Research Article

IMPLICIT ATTITUDE MEASURES: Consistency, Stability, and Convergent Validity

William A. Cunningham,[1] Kristopher J. Preacher,[2] and Mahzarin R. Banaji[1]

[1]Yale University and [2]The Ohio State University

Abstract—*In recent years, several techniques have been developed to measure implicit social cognition. Despite their increased use, little attention has been devoted to their reliability and validity. This article undertakes a direct assessment of the interitem consistency, stability, and convergent validity of some implicit attitude measures. Attitudes toward blacks and whites were measured on four separate occasions, each 2 weeks apart, using three relatively implicit measures (response-window evaluative priming, the Implicit Association Test, and the response-window Implicit Association Test) and one explicit measure (Modern Racism Scale). After correcting for interitem inconsistency with latent variable analyses, we found that (a) stability indices improved and (b) implicit measures were substantially correlated with each other, forming a single latent factor. The psychometric properties of response-latency implicit measures have greater integrity than recently suggested.*

For most of psychology's history, the measurement of attitudes and beliefs has been limited to self-report questionnaires, semantic differential scales, and feeling thermometers. The use of such measures necessarily assumes that individuals have both the ability and the motivation to report attitudes and beliefs accurately, assumptions that do not stand up to scrutiny (Greenwald & Banaji, 1995; Nisbett & Wilson, 1977). In recent years, acknowledging these limits has prompted the development of new methods to assess attitudes and beliefs that reside outside conscious awareness and control. Among the most influential of these methods has been Fazio, Sanbonmatsu, Powell, and Kardes's (1986) *evaluative priming* technique, a variation of a well-established semantic priming technique (Neely, 1977). More recently, Greenwald, McGhee, and Schwartz (1998) developed the Implicit Association Test (IAT) to examine unconscious attitudes by also measuring the relative strength of the association between the attitude object and the concept "good" versus the association between the attitude object and the concept "bad." These measures provide an appealing way to investigate attitudes not only because they potentially circumvent social desirability constraints on interpretation, but also because they can potentially reveal unique components of attitudes that lie outside conscious awareness and control (Banaji, 2001).

Despite the wide usage of response-latency measures of implicit attitudes, their construct validity remains largely untested. Specifically, early reports suggest that measures of implicit attitudes do not correlate strongly either across time (i.e., poor test-retest reliability) or with performance on conceptually similar measures (i.e., poor convergent validity). For example, Sherman, Presson, Chassin, and Rose (1999) found low correlations between measures of evaluative priming and the IAT (mean $r = .06$). Similarly, Brauer, Wasel, and Niedenthal (2000) reported low correspondence between lexical decision and evaluative priming measures ($r = .27$), and Fazio (1999) found a .05 correlation between evaluative priming and IAT measures of automatic race attitude. Such observations have led to skepticism regarding the validity of implicit attitude measures and suggested that this new generation of measures fails to assess attitudes accurately (Cameron, Alvarez, & Bargh, 2000).

We suggest that such conclusions may be premature. Without the systematic evaluation of interitem inconsistencies (e.g., measurement error) of implicit attitude measures, initial estimates of stability and convergent validity may be misleading. The degree of measurement error associated with any measure sets an upper limit for correlations. Analyses of our own data suggest that implicit attitude measures may be somewhat lower in interitem consistency than are self-report measures. For example, using multiple IATs, we found that a substantial proportion of the variance in each IAT can be attributed to measurement error (Cunningham, Nezlek, & Banaji, 2001). It appears that low reliability is not restricted to measures of implicit attitude; similar concerns about reliability are echoed in research on semantic priming. On the rare occasions on which it has been measured, reliability has proven to be embarrassingly low (Buchner & Wippich, 2000; Madden, Pierce, & Allen, 1993), suggesting that low interitem consistency may be a characteristic of response-latency measures more generally.

Contrary to popular opinion, low reliability (high measurement error) need not be a threat to construct validity (see Bollen, 1989). That is, although correlations between measured variables can be only as valid as their reliabilities will allow, analyses that utilize latent variable models, with multiple measures of each construct, circumvent this problem. In several simulation studies, Little, Lindenberger, and Nesselroade (1999) found that such analyses provided unbiased estimates of the true population correlations and, perhaps more important, did not overcorrect for measurement error. Reliability does not constrain validity in latent variable analyses.

The potential gain from using response-latency measures to study implicit attitudes is great, and it would be unfortunate if lack of evidence or weak tests biased assessments of their value. To address the psychometric properties of implicit measures, we investigated the construct validity of such measures, choosing two primary measures: evaluative priming and the IAT. Many variations of both procedures exist (e.g., Bargh, Chaiken, Govender, & Pratto, 1992; Blair & Banaji, 1996; Cunningham et al., 2001; Dovidio, Evans, & Tyler, 1986; Fazio, Jackson, Dunton, & Williams, 1995; Kawakami, Dion, & Dovidio, 1998; Wittenbrink, Judd, & Park, 1997), and we used versions of the techniques known to produce replicable effects.

We assessed implicit attitudes using three versions of the two measures. The first was response-window evaluative priming, a version of evaluative priming that imposed a short response window of 200 to 600 ms within which a person was required to respond to the target word. The response-window procedure is known to significantly enhance the magnitude of both subliminal and supraliminal priming effects (see Draine & Greenwald, 1998). The second measure was the

Address correspondence to William A. Cunningham or Mahzarin R. Banaji, Department of Psychology, Yale University, P.O. Box 208205, New Haven, CT 06520; e-mail: william.cunningham@yale.edu or mahzarin.banaji@yale.edu; or e-mail Kristopher J. Preacher: preacher2@osu.edu.

