Measuring Task-Switching Ability in the Implicit Association Test

Mitja D. Back, Stefan C. Schmukle, and Boris Egloff

Johannes Gutenberg University Mainz, Germany

Abstract. Recently, the role of method-specific variance in the Implicit Association Test (IAT) was examined (McFarland & Crouch, 2002; Mierke & Klauer, 2003). This article presents a new content-unspecific control task for the assessment of task-switching ability within the IAT methodology. Study 1 showed that this task exhibited good internal consistency and stability. Studies 2–4 examined method-specific variance in the IAT and showed that the control task is significantly associated with conventionally scored IAT effects of the IAT-Anxiety. Using the D measures proposed by Greenwald, Nosek, and Banaji (2003), the amount of method-specific variance in the IAT-Anxiety could be reduced. Possible directions for future research are outlined.

Keywords: IAT, method-specific variance, task-switching

The Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998) measures strengths of associations between concepts by comparing response times in 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. Developed in the field of attitude research, the IAT was adapted to measure personality traits such as self-esteem (Greenwald & Farnham, 2000), shyness (Asendorpf, Banse, & Mücke, 2002), and anxiety (Egloff & Schmukle, 2002). The IAT-Anxiety, for example, combines the task of categorizing *anxiety* versus *calmness* words with classifying items (e.g., my or they) into *self* and *other* categories. An IAT effect indicating implicit anxiety is computed as the difference in mean categorization latency when self and anxiety share the same response key (self + anxiety) as compared to self + calmness. The IAT effect thus measures how much easier it is for participants to categorize self items with anxiety items than self items with calmness items (Egloff & Schmukle, 2002).

The IAT exhibited – in contrast to other implicit measures – good internal consistency (e.g., Banse, Seise, & Zerbes, 2001; Bosson, Swann, & Pennebaker, 2000; Egloff & Schmukle, 2002). Even more important, the IAT showed incremental validity in the prediction of spontaneous behavior (Asendorpf et al., 2002; Egloff & Schmukle, 2002; McConnell & Leibold, 2001; see Fazio & Olson, 2003, for a recent review). These impressive demonstrations of reliability and validity notwithstanding, several important issues concerning the processes that underlie the IAT effect await further scrutiny. Hypotheses concerning how the IAT works and what it measures include assumptions of automatic associations (Greenwald et al., 1998; Greenwald & Nosek, 2001), response criterion shifts (Brendl, Markman, & Messer, 2001), response conflicts (De Houwer, 2001), environmental associations (Karpinski & Hilton, 2001), figure-ground asymmetries (Rothermund & Wentura, 2001, 2004), and task-set switching (Mierke & Klauer, 2001, 2003). The latter theoretical account will be used in this study and is explained in detail below.

Task-Switching and Method-Specific Variance in the IAT

In their task-switching cost model, Mierke and Klauer (2001) assume that the switching between different

Experimental Psychology 2005; Vol. 52(3):167–179
DOI: 10.1027/1618-3169.52.3.167

task sets – the categorization of items into target concepts vs. the categorization of items into attribute concepts – is the crucial mechanism underlying the IAT effect. A task-set is a complex of cognitive settings required for performing a given task, including "which attribute of the stimulus to attend to, which response mode and value to get ready, what classification of the relevant stimulus attribute to perform, how to map those classes to response values, with what degree of caution to set one's criterion for response etc." (Monsell, Yeung, & Azuma, 2000, p. 252).

Switching between task sets requires time and, thus, leads to a performance cost (Meiran, 1996; Monsell et al., 2000; Rogers & Monsell, 1995; Rubinstein, Meyer, & Evans, 2001). It is assumed that task-switching involves executive control processes (see Rubinstein et al. 2001, for an overview and Wylie and Allport, 2000, for an alternative conceptualization of task-switching). According to Rubinstein et al., these executive control processes involve goal-shifting, rule-activation and rule-inhibition. In order to switch between task sets, participants need to actualize the goal of the trial (e.g., "goal is to discriminate between anxiety vs. calmness words"), to activate the appropriate rules for the task after the identification of the stimulus (e.g., "anxiety word requires reaction with the left response key, calmness word requires reaction with the right key"), and to inhibit rules that belong to the other task and whose activation could interfere with the processing of the actual task (e.g., "anxious has nothing to do with the category *self* (maybe even though I am anxious)"). Task-switching costs reflect the duration of these executive control processes.

Crucial for the understanding of the IAT is that these executive control processes are assumed to be needed more in the incompatible than in the compatible condition of the IAT. Consequently, performance costs associated with task-set switching affect the IAT conditions asymmetrically, and thus affect the IAT score itself. In the compatible condition of an IAT-Anxiety, for example, in the case of a low anxious individual in the condition where *self* and *calmness* share one and *others* and *anxiety* share the other response key, the task is very easy. Concepts that share the same response key overlap to a high degree, so that the task could be easily simplified by reacting according to a *self–other* or a *positive–negative* rule in every trial. In this case, few executive control processes are needed, resulting in lower task-switching costs and faster responses in the compatible condition.

However, in the incompatible condition, for example, in the case of a low anxious individual when *self* and *anxious* share one and *others* and *calmness* share the other response key, the task is much more difficult. The concepts that share one response key are not at all or only slightly associated. Furthermore, each concept is strongly associated with another concept that is assigned to the opposite response key. As participants have no possibility to simplify the exercise, they have to distinguish between two different tasks. Executive control processes are needed to realize which task is to be executed and to ignore or actively inhibit information that is related to the other task. As more executive control processes are needed in the incompatible condition, task-switching costs are higher, leading to slower responses.

