendorf, 1992; McCrae & Costa, 1983).

Independent of this controversy, we believe that the location of socially desirable responding in the suggested structure of unaware (introspective limits) and aware (response factors) error tendencies in self-reports is uncertain. This assumption is corroborated by factor analytic studies (Paulhus, 1984) that have suggested the structural partitioning of social desirability into two clusters, labeled *self-deception* (the tendency to give favorably biased but honestly held self-descriptions) and *impression management* (the tendency to give favorable self-descriptions to others). Thus, according to this structure, social desirability comprises both the willingness (impression management) and the ability (self-deception) to accurately describe oneself.

cesses are known to shape a wide variety of behavioral responses (Kihlstrom, 1987). Implicit self-esteem, for example, is defined as an automatic, overlearned, and nonconscious evaluation that guides spontaneous behavior in self-relevant situations (Greenwald & Banaji, 1995).

Measurement of Individual Differences in Implicit Cognition

The measurement of implicit personality constructs is necessarily of an indirect nature, because these constructs operate in an automatic (nonconscious) mode. Among the first attempts to assess implicit self-esteem and self-concept were projective measures like the Thematic Apperception Test (TAT; Murray, 1943). Another research tradition tried to gather information about the person's self-concept by means of cognitive measures, such as memory scores or response latencies (Cantor & Mischel, 1977; Markus, 1977; Rogers, Kuiper, & Kirker, 1977). However, these measures were only of limited sensitivity to individual differences (Greenwald & Farnham, 2000). Renewed attention to implicit measures in the last decade led to a variety of procedures for assessing implicit self-esteem (see Bosson, Swann, & Pennebaker, 2000, for an overview). Unfortunately, only a few of these measures show adequate internal consistency and stability (Bosson et al., 2000).

The Implicit Association Test

Recently, a promising new measurement tool for assessing implicit processes was introduced: the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998). The IAT measures strengths of associations between concepts by comparing response times on two combined discrimination tasks. Participants are required to sort stimuli representing four concepts using just two responses, each assigned to two of the four concepts. The basic assumption of the IAT is that if two concepts are highly associated, the sorting task will be easier (i.e. faster) when the two associated concepts share the same response key than when they share different response keys. For instance, researchers can assess implicit attitudes toward Blacks and Whites by combining the tasks of classifying Black versus White names and discriminating pleasant versus unpleasant word meanings. Individuals with implicit prejudices against Blacks should react more slowly when Black names and pleasant attributes (Black + pleasant) share the same response key, as compared with the reversed configuration (Black + unpleasant).

A recent important adaptation of the IAT was the development of a variant to measure personality traits such as self-esteem and self-concept (Greenwald & Farnham, 2000). The researchers did this by combining the task of categorizing pleasant versus unpleasant word meanings and classifying items (e.g., *my* or *they*) into self and other categories. An IAT measure of implicit self-esteem is computed as the difference in mean categorization latency when self and pleasant share the same response key (self + pleasant), as compared with self + unpleasant. The IAT effect thus measures how much easier it is for participants to categorize self items with pleasant items than self items with unpleasant items.

In the last few years, an impressive number of studies demonstrating reliability and validity of the IAT have accumulated (overview in Greenwald & Nosek, 2001). Initial research established that the IAT is able to detect valence differences that are associated with nonsocial objects (flowers vs. insects) and participant populations (Black and White races). Furthermore, the IAT effect was not influenced by procedural variations nor by outliers and error treatment (Greenwald et al., 1998). Variation in stimulus familiarity could also be ruled out as an alternative explanation for the observed IAT effects (Dasgupta, McGhee, Greenwald, & Banaji, 2000; Ottaway, Hayden, & Oakes, 2001). IAT measures and self-report measures of the same construct show a mixture of associations and dissociations (mean $r = .24$; Nosek, Banaji, & Greenwald, 2002a), depending on the domain assessed. High agreement was found in attitudes toward homosexuality (Banse, Seise, & Zerbes, 2001) and political candidates (Nosek et al., 2002a). In contrast, divergence between implicit and explicit measures was observed in attitudes toward gender (Greenwald & Farnham, 2000) and race (Greenwald et al., 1998). Concerning predictive validity, McConnell and Leibold (2001) showed that White participants' implicit racial attitudes were associated with nonverbal indicators of apparent discomfort during an interaction with an African American. Rudman and Glick (2001) found that prejudice against female job applicants was predicted by IAT-assessed gender stereotypes. Furthermore, implicit race prejudice was associated with amygdala activation as assessed by functional magnetic resonance imaging (fMRI) in White participants exposed to unfamiliar Black faces (Phelps et al., 2000).

IAT–Anxiety

Using the same rationale as was used by Greenwald and Farnham (2000) in their self-esteem IAT, we adapted the IAT to provide an implicit measure of the self-concept of anxiety. For this aim, the categorization into self and other categories was combined with the classification of items into anxiety and calmness categories. The IAT–Anxiety comprises a sequence of five blocks (see Figure 1 for an overview). Participants were instructed that they would be making a series of category judgments. On each trial, a stimulus word was presented in the center of a computer screen, and participants pressed a left or right key to categorize the stimulus as quickly and correctly as possible into one of the categories.

In Block 1, participants practiced the target concept discrimination by categorizing stimuli into self and other categories. In Block 2, they did the same for the attribute discrimination by sorting items into anxiety and calmness categories. Block 3 introduced the first of the two series of critical trials: Participants categorized items into two combined categories, each including the attribute and the target concept that were assigned to the same key in the preceding two blocks (e.g., self + anxiety for the left key, and other + calmness for the right key). In Block 4, key assignment for the attribute dimension was switched. Block 5 (the second critical block) was complementary to Block 3: The only difference was that it used the reversed key assignment that participants practiced in the previous block (e.g., self + calmness for the left key and other + anxiety for the right key). An IAT effect indicating implicit anxiety was then computed as the difference between

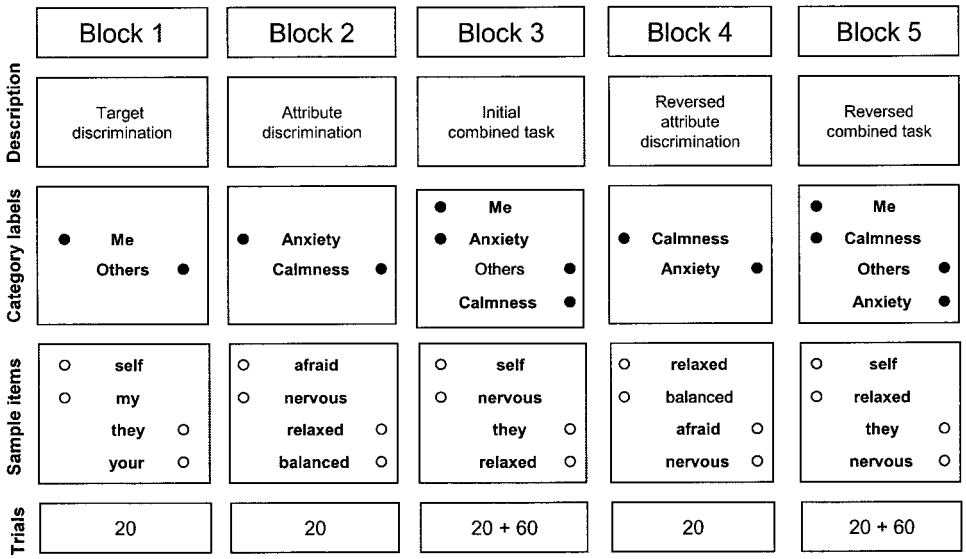

Figure 1. Illustration of the Implicit Association Test—Anxiety (IAT–Anxiety). The IAT–Anxiety procedure comprises five blocks. Blocks 1 and 2 are practice blocks: They introduce the discriminations of the target concepts and of the attributes. Category labels for these discriminations are assigned to a left or right response key, as indicated by the black circles. In the fourth row, sample items of each category are shown, with correct responses indicated as open circles. Block 3 (first critical block) combines target and attribute discriminations. In Block 4, the response assignment for the attribute discrimination is reversed, and in Block 5 (second critical block) the target and attribute discriminations are recombined. Both critical blocks (Blocks 3 and 5) comprise 20 practice trials and 60 critical trials. Adapted from "Measuring Individual Differences in Implicit Cognition: The Implicit Association Test," by A. G. Greenwald, D. E. McGhee, & J. L. K. Schwartz, 1998, *Journal of Personality and Social Psychology, 74,* Figure 1. Copyright 1998 by the American Psychological Association.

self + anxiety and self + calmness.[2] Thus, because it measures the relative ease of categorizing self items with anxiety items as compared with self items with calmness items, the IAT–Anxiety effect is an indicator of the implicit self-concept of anxiety.[3]

Goals of This Research

The problem of introspective limits and response factors in self-report data is—to a certain extent—inherent in every domain of personality. Individual differences in the tendency to become anxious are certainly no exception to this rule. On the contrary, discrepancies between reports on trait anxiety scales and actual behavioral and physiological responding early led researchers to question the predictive validity of self-report data in the field of anxiety measurement (Hodges, 1976; Levitt, 1967).

One way of coping with this problem was to suggest the joint consideration of scores in tests of trait anxiety and social desirability to distinguish truly low-anxious individuals (i.e., low anxiety, low social desirability) from so-called repressors (i.e., low anxiety, high social desirability; Weinberger, Schwartz, & Davidson, 1979). The underlying rationale of this approach was that (a) certain individuals are inaccurate in their responses to trait anxiety questionnaires, (b) these individuals could be detected by high scores on social desirability scales, and (c) thus, the use of social desirability scores in addition to anxiety scores would enhance the prediction of anxiety-related behaviors.