IAT (Greenwald et al., 1998), and the third was a response-window version of the IAT that contained overlapping components of the first two measures—it used the IAT procedure but included a response window. Participants completed each of the three measures at four separate testing sessions, with sessions separated by 2 weeks.

In this article, we mainly focus on estimates of *interitem consistency* (the extent to which items within a single measure at a single measurement occasion correlate with each other), *stability* (the extent to which a measure at one measurement occasion correlates with the same measure at other times), and *convergent validity* (the extent to which different measures that are designed to tap the same construct correlate with each other).

Secondarily, we examine the relationship between implicit and explicit attitudes. Models of implicit attitude, like models of implicit memory, assume that implicit and explicit attitudes reflect separate processes and, in theory, that measures of the two ought to be unrelated (Greenwald & Banaji, 1995). Research on relationships between implicit and explicit attitudes, however, has resulted in inconclusive observations: As many studies have reported relationships as have not (for reviews, see Blair, in press; Brauer et al., 2000). It is possible that just as the unreliability of implicit measures can attenuate correlations among implicit measures, the same unreliability can obscure relationships between implicit and explicit attitudes. In this study, a measure of explicit race attitude, the Modern Racism Scale (McConahay, 1986), was also administered at each testing session.

Explicit attitude measures have been used for more than 70 years (Thurstone, 1928), and their psychometric properties are not in question. We therefore do not include multiple measures of explicit attitudes. Additionally, this study focuses on convergent validity and not divergent validity because we have addressed that issue in greater depth elsewhere (Cunningham et al., 2001). In this article, we focus on the interitem consistency within each implicit measure at each testing occasion, account for potential measurement error, and then estimate the stability within implicit measures and convergent validity across implicit measures.

METHOD

Participants

Ninety-nine Ohio State University students participated in partial fulfillment of a course requirement. Data from 3 participants were removed because they attended only the first session, and data from 3 additional participants were removed because of consistently excessive error rates (in excess of 70%), leaving a final sample of 93 (52 female and 41 male) participants.

Materials

The same stimulus set was used for each of the three implicit measures. Twelve (six male and six female) morphed human faces representing black and white social groups were used. Evaluative words were taken from the Bellezza, Greenwald, and Banaji (1986) norms. All stimuli are presented in Figure 1.

Procedure

At each of the four sessions, participants completed all four measures (three relatively implicit attitude measures and one explicit attitude measure) on a computer. The order of the implicit measures was randomized for each participant at each session. The order of implicit and explicit measures was counterbalanced. Testing sessions were separated by 2 weeks.

IAT

Participants responded to black or white faces representing social groups and positive or negative words representing evaluative attributes. Participants used separate computer keys to indicate whether each face was black or white and used the same two keys to indicate whether each word was good or bad. For half the trials (the *white + good* block), participants classified white faces and good words on one key and black faces and bad words on another key. For the other half (the *white + bad* block), participants classified black faces and good words on one key and white faces and bad words on another key. The order of blocks was counterbalanced. Each block consisted of 70 trials. If a participant responded incorrectly, a red *X* appeared until the response was corrected.

Response-window evaluative priming

Participants responded to words that had an evaluatively negative or positive meaning by pressing one key for bad words and another key for good words. Immediately preceding each target word, a black or white face appeared for 200 ms, followed by a blank screen for 100 ms. Participants were required to respond within a response window (200 to 600 ms) after the target was presented. A white exclamation point appeared during that time period and turned red if the participant responded within the window. The exclamation point disappeared if the participant did not respond in time (see Draine & Greenwald, 1998). After 50 practice trials, participants completed 200 critical trials. After each block of 50 trials, they received feedback about their accuracy and the percentage of their responses falling within the window. Participants were instructed to make at least 70% of their responses within the window, and if they performed below that level, to try harder. No error feedback was provided until the end of each 50-trial block.

Response-window IAT

The response-window IAT was identical to the IAT with one significant difference: Participants were required to respond to stimuli within a 225- to 675-ms window after the stimulus was presented. No error feedback was provided.

Modern Racism Scale

Questions from the Modern Racism Scale (McConahay, 1986) were administered by computer. Participants responded on 6-point scales (1 = *strongly disagree*, 6 = *strongly agree*).

RESULTS

IAT

Following Greenwald et al. (1998), we deleted extreme outlier trials (i.e., response latencies greater than 3,000 or less than 300 ms; 1.7% of trials), as well as practice trials (first 20). Mean response latencies were calculated for both the white + good and the white + bad blocks. The results were consistent with those of previous research