Mierke and Klauer (2001) have shown that the difference in task-switching costs is an appropriate model to explain the difference in reaction times between the compatible and the incompatible condition. As the compatibility of the two critical conditions in a personality IAT varies as a function of the participant's personality, the difference of both blocks in the need for task-switching varies, resulting in different IAT effects. A central question for understanding the processes underlying the IAT is whether task-switching costs play a *mediational* or a *contaminating* role (Mierke & Klauer, 2003). In the first case, differences in task-switching costs are just a consequence of differences in the construct to be measured. In the latter case, in contrast, task-switching costs vary independently of the construct to be measured. This contaminating role of task-switching ability constitutes a possible source of method-specific variance in the IAT. Imagine two equally non-anxious participants who differ in their task-switching abilities, e.g., in the ability to inhibit irrelevant information: As they possess the same implicit anxiety-relevant associations, they have to deal with the same degree of incompatibility in the self + anxiety condition. Nevertheless, the participant with higher task-switching abilities will reach a smaller absolute IAT effect. Thus, content-specific IAT scores are contaminated with individual differences in the ability to switch between task sets. McFarland and Crouch (2002) refer to this issue as a cognitive skill confound on the IAT, whereas Mierke and Klauer (2003) call it reliable contamination.

Measuring Task-Switching Ability

In their initial IAT study, Greenwald et al. (1998; Experiment 1) tested two nonsocial (irrelevant) IATs – a flower-insect and a musical instrument-weapon IAT. They reported that these two nonsocial IATs showed a correlation of .58 and suggested that this may reflect systematic method variance. McFarland and Crouch (2002) argued that an association between two content

IATs that are independent with respect to the construct to be measured is due to a general cognitive skill confound. Following this line of reasoning, they used two irrelevant content IATs as control tasks to measure the "general cognitive skill of how quickly one can correctly categorize exemplars presented in incongruent categories as compared to when they are presented in congruent ones" (p. 483). The first control task used the categories delicious (e.g., "tasty") – not-delicious ("rancid") and happiness ("optimism") – unhappiness ("hopeless"). The well-known flower-insect IAT was chosen as a second control task. McFarland and Crouch found significant correlations between both control tasks and IATs assessing prejudice and self-esteem. In general, individuals with higher IAT effects in the control IATs showed also higher implicit prejudice and self-esteem scores, indicating lower task-switching ability and, thus, confirming the assumption of a cognitive skill confound on the IAT.

To analyze the same issue, Mierke and Klauer (2003) developed an IAT which – in contrast to the control IATs of McFarland and Crouch (2002) – is not based on preexisting associations between target categories and attributes. Thus, a resulting IAT effect could not be explained by individual differences in underlying associations. In their geometrical IAT, simple geometrical objects (rectangles, triangles, and circles) were used as stimuli. Participants had to distinguish between red and blue objects (target categories) and between small and large objects (attribute categories). The stimuli representing the attribute thereby belonged to neither of the target categories, that is, they were neither red nor blue, but colored in one of three alternative colors. An association between target and attribute categories was created experimentally by imposing a contingency between size and color of the stimuli: Whereas all red objects were small, all objects of the target category "blue" were large.

Mierke and Klauer were able to show that task-switching costs play, at least partly, a contaminating role. Substantial and significant correlations of the geometrical IAT with a conventional flower-insect attitude IAT and an extraversion IAT showed that standard content IATs are reliably contaminated with the effect of task-switching costs. Mierke and Klauer concluded that conventionally scored IAT effects generally contain both stable content-specific and stable method-specific variance. The latter variance is assumed to be due to stable individual differences in task-switching costs, i.e., task-switching ability. Surprisingly, the amount of method-specific variance could be substantially reduced by using the new scoring procedures (D measures) that are based on scaling the IAT effects in units of the individuals' standard deviations (Greenwald, Nosek, & Banaji, 2003). In fact, no significant associations between the control task and both the extraversion and the flower-insect IAT were found when the content-specific IATs were scored according to the improved algorithm.

Interestingly, Mierke and Klauer (2003) used the flower-insect IAT as a criterion for their geometrical control IAT whereas McFarland and Crouch (2002) used the flower-insect IAT as a control task itself. This indicates the difficulty to define and create "pure" control IATs. On the one hand, the flower-insect IAT constitutes a control task because its IAT effect incorporates variance due to task-switching ability. On the other hand, it can be regarded as a content-specific IAT because there are individual differences in the preference for flowers as compared to insects. Generally speaking, a control IAT should measure content-unspecific abilities that are relevant during the completion of IATs (e.g., task-switching ability). To the extent that the effects of a control IAT are based on individual differences in associations between the categories, this task is itself contaminated.

With their geometrical IAT, Mierke and Klauer (2003) solved this problem thoughtfully by experimentally inducing associations not existing a priori. The question here is if participants differ in the strength of associations they build during the IAT procedure. The more they do so, the less the control IAT effect is solely dependent on task-switching ability. Another aim is to parallelize the procedure of a standard IAT more directly by (a) using verbal material or characters, respectively, instead of geometrical objects and (b) relying on preexisting associations.