Another approach to deal with introspective limits and response factors in anxiety measurement is to use implicit tests such as the IAT. In this article, we present a series of studies that rose out of a research program that aims at establishing the IAT–Anxiety as a reliable and valid indirect measure of trait anxiety. Study 1 analyzes internal consistency and stability of the IAT–Anxiety because reliability of measurement is a necessary prerequisite for every test—whether implicit or explicit. According to their theo-

[2] Note that we did not counterbalance the order of Block 3 (self + anxiety) and Block 5 (self + calmness). We believe that—in an individual-differences perspective—the advantages of our procedure outweigh the disadvantages. Certainly a fixed presentation order does not permit an interpretation of the IAT score in absolute terms (e.g., "In general, it is easier to associate self with calmness than self with anxiety"), but it makes possible an optimal comparison between participants and thus generates an ordering according to the (relative) size of the IAT effect. Because all participants respond to the identical task, the measurement of individual differences is maximized. This should lead to higher validity coefficients.

[3] A reviewer raised the question of whether our IAT measures the construct of anxiety, because in the reviewer's opinion, our IAT could also be named IAT–Calmness. We agree with the latter (but not with the former). We assume that our IAT is an empirical indicator of a one-dimensional construct with bipolar opposites (anxiety vs. calmness). The name IAT–Anxiety represents one pole of this bipolar construct. This procedure is mirrored in explicit anxiety measures: For example, the STAI includes items representing calmness, such as "I am calm, cool, and collected." These items are reverse scored and then summed up with items measuring the anxiety pole (e.g., "I feel nervous and restless") to yield an empirical indicator of the anxiety construct.

retical foundation, implicit tests should not be affected by conscious distortion of the participant. To analyze this claim empirically, we had participants of Study 2 respond to the IAT–Anxiety with a faking instruction. Increments in predictive validity—as compared with explicit measures of trait anxiety and social desirability—would constitute, in our view, the most convincing argument for the wider use of implicit tests in general and the IAT in particular. For this reason, Studies 3 and 4 analyze whether the IAT–Anxiety is able to predict relevant criteria above questionnaire measures of trait anxiety and social desirability.

Study 1: Reliability of the IAT–Anxiety

This study examines internal consistency and stability of the IAT–Anxiety. Reliability of measurement is an important but often underestimated issue in research using variables derived from cognitive–experimental procedures (Bosson et al., 2000; Buchner & Wippich, 2000; Byrne & Holcomb, 1962). Every measure, whether implicit or explicit, whether predictor or criterion, should show adequate reliability as a necessary prerequisite for further consideration in research that is concerned with individual differences.

Fortunately, research with the IAT was concerned with the topic of reliability right from the beginning (Greenwald et al., 1998). Internal consistency coefficients (split half or Cronbach's alpha) of various variants of the IAT usually show values in the range of .80 (Banse et al., 2001; Bosson et al., 2000). To exhibit good internal consistency alone would be sufficient for a state measure. To be regarded as a trait measure, however, the measure must show stability over time in addition to internal consistency. We know of four studies that analyzed the IAT in a test–retest design: For a race IAT, Dasgupta and Greenwald (2001) reported a correlation of .65 (time interval = 1 day). Cunningham, Preacher, and Banaji (2001)—also using a race IAT—found a mean test–retest correlation of .32 (range between .16 and .51; mean time interval = 28 days). For a self-esteem IAT, Bosson et al. (2000) reported a test–retest correlation of .69 (time interval = 31 days), and Greenwald and Farnham (2000) found a correlation of .52 (time interval = 8 days). Taken together, stability coefficients of the IAT are lower than are internal consistency coefficients. On the other hand, average stability coefficients in the area of .60 (with the exception of Cunningham et al., 2001) are considerably higher than what is reported of other implicit measures (Bosson et al., 2000).

In addition to the examination of the IAT–Anxiety's reliability, this study analyzes the association of the IAT–Anxiety with self-report measures of trait anxiety and social desirability in an exploratory manner. Previous research has shown that the relation of the IAT to parallel explicit measures depends on the construct under investigation: In their overview on Web-based IATs, Nosek et al. (2002a) reported an average correlation of .24, with a range between .08 (age attitude) and .52 (political candidates). Banse et al. (2001) found for attitudes toward homosexuality an even higher association ($r = .62$). Concerning the association between implicit and explicit self-esteem, correlation coefficients of .21 (Bosson et al., 2000) and of .17 (Greenwald & Farnham, 2000; Nosek et al., 2002a) were reported.

Method

Participants. Forty-one introductory psychology students (33 women and 8 men) of Johannes Gutenberg-University Mainz participated in this study in exchange for research participation credit. Their average age was 22.1 years ($SD = 3.1$).

Procedure. As part of a larger project, the IAT–Anxiety was administered twice, with a time lag of 1 week. Participants responded to the IAT in individual experimental sessions. In Session 2, questionnaire measures of trait anxiety and social desirability were completed in addition to the IAT.[4]

Explicit measures. Trait anxiety was assessed by means of the Trait form of the State–Trait Anxiety Inventory (STAI; Spielberger, Gorsuch, & Lushene, 1970; for the German version, see Laux, Glanzmann, Schaffner, & Spielberger, 1981). This widely used questionnaire contains 20 items that assess enduring symptoms of anxiety on a 4-point Likert scale (1 = *almost never*, 4 = *almost always*). The revised form of the Social Desirability Scale-17 (SDS-17R; Stöber, 2001) was used to measure social desirability. This scale contains 16 items and provides an update of the approach introduced by Crowne and Marlowe (1960) by using socially desirable but infrequent or socially undesirable but frequent behaviors to which the respondent answers on a true-false format.

IAT. The IATs were administered on personal computers with the program FIAT for Windows 2.3 (Farnham, 1998). Stimuli from the self (e.g., *me*, *my*) and other (e.g., *they*, *your*) categories were presented, as well as stimuli from the anxiety (e.g., *nervous*, *afraid*) and calmness (e.g., *relaxed*, *balanced*) categories (see the Appendix for the complete stimulus set, which consisted of five items per category). The IAT procedure comprises five blocks (see Figure 1). Participants practiced the discrimination of self and other items (target discrimination) in the first block, which comprised 20 trials (each item was presented twice). They did the same for the attribute discrimination by sorting items into anxiety and calmness categories in Block 2 and by practicing the switched key assignment in Block 4 (20 items each). The critical Blocks 3 and 5 consisted of 20 practice trials and 60 critical trials. In these trials, participants categorized items into two combined categories, each including the attribute and the target concept that were assigned to the same key.

Participants were told they would be making a series of category judgments. On each trial, a stimulus word was displayed in the center of a computer screen. Category labels were displayed on the left and right sides of the screen. Participants used the letter *A* on the left side of the keyboard and the number *5* on the right-side numeric keypad for their responses. They were told, "Please try to be as accurate as possible, while also going as quickly as possible. If your selection is incorrect, you will see a red *X*. To continue to the next judgment, you must make the correct selection." Participants were told to keep their index fingers on the *A* and *5* keys

[4] It seems that there is an implicit convention (which was made explicit to us by a reviewer) in IAT studies to assess explicit measures prior to the IAT. In contrast to this procedure, in our Studies 1, 3, and 4, the IAT was assessed prior to explicit measures. We believe that presenting the IAT first has the advantage of smaller (if any) carry-over effects from one measure to the other because implicit measures are less subject to conscious engagement. Thus, working on the STAI first might have effects on the IAT through cognitions such as, "Okay, these 20 items were all concerned with anxiety. It seems that these researchers are investigating this topic. Let's see whether this reaction time task is also about anxiety." Thoughts like these might make the assessment of personality construct with the IAT less implicit. In contrast, we assume that working on the IAT first does not involve this kind of conscious reasoning. Anyway, Study 2 shows that there were no effects of presentation mode (STAI first vs. IAT first) on the means of both measures. We acknowledge that this finding does not exclude the possibility that presentation mode has an effect on predictive validity by changing the rank order of participants on each of the measures.

throughout the experiment to facilitate fast responding. An intertrial interval of 150 ms was used. The computer recorded elapsed time between the start of each stimulus presentation and the correct response. Mean latencies and error rates were displayed after each block. We computed the IAT effect for anxiety by subtracting the mean latency in the critical trials of Block 3 (self + anxiety) from the critical trials of Block 5 (self + calmness).

Data reduction. IAT data were treated in accordance with the procedure outlined by Greenwald et al. (1998): (a) The first two trials of each block of critical trials were dropped, (b) trials with latencies less than 300 ms or greater than 3,000 ms were recoded to 300 ms or 3,000 ms, respectively, and (c) the resulting values were log transformed before they were averaged. Throughout this article, we performed all analyses with log-transformed values. For presentation purposes, average IAT effects were reported in milliseconds. There were no participants with error rates that exceeded 20% or with mean latencies above 2,000 ms (participants with these values are usually dropped from analyses).

Results and Discussion

At the first measurement occasion, the IAT–Anxiety showed an average value of −112 ms (*SD* = 109 ms).[5] In Session 2, there was a mean of −82 ms (*SD* = 106 ms). To compute internal consistency, we first separately subtracted each trial's (log-transformed) response latency of the self + anxiety block from the response latency of the corresponding trial of the self + calmness block (first latency in Block 5 minus the first latency in Block 3, second latency in Block 5 minus the second latency in Block 3, etc.). We then computed Cronbach's alpha of these 58 difference scores (cf. Bosson et al., 2000). Alpha thus reflects the internal consistency in the tendency to associate anxiety-related stimuli—relative to calmness-related stimuli—with the self. Cronbach's alpha was .77 for the first occasion of measurement and .80 for second session. The stability (test–retest correlation) of the IAT–Anxiety was .57.

Self-reported trait anxiety (as measured by the STAI) showed a mean of 37.76 (*SD* = 8.25) and an internal consistency of .90. Social desirability (as measured by the SDS-17R) had a mean of 7.63 (*SD* = 3.26) and an alpha of .70. The IAT–Anxiety effect did not correlate significantly with explicit anxiety ($r = .24$, $p = .14$) or social desirability ($r = -.18$, $p = .26$).