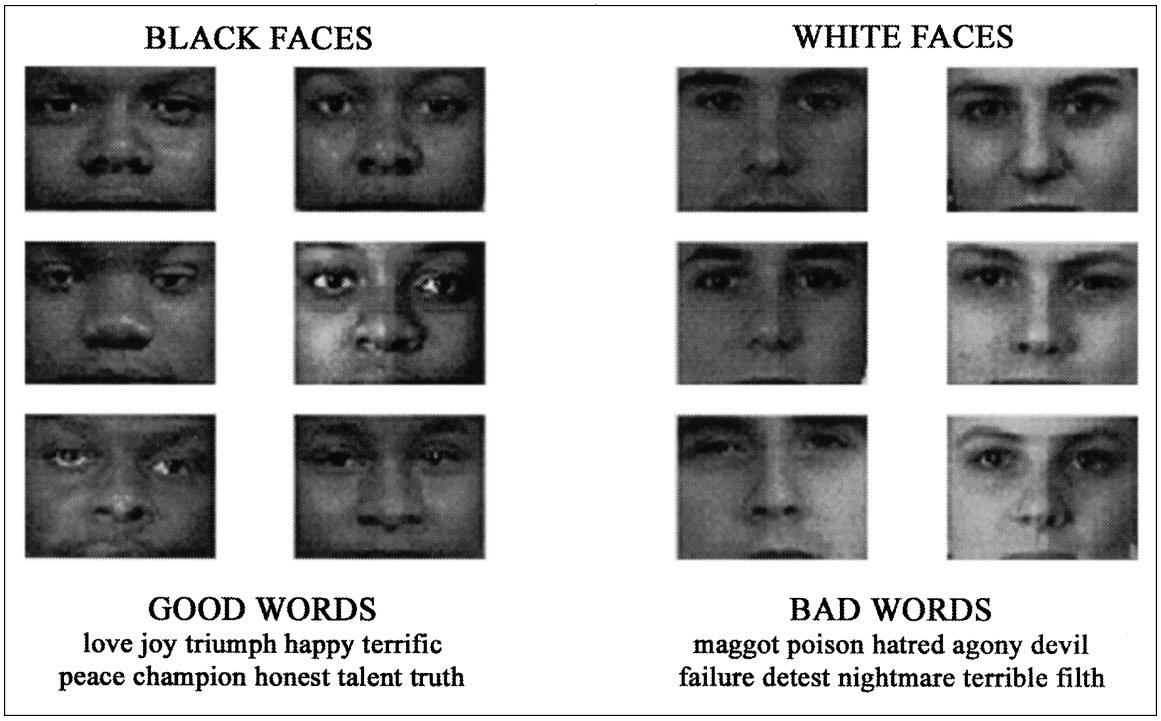

Fig. 1. Stimuli used for implicit measures.

(Greenwald et al., 1998): Participants took longer to respond in the white + bad condition (M = 743.90 ms) than in the white + good condition (M = 612.55 ms), indicating stronger associations between "white" and "good" (and between "black" and "bad") than between "white" and "bad" (and between "black" and "good"), $t(92) = 13.07$, $p < .0001$, $d = 1.36$ (see Fig. 2, left panel).[1]

Response-Window Priming

Because participants were required to respond within a response window of 200 to 600 ms (with a center of 400 ms), responses with latencies greater than 800 ms (twice the acceptable deviation from the center of the response window) were deleted. Mean accuracy was then calculated for white + good (black + bad) trials and for white + bad (black + good) trials. The results were consistent with previous research measuring race bias with evaluative priming (Dovidio et al., 1986; Fazio et al., 1995; Wittenbrink et al., 1997): Fewer errors were made for the white + good (black + bad) trials (M = 76% correct) than the white + bad (black + good) trials (M = 71% correct), indicating that participants more easily paired "black" with "bad" and "white" with "good" than they paired "black" with "good" and "white" with "bad," $t(92) = 8.78$, $p < .0001$, $d = 0.91$ (see Fig. 2, middle panel). Individual differences in race bias were calculated as the difference in mean accuracy between these pairings, so higher scores indicate stronger associations between "black" and "bad" and between "white" and "good" than between "black" and "good" and between "white" and "bad."

Response-Window IAT

As with the previous measures, responses that fell too far outside the response window (in this case, greater than 900 ms) and practice trials were deleted. Mean accuracy was calculated for both the white + good and the white + bad blocks. Participants made fewer errors in the white + good blocks (M = 83% correct) than in the white + bad blocks (M = 76% correct), $t(92) = 9.67$, $p < .0001$, $d = 1.00$ (see Fig. 2, right panel). These results mirror those found for the standard IAT.

Modern Racism Scale

The Modern Racism Scale was scored according to published protocols (McConahay, 1986). The mean was 2.78 (SD = 0.78), which is significantly lower than the midpoint of the scale (3.5), $t(92) = -8.94$, $p < .0001$, $d = 0.93$. In contrast to the strong pro-white attitude revealed on the implicit measures, participants reported disagreement with explicit statements of prejudice and negative beliefs about black Americans, showing a dissociation between implicit and explicit attitudes.

Psychometric Properties of Implicit Attitude Measures

Of primary interest for the present research are the estimates of interitem consistency, stability, and convergent validity for these mea-

1. We conducted tests to identify multivariate outliers and recode missing data. Multivariate outliers were defined as scores with Student t residuals greater than 2.58 after being regressed on all other variables. Each multivariate outlier was recoded to the value corresponding to the square root of its Student t residual. Missing data were estimated from the same regression equations. Outliers constituted 1.68% of the data, and 4.3% of the data were missing.