Independently of the control IATs described above, we developed an IAT that aims at measuring task-switching ability, thereby structurally resembling the critical mechanism of a content-specific IAT very closely. In this task-switching ability IAT (TSA-IAT) participants have to sort stimuli from *letter* (e.g., N, B) and *number* categories (e.g., 5, 8) as well as stimuli from *word* (e.g., shirt, pen) and *calculation* categories (e.g., 7 – 4 = 3, 4 + 5 = 9). Thus, each target category is clearly associated with one and independent of the other attribute category: Letter is associated with word but independent of calculation whereas number is associated with calculation but independent of word. TSA-IAT effects are computed by subtracting performance in the letter + word condition from performance in the letter + calculation condition. These scores are highly polarized: The letter + word condition is the compatible one, whereas the letter + calculation condition is the incompatible one for each participant. For these reasons, abilities concerning the use of the necessary executive control processes for task-switching have a unidirectional effect on TSA-IAT scores. Categories of the TSA-IAT were chosen to minimize individual differences in the strength of preexisting associations. Thus, the task should mainly measure task-switching ability. Furthermore, as for other control IATs, scores on the TSA-IAT should be independent of individual differences in personality self-concepts. Table 1 summarizes relevant features of different control IATs.

Goals of this Research

The hypothesis of reliable contamination bases on the assumption that a general ability is confounded with the IAT effect. According to the task-switching account (Mierke & Klauer, 2001), this general ability is the ability to switch between task sets. In this article, we first examine internal consistency and retest stability of an IAT for measuring this task-switching ability. Studies 2–4 then analyze whether the TSA-IAT correlates with the IAT-Anxiety. Furthermore, we examine whether the association between content-specific and control IATs is reduced when the improved scoring algorithm (Greenwald et al., 2003) is used.

Study 1

Study 1 aimed at examining internal consistency and stability of the TSA-IAT. Because the TSA-IAT is assumed to measure an ability, we used a comparatively long interval of five months between test and retest.

Method

Participants

Thirty-two volunteers (28 women and 4 men) participated in exchange for research participation credit or monetary compensation. Their average age at the first occasion of measurement was 23.13 years (SD = 5.06).

Procedure and Design

TSA-IAT. Each IAT was administered on a personal computer with the program Inquisit (Draine, 2001). It consisted of stimuli from *letter* (N, K, B, R, E) and *number* (7, 3, 8, 5, 2) categories as well as items from *word* (pen, shirt, telephone, spoon, wall) and *calculation* (1 + 3 = 4, 7 – 4 = 3, 2 + 2 = 4, 4 + 5 = 9, 8 – 6 = 2) categories. The IAT procedure comprises the usual five blocks. Participants practiced the discrimination of letter and number items (target discrimination) in the first block that comprised 20 trials (each item was presented twice). The same was done for the attribute discrimination by sorting items into *word* and *calculation* categories in Block 2 and for practicing the switched key assignment in Block 4

Table 1. Features of Control IATs.

Feature	"irrelevant IATs" (McFarland & Crouch, 2002)	"geometrical IAT" (Mierke & Klauer, 2003)	"TSA-IAT"
Material	Words with different valence (e.g., "cookies," "liver," "joy," "misery")	Geometrical objects (e.g., large blue circle, small red rectangle, large pink triangle, small green circle)	Neutral characters and words (e.g., "B," "5," "spoon," "4 + 5 = 9")
Target concepts	1: Delicious – Not delicious 2: Flower – Insects	Blue objects – Red objects (all blue objects are large, and all red objects are small)	Letter – Number
Attribute concepts	1: Happy – Unhappy 2: Pleasant – Unpleasant	Large objects – Small objects (other colors than red and blue)	Word – Calculation
Associational mechanism	Preexisting associations between concrete concepts with moderate strength; preferred associations between each target with one attribute concept	Experimentally imposed associations; no preexisting associations	Preexisting associations between abstract concepts with maximal strength; clear association between each target concept with one but not the other attribute concept
Individual differences in associations	Moderate differences in preexisting associations	Individual differences in the strength of experimentally imposed associations?	Minimal individual differences in the strength and no differences in the direction of the preexisting associations

(20 trials 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. The block in which letter and word share one response key was considered to be compatible, whereas we refer to the letter-calculation block as the incompatible block. Compatibility order (compatible vs. incompatible mapping first) for the first and second IAT was counterbalanced between participants.

Participants were told they would be making a series of category judgments. On each trial, a stimulus was displayed in the center of a computer screen. Category labels were displayed on the left and right sides of the window. Participants used the letter "A" on the left side of the keyboard and the number "5" of the right-side numeric keypad for their responses. They were told, "Please try to be as accurate as possible, while also being as quick 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 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.

Data reduction. We used two different procedures for computing IAT scores, the conventional algorithm (Greenwald et al., 1998) and the improved scoring algorithm (Greenwald et al., 2003). The *conventional algorithm* included (a) dropping the first two trials of the 60 critical trials in Blocks 3 and 5, leading to 58 trials per block, (b) recoding trials with latencies less than 300 ms or greater than 3,000 ms to 300 ms or 3,000 ms, respectively, (c) log-transforming the resulting values prior to averaging, and d) computing the IAT effect for task-switching ability by subtracting the mean score in the critical test trials of the letter + word condition from the critical test trials of the letter + calculation condition.

The first variation of the *improved scoring algorithm* (so-called D_1 measure) consists of (a) eliminating trials with latencies greater than 10,000 ms, (b) including error trials in the analysis by using the latency until the correct response was given (built-in error penalty), (c) subtracting the mean latency in the critical trials of the letter + word condition from the critical trials of the letter + calculation condition separately for test and practice trials, (d) dividing these differences by the individual-respondent standard deviations of reaction times in test and practice trials, respectively, (e) computing the IAT effect as the weighted average of these two scores. Note that Greenwald et al. (2003) recommended using an unweighted average of both scores. We preferred to use the weighted average in our analyses because otherwise the (20) practice trials would have been three times more important than the (60) test trials. Throughout this article, we performed all analyses with log transformed values and the new D_1 measure.[1] For presentation purposes, average IAT effects were also reported in milliseconds.