Results of this initial study document that the IAT–Anxiety exhibited good internal consistency: Alphas in the range of .80 are comparable to consistency coefficients shown by questionnaire measures. Test–retest reliability of the IAT, however, was lower than the stability coefficients in the .80s that are usually found for explicit anxiety measures (Spielberger et al., 1970). On the other hand, a test–retest correlation of .57 is considerably higher than that reported for other implicit measures (Bosson et al., 2000). Implicit and explicit measures were only weakly related. In sum, this pattern of results is very similar to the one reported for a self-esteem IAT (Bosson et al., 2000; Greenwald & Farnham, 2000) and provides a promising base for further research with the IAT–Anxiety.

Study 2: Fakability of the IAT–Anxiety

Conscious distortion of self-report scales is an important issue in many contexts (Rosse, Stecher, Miller, & Levin, 1998). For this reason, the provision of an assessment technique that is less (or not at all) susceptible to faking would be a great step toward a more valid measurement of the respective construct. The IAT was developed in part to fill this gap (Greenwald et al., 1998). Implicit measures should be robust against conscious distortion because they function in an indirect mode that is not (or at least not readily) understandable for participants.

This claim was empirically analyzed by Banse et al. (2001). The authors instructed a group of participants—who were naive with respect to the functioning of the IAT—to fake an extremely positive attitude toward homosexuals. A control group responded to the IAT without a manipulation instruction. Both groups also completed explicit measures of attitudes toward homosexuality. Banse et al. expected and found that participants could deliberately manipulate explicit but not implicit attitudes.

Kim (2001) showed that even participants who were fully informed about the functioning of the IAT were not able to fake positive implicit attitudes toward Blacks. Although these participants were able to slow down their responses in the blocks compatible with prejudice (Black + unpleasant), they could not voluntarily speed their responses up in the incompatible blocks (Black + pleasant). The authors claimed that a slowing strategy can be detected quite easily and, thus, that this kind of faking would not remain undiscovered.

In this study, we instructed naive participants to make a very good impression during a job application scenario. Thus, we used a faking instruction that is more subtle than that of Banse et al. (2001), as we did not explicitly mention the reduction of anxiety scores as a goal of this study. We believe that our approach mirrors most applied (and most research) settings: Participants usually do not exactly know the dimensions that are assessed. They might have hypotheses about the constructs of interest, but usually they have the comparatively unspecific intention to present themselves in a favorable light and the specific goal to pass the test (to get a job, a promotion, the driver's license, diminished culpability, etc.). In sum, we expected that participants were not able to voluntarily fake a nonanxious impression on the IAT. In contrast, explicit measures of personality should be subject to conscious distortion (Häcker, Schwenkmezger, & Utz, 1979; Smith, 1974; Stöber, 2001).

Method

Participants. Forty students (31 women, 9 men) of Johannes Gutenberg-University Mainz participated in this study in exchange for research participation credit. Their average age was 22.2 years (*SD* = 2.7).

Measures. The IAT–Anxiety used in this study was identical to that of Study 1. In addition, participants completed the German version (Laux et al., 1981) of the Trait form of the STAI (Spielberger et al., 1970) and the SDS-17R (Stöber, 2001).

Procedure. Participants were randomly assigned to a faking condition and a control condition. Participants received the respective instruction and then responded to the IAT and the questionnaires. Presentation order of the IAT and the STAI was counterbalanced across conditions. The IAT and the STAI were presented directly after the treatment, whereas the SDS-17R was always presented at the last position. This was done to maximize the instruction effects on both anxiety measures. In the control condition,

[5] Note that high IAT–Anxiety scores indicate high anxiety. Because of the fixed presentation order of Block 3 and Block 5 (see Footnote 2) we refrained from computing effect sizes for the IAT effect.

participants received the standard instructions (see Study 1) and then responded to the respective measure.

In the faking condition, participants received the following instruction:

> Imagine that you were applying for a job in a large international consulting company. To get this job, you should try to make a very good impression. Please try to present yourself in a favorable light without exaggerating too much or being implausible.

Reduction of the IAT data was done as described in Study 1. There were no participants with error rates that exceeded 20% or with mean latencies above 2,000 ms. Presentation order of the anxiety measures (explicit measure first vs. implicit measure first) showed no main effects nor interactions with the experimental instruction on anxiety scores. Thus, this factor was dropped from analysis.

Results and Discussion

Internal consistencies of all measures were unaffected by the experimental manipulation. For the IAT, Cronbach's alphas were .80 (control condition) and .71 (faking condition), for the STAI, alphas were .79 (control condition) and .85 (faking condition), and for the SDS-17R, alphas were .79 (control condition) and .73 (faking condition). The main results of this study are shown in Table 1. The elevated social desirability scores in the faking condition documented that participants complied with the instruction, $t(38) = 2.29$, $p = .027$ (Cohen's $d = 0.73$). As expected, participants displayed lower explicit anxiety scores in the faking condition as compared with the control condition, $t(38) = 1.98$, $p = .055$ ($d = 0.63$). In contrast, IAT scores were not significantly affected by this treatment, $t(38) = 0.52$, *ns* ($d = 0.16$).

Taken together, as compared with an explicit anxiety measure, the IAT was considerably less subject to conscious distortion. On the other hand, one should not conclude from the nonsignificant results of the *t* tests reported above that the IAT–Anxiety cannot be faked at all, because an effect size of 0.16 could indicate a small effect. Future studies that analyze this important issue in greater detail should use larger sample sizes and could also use within-subject designs. In these studies, several additional factors could be investigated concerning their impact on the fakability of the IAT: (a) The reduction of the anxiety score could be explicitly mentioned as primary goal of the faking condition (cf. Banse et al., 2001), (b) it would be interesting to contrast naive and informed participants with respect to the functioning of the IAT–Anxiety (cf. Kim, 2001), and (c) an examination of the effects of an instruction to fake higher anxiety scores would also be very interesting, because in some clinical assessment contexts, a simulation of anxiety might be expected.

Study 3: Prediction of Experimenter-Rated Anxiety and Performance After Failure

Both studies described so far have dealt with necessary requirements of an implicit personality test: reliability of measurement and resistance to faking. Although both factors are certainly of great relevance for establishing the IAT as a measure of choice, in our view the most important issue for a test is to show *predictive validity*. In other words, the prediction of empirical indicators of the respective construct constitutes the cornerstone of every assessment tool.

The IAT aims at assessing attitudes and personality dispositions. Of course, there are a wide variety of established self-report measures available for this task. Thus, a new measure must exhibit *incremental validity* (i.e., it should be shown that the IAT predicts variance in relevant criteria in addition to explicit tests of the same construct). Furthermore, we suggest that validation studies should also include a measure of social desirability to control for the tendency to present oneself in a favorable light. Thus, according to the proposed incremental validation strategy, zero-order correlations between the IAT and criterion variables are not sufficient for demonstrating validity. In contrast, the IAT should be a significant predictor even when the effects of an explicit measure of the same construct and a measure of social desirability are controlled for. In the case of anxiety, it also seems necessary to control for the interaction term of anxiety and social desirability, because it was postulated that specific configurations of both variables could explain variance in addition to the main effects (Weinberger et al., 1979).

Implicit tests seem especially promising for the prediction of behaviors that are not normally subject to conscious control (Fazio, Jackson, Dunton, & Williams, 1995). Nonverbal behavior, for example, is a well-suited and often-used criterion variable in validation studies (Dovidio, Kawakami, & Gaertner, 2002; Dovidio, Kawakami, Johnson, Johnson, & Howard, 1997; Fazio et al., 1995; McConnell & Leibold, 2001; Spalding & Hardin, 1999). A basic assumption of this research tradition is that nonverbal behavior represents a relatively spontaneous form of behavior because it is less frequently monitored and less effectively con-

Table 1
Effects of the Faking Instruction on Implicit and Explicit Measures (Study 2)

	Experimental condition						
	Control		Faking				
Measure	*M*	*SD*	*M*	*SD*	*t*(38)	*p*	Cohen's *d*
IAT (ms)	−140	143	−166	117	0.52[a]	.609[a]	0.16[a]
STAI	35.40	5.75	31.70	6.05	1.98	.055	0.63
SDS-17R	6.60	3.76	9.20	3.40	2.29	.027	0.73

Note. $N = 20$ in each experimental condition. All *p*s are two-tailed. IAT = Implicit Association Test—Anxiety; STAI = Trait form of the State–Trait Anxiety Inventory; SDS-17R = Social Desirability Scale 17 (revised form).
[a] These values are based on log-transformed IAT measures.

trolled. Thus, nonverbal behavior might be subject to " 'leakage' of negativity" (Fazio et al., 1995, p. 1,026) that someone is actually experiencing—despite all individual efforts to appear in a nonnegative (e.g., less prejudiced or less anxious) way. Consequently, indices of nonverbal behavior should be associated with implicit tests because neither measure is subject (or to a lesser extent than questionnaire measures) to response factors and introspective limits (Greenwald et al., 2002).

Results of previous research are generally supportive of this assumption: Dovidio et al. (1997) found that an implicit measure (a priming task) predicted eye blinking and gaze avoidance with a Black relative to a White interviewer better than did self-report measures. McConnell and Leibold (2001) replicated and extended this finding with the IAT as an implicit test and used several additional behaviors and experimenter ratings as criteria of prejudice. Additionally, Fazio et al. (1995) were able to show that a priming measure of prejudice was associated with experimenter-rated prejudice during a debriefing procedure. Similarly, Spalding and Hardin (1999) found that implicit self-esteem (also assessed by a priming procedure) predicted experimenter-rated anxiety during a self-relevant interview.