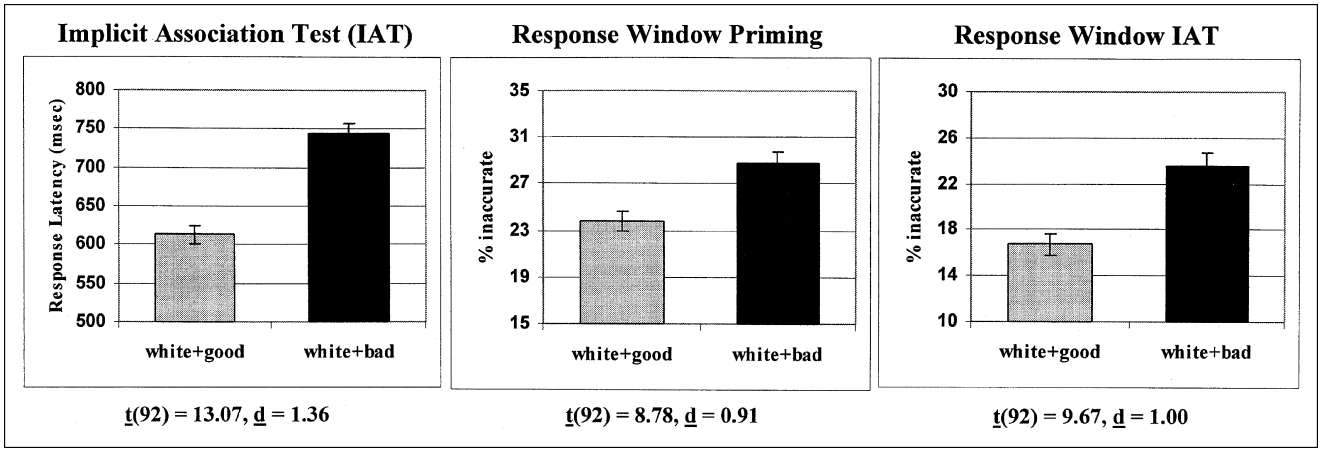

Fig. 2. Mean differences between white + good and white + bad trials. The dependent measure was response latency for the Implicit Association Test (left panel) and percentage errors for response-window priming (middle panel) and the response-window Implicit Association Test (right panel).

sures of implicit attitudes. Because these properties are conceptually discrete, they are discussed separately.

Interitem consistency of measures

In essence, interitem consistency is the homogeneity of responses to all items within a particular measure. The more heterogeneous (or internally inconsistent) a measure is, the greater the attenuation of correlations involving that measure. Initial estimates of interitem consistency for response-latency measures have been somewhat lower than is typical for self-report measures (Cunningham et al., 2001).

Interitem consistency is defined as the proportion of total variance that reflects consistent variance, with 1.0 representing measurement without random error and 0.0 representing measurement consisting of only random error. Current conventions suggest that interitem consistencies of .80 (20% error) or higher represent good reliability (Cronbach, 1951), although many widely used scales remain in the range of .70 (see Robinson, Shaver, & Wrightsman, 1991).

To generate estimates of the interitem consistencies for implicit attitude measures, we calculated Cronbach's alpha for each of the measures at each occasion.[2] The four interitem consistencies calculated for each measure, although not consistently high, were within a liberally acceptable range (mean alphas were .78 for the IAT, .63 for the response-window IAT, and .64 for response-window evaluative priming, for an overall mean alpha of .69). Taken together, the interitem consistency of these implicit measures is indeed lower than the interitem consistency of most standard measures of attitudes and beliefs, and such measurement error has the potential to attenuate estimates of stability and convergent validity. We believe that the lower reliability of implicit attitude measures in this study, compared with the reliability of traditional explicit measures, derives from two sources. First, individual millisecond reaction times fluctuate across trials. Second, difference scores inherently remove reliable variance that enters into this computation; response latencies for compatible and incompatible trials are highly correlated, thereby especially reducing total reliable variance (see Cohen & Cohen, 1982). In subsequent analyses, we provide estimates of stability and convergent validity that overcome these problems.

Stability of measures

In addition to being internally consistent, a measure of individual differences that measures a stable construct should be consistent over time. Failure to demonstrate stability over time suggests that the measure may be invalid. An examination of the correlation matrix presented in Table 1 suggests that response-latency attitude measures may not be stable: The average test-retest correlation is only .27. However, these low initial estimates of stability for the implicit attitude measures are confounded with measurement error (see Tisak & Tisak, 1996), and therefore are underestimates.

To circumvent the problem of attenuated stability estimates, we used a latent variable approach to separate measurement error from estimates of stability.[3] We first divided each implicit attitude measure into three parcels at each measurement occasion, and then, using a latent-growth-curve framework established by Tisak and Tisak (2000), we decomposed the parcel variances into constituent components representing permanent (enduring) variance, variance due to temporal effects, variance unique to each parcel, and error variance. An estimate of stability was computed for each implicit attitude measure by dividing the sum of the permanent variance components by the sum of the permanent and temporal components over the three parcels at each

2. Because each of these measures uses a difference score as an index of implicit attitude, we computed alpha using Cohen and Cohen's (1982) equation for difference scores and using each response latency as an "item." Each of the two components of the difference score (white + good vs. white + bad) had remarkable reliability, but a strong correlation between these components resulted in reduced reliability of the difference score.

3. Another method for correcting measurement error is the correction-for-attenuation formula. Although this adjustment provides similar estimates, latent variable approaches are more accurate.

Table 1. *Means, standard deviations, reliabilities, and correlation matrix of measured variables*

	MRS1	MRS2	MRS3	MRS4	IAT1	IAT2	IAT3	IAT4	RW-IAT1	RW-IAT2	RW-IAT3	RW-IAT4	RW-P1	RW-P2	RW-P3	RW-P4
MRS1	(.74)															
MRS2	.80	(.75)														
MRS3	.78	.82	(.82)													
MRS4	.76	.77	.86	(.79)												
IAT1	.21	.15	.15	.14	(.88)											
IAT2	.13	.14	.10	.08	.31	(.78)										
IAT3	.16	.26	.23	.20	.42	.50	(.75)									
IAT4	.14	.17	.16	.13	.16	.33	.17	(.68)								
RW-IAT1	.20	.16	.19	.26	.33	.11	.23	.07	(.68)							
RW-IAT2	.26	.29	.18	.19	.20	.27	.36	.29	.26	(.59)						
RW-IAT3	.35	.33	.34	.25	.28	.29	.34	.33	.36	.39	(.71)					
RW-IAT4	.19	.17	.08	.07	.12	.25	.30	.14	.01	.17	.24	(.51)				
RW-P1	.00	.11	−.07	−.04	.27	.18	.19	.02	.03	.01	.02	.07	(.52)			
RW-P2	.16	.08	.04	.08	.26	.27	.24	.22	.14	.32	.32	.17	.13	(.66)		
RW-P3	.12	.01	.02	.07	.13	.19	.18	.00	.02	.00	.11	.04	.17	.30	(.60)	
RW-P4	.33	.18	.26	.31	.14	.24	.31	.15	.22	.20	.27	.04	.01	.48	.42	(.74)
Mean	2.78	2.74	2.72	2.86	159.86	163.01	88.05	115.98	0.06	0.08	0.06	0.08	0.04	0.06	0.05	0.05
SD	0.84	0.82	0.88	0.86	220.66	129.51	101.30	90.76	0.12	0.10	0.11	0.10	0.07	0.08	0.08	0.09