The TSA-IAT was administered twice with a time lag of approximately five months (M = 164 days, SD = 13 days). Self-generated unique code numbers of each participant assured proper combination of the data from both measurement occasions.

Results

Mean IAT scores were positive for the first (M = 233 ms, SD = 101 ms) and the second (M = 244 ms, SD = 111 ms) measurement occasion (range: 35 ms to 502 ms). To compute internal consistencies of log-scored TSA-IAT effects, we first separately subtracted each trial's (log transformed) response latency of the second critical IAT condition from the response latency of the corresponding trial of the first critical IAT condition (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) for each IAT. Internal consistency was satisfactory for the first (α = .76) and the second (α = .80) IAT. We estimated the reliability of improved scored TSA-IAT effects by applying the algorithm separately to two mutually exclusive subsets of the IATs combined-task trials. The Spearman-Brown adjusted split-half correlation was .74 for the first and .76 for the second measurement occasion.

As expected, the correlation between both IATs was significant and comparatively high (r = .63, $p < .001$ for log scored and r = .62, $p < .001$ for improved scored IAT effects). Interestingly, the retest reliability of the TSA-IAT over a time period of five months was similar to that found after an immediate retest (Mierke & Klauer, 2003). Additionally, this finding is

[1] In comparison Mierke and Klauer (2003) scored their control task only conventionally. They argued that a control task should maximize the amount of method-specific variance. Indeed, the use of a scoring procedure that is thought to reduce method-specific variance seems to be at odds with that goal. Nevertheless we use both scoring procedures for explorative purposes.

comparable to retest coefficients reported for content-specific IATs (Egloff, Schwerdtfeger, & Schmukle, in press).

Study 2

If task switching constitutes an essential mechanism for the occurrence of content-specific IAT effects, one should expect correlations between content-specific IATs and IATs that measure task-switching ability. Mierke and Klauer (2003) showed that a content-unspecific geometrical IAT shared variance with a flower-insect IAT and a self-extraversion IAT, if the latter ones were conventionally scored. However, method-specific variance could be remarkably reduced by scoring the content IAT according to the improved scoring algorithm. In this study, we analyzed the association among the TSA-IAT and a personality IAT, the IAT-Anxiety (Egloff & Schmukle, 2002).

Method

Participants

Eighty-three students (63 women and 20 men) of Johannes Gutenberg University Mainz participated in this study in exchange for research participation credit or monetary compensation. Their average age was 23.78 years (*SD* = 4.82).

Procedure and Design

TSA-IAT. Procedure, design, and data reduction of the TSA-IAT were identical to those of Study 1, with the exception that a fixed compatibility order was used in Study 2 (Block 3: letter–word; Block 5: letter–calculation).

IAT-Anxiety. The IAT-Anxiety consisted of stimuli from *self* (me, my, own, I, self) and *other* (they, your, them, you, others) categories as well as of items from *anxiety* (nervous, afraid, fearful, anxious, uncertain) and *calmness* (relaxed, balanced, at ease, calm, restful) categories. The IAT procedure comprises five blocks. Participants practiced the discrimination of self and other items (target discrimination) in the first block that comprised 20 trials (each item was presented twice). The same was done for the attribute discrimination by sorting items into *anxiety* and *calmness* categories in Block 2 and for practicing the switched key assignment in Block 4 (20 trials 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. The order in which the self + anxiety and self + calmness conditions were to be performed was fixed (Block 3: self + calmness; Block 5: self + anxiety). Instruction, procedure, and data reduction of the IAT-Anxiety were the same as for the TSA-IAT in Study 1. An IAT-Anxiety score was computed by subtracting the performance in the self + anxiety condition from the performance in the self + calmness condition. To test for method-specific variance, we used the absolute magnitude of the IAT-Anxiety scores (see Mierke & Klauer, 2003). This is necessary because the compatibility of the two critical conditions in a personality IAT – and thus the direction of possible correlations with control task measures – varies as a function of the participant's personality.

All tasks were administered in individual experimental sessions with a fixed presentation order (IAT-Anxiety, TSA-IAT).

Results

Descriptive Statistics

Means, standard deviations, and internal consistencies of each measure are displayed in Table 2. Internal consistencies were computed as described in Study 1.

Main Analysis

To test for method-specific variance, the absolute magnitude of the IAT-Anxiety effect was correlated with the conventionally scored TSA-IAT effect. As can be seen in Table 3, this correlation was significantly positive, $r = .35$, $p = .001$, documenting that participants with larger effects in the control task also showed larger effects in the IAT-Anxiety. The magnitude of this correlation is comparable to that reported by Mierke and Klauer (2003), replicating the finding of substantial method-specific variance in a conventionally scored trait IAT. Interestingly, the association between both IATs was only slightly reduced when the D_1 measure was used ($r = .31$, $p = .004$). Both correlation coefficients did not differ significantly ($z = .69$; see Steiger, 1980). When both IATs were scored according to the new scoring algorithm, a correlation of $r = .32$, $p = .004$ was found.

Table 2. Descriptive Statistics of the IAT Measures (Studies 2, 3 and 4).

	IAT effect							
	Milliseconds		log transformed			D_1 measure		
Measure	*M*	(*SD*)	*M*	(*SD*)	α	*M*	(*SD*)	α
Study 2								
TSA-IAT	311	(147)	.425	(.155)	.88	1.000	(.228)	.84
IAT-Anxiety	–175	(131)	–.187	(.127)	.79	–.477	(.274)	.84
Study 3								
TSA-IAT	222	(113)	.326	(.131)	.82	.885	(.248)	.74
IAT-Anxiety	–67	(86)	–.088	(.094)	.69	–.239	(.287)	.73
Study 4								
TSA-IAT	339	(165)	.436	(.177)	.88	.909	(.241)	.77
IAT-Anxiety	–97	(121)	–.103	(.112)	.73	–.282	(.276)	.74

Note. $N = 83$ for Study 2, $N = 57$ for Study 3, $N = 62$ for Study 4. IAT = Implicit Association Test. TSA = task-switching ability. For details of the respective computation procedures see text.