Our validation studies of the IAT–Anxiety follow this line of reasoning. Consequently, Study 3 examines the predictive validity of implicit and explicit anxiety measures with respect to experimenter-rated anxiety and performance after failure stress. We included a performance measure as an additional criterion because research on the anxiety–performance relationship has a long tradition (Eysenck, 1992; Krohne & Laux, 1982; Seipp, 1991; Zeidner, 1998). Usually, there is a moderate, negative relationship between explicit anxiety measures and performance, although there are also some inconsistent findings (Seipp, 1991). We reasoned that stress-related changes in performance of simple tasks are good candidates for an association with implicit tests because cognitive, emotional, and physiological components of the anxiety response interfere with good performance in these tasks.

In sum, we contrasted explicit and implicit anxiety measures in the prediction of self-reported state anxiety, experimenter-rated anxiety, and changes in performance after failure stress. We expected that state anxiety would be predicted by trait anxiety. Concerning experimenter-rated anxiety and performance, we hypothesized that the IAT shows incremental validity (i.e., it should explain variance in these criteria even when the effects of explicit anxiety and social desirability are controlled for). A second aim of this study was to further examine the relationship of the IAT with explicit anxiety measures. For this reason, participants responded to the STAI and an anxiety thermometer. Additionally, they provided an explicit rating of the IAT stimuli. We expected implicit and explicit measures to show only weak associations, whereas the explicit measures should be highly correlated.

Method

Participants. Sixty-two students (35 women, 27 men) of Johannes Gutenberg-University Mainz participated in this study in exchange for research participation credit. Their average age was 24.2 years ($SD = 5.2$).

Measures. The IAT–Anxiety used in this study was identical to that of Studies 1 and 2. In addition, participants completed three explicit trait anxiety measures: (a) the German version (Laux et al., 1981) of the Trait form of the STAI (Spielberger et al., 1970), (b) an anxiety thermometer (i.e., "How anxious are you? Please indicate your anxiety on a scale from 0 [not at all] to 100 [very high]"), and (c) an explicit rating of the five anxiety and the five calmness stimuli of the IAT (i.e., "Please indicate on a scale from 0 [not at all] to 5 [very high] the extent to which the following attributes apply to you"). The five calmness items were reverse scored. Thus, the explicit rating of the IAT stimuli could vary between 0 (*no anxiety*) and 50 (*high anxiety*). Furthermore, participants completed the German version (Musch, Brockhaus, & Bröder, 2002) of the Balanced Inventory of Desirable Responding (BIDR; Paulhus, 1998). This inventory assesses two components of socially desirable responding, self-deceptive enhancement and impression management, with 10 items each. We used the BIDR in this study to examine the possibility that it was just one component of social desirability (rather than the composite score) that shared variance with the IAT, the explicit anxiety measures, and the criteria.

Participants indicated their state anxiety at baseline and after the stress induction on an eight-item scale (e.g., "worry," "nervous," "tense") ranging from 0 (*not at all*) to 5 (*very*). Both experimenters rated the apparent anxiety of the participants on a two-item scale ("worried," "tensed") ranging from 0 (*not at all*) to 5 (*very*). Item averages of the state measures are reported throughout this article.

Performance was measured by means of the d2 Test of Attention (d2 Test; Brickenkamp, 1994). In this test, participants are required to perform a simple discrimination task by crossing out relevant stimuli and by ignoring irrelevant stimuli. Relevant stimuli consist of the letter *d* and two lines above or below the letter. Thus, there are three relevant stimuli: *d*s with two lines above the letter, *d*s with two lines below the letter, and *d*s with one line above and one line below the letter. Participants are required to ignore *d*s with more or fewer than two lines as well as the letter *p* (irrespective of the number of lines and their location). Stimuli are presented on a piece of paper in 14 rows of 47 stimuli each. We calculated a d2 Test performance index by subtracting errors (misses and false alarms) from the total number of processed stimuli. We computed change scores for self-rated state anxiety, experimenter-rated state anxiety, and performance by subtracting baseline values from those after the stress induction.

Procedure. On arrival at the laboratory, participants were greeted by two experimenters and completed the IAT–Anxiety and the explicit trait measures. Then participants indicated their state anxiety. Furthermore, apparent anxiety of the participants was rated by the experimenters, who were unaware of the explicit and implicit anxiety measures. Afterward, participants received the instructions for the d2 Test. It was emphasized that this test constitutes a performance test that measures how well one is able to concentrate on a task. Then participants worked on some practice stimuli, and emerging questions were answered to ensure that everyone understood the nature of this test. Participants were told that they would work on this task for 2 min. They were instructed to work "as quickly and as accurately as possible." After having completed the task, participants received negative feedback about their performance to induce stress: After an inspection of the results, an experimenter told them that their performance was "not very good. In average, participants get two rows further." Participants were then told that they would work on the same task for a second time. The experimenters rated the participants' apparent anxiety during this second trial. Additionally, participants retrospectively indicated their state anxiety during the second task.

An elaborate debriefing was carried out for every participant. They were informed in detail about the purpose of the study. Specific care was taken to ensure that every participant understood the deceptive nature of the negative feedback. It was particularly emphasized that they actually had not failed. All participants seemed satisfied with this explanation. The experimenters then asked the participants' cooperation in not discussing the study with others.

Results

Descriptive statistics. Means, standard deviations, and internal consistencies of the trait measures are displayed in Table 2. The correlation matrix of these measures can be seen in Table 3. The pattern of results can be described as follows: (a) The IAT–Anxiety was independent of all explicit anxiety measures as well as of both components of social desirability, (b) the explicit anxiety measures were highly associated, and (c) impression management was independent of the explicit anxiety measures, whereas self-deceptive enhancement showed a moderate negative relationship with these scales.

The descriptive statistics of the state measures are displayed in Table 4. Self-reported state anxiety showed only a modest increase from baseline to the stress situation, $t(61) = 1.73$, $p = .09$. In contrast, experimenter ratings of participants' anxiety rose to a larger extent, $t(61) = 11.21$, $p < .001$. Performance scores also increased from baseline to stress, $t(61) = 8.04$, $p < .001$. To evaluate the effectiveness of our stress induction on performance, one should take into account that participants in neutral situations usually show in these kinds of tests an increase of 25% from the first to the second assessment (Brickenkamp, 1994; Westhoff, 1989). Thus, an increase of 8%, as shown by our participants, indicates, in fact, a performance decrease due to stress.

Main analyses. In all analyses, our focus was on the difference scores because they reflect the dynamic change in the criterion variables from baseline to stress, which should be an indicator of anxiety. We used the raw scores at the second measurement point as additional criteria because they indicate the status after the stress induction.

Results of the main analyses are displayed in Table 5. We first computed zero-order correlations between each predictor (STAI and IAT, respectively) and each criterion. The explicit anxiety measure significantly explained variance in self-rated state anxiety and in experimenter-rated state anxiety after the stress induction (Time 2). According to our incremental validation strategy, we then performed hierarchical regressions with change scores[6] and levels after the stress induction of self-rated state anxiety, experimenter-rated state anxiety, and performance as separate criterion. In each of these six regressions, the explicit anxiety measure was entered in Step 1; impression management, self-deceptive enhancement, and the interaction terms of both social desirability components with the STAI[7] (STAI × Impression Management and STAI × Self-Deceptive Enhancement) were entered in Step 2; and the IAT was entered in Step 3. Thus, we were able to estimate the portion of variance that the implicit anxiety measure shared with the state anxiety indicators when all other variables (main effects and interactions) were controlled for. As can be seen from the last column of Table 5, the IAT predicted changes in experimenter-rated anxiety and changes in performance after stress that explicit measures were unable to predict.[8]

Table 2
Descriptive Statistics of the Trait Measures (Study 3)

Measure	*M*	*SD*	α
Implicit anxiety			
IAT (ms)	−141	144	.78
Explicit anxiety			
STAI	38.94	7.94	.89
Thermometer	42	22	
IAT stimuli	18.21	7.37	.89
Social desirability (BIDR)			
Impression management	31.89	10.57	.75
Self-deceptive enhancement	40.24	7.31	.64

Note. $N = 62$. Possible ranges of the explicit measures: STAI = 20–80; Thermometer = 0–100; IAT stimuli = 0–50; BIDR impression management and self-deceptive enhancement = 10–70. IAT = Implicit Association Test—Anxiety; STAI = Trait form of the State–Trait Anxiety Inventory; IAT stimuli = explicit rating of the Implicit Association Test stimuli; BIDR = Balanced Inventory of Desirable Responding.

Table 3
Correlations Among the Trait Measures (Study 3)

Measure	1	2	3	4	5	6
Implicit anxiety						
1. IAT	—	−.06	.04	−.03	.14	.16
Explicit anxiety						
2. STAI		—	.56***	.77***	−.46***	.09
3. Thermometer			—	.80***	−.19	.11
4. IAT stimuli				—	−.36**	.09
Social desirability (BIDR)						
5. Self-deceptive enhancement					—	.14
6. Impression management						—

Note. $N = 62$. IAT = Implicit Association Test—Anxiety; STAI = Trait form of the State–Trait Anxiety Inventory; IAT stimuli = explicit rating of the Implicit Association Test stimuli; BIDR = Balanced Inventory of Desirable Responding.
** $p < .01$. *** $p < .001$ (two-tailed).

Supplementary analysis. To examine the impact of speed of responding on the IAT score, we analyzed the association of the mean reaction times in Block 3 ($M = 1{,}031$ ms, $SD = 202$ ms) and Block 5 of the IAT ($M = 890$ ms, $SD = 191$ ms) with the number of correct responses in the d2 Test during baseline and stress (see Table 4 for the descriptive statistics of these variables). As can be seen from Table 6, speed of responding in the IAT was associated with the number of correct responses in the d2 Test (correlations between −.53 and −.64, ps $< .001$). This indicates the existence of a general speed factor that determines fast responding in both tasks. It is important to note that the IAT effect (i.e., the difference score of Block 5 and Block 3) did not correlate with the perfor-

[6] Instead of using the respective change score as dependent variable, a reviewer suggested that we use the Time 2 measure as dependent variable and enter the Time 1 measure as a covariate in a new Step 1 of the hierarchical regression. This procedure led to an identical pattern of results in all analyses involving change scores. Similarly, the use of the mean latency in the IAT as a covariate did not change significance levels in all regression analyses.