Note. MRS = Modern Racism Scale; IAT = Implicit Association Test; RW-IAT = response-window IAT; RW-P = response-window evaluative priming. Numbers following the variable labels denote testing occasions. Numbers on the main diagonal denote Cronbach's alphas.

measurement occasion.[4] These estimates provided the proportion of consistent, substantively meaningful variance that remained stable over time. The stability estimates were .46 for the IAT, .68 for response-window priming, and .36 for the response-window IAT. From these proportions, we calculated a *stability index* that represents the correlation between enduring attitude and consistent variance at each occasion: The value of the index was .68 for the IAT, .83 for response-window priming, and .60 for the response-window IAT. Although other studies have indicated that implicit attitude measures fail to correlate significantly across time, we found that such measures, when analyzed as latent variables, are quite stable.[5]

Convergent validity of implicit measures

Conceptually, differing measures of implicit attitudes are assumed to tap the same implicit attitude, and failures to find correlations among measures naturally elevate concerns about the validity of implicit attitude measurement. As can be observed in Table 1, bivariate correlations among implicit attitude measures can be surprisingly low (the mean r was .19). Again, we suggest that measurement error masks existing relationships among these particular latency-based implicit attitude measures.

To examine convergent validity more accurately, we performed a confirmatory factor analysis in which the four administrations of each measure were modeled as indicators of their respective latent constructs, and the correlations among these latent variables were estimated (see Fig. 3).[6] The model presented in Figure 3 provides an adequate representation of relationships among the measured variables, $\chi^2(98, N = 93) = 109.95$, $p = .19$; *NNFI* = .96; *CFI* = .97; *RMSEA* = 0.042 (90% confidence interval: 0.00, 0.072).[7] All of the latent variable correlations were statistically significant and were substantially greater (mean $r = .63$) than the correlations estimated by the simple bivariate correlations. These estimates provide evidence for the convergent validity of widely used implicit attitude measures.

In addition, each of the measures was significantly correlated with the Modern Racism Scale, mean $r = .35$. This significant correlation is consistent with our previous finding (Cunningham et al., 2001), using multiple measures of both implicit and explicit attitudes, that the two sets of measures are correlated, yet distinct.

A single implicit latent variable?

Thus far, we have established that implicit attitude measures are robustly correlated. However, it is unclear whether different implicit

4. We used LISREL 8.30 (Jöreskog & Sörbom, 1996), maximum likelihood estimation, and covariances as input.

5. Implicit attitude measures appear to be less stable than explicit attitude measures. The greater stability of explicit measures, however, may well be a measurement artifact reflecting more a conscious belief in consistency and consistent responding than actual underlying consistency.

6. An a priori power analysis (MacCallum, Browne, & Sugawara, 1996) was conducted to determine the minimum sample size required for a power level of .80, assuming a null hypothesis of close fit (Ho: *RMSEA* = 0.05) and an alternative hypothesis of unacceptable fit (Ha: *RMSEA* = 0.10). The minimum acceptable sample size was determined to be $N = 63$ for the measurement confirmatory factor analysis and $N = 62$ for the more restricted model. We obtained $N = 93$, yielding a power estimate of .96.

7. Model fit was assessed with several fit indices: nonnormed fit index (NNFI; Bentler & Bonett, 1980), comparative-fit index (CFI; Bentler, 1990), and root-mean-square error of approximation (RMSEA; Browne & Cudeck, 1992; Steiger & Lind, 1980). Criteria for acceptable model fit include NNFI and CFI values above .90 and RMSEA values below 0.08.