Table 3. Correlations of task-switching ability with differently scored IAT-anxiety measures.

	IAT-Anxiety			
Task-switching ability (method-specific variance)	log	D_1	*z*	*p*
Study 2	.35**	.31**	0.69	.490
Study 3	.23⁺	.07	2.75	.006
Study 4	.27*	.08	2.21	.027
Studies 2–4 combined	.29***	.17**	3.09	.002

Note. $N = 83$ for Study 2, $N = 57$ for Study 3, $N = 62$ for Study 4. IAT = Implicit Association Test. IAT-Anxiety scores refer to the absolute, unsigned values. log = values based on log-transformed IAT measures. D_1 = values based on IAT measures scored according to the first variation of the improved scoring algorithms (D_1 measure). For details of the respective computation procedures, see text.
$^+ p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$ (two-tailed).

Study 3

As pointed out by Mierke and Klauer (2003), the finding of method-specific variance in the IAT is of importance with regard to several issues: First, the high internal consistency and the medium-sized stability coefficients of content IATs might be partly due to individual ability differences. Thus, psychometric properties of content IATs might be overestimated because of reliable content-unspecific variance due to method factors (see Egloff et al., in press; Schmukle & Egloff, in press; Steffens & Buchner, 2003). Second, associations among different content IATs (see Gawronski, 2002) might be inflated due to method variance: If there is a base rate correlation of about .30 among IATs, irrespective of content, convergent validity might be overestimated and discriminant validity might be underestimated (see also McFarland & Crouch, 2002). Third, associations with criterion variables such as behavior observations or self-report measures might be enhanced by controlling for or reducing method-specific variance. For these reasons, a scoring procedure that is able to reduce method specific variance would be very helpful. However, in contrast to the results reported by Mierke and Klauer, the D_1 measure was not able to remove a substantial amount of method-specific variance in a trait IAT in Study 2. Study 3 aimed at replicating this finding.

Method

Participants

Fifty-seven students (46 women and 11 men) of Johannes Gutenberg University Mainz participated in this study in exchange for research participation credit or monetary compensation. Their average age was 23.14 years ($SD = 5.01$).

Procedure and Design

Each participant performed the TSA-IAT and the IAT-Anxiety. Procedure and design of the IATs were identical to those of Study 2 with the exception that the compatibility order was counterbalanced between participants for both IATs. The measures were administered in a fixed order (IAT-Anxiety, TSA-IAT).

Results

Descriptive Statistics

For each IAT, means, standard deviations, and internal consistencies are displayed in Table 2.

Main Analysis

Reliable contamination was tested by correlating the absolute magnitude of the IAT-Anxiety with the conventionally scored TSA-IAT (see Table 3). This correlation was marginally significant when conventional scoring was used ($r = .23$, $p = .088$). This time, however, the improved scoring algorithm removed method-specific variance as indicated by the low correlation between the TSA-IAT and the improved scored IAT-Anxiety ($r = .07$, *ns.*). The correlation between TSA-IAT and improved scored IAT-Anxiety was significantly lower than that between TSA-IAT and log-scored IAT-Anxiety ($z = 2.75$). The correlation between both improved scored IAT measures was also very low, $r = .02$, *ns*.

Study 4

The results of Studies 2 and 3 converge by showing evidence for method-specific variance in conventionally scored trait IATs. However, both studies diverge with respect to the effects of the improved scoring procedure on reducing this contamination: Whereas the D_1 measure was able to remove method-specific variance in Study 3, this effect was absent in Study 2. For this reason, we conducted another study to further examine this issue. Additionally, Study 4 explored if the validity of the IAT can be enhanced by controlling for method-specific variance. This should be the case under the (in most cases very plausible) assumption that method-specific variance in the IAT is unrelated to variance in the respective criterion. Accordingly, Mierke and Klauer (2003) could show that statistically controlling for task-switching ability – as measured by their geometrical IAT – increased the association between an extraversion IAT and an explicit extraversion measure. However, the improvement was rather small. To examine a possible enhancement of the validity of the IAT-Anxiety by using the TSA-IAT, participants also explicitly rated the IAT-Anxiety items in Study 4. We then analyzed the association among implicit and explicit anxiety measures with and without controlling for method-specific variance.

Method

Participants

Sixty-two students (34 women and 28 men) of Johannes Gutenberg University Mainz participated in this study in exchange for research participation credit or monetary compensation. Their average age was 22.05 years (*SD* = 2.29).

Procedure and Design

Each participant completed the TSA-IAT and the IAT-Anxiety. Procedure and design of the IATs were identical to those of Study 3. Additionally, participants completed 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 reversed scored. Thus, the explicit rating of the IAT stimuli could vary between 0 (no anxiety) and 50 (high anxiety). The measures were administered in a fixed order (IAT-Anxiety, TSA-IAT, rating of IAT items).

Results

Descriptive Statistics

For each IAT measure, means, standard deviations, and internal consistencies are displayed in Table 2. Self-reported trait-anxiety showed a mean of 20.45 (*SD* = 7.25) and an internal consistency of d = .84.