[7] Following standard practice, all variables were centered before the interaction terms were computed.

[8] Column 4 of Table 5 also shows two significant ΔR^2s indicating effects of social desirability (Step 2). For changes in self-reported state anxiety, impression management, $\beta = .40$, $p = .001$, and Impression Management × STAI, $\beta = .29$, $p = .019$, were significant predictors in Step 2 of the hierarchical regressions. For self-reported state anxiety after the stress induction, impression management was a significant predictor, $\beta = .35$, $p = .007$.

mance raw scores in the d2 Test. Similarly, the difference score in the d2 Test (change score from baseline to stress) was unrelated to the reaction times in both IAT blocks. Thus, both difference scores did not reflect general speed of responding.

Discussion

This study provides compelling evidence for the predictive validity of the IAT–Anxiety. In line with our incremental validation strategy, the IAT accounted for variance in performance changes due to stress as well as in changes in experimenter-rated anxiety even when the effect of explicit measures of anxiety and social desirability were controlled for. In contrast, the explicit anxiety measure was associated with levels of self-rated and observer-rated anxiety. Taken together, in this study implicit and explicit anxiety measures functioned in a complementary manner in two ways: First, there were different *content domains* of predictive power—self-reported state anxiety (explicit anxiety) versus performance changes (implicit anxiety). Second, there were different *process domains* of predictive power—level (explicit anxiety) versus change (implicit anxiety). Thus, it might be a good idea to use both implicit and explicit measures to enhance the portion of variance in human behavior that can be accounted for by trait measures.

Several limitations to this initial validation study of the IAT–Anxiety should also be noted: Although we regard the induced failure task as a well-suited paradigm for investigating anxiety-related processes, replications using other stressors are clearly warranted. Because self-rated state anxiety only rose to a moderate extent, future studies might use stressors with higher impact on participants' subjective feeling state (Egloff & Krohne, 1996). Additionally, more complex cognitive tasks could be used as performance indices: Hock, Kohlmann, and Egloff (2002), for example, were able to show that performance in an exam was predicted by an implicit anxiety measure (a measure of attention allocation toward threat; MacLeod, Mathews, & Tata, 1986). Similarly, Nosek, Banaji, and Greenwald (2002b) demonstrated that math performance was independently predicted by implicit (IAT) and explicit math attitudes. Furthermore, it is not exactly clear why the IAT predicts impaired performance. One possible mechanism might be that implicitly measured anxiety leads to increases in worry, which, in turn, result in impaired performance. However, state anxiety (level and change score) was unrelated to the IAT in this study. Similarly, the Hock et al. (2002) study showed, by using path analysis, that the negative effects of the implicit anxiety measure on performance were independent of worry (which did also show direct negative effects on performance). Finally, the experimenter rating of anxiety used in this study constitutes a rather global (although often used) index of anxious behavior. Thus, in the second validation study presented in this article, trained coders rated different behavioral indicators of anxiety to provide a more thorough assessment of this important criterion variable.

Table 4
Descriptive Statistics of the State Measures (Study 3)

Measure	*M*	*SD*
Self-rated state anxiety		
Baseline (T1)	1.33	0.95
Stress (T2)	1.57	1.04
Change (T2 − T1)	0.24	1.09
Experimenter-rated state anxiety		
Baseline (T1)	1.19	0.57
Stress (T2)	2.17	0.75
Change (T2 − T1)	0.98	0.69
Performance		
Baseline (T1)	199.2	30.4
Stress (T2)	215.1	30.4
Change (T2 − T1)	15.9	15.6

Note. $N = 62$. Possible ranges of the state anxiety measures = 0–5. T = time.

Table 5
Predictions of the State Measures by the Trait Measures (Study 3)

	Zero-order correlation		Hierarchical regression		
Measure	STAI *r*	IAT *r*	Step 1 STAI R^2	Step 2 BIDR[a] ΔR^2	Step 3 IAT ΔR^2
Self-rated state anxiety					
Stress (T2)	.33**	−.01	.112**	.116†	.001
Change (T2 − T1)	.06	.05	.004	.291**	.004
Experimenter-rated state anxiety					
Stress (T2)	.26*	.12	.065*	.044	.031
Change (T2 − T1)	.08	.26*	.006	.094	.085*
Performance					
Stress (T2)	.15	−.10	.022	.039	.016
Change (T2 − T1)	.04	−.27*	.002	.015	.085*

Note. $N = 62$. STAI = Trait form of the State–Trait Anxiety Inventory; IAT = Implicit Association Test—Anxiety; BIDR = Balanced Inventory of Desirable Responding; T = time; T1 = baseline.
[a] In Step 2 of the hierarchical regression, impression management, self-deceptive enhancement, and the interaction terms of both social desirability components with the STAI (STAI × Impression Management and STAI × Self-Deceptive Enhancement) were entered into the equation.
† $p < .1$. * $p < .05$. ** $p < .01$ (two-tailed).

Study 4: Prediction of Behavioral Anxiety Indicators During a Stressful Speech

This study examines the utility of implicit and explicit anxiety measures in the prediction of behavior during a stressful task. Following the rationale of Study 3, indices of nonverbal behavior are assumed to be associated with implicit tests because both types of measures are quite difficult to control and, thus, should be comparatively pure indicators of the anxiety construct. Complementary to Study 3, this study highlights two additional features. First, we implemented another anxiety-arousing situation, an evaluated speaking task. This task is closely related to anxiety-inducing stressors that occur in everyday life, and it is known to produce substantial increases in subjective and objective state anxiety indicators (al'Absi et al., 1997; Egloff, Wilhelm, Neubauer, Mauss, & Gross, 2002). Second, several behavioral indicators of anxiety were rated by trained coders to provide a

Table 6
Correlations Among the Mean Reaction Times in the IAT and the Number of Correct Responses in the d2 Test (Study 3)

Measure	1	2	3	4	5	6
IAT						
1. Mean RT Block 3 (self + anxiety)[a]	—	.76***	−.35**	−.56***	−.57***	−.02
2. Mean RT Block 5 (self + calmness)[a]		—	.35**	−.53***	−.64***	−.20
3. Difference score (IAT effect)[a]			—	.04	−.10	−.27*
d2 Test						
4. No. correct responses: baseline				—	.87***	−.26*
5. No. correct responses: stress					—	.26*
6. Difference score						—

Note. $N = 62$. IAT = Implicit Association Test—Anxiety; RT = reaction time.
[a] Log transformed.
* $p < .05$. ** $p < .01$. *** $p < .001$ (two-tailed).

more detailed picture of the participants' anxiety response to the speaking task.

In sum, the aim of this study is to provide a second test of the IAT–Anxiety's predictive power by using another stressful situation and additional anxiety indicators. We expected that the explicit anxiety measure would predict state anxiety during the speaking task. We hypothesized that the implicit anxiety measure would predict behavioral anxiety indicators even when questionnaires measures were controlled for.

Method

Participants. Thirty-three introductory psychology students (29 women, 4 men) of Johannes Gutenberg-University Mainz participated in this study in exchange for research participation credit. Their average age was 22.0 years ($SD = 3.1$).

Measures. The IAT–Anxiety used in this study was identical to that of the other three studies. In addition, participants completed the German version (Laux et al., 1981) of the Trait form of the STAI (Spielberger et al., 1970) and the SDS-17R (Stöber, 2001). State anxiety at baseline and during the speech was assessed by means of a three-item scale ("anxious," "nervous," "uncertain") ranging from 0 (*not at all*) to 3 (*very*). We report item averages of this state anxiety measure throughout this article.

Behavioral measures of anxiety were obtained by means of two trained judges who were unaware of the scores on the anxiety measures and who rated the videotapes of the speeches on five behavioral expressions of anxiety: number of nervous mouth movements, number of eye blinks, hand position and movements, speech dysfluency, and a global rating of anxiety. These indicators of anxiety have been validated in previous studies (e.g., Dow, 1985; Fydrich, Chambless, Perry, Buergner, & Beazley, 1998; Monti et al., 1984). The judges started by counting the *number of nervous mouth movements*, defined as lip biting, lip licking, twitches of the mouth, and pressing of the lips. In a second trial, the *number of eye blinks* was counted. Then judges coded for *hand position and movements* by rating on a 5-point scale the amount of anxiety that was signaled through this channel (1 = *not at all*, 5 = *extremely*). We used a rating scale (rather than counting hand movements) because a high number of hand movements can indicate high anxiety (e.g., nervous face or hair touching) or low anxiety (e.g., expressive but relaxed underpinning of the speech). Similarly, a lack of hand movements can indicate any level of anxiety, depending on the position of the hand (e.g., pressed on the legs or put under the legs vs. a relaxed position). *Speech dysfluency* was defined as number of pauses, length of pauses, and number of verbal dysfluencies (1 = *not at all dysfluent*, 5 = *extremely dysfluent*).[9] In the last trial, judges rated their overall impression of the speaker's *anxiety*. Interrater reliabilities were satisfactory (see *Results* section), and, thus, means of the two raters were used for subsequent analyses.

Procedure. The three trait measures (IAT, STAI, and SDS-17R) were assessed in a separate session that was ostensibly unrelated to the speaking task. At Session 2, participants first worked on some questionnaires not relevant to this research to get accustomed to the laboratory. Then they indicated their state anxiety (baseline score). Afterward, participants received the following instructions for the speech:

> This experiment analyzes how well you are able to comprehend and present a scientific text under time pressure. This ability is an important prerequisite for a successful completion of your courses and the oral examinations during the pre-diploma and diploma. The experiment will proceed as follows: You will first read and prepare a scientific text for 10 minutes. Then you will orally present the content of this text for 3 minutes. Your speech will be videotaped and later scored by a panel of judges who will rate and compare your speech to others given under the same circumstances. Please try to deliver a comprehensive and well-structured speech, talking for the full 3 minutes.