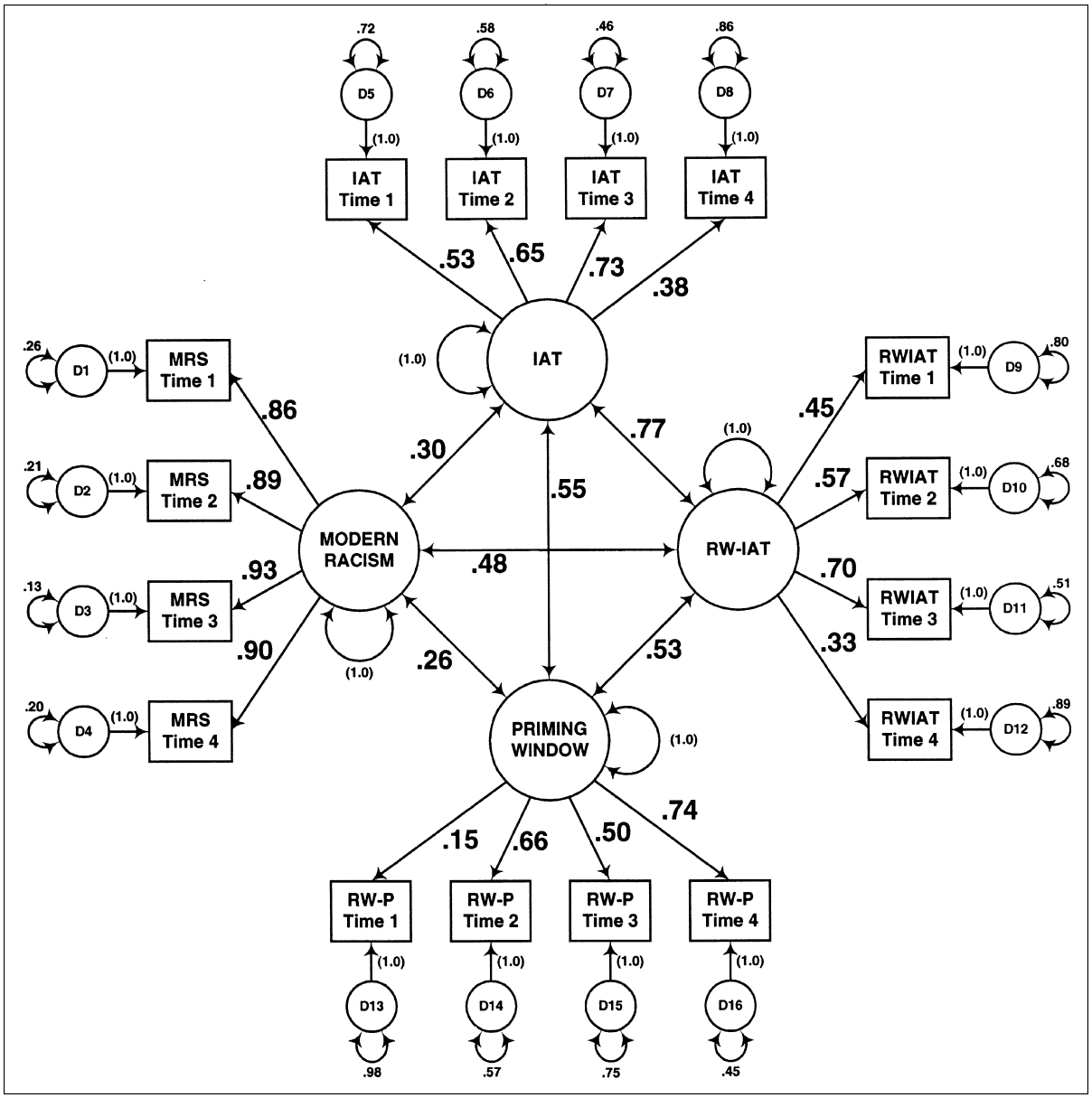

Fig. 3. Results of the confirmatory factor analysis. Numbers in parentheses are fixed parameters. All other numbers are estimates of free parameters. IAT = Implicit Association Test; MRS = Modern Racism Scale; RWIAT = response-window IAT; RW-P = response-window evaluative priming.

measures tap the same latent implicit attitude. If implicit attitude measures are not simply correlated, but tap the same underlying construct of implicit attitude, the variance of each measure can be conceptualized as follows:

$$\sigma^2_{\text{implicit measure}} = \sigma^2_{\text{implicit attitude}} + \sigma^2_{\text{specific to measure}} + \sigma^2_{\text{error}}.$$

To evaluate the possibility that each of the implicit measures taps the same underlying implicit attitude, we performed a second confirmatory factor analysis in which each of the three implicit constructs was modeled as a single second-order implicit attitude construct (see Fig. 4). In addition, the correlation between the explicit attitude construct and this general implicit attitude construct was estimated.

This model fit the data well, $\chi^2(100, N = 93) = 111.58$, $p = .20$; *NNFI* = .96; *CFI* = .97; *RMSEA* = 0.041 (90% confidence interval: 0.00, 0.071). A nested chi-square test indicated that this model fit the data as well as the previous confirmatory factor analysis, $\chi^2(2, N = 93) = 1.63$, n.s. Each of the loadings between the first-order implicit constructs and the second-order latent construct was substantial and significant (mean $\beta = .79$), suggesting that each of the methods for measuring implicit attitudes taps the same implicit attitude, with some unique components. Furthermore, the general implicit attitude latent construct was correlated with the explicit attitude construct, $r = .45$, $p < .01$. It is important to note that the implicit-explicit relationship was significantly weaker than the implicit-implicit relationships,