Main Analysis

Reliable contamination was tested by correlating the absolute magnitude of the IAT-Anxiety with the conventionally scored TSA-IAT (see Table 3). As in Studies 2 and 3, both measures were significantly associated when the conventional scoring algorithm was used ($r = .27$, $p = .033$). TSA-IAT and improved scored IAT-Anxiety correlated $r = .08$, *ns*. Thus, improved scoring removed method-specific variance: The correlation between conventionally scored TSA-IAT and improved scored IAT-Anxiety was again significantly lower than that between log scored IAT-Anxiety and TSA-IAT, $z = 2.21$. The correlation between both improved scored measures was also very low, $r = -.02$, *ns*.

To test for construct-specific variance, we correlated the signed IAT-Anxiety scores with the explicit anxiety measure. A small but significant correlation was found for conventional scoring ($r = .27$, $p = .033$) confirming the findings of a small amount of common variance between explicit and implicit trait measures (e.g., Bosson et al., 2000; Egloff & Schmukle, 2002, 2003; Mierke & Klauer, 2003). As they share no content- or method-specific variance, the TSA-IAT and the explicit measure were not significantly related ($r = -.07$). Thus, one can assume that controlling for method-specific variance should increase the association between the explicit anxiety measure and the log-scored IAT-Anxiety due to a smaller amount of error variance in the implicit anxiety measure. However, construct-specific variance was not higher when improved scoring was used ($r = .26$, $p = .045$). Similarly, controlling for task-switching ability did not markedly increase the association among the explicit and the implicit anxiety measure ($r = .28$, $p = .027$).[2]

Thus, Study 4 confirms the finding of Mierke and Klauer (2003) that the amount of additional criterion variance that could be explained by controlling for or reducing method-specific variance might be quite limited. In our view, the evaluation of the improved scoring algorithm and of the effect of controlling for method-specific variance with respect to predictive validity would benefit from including other external criteria like, e. g., behavioral data. This approach seems even more necessary when considering the debate about the relation of implicit and explicit personality measures and, thus, about the appropriateness of using explicit measures as criteria for implicit measures.

Combined Analysis of Studies 2–4

The results of Studies 2–4 agree in that medium-sized correlations between the absolute effect of a conventionally scored IAT-Anxiety and the TSA-IAT were found in all studies. However, there seems to be an apparent inconsistency among the findings regarding the suitability of the D measure for removing method-specific variance: When applying improved scoring, this contamination remained in one (Study 2) but not in the other two studies (Studies 3 and 4).

To analyze possible differences and communalities of the results across studies, we first examined the homogeneity of the correlation coefficients among the TSA-IAT and the IAT-Anxiety by applying a formula proposed by Hedges and Olkin (1985). This analysis showed that the three correlation coefficients among the TSA-IAT and the log-scored IAT-Anxiety (r values = .35, .23, and .27, see Table 3) did not differ from each other, $Q = 0.60$, *ns*. More interestingly, also the three correlations among the improved scored IAT-Anxiety and the TSA-IAT (r values = .31, .07, and .08) were homogenous, $Q = 2.95$, *ns*. We then estimated the *mean association* among method- and content-specific IATs (Hedges & Olkin, 1985): For the conventionally scored IAT-Anxiety, this correlation coefficient was $r = .29$, $p < .001$. For the improved scored IAT-Anxiety, the coefficient was $r = .17$, $p < .01$ (see last row of Table 3).

By applying Steiger's (1980) formula, we then analyzed if these average correlation coefficients differed from each other: This was indeed the case, $z = 3.09$, $p < .01$. Thus, it seems safe to conclude that IAT effects scored according to the improved algorithm are less contaminated with ability-related factors than those scored with the conventional algorithm. This pattern of results nicely resembles the findings of Mierke and Klauer (2003, Study 3), who report r values of .29 (log scored Extraversion IAT and control task) and of .12 (improved scored Extraversion IAT and control task).[3]

General Discussion

Taken together, the results of the four studies presented in the article allow the following conclusions: First, there are individual differences in task-switching performance that can be reliably measured by means

[2] According to Mierke and Klauer (2003), to control the implicit-explicit correlation (signed IAT-Anxiety score and IAT-Anxiety items) for method-specific variance one should not use the partial correlation because the correlation between signed IAT-Anxiety score and TSA score is not a suitable measure of method-specific variance. Sign differences in the IAT-Anxiety score would conceal the amount of method-specific variance. For these reasons, we regressed the absolute values of the IAT-Anxiety scores on the scores in the TSA-IAT. Regression residuals should be free from method-specific variance. To ensure the integrity of the sign information – that is removed in the regression analysis of absolute IAT-Anxiety scores – the regression residuals were shifted so that the smallest residual was zero, and multiplied by plus or minus one, depending on the sign of the original effect in the IAT-Anxiety.

[3] Consistent with findings reported by Mierke and Klauer (2003) and McFarland and Crouch (2002), the millisecond-unit measure was most strongly confounded with task switching ability (mean correlation, $r = .31$, $p < .001$). This association differed significantly from the one between task-switching ability and improved scored IAT-Anxiety ($z = 2.60$, $p = .009$), but not from the one between task-switching ability and log-scored IAT-Anxiety ($z = .50$, $p = .617$).

of the TSA-IAT. Second, we found medium-sized correlations between the absolute effect of a conventionally scored IAT-Anxiety and a content-unspecific control measure. These correlations cannot be explained by individual differences in anxiety. Thus, replicating the research of Mierke and Klauer (2003), the existence of a substantial amount of method-specific variance in a conventionally scored Trait-IAT could be confirmed. Third, when combining the data from Studies 2–4, we were able to show that the improved algorithm is able to significantly reduce this contamination. However, fourth, even improved scored trait IATs still share some variance with content-unspecific IATs, and thus contain some method variance.