The scientific text was concerned with the composition and the function of the blood and was compiled from a physiology textbook (Schmidt & Thews, 1987). Pretests had shown that it was a very difficult task to deliver a speech based on the contents of this text because of the number of details and technical terms as well as because of the time constraints. Participants were told that they were not allowed to use the text or their notes during the speech. A video camera was positioned directly in front of the participants, and care was taken to maximize the evaluative nature of this task. The experimenter remained behind the camera during the speaking task. Participants remained seated throughout the entire task. After delivering their speech, participants indicated their state anxiety during the speech.

An elaborate debriefing was carried out for every participant. They were informed in detail about the purpose of the study. Participants were told that this study was not concerned with analyzing the ability to successfully complete courses and prediploma or diploma. Specific care was taken to ensure that every participant understood that the performance on the speech task was actually not an indicator of this ability. It was further emphasized that the videos of the speech would only be viewed by trained coders for

[9] This category was originally coded as speech fluency (1 = *not at all fluent*, 5 = *extremely fluent*). A reviewer suggested that we reverse the scoring of this measure to give all five anxiety indicators the same direction of scoring.

scientific purposes. The experimenter then asked the participants' cooperation in not discussing the study with others.

Results

Descriptive statistics. Means, standard deviations, and internal consistencies of the trait measures and of self-reported state anxiety are shown in Table 7. Once again, the IAT–Anxiety did not correlate significantly with explicit anxiety ($r = -.08$, $p = .66$). State anxiety scores rose from baseline to speech, $t(32) = 8.44$, $p < .001$, indicating that participants were subjectively affected by the experimental manipulation. Table 8 displays the descriptive statistics of the anxiety-related behaviors during the speech. It can also be seen from Table 8 that the interrater correlation was high for the two counted categories and adequate for the three ratings.

Main analyses. We first computed zero-order correlations between the two predictors (STAI and IAT, respectively) and each criterion (see Table 9). The explicit anxiety measure was significantly associated with self-rated state anxiety during the speech (Time 2). We then performed hierarchical regressions with state anxiety and each behavior rating as a separate criterion. In each of these seven regressions, the explicit anxiety measure was entered in Step 1, social desirability and the interaction terms of social desirability with the STAI were entered in Step 2, and the IAT was entered in Step 3. Thus, according to our incremental validation strategy, we were able to estimate the portion of variance that the implicit anxiety measure shared with the anxiety indicators during the speech when all other variables (main effects and interactions) were controlled for. As can be seen from the last column of Table 9,[10] with the exception of eye blinks, the IAT predicted behaviors during the stressful speech that explicit measures were unable to predict.[11]

Discussion

This study provides a second demonstration of the IAT–Anxiety's predictive validity. Whereas the explicit anxiety measure only accounted for self-reported anxiety level during the speech, the implicit measure predicted four of the five behavioral anxiety indicators. This study thus replicates and extends the results of Study 3 by using another kind of stressor and by providing a more fine-graded analysis of anxiety-related behaviors.

We further examined the status of eye blinks as an anxiety indicator during a stressful speech by regressing the four specific behavior codes (nervous mouth movements, eye blinks, hand position and movements, and speech dysfluency) on the global anxiety rating. Thus, we analyzed which of the four specific behaviors was related to the judges' global anxiety rating. This analysis showed a highly significant portion of explained variance in the global anxiety rating by the four specific behaviors, $R^2 = .67$, $F(4, 28) = 13.98$, $p < .001$. The most important predictors were speech dysfluency, $\beta = .56$, $p < .001$, and hand position and movements, $\beta = .39$, $p = .004$, followed by nervous mouth movements, $\beta = .24$, $p = .082$. In contrast, eye blinks were unrelated to the global anxiety rating, $\beta = -.10$, $p = .37$. Thus, in the context of our speech, the number of eye blinks seems to indicate a process unrelated to judges' anxiety ratings. A possible explanation is that there are large interindividual differences in the base rate of eye blinks that mask possible differential changes due to the speech. This argues for the inclusion of an eye blink baseline to separate general level from anxiety-related changes in eye blinks.

Table 7
Descriptive Statistics of the Trait and State Measures (Study 4)

Measure	*M*	*SD*	α
Trait measures			
IAT (ms)	−108	101	.76
STAI	37.85	8.22	.90
SDS-17R	7.84	2.82	.59
State anxiety			
Baseline (T1)	0.30	0.45	.74
Speech (T2)	1.17	0.64	.74
Change (T2 − T1)	0.86	0.58	.67

Note. $N = 33$. Possible ranges of the measures: STAI = 20–80; SDS-17R = 0–16; self-report state measures = 0–3. IAT = Implicit Association Test—Anxiety; STAI = Trait form of the State–Trait Anxiety Inventory; SDS-17R = Social Desirability Scale 17 (revised form).

Table 8
Descriptive Statistics of the Behavioral Anxiety Indicators (Study 4)

Behavior	*M*	*SD*	*r*
Nervous mouth movements (per minute)	6.92	2.20	.81
Eye blinks (per minute)	35.35	14.55	.97
Hand position and movements (rating)	2.85	0.88	.68
Speech dysfluency (rating)	3.23	0.88	.71
Anxiety (rating)	3.03	0.85	.67

Note. $N = 33$. r = interrater correlation. Possible ranges of the behavior ratings = 1–5.

General Discussion

In this article, we have presented a series of four studies that aim at establishing the IAT–Anxiety as a reliable and valid indirect measure of the anxiety construct. We started by analyzing internal consistency and stability because reliability of measurement con-

[10] For speech dysfluency, there was also a significant effect of social desirability in Step 2 of the hierarchical regression, $\beta = .47$, $p = .014$.

[11] As hypothesized by a reviewer, a combined behavioral anxiety index (average of the five behavioral anxiety measures after standardization) showed a stronger association with the IAT than did any of the five measures individually (cf. Table 9; $r = .44$, $p = .011$). A hierarchical regression using this combined index as criterion yielded the following results: Step 1 (STAI), $R^2 = .002$, *ns*; Step 2 (SDS and STAI × SDS), $\Delta R^2 = .032$, *ns*; Step 3 (IAT), $\Delta R^2 = .185$, $p = .016$. The exclusion of the number of eye blinks from this index (see *Discussion* section of Study 4 for the justification of this procedure) resulted in an even higher association between the IAT and the combined behavioral anxiety index ($r = .47$, $p = .006$). A hierarchical regression using this revised combined index as criterion yielded the following results: Step 1 (STAI), $R^2 = .010$, *ns*; Step 2 (SDS and STAI × SDS), $\Delta R^2 = .032$, *ns*; Step 3 (IAT), $\Delta R^2 = .209$, $p = .009$.

Table 9
Predictions of the State Measures by the Trait Measures (Study 4)

Measure	Zero-order correlation: STAI *r*	Zero-order correlation: IAT *r*	Hierarchical regression: Step 1 STAI R^2	Step 2 SDS-17R[a] ΔR^2	Step 3 IAT ΔR^2
Self-reported state anxiety					
Speech (T2)	.42*	.03	.178*	.044	.001
Change (T2 − T1)	.28	−.05	.078	.034	.010
Behavioral anxiety indicators					
Nervous mouth movements	.01	.33†	.000	.016	.104†
Eye blinks	−.17	−.01	.029	.010	.000
Hand position and movements (rating)	.19	.39*	.037	.019	.147*
Speech dysfluency (rating)	.02	.29†	.000	.190*	.075†
Anxiety (rating)	.12	.38*	.014	.026	.136*

Note. $N = 33$. STAI = Trait form of the State–Trait Anxiety Inventory; IAT = Implicit Association Test—Anxiety; SDS-17R = Social Desirability Scale 17 (revised form); T = time; T1 = baseline.
[a] In Step 2 of the hierarchical regression, social desirability and the interaction term of social desirability with the STAI (STAI × SDS-17R) were entered into the equation.
† $p < .1$. * $p < .05$ (two-tailed).

stitutes a necessary prerequisite for every test. The results of Study 1 demonstrate that the IAT–Anxiety indeed shows satisfactory internal consistency and adequate stability. In a next step, we examined the extent to which implicit and explicit tests are vulnerable to conscious distortion (Study 2). We found that a faking instruction led only to very small (and nonsignificant) effects on the implicit measure, whereas the explicit anxiety measure was affected by this experimental manipulation to a higher degree. The remaining two studies were concerned with the most important property of every assessment tool, the issue of predictive validity. We used an incremental validation strategy by showing that the IAT predicted changes in experimenter-rated anxiety and performance after failure (Study 3) as well as global and specific behavioral anxiety indicators during a stressful speech (Study 4) when the main and interactive effects of explicit anxiety and social desirability were controlled for. This observed pattern of results conceptually replicates and extends previous findings (e.g., McConnell & Leibold, 2001; Spalding & Hardin, 1999) in that Studies 3 and 4 both show a dissociation involving state anxiety—which was predicted by the explicit anxiety measure—and behavioral manifestations of anxiety—which were predicted by the implicit anxiety measure (cf. Asendorpf, Banse, & Mücke, 2002, for a discussion of different forms of dissociations). Furthermore, we found low or no correlations between the IAT and explicit anxiety measures across all studies. Taken together, the results of these studies provide compelling evidence for establishing the IAT–Anxiety as an indirect test of anxiety.