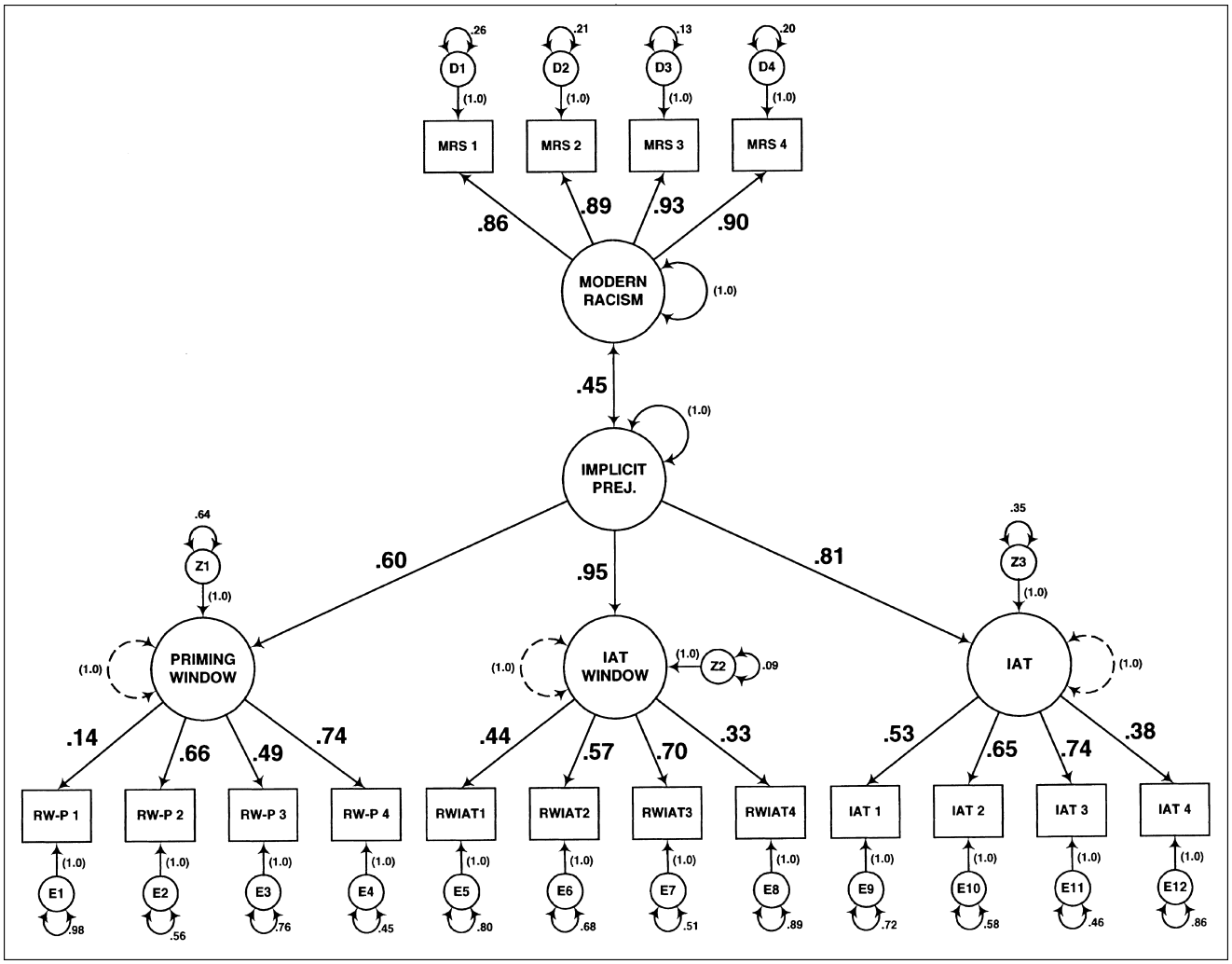

Fig. 4. Results of the second-order confirmatory factor analysis. Numbers in parentheses are fixed parameters. All other numbers are estimates of free parameters. IAT = Implicit Association Test; MRS = Modern Racism Scale; RWIAT = response-window IAT; RW-P = response-window evaluative priming; PREJ. = prejudice.

$\chi^2(1) = 9.85$, $p < .001$, a finding that supports hypotheses of attitude dissociation.

DISCUSSION

In a replication of previous research, each of three implicit attitude measures revealed greater association between "white" and "good" and between "black" and "bad" than between "white" and "bad" and between "black" and "good." In addition, the overall magnitude of this association was consistently strong across the four testing sessions. Furthermore, we found a dissociation between implicit and explicit measures of race attitude: Participants simultaneously self-reported nonprejudiced explicit attitudes toward black Americans while showing an implicit difficulty in associating black with positive attributes.

The main purpose of the present investigation, however, was to address a recurring question regarding the stability and convergent validity of implicit measures. We first demonstrated that after correction for measurement error, implicit attitude measures proved consistent across time and across measures, so that a person who scored high on one measure generally scored high on others. Second, we showed that all three implicit attitude measures are not only correlated with each other, but also form a single latent construct.

Our analyses of implicit attitude measures suggest that the degree of measurement error in response-latency measures can be substantial—estimates of Cronbach's alpha indicated that, on average, more than 30% of the variance associated with the measurements was random error. These low estimates of reliability are mirrored in an analysis of the test-retest reliability of implicit stereotyping measures (mean $r = .52$; Kawakami & Dovidio, in press). When using latency-based measures as indices of individual differences, it may be essential to employ analytic techniques, such as covariance structure modeling, that can separate measurement error from a measure of individual differences. Without such analyses, estimates of relationships involving implicit measures may produce misleading null results.

The present research provides support for an association between implicit and explicit measures of race (black-white) attitude. Specifi-

cally, we found that the Modern Racism Scale was correlated with all three measures of implicit race attitude. Furthermore, it was correlated with the general implicit prejudice construct, indicating that this relationship cannot be explained as an artifact of a particular measure. Thus, these results contradict the idea of a complete dissociation between implicit and explicit attitude. This association does not imply that implicit and explicit attitude measures are identical by any means. Although multiple measures of both implicit and explicit attitudes are robustly correlated, the two kinds of attitude measures also tap unique sources of variance (Cunningham et al., 2001); a single-factor solution does not fit the data.

In two confirmatory factor analyses, each of the implicit measures substantially and reliably correlated with the others, demonstrating convergent validity for implicit attitude measures. Moreover, the different implicit attitude measures were measures of a single latent variable, suggesting that each measure taps the same representation. Early failures to find relationships among implicit measures may be due more to measurement error inherent in latency-based measures than to lack of convergent validity. These findings have important implications for the use of response-latency measures as indicators of implicit attitudes and implicit memory: When subjected to appropriate statistical tests, different techniques of implicit cognition are robustly related to each other. The next phase of implicit attitude measurement will no doubt address the predictive validity of such measures. We suggest that it is especially important to use latent variable analyses when studying individual differences in attitude and behavior. Insofar as convergent validity is concerned, rumors of the death of implicit measures have been greatly exaggerated. In the future, it would be wise to base claims regarding the validity of implicit measures on appropriately rigorous procedures of measurement and analysis.