Important questions for future research concern the way the new scoring algorithm works (why is it less susceptible to ability-related individual differences?), possible moderators of its effectiveness (under which circumstances does this algorithm work?), and the mechanisms that contribute to content- and method-specific portions of variance in the IAT effect. Mierke and Klauer (2003) concluded that "there is currently no clear-cut account for how the new algorithms work" (p. 1190). Thus, features of improved scoring like its greater independence from speed of responding (Greenwald et al., 2003) and age (Hummert, Garstka, O'Brien, Greenwald, & Mellott, 2002) should be further analyzed with respect to its ability to reduce contamination with task-switching ability. In a reanalysis and an additional replication of the McFarland and Crouch (2002) study, Cai, Sriram, Greenwald, and McFarland (in press) recently demonstrated the effectiveness of the improved scoring algorithm in reducing the influence of speed of responding on IAT scores. Furthermore, the correlation between control IATs and content-specific IATs could be reduced in some but not in other cases. The authors conclude that the D measures are very effective in eliminating the speed of responding confound but that there may be other sources of cognitive confounds – such as suppression skills – that the D scores may not reduce.

Using our data, we could confirm the speed-related findings of Cai et al. (in press): A measure of average response latency was computed for each respondent as an equal-weight average of mean latencies from each of the two critical data blocks of both the TSA-IAT and the IAT-Anxiety (practice as well as critical trials – involving a total of 320 trials for each study). This measure correlated with the absolute magnitude of the TSA-IAT effect and the IAT-Anxiety effect. However, the results varied between the different scoring algorithms: The mean correlations across all studies with speed of responding were $r = .58$, $p < .001$, for the TSA-IAT and $r = .48$, $p < .001$, for the IAT-Anxiety, when the IAT effect was scored in millisecond units. Whereas the conventional log-scored algorithm was still confounded with speed of responding ($r = .37$, $p < .001$, for the TSA-IAT and $r = .29$, $p < .001$, for the IAT-Anxiety), improved scoring eliminated this association ($r = -.03$, *ns,* for the TSA-IAT and $r = .05$, *ns*, for the IAT-Anxiety). The differences between the scoring algorithms concerning their correlation with speed were all highly significant for both the TSA-IAT and the IAT-Anxiety (z between 5.91 and 10.22, all p values $< .001$).

Furthermore, future research could explore the core mechanisms of the IAT in more detail by measuring the assumed processes and the associated individual differences independently of the IAT-paradigm. Concerning task-switching ability, one could search for associations between tests of executive abilities (e.g., general intelligence, measures of central executive functions, working-memory span) and the IAT effect. Another possibility would be to analyze the influence of ability differences in more specific executive processes like inhibition. For example, rule inhibition (Rubinstein et al., 2001) – to inhibit rules that belong to the other task and whose activation could interfere with the processing of the actual task – is certainly an important executive process during an IAT. Using magnetic resonance imaging data, Chee, Sriram, Soon, and Lee (2000) showed that inhibitory processes might play a role in IAT responses.

Similarly, one might use the IAT methodology and other paradigms to measure and examine the influence of other possibly contributing processes. Beside the important role of executive control processes, we assume that *implicit learning of category associations* might also influence the IAT effect. Structurally, the IAT can be interpreted as a learning task: Participants are required to learn the assignment of each category to a response key and, thus, to build up or strengthen an association between the two categories that share one response key (e.g., to strengthen the association between self and anxiety). The better participants learn this association, the faster they will react on average in each condition. On the other hand, better learned associations can be obstructive in the next critical IAT condition as the assignment of categories changes. For example, when participants have to react to self and calmness stimuli with the same response key, the task is the more difficult the more they strengthened the self–anxiety association in the preceding critical IAT condition. For these reasons, learning influences the degree of compatibility, i.e., the ease of both critical IAT conditions. Accordingly, implicit learning of category association also influences the need for executive control processes, the time required for executive control processes, and ultimately the IAT effect itself.

As in the case of executive control processes, implicit learning of category associations might also play a contaminating role: Individuals differ in various general aspects of learning, for example, how fast they build associations and how flexible or rigid they adhere to these associations. These individual differences are independent of the construct to be measured and therefore may reflect another source of method-specific variance in the IAT. Using two equally non-anxious participants as an example, they can reach different IAT-Anxiety scores – in the absence of differences in task-switching ability – if they show different learning characteristics. Consider a participant who is a fast learner and thus builds associations very quickly in Block 3 (e.g., in the self + calmness condition) and/or who is very rigid in sticking to the once learned associations in Block 5 (self + anxiety condition): S/he will receive a higher absolute IAT effect than a participant with rather slow and/or flexible learning characteristics. As a consequence, the first participant will be rated as less anxious than the second one.

Furthermore, implicit learning of category association may help to explain some of the effects involving procedural variables identified in prior IAT research (see Greenwald & Nosek, 2001), especially the order effect and the effect of previously taken IATs. The IAT effect is generally smaller when the incongruent condition precedes the congruent task than when it follows the congruent condition (order effect). In the latter case, strong preexisting associations are actualized and strengthened in Block 3 (e.g., in the self + calmness condition in the case of a low anxious participant), which makes it even more difficult to manage the task (i.e., to learn new associations) when incongruent categories share one response key in Block 5 (e.g., in the self + anxiety condition). In the case when the incongruent condition is to be performed in first place, participants have to learn the incongruent association without being opposed to the congruent association before. This should make the incongruent task easier and the IAT effect smaller. Performing the congruent task second should make it more difficult, as participants have to inhibit the previously learned new associations, also leading to a smaller IAT effect.