We interpret our findings as indicating that the IAT taps sources of common variance with the criterion variables that self-report measures do not. This leads to two questions: (a) What kind of variables is the IAT especially suited to predict? and (b) Why does the IAT predict these variables better than do self-report measures? Concerning the first question, implicit tests seem especially promising for the prediction of spontaneous behaviors that are not normally subject to conscious control: Our studies show that the IAT predicted *nonverbal behaviors* indicating anxiety. Similarly, McConnell and Leibold (2001) found that behaviors indicating implicit prejudice were predicted by a race IAT. Asendorpf et al. (2002) showed that especially spontaneous behavioral manifestations (in contrast to more controlled forms of behavior) of shyness were predicted by an IAT measuring shyness. These findings are in accordance with results obtained using priming tasks (Dovidio et al., 1997; Fazio et al., 1995; Spalding & Hardin, 1999) and projective tests (Bornstein, 1998) as implicit measures. *Physiological reactivity* constitutes another form of behavior that is usually not effectively controlled. Consequently, Phelps et al. (2000) showed that IAT-measured implicit prejudice correlated with fMRI-assessed amygdala activation and eye blink startle responses when participants were shown Black compared with White faces. Similarly, Egloff et al. (2002) found that an implicit anxiety measure (a measure of attention allocation toward threat; MacLeod et al., 1986) predicted blood pressure and heart rate changes due to a speaking task. A third promising criterion variable seems to be *task performance*, as our study shows that the IAT was able to predict performance changes after failure in a simple discrimination task. It is interesting that more complex cognitive tasks were also predicted by implicit measures: Hock et al. (2002) found that a measure of attention allocation toward threat (MacLeod et al., 1986) predicted performance in an important exam, and Nosek et al. (2002b) reported an effect of implicit math attitudes on math performance (cf. also the TAT literature on achievement; McClelland, Koestner, & Weinberger, 1989).

Why is the IAT able to predict these variables? According to Greenwald et al. (1998), the IAT measures relative strengths of automatic associations between pairs of concepts. Defined operationally, the IAT–Anxiety, for example, measures how much easier it is for participants to categorize self items with anxiety-related items than self items with calmness-related items. It is further assumed that implicit measures are not subject to conscious control, and, thus, response factors and introspective limits (Greenwald et al., 2002) do not diminish their validity. Consequently, relatively spontaneous forms of behavior, such as nonverbal behavior or physiological reactivity, should be associated with implicit tests because all measures rule out individual efforts to appear in a nonnegative (e.g., less prejudiced or less anxious) way.

Of course, these are neither completely satisfactory nor sufficient explanations of how the IAT functions and why it predicts behavior independently of self-report measures. Consequently, the development of a theoretical model of components of the IAT effect reached the first place in a recent Top Ten to-do list (Greenwald, 2001; cf. also Devine, 2001). Similarly, recent research has produced several—sometimes contradictory, sometimes supplementary—theoretical accounts of mechanisms (see Mierke & Klauer, 2001, for an overview) that are assumed to constitute the foundation of the IAT effect, such as, for example, response criterion shifts (Brendl, Markman, & Messner, 2001),

response conflicts (De Houwer, 2001), environmental associations (Karpinski & Hilton, 2001), task-set switching (Mierke & Klauer, 2001), and figure-ground asymmetries (Rothermund & Wentura, 2001). Perhaps at this stage of knowledge, an analogy to a new medicine might be appropriate: The exact way this medicine works is currently unknown, but there is evidence that it contributes to the curing of the illness (i.e., the sometimes insufficient behavior prediction). Furthermore, the new medicine has no known side effects. Thus, it might be advisable and worthwhile to complement (not to replace) traditional explicit measures with implicit ones.

Another related question of great theoretical and practical interest is whether we should talk about implicit and explicit anxiety constructs or about implicit and explicit anxiety measures. The former implies two different modes of functioning that, consequently, lead to two different and independent constructs. The latter way of thinking assumes one construct (namely, anxiety) and two ways of measurement, an implicit and an explicit one. This would mirror Cattell's (1957) distinction of Q-data (questionnaire data) and T-data (test data). According to this distinction, the IAT–Anxiety is classified as a so-called objective test of personality. The low correlations (and the concomitant high internal consistencies) argue for two different constructs, because if one views explicit and implicit anxiety as a monotrait–heteromethod block of a multitrait–multimethod matrix (Campbell & Fiske, 1959), the empirical findings indicate low convergent validity.

Apart from these theoretical issues, we see several promising areas in which further research using implicit measures such as the IAT is especially warranted. As mentioned before, further demonstrations of the predictive validity of the IAT in experimental and applied settings (cf. Teachman, Gregg, & Woody, 2001) are certainly welcome. Another important issue is to further investigate trait and state portions of the IAT effect. As reported in Study 1 of this article and by other research groups (Bosson et al., 2000; Dasgupta & Greenwald, 2001; Greenwald & Farnham, 2000), internal consistency coefficients of the IAT are consistently larger than stability coefficients. These findings indicate that the observed variance of the IAT might capture state variance in addition to trait variance. Recent studies on context effects on the IAT as well as studies that experimentally induced changes in IAT-assessed attitudes corroborate this line of reasoning (Blair, Ma, & Lenton, 2001; Dasgupta & Greenwald, 2001; Lowery, Hardin, & Sinclair, 2001; Rudman, Ashmore, & Gary, 2001; Wittenbrink, Judd, & Park, 2001). How stable over time and how consistent across situations are IAT-assessed personality constructs? These questions could be tackled with different research designs, correlational and experimental ones. Latent state–trait theory is a methodological tool that might be helpful in disentangling and quantifying portions of state and trait variance in the IAT effect (Steyer, Schmitt, & Eid, 1999). Additionally, experimental manipulation of anxiety states is necessary to examine the nature of the IAT effect: Does IAT-measured anxiety change in an anxiety-arousing situation?

To conclude, there are several important open research questions concerning the IAT. At this stage of knowledge, the number of questions might exceed the number of answers by far. Nevertheless, we are confident that future research will answer most, if not all, of these questions. What can be said thus far is that the IAT is certainly a worthwhile and perhaps a necessary addition to explicit measures in the prediction of behavior.

References

al'Absi, M., Bongard, S., Buchanan, T., Pincomb, G. A., Licinio, J., & Lovallo, W. R. (1997). Cardiovascular and neuroendocrine adjustment to public speaking and mental arithmetic stressors. *Psychophysiology, 34,* 266–275.

Asendorpf, J. B., Banse, R., & Mücke, D. (2002). Double dissociation between implicit and explicit personality self-concept: The case of shy behavior. *Journal of Personality and Social Psychology, 83,* 380–393.

Banse, R., Seise, J., & Zerbes, N. (2001). Implicit attitudes towards homosexuality: Reliability, validity, and controllability of the IAT. *Zeitschrift für Experimentelle Psychologie, 48,* 145–160.

Blair, I. V., Ma, J. E., & Lenton, A. P. (2001). Imagining stereotypes away: The moderation of implicit stereotypes through mental imagery. *Journal of Personality and Social Psychology, 81,* 828–841.

Borkenau, P., & Ostendorf, F. (1992). Social desirability scales as moderator and suppressor variables. *European Journal of Personality, 6,* 199–214.

Bornstein, R. F. (1998). Implicit and self-attributed dependency strivings: Differential relationships to laboratory and field measures of help seeking. *Journal of Personality and Social Psychology, 75,* 778–787.

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

Brendl, C. M., Markman, A. B., & Messner, C. (2001). How do indirect measures of evaluation work? Evaluating the inference of prejudice in the Implicit Association Test. *Journal of Personality and Social Psychology, 81,* 760–773.

Brickenkamp, R. (1994). *Test d2 Aufmerksamkeits-Belastungs-Test* (8th ed.) [d2 Test of Attention]. Göttingen, Germany: Hogrefe.

Buchner, A., & Wippich, W. (2000). On the reliability of implicit and explicit memory measures. *Cognitive Psychology, 40,* 227–259.

Byrne, D., & Holcomb, J. (1962). The reliability of a response measure: Differential recognition-threshold scores. *Psychological Bulletin, 59,* 70–72.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait–multimethod matrix. *Psychological Bulletin, 56,* 81–105.

Cantor, N., & Mischel, W. (1977). Traits as prototypes: Effects on recognition memory. *Journal of Personality and Social Psychology, 35,* 38–48.

Cattell, R. B. (1957). *Personality and motivation: Structure and measurement.* New York: World Book.

Cronbach, L. J. (1990). *Essentials of psychological testing* (5th ed.). New York: Harper & Row.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24,* 349–354.

Cunningham, W. A., Preacher, K. J., & Banaji, M. R. (2001). Implicit attitude measures: Consistency, stability, and convergent validity. *Psychological Science, 12,* 163–170.

Dasgupta, N., & Greenwald, A. G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality and Social Psychology, 81,* 800–814.

Dasgupta, N., McGhee, D. E., Greenwald, A. G., & Banaji, M. R. (2000). Automatic preference for White Americans: Eliminating the familiarity explanation. *Journal of Experimental Social Psychology, 36,* 316–328.

De Houwer, J. (2001). A structural and a process analysis of the Implicit Association Test. *Journal of Experimental Social Psychology, 37,* 443–451.

Devine, P. G. (2001). Implicit prejudice and stereotyping: How automatic are they? Introduction to the special section. *Journal of Personality and Social Psychology, 81,* 757–759.

Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and

explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology, 82,* 62–68.

Dovidio, J. F., Kawakami, K., Johnson, C., Johnson, B., & Howard, A. (1997). On the nature of prejudice: Automatic and controlled processes. *Journal of Experimental Social Psychology, 33,* 510–540.

Dow, M. G. (1985). Peer validation and idiographic analysis of social skills deficits. *Behavior Therapy, 16,* 76–86.

Edwards, A. L. (1957). *The social desirability variable in personality assessment and research.* New York: Dryden.

Egloff, B., & Krohne, H. W. (1996). Repressive emotional discreteness after failure. *Journal of Personality and Social Psychology, 70,* 1318–1326.

Egloff, B., Wilhelm, F. H., Neubauer, D. H., Mauss, I. B., & Gross, J. J. (2002). Implicit anxiety measure predicts cardiovascular reactivity to an evaluated speaking task. *Emotion, 2,* 3–11.