Acknowledgments—For discussions and advice about statistical analyses, we thank Todd Little, Robert C. MacCallum, John Tisak, and the Ohio State University "quantitative brown bag group." For comments on earlier versions of this article, we thank R. Bhaskar, Siri Carpenter, Geoff Cohen, Kathleen Brophy Cunningham, Jack Glaser, Richard Hackman, Franciszek Lachowicz, Kristi Lemm, and Laurie Rudman. For data collection and management, we thank Ann Marie Altman. This research was supported by grants from the National Institute of Mental Health (MH-57672) and the National Science Foundation (SBR-9709924) to Mahzarin R. Banaji.

REFERENCES

Banaji, M.R. (2001). Implicit attitudes can be measured. In H.L. Roediger, J.S. Nairne, I. Neath, & A. Surprenant (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 117–150). Washington, DC: American Psychological Association.

Bargh, J.A., Chaiken, S., Govender, R., & Pratto, F. (1992). The generality of the automatic attitude activation effect. *Journal of Personality and Social Psychology, 62*, 893–912.

Bellezza, F.S., Greenwald, A.G., & Banaji, M.R. (1986). Words high and low in pleasantness as rated by male and female college students. *Behavior Research Methods, Instruments, & Computers, 18*, 299–303.

Bentler, P.M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*, 238–246.

Bentler, P.M., & Bonett, D.G. (1980). Significance tests and goodness-of-fit in the analysis of covariance structures. *Psychological Bulletin, 88*, 588–606.

Blair, I.V. (in press). Implicit stereotypes and prejudice. In G. Moskowitz (Ed.), *Cognitive social psychology: On the future of social cognition*. Mahwah, NJ: Erlbaum.

Blair, I.V., & Banaji, M.R. (1996). Automatic and controlled processes in stereotype priming. *Journal of Personality and Social Psychology, 70*, 1142–1163.

Bollen, K.A. (1989). *Structural equations with latent variables*. New York: John Wiley & Sons.

Brauer, M., Wasel, W., & Niedenthal, P. (2000). Implicit and explicit components of prejudice. *Review of General Psychology, 4*, 79–101.

Browne, M.W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods and Research, 21*, 230–258.

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

Cameron, J.A., Alvarez, J.M., & Bargh, J.A. (2000, February). *Examining the validity of implicit measures of prejudice*. Poster presented at the annual meeting of the Society for Personality and Social Psychology, Nashville, TN.

Cohen, J., & Cohen, P. (1982). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 6*, 297–334.

Cunningham, W.A., Nezlek, J.B., & Banaji, M.R. (2001). *Conscious and unconscious ethnocentrism: Revisiting the ideologies of prejudice*. Unpublished manuscript, Yale University, New Haven, CT.

Dovidio, J.F., Evans, N., & Tyler, R.B. (1986). Racial stereotypes: The contents of their cognitive representations. *Journal of Experimental Social Psychology, 22*, 22–37.

Draine, S.C., & Greenwald, A.G. (1998). Replicable unconscious semantic priming. *Journal of Experimental Psychology: General, 127*, 286–303.

Fazio, R.H. (1999). [Correlation between the IAT and the Bona Fide Pipeline]. Unpublished raw data.

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.

Fazio, R.H., Sanbonmatsu, D.M., Powell, M.C., & Kardes, F.R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology, 50*, 229–238.

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

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

Jöreskog, K.G., & Sörbom, D. (1996). *LISREL 8: User's reference guide*. Chicago: Scientific Software, International.

Kawakami, K., Dion, K.L., & Dovidio, J.F. (1998). Racial prejudice and stereotype activation. *Personality and Social Psychology Bulletin, 24*, 407–416.

Kawakami, K., & Dovidio, J.F. (in press). Implicit stereotyping: How reliable is it? *Personality and Social Psychology Bulletin*.

Little, T.D., Lindenberger, U., & Nesselroade, J.R. (1999). On selecting indicators for multivariate measurement and modeling with latent variables: When "good" indicators are bad and "bad" indicators are good. *Psychological Methods, 7*, 192–211.

MacCallum, R.C., Browne, M.W., & Sugawara, H.M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*, 130–149.

Madden, D.J., Pierce, T.W., & Allen, P.A. (1993). Age-related slowing and the time course of semantic priming in visual word identification. *Psychology & Aging, 8*, 490–507.

McConahay, J.B. (1986). Modern racism, ambivalence, and the modern racism scale. In J.F. Dovidio & S.L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 91–125). Orlando, FL: Academic Press.

Neely, J.H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General, 106*, 226–254.

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

Robinson, J.P., Shaver, P.R., & Wrightsman, L.S. (1991). *Measures of personality and social psychological attitudes, Vol. 1*. San Diego: Academic Press.

Sherman, S.J., Presson, C.C., Chassin, L., & Rose, J.S. (1999). *Implicit and explicit attitudes toward cigarette smoking: The effects of context and motivation*. Unpublished manuscript, Indiana University, Bloomington.

Steiger, J.H., & Lind, J. (1980, June). *Statistically based tests for the number of common factors*. Paper presented at the annual meeting of the Psychometric Society, Iowa City, Iowa.

Thurstone, L.L. (1928). Attitudes can be measured. *American Journal of Sociology, 33*, 529–554.

Tisak, J., & Tisak, M.S. (1996). Longitudinal models of reliability and validity: A latent curve approach. *Applied Psychological Measurement, 20*, 275–288.

Tisak, J., & Tisak, M.S. (2000). Permanency and ephemerality of psychological measures with application to organizational commitment. *Psychological Methods, 5*, 175–198.

Wittenbrink, B., Judd, C.M., & Park, R. (1997). Evidence for racial prejudice at the implicit level and its relationship with questionnaire measures. *Journal of Personality and Social Psychology, 72*, 262–274.

(RECEIVED 5/22/00; REVISION ACCEPTED 8/20/00)