Recently Klauer and Mierke (in press) proposed a related explanation for compatibility-order effects. In line with their task-switching account, attitude accessibility in a compatible IAT block should be affected due to task-set inertia – a form of enduring activation and inhibition of task sets – when it follows an incompatible IAT block. They could indeed show that speed in the evaluation of stimuli belonging to the IAT target categories was reduced following an incompatible but not following a compatible flower-insect IAT block.[4]

As a second procedural effect, participants were found to show smaller IAT effects in retests. Whereas congruent (i.e., already well-learned) associations can only be strengthened until a certain degree (a ceiling effect), the amount of learning regarding the new incongruent associations is much greater. Thus, the building and strengthening of associations over the course of IAT sessions affects the incongruent associations more strongly than the congruent ones, leading to smaller IAT effects. One reviewer suggested that IAT magnitudes may be reduced on second IATs even when they are not the same IAT as the first one. Thus, learning during an IAT session may function on a more general level – and may not be restricted to the specific associations in an IAT.

To conclude, method-specific variance plays a critical role in conventionally scored IATs. This influence can be reduced when the improved scoring algorithm is used. Examining the mechanism of how the improved algorithm reduces variance due to task-switching ability is certainly an important goal for future research. Eventually, the analysis of sources of method-specific variance in the IAT might lead to a better understanding of how the IAT functions.

Acknowledgement

This research was supported by Deutsche Forschungsgemeinschaft (German Research Foundation) Grant EG 143/2-1. We thank Jennifer Dickes for her help with data collection.

[4] In our view it might be difficult to distinguish empirically between these two conceptualizations ("implicit learning of category associations" and "modified attitude accessibility due to task-set inertia"). At least, the assumption of an "implicit learning of category associations" is in line with the recent findings of Klauer and Mierke (in press). It would as well predict slower responses in evaluations that occur after incompatible IAT blocks including regular task-switches. Furthermore, it neither involves a shift in response criterion, nor the learning of specific stimulus-response associations nor effects on the explicit measurement of category association strength, all of which are alternative explanations ruled out by Klauer and Mierke.

References

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.

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.

Cai, H., Sriram, N., Greenwald, A. G., & McFarland, S. G. (in press). The Implicit Association Test's *D* measure can minimize a cognitive skill confound: Comment on McFarland and Crouch (2002). *Social Cognition*.

Chee, M. W. L., Sriram, N., Soon, C. S., & Lee, K. M. (2000). Dorsolateral prefrontal cortex and the implicit association of concepts and attributes. *NeuroReport, 11,* 135–140.

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

Draine, S. C. (2001). Inquisit (Version 1.31) [Computer software]. Seattle, WA: Millisecond Software.

Egloff, B., & Schmukle, S. C. (2002). Predictive validity of an Implicit Association Test for assessing anxiety. *Journal of Personality and Social Psychology, 83,* 1441–1455.

Egloff, B., & Schmukle, S. C. (2003). Does social desirability moderate the relationship between implicit and explicit anxiety measures? *Personality and Individual Differences, 35,* 1697–1706.

Egloff, B., Schwerdtfeger, A., & Schmukle, S. C. (in press). Temporal stability of the Implicit Association Test-Anxiety. *Journal of Personality Assessment*.

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology, 54,* 297–327.

Gawronski, B. (2002). What does the Implicit Association Test measure? A test of the convergent and discriminant validity of prejudice-related IATs. *Experimental Psychology, 49,* 171–180.

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.

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

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* Orlando, FL: Academic Press.

Hummert, M. L., Garstka, T. A., O'Brien, L. T., Greenwald, A. G., & Mellott, D. S. (2002). Using the Implicit Association Test to measure age differences in implicit social cognitions. *Psychology and Aging, 17,* 482–495.

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

Klauer, K. C., & Mierke, J. (in press). Task-set inertia, attitude accessibility, and compatibility-order effects: New evidence for a task-set switching account of the IAT effect. *Personality and Social Psychology Bulletin*.

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.

McFarland, S. G., & Crouch, Z. (2002). A cognitive skills confound on the Implicit Association Test. *Social Cognition, 20,* 481–506.

Meiran, N. (1996). Reconfiguration of processing mode prior to task performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 1423–1442.

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

Mierke, J., & Klauer, K. C. (2003). Method-specific variance in the Implicit Association Test. *Journal of Personality and Social Psychology, 85,* 1180–1192.

Monsell, S., Yeung, N., & Azuma, R. (2000). Reconfiguration of task-set: Is it easier to switch to the weaker task? *Psychological Research, 63,* 250–264.

Rogers, R., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General, 124,* 207–231.

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

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

Rubinstein, J. S., Meyer, D. E., & Evans, J. E. (2001). Executive control of cognitive processes in task switching. *Journal of Experimental Psychology: Human Perception and Performance, 27,* 763–797.

Schmukle, S. C., & Egloff, B. (in press). A latent state-trait analysis of explicit and implicit personality measures. *European Journal of Psychological Assessment*.

Steffens, M. C., & Buchner, A. (2003). Implicit Association Test: Separating transsituationally stable and variable components of attitudes toward gay men. *Experimental Psychology, 50,* 33–48.

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin, 87,* 245–251.

Wylie, G., & Allport, A. (2000). Task switching and the measurement of "switch costs". *Psychological Research, 63,* 212–233.

Received May 26, 2004
Revision received August 1, 2004
Accepted August 2, 2004

Address for correspondence

Boris Egloff
Department of Psychology
Johannes Gutenberg University Mainz
D-55099 Mainz
Germany
Tel. +49 6131 392 2485
Fax +49 6131 392 2483
Email egloff@mail.uni-mainz.de