Epstein, S. (1994). Integration of the cognitive and psychodynamic unconscious. *American Psychologist, 49,* 709–724.

Eysenck, M. W. (1992). *Anxiety: The cognitive perspective.* Hove, England: Erlbaum.

Farnham, S. D. (1998). FIAT for Windows (Version 2.3) [Computer software]. Seattle, WA: Author.

Fazio, R. H. (1990). Multiple processes by which attitudes guide behavior: The MODE model as an integrative framework. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 23, pp. 75–109). San Diego, CA: Academic Press.

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology, 69,* 1013–1027.

Funder, D. C. (2001). Personality. *Annual Review of Psychology, 52,* 197–221.

Fydrich, T., Chambless, D. L., Perry, K. J., Buergner, F., & Beazley, M. B. (1998). Behavioral assessment of social performance: A rating system for social phobia. *Behaviour Research and Therapy, 36,* 995–1010.

Greenwald, A. G. (2001, October). *What's wrong with the Implicit Association Test?* Paper presented at the Society of Experimental Social Psychology Workshop, Spokane, WA.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review, 102,* 4–27.

Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., & Mellott, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review, 109,* 3–25.

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

Greenwald, A. G., McGhee, D. E., & Schwartz, J. 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., & Nosek, B. A. (2001). Health of the Implicit Association Test at age 3. *Zeitschrift für Experimentelle Psychologie, 48,* 85–93.

Häcker, H., Schwenkmezger, P., & Utz, H. (1979). Über die Verfälschbarkeit von Persönlichkeitsfragebogen und objektiven Persönlichkeitstests unter SD-Instruktion und in einer Auslesesituation [The effects of S-D instructions and selection conditions on the fakability of personality questionnaires and objective personality tests]. *Diagnostica, 25,* 7–23.

Hock, M., Kohlmann, C.-W., & Egloff, B. (2002). *Anxiety, coping, and performance: The role of trait and state variables.* Manuscript in preparation.

Hodges, W. F. (1976). The psychophysiology of anxiety. In M. Zuckerman & C. D. Spielberger (Eds.), *Emotions and anxiety: New concepts, methods, and applications* (pp. 175–194). Hillsdale, NJ: Erlbaum.

Holden, R. R., Wood, L. L., & Tomashewski, L. (2001). Do response time limitations counteract the effect of faking on personality inventory validity? *Journal of Personality and Social Psychology, 81,* 160–169.

Karpinski, A., & Hilton, J. L. (2001). Attitudes and the Implicit Association Test. *Journal of Personality and Social Psychology, 81,* 774–788.

Kihlstrom, J. F. (1987, September 18). The cognitive unconscious. *Science, 237,* 1445–1452.

Kim, D.-Y. (2001). *Voluntary controllability of implicit cognition: Can an implicit measure (the IAT) of attitudes be faked?* Manuscript submitted for publication.

Krohne, H. W., & Laux, L. (Eds.). (1982). *Achievement, stress, and anxiety.* Washington, DC: Hemisphere/McGraw-Hill.

Laux, L., Glanzmann, P., Schaffner, P., & Spielberger, C. D. (1981). *Das State-Trait-Angstinventar (STAI)* [The State–Trait Anxiety Inventory]. Weinheim, Germany: Beltz.

Levitt, E. E. (1967). *The psychology of anxiety.* New York: Bobbs-Merill.

Lowery, B. S., Hardin, C. D., & Sinclair, S. (2001). Social influence effects on automatic racial prejudice. *Journal of Personality and Social Psychology, 81,* 842–855.

MacLeod, C., Mathews, A., & Tata, P. (1986). Attentional bias in emotional disorders. *Journal of Abnormal Psychology, 95,* 15–20.

Markus, H. (1977). Self-schemata and processing information about the self. *Journal of Personality and Social Psychology, 35,* 63–78.

McClelland, D. C., Koestner, R., & Weinberger, J. (1989). How do self-attributed and implicit motives differ? *Psychological Review, 96,* 690–702.

McConnell, A. R., & Leibold, J. M. (2001). Relations among the Implicit Association Test, discriminatory behavior, and explicit measures of racial attitudes. *Journal of Experimental Social Psychology, 37,* 435–442.

McCrae, R. R., & Costa, P. T. (1983). Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology, 51,* 882–888.

Mierke, J., & Klauer, K. C. (2001). Implicit association measurement with the IAT: Evidence for effects of executive control processes. *Zeitschrift für Experimentelle Psychologie, 48,* 107–122.

Monti, P. M., Boice, R., Fingeret, A. L., Zwick, W. R., Kolko, D., Munroe, S., & Grunberger, A. (1984). Midi-level measurement of social anxiety in psychiatric and non-psychiatric samples. *Behaviour Research and Therapy, 22,* 651–660.

Murray, H. A. (1943). *Thematic Apperception Test manual.* Cambridge, MA: Harvard University Press.

Musch, J., Brockhaus, R., & Bröder, A. (2002). Ein Inventar zur Erfassung von zwei Faktoren sozialer Erwünschtheit [An inventory for the assessment of two factors of social desirability]. *Diagnostica, 48,* 121–129.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84,* 231–259.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002a). Harvesting implicit group attitudes and beliefs from a demonstration website. *Group Dynamics, 6,* 101–115.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002b). Math = male, me = female, therefore math ≠ me. *Journal of Personality and Social Psychology, 83,* 44–59.

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist, 17,* 776–783.

Ottaway, S. A., Hayden, D. C., & Oakes, M. A. (2001). Implicit attitudes and racism: Effects of word familiarity and frequency on the Implicit Association Test. *Social Cognition, 19,* 97–144.

Paulhus, D. L. (1984). Two component models of socially desirable responding. *Journal of Personality and Social Psychology, 46,* 598–609.

Paulhus, D. L. (1998). *The Balanced Inventory of Desirable Responding.* Buffalo, NY: Multi-Health Systems.

Phelps, E. A., O'Connor, K. J., Cunningham, W. A., Funayama, E. S.,

Gatenby, J. C., Gore, J. C., & Banaji, M. R. (2000). Performance on indirect measures of race evaluation predicts amygdala activation. *Journal of Cognitive Neuroscience, 12,* 729–738.

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

Rosenberg, M. J. (1969). *The conditions and consequences of evaluation apprehension.* In R. Rosenthal & R. L. Rosnow (Eds.), Artifact in behavioral research (pp. 279–349). New York: Academic Press.

Rosse, J. G., Stecher, M. D., Miller, J. L., & Levin, R. A. (1998). The impact of response distortion on preemployment personality testing and hiring decisions. *Journal of Applied Psychology, 83,* 634–644.

Rothermund, K., & Wentura, D. (2001). Figure-ground-asymmetries in the Implicit Association Test (IAT). *Zeitschrift für Experimentelle Psychologie, 48,* 94–106.

Rudman, L. A., Ashmore, R. D., & Gary, M. L. (2001). "Unlearning" automatic biases: The malleability of implicit prejudice and stereotypes. *Journal of Personality and Social Psychology, 81,* 856–868.

Rudman, L. A., & Glick, P. (2001). Prescriptive gender stereotypes and backlash toward agentic women. *Journal of Social Issues, 57,* 743–762.

Schmidt, R. F., & Thews, G. (Eds.). (1987). *Physiologie des Menschen* [Human physiology] (23rd ed.). Berlin, Germany: Springer-Verlag.

Seipp, B. (1991). Anxiety and academic performance: A meta-analysis of findings. *Anxiety Research, 4,* 27–41.

Smith, R. C. (1974). Response bias in the State–Trait Anxiety Inventory: Detecting the exaggeration of stress. *Journal of Psychology, 86,* 241–246.

Spalding, L. R., & Hardin, C. D. (1999). Unconscious unease and self-handicapping: Behavioral consequences of individual differences in implicit and explicit self-esteem. *Psychological Science, 10,* 535–539.

Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. (1970). *STAI Manual for the State–Trait Anxiety Inventory.* Palo Alto, CA: Consulting Psychologists Press.

Steyer, R., Schmitt, M., & Eid., M. (1999). Latent state–trait theory and research in personality and individual differences. *European Journal of Personality, 13,* 389–408.

Stöber, J. (2001). The Social Desirability Scale-17 (SDS-17): Convergent validity, discriminant validity, and relationship with age. *European Journal of Psychological Assessment, 17,* 222–232.

Teachman, B. A., Gregg, A. P., & Woody, S. R. (2001). Implicit associations for fear-relevant stimuli among individuals with snake and spider fears. *Journal of Abnormal Psychology, 110,* 226–235.

Viswesvaran, C., & Ones, D. S. (1999). Meta-analyses of fakability estimates: Implications for personality measurement. *Educational and Psychological Measurement, 59,* 197–210.

Weinberger, D. A., Schwartz, G. E., & Davidson, R. J. (1979). Low-anxious, high-anxious, and repressive coping styles: Psychometric patterns and behavioral and physiological responses to stress. *Journal of Abnormal Psychology, 88,* 369–380.

Westhoff, K. (1989). Übungsabhängigkeit von Leistungen in Konzentrationstests. [Effect of practice on performance in concentration tests] *Diagnostica, 35,* 122–130.

Wilson, T. D., Lindsley, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review, 107,* 101–126.

Wittenbrink, B., Judd, C. M., & Park, B. (2001). Spontaneous prejudice in context: Variability in automatically activated attitudes. *Journal of Personality and Social Psychology, 81,* 815–827.

Zeidner, M. (1998). *Test anxiety: The state of the art.* New York: Plenum.

Appendix

Items of the IAT–Anxiety

Category label			
Me	Others	Anxiety	Calmness
I	They	Nervous	Relaxed
Self	Them	Afraid	Balanced
My	Your	Fearful	At ease
Me	You	Anxious	Calm
Own	Others	Uncertain	Restful

Note. The original German items can be obtained from us on request.

Received February 27, 2002
Revision received May 28, 2002
Accepted May 30, 2002 ■