More Error than Attitude in Implicit Association Tests (IATs), a CFA-MTMM analysis of measurement error.

Authors

Dr Susan Chequer[1]

University of Tasmania, Churchill Avenue, Sandy Bay, Tasmania, Australia, 7005

Dr Michael G. Quinn

Division of Psychology, School of Medicine, College of Health & Medicine, University of Tasmania, Private Bag 30, Hobart, Tasmania, Australia, 7001

Highlights

- The Implicit Association Test (IAT) has many design characteristics that make it susceptible to random and systematic error variance.
- A CFA-MTMM analysis was applied to partial random and systematic error variance from trait variance for four IATs, revealing significant error variance in IAT scores.
- On average, random error accounted for over 50% of the variance in the IAT data, and method variance a further 30%, with less than 20% identifiable trait variance.
- These findings have significant implications for use of IATs in applied research.

Abstract

Many design characteristics of the popular Implicit Association Test (IAT) appear to make the task highly susceptible to measurement error. This study examined potential sources of measurement error for two types of IAT, the classic verbal IAT (VIAT) and a fully pictorial IAT (PIAT). A CFA-MTMM analytical approach was used to estimate the influence of both random error and method variance on the IAT scores. Four empirical IATs were employed to assess implicit bias towards Middle Eastern and European people ('Racial' VIAT and PIAT) and countries ('Country' VIAT and PIAT). They were completed by 198 student participants from an

[1] Contact Susan on Suse.Chequer@gmail.com; Current Affiliation for Susan Chequer is: Australian Childhood Foundation, 39 Tasma Street, North Hobart, Tasmania, Australia, 7000

Australian University. The CFA-MTMM analysis provided clear evidence of measurement error confounding IAT scores. Specifically, IAT data was shown to be, on average, comprised of just over 50% random error variance, nearly 30% method variance and under 20% trait variance. These results demonstrate unequivocally that IAT scores are predominantly composed of measurement error not implicit attitudes. These findings have significant implications for the use of IATs in applied research. Options for minimising the impact of high error variance in future implicit attitudinal research are considered.

Keywords

Implicit Attitude Test, PIAT, Measurement Error, Method Variance, Validity, Confirmatory Factor Analysis, CFA-MTM

1. Introduction

The Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998; see also https://implicit.harvard.edu/implicit/) is a highly popular attitude measure that claim to provide a 'truer' representation of a participant's attitude by minimising opportunity for self-censorship. As conceptually attractive as the IAT is, nearly two decades of research has demonstrated at best questionable, and often poor reliability (e.g. Bar-Anan & Nosek, 2014; Buchner & Wippich, 2000; Cunningham, Preacher, & Banaji, 2001; Dasgupta, McGhee, Greenwald, & Banaji, 2000; Gawronski, Morrison, Phills, & Galdi, 2017; Krause, Back, Egloff, & Schmukle, 2010; Rae & Olson, 2018). Most critics of the IAT have largely focused on its' lack of predictive validity[2] (e.g. Blanton et al., 2009; Carlsson & Agerstrom, 2016; Oswald, Mitchell, Blanton, Jaccard, & Tetlock, 2013, 2015; Rae & Olson, 2018; Singal, 2017), with little discussion about the test's

---

[2] Although these views have been contested (e.g. Greenwald, Banaji, & Nosek, 2015; Lane, Banaji, Nosek, & Greenwald, 2007).

2

construct validity (see Feest, 2020)[3]. It has been suggested that measurement error is a potential cause of the issues surrounding low reliability and validity for implicit attitude measures and may have been providing an upper limit for construct validity estimates (e.g. Cunningham et al., 2001; Oswald et al., 2013; Siers & Christiansen, 2013). There are multiple aspects of the implicit assessment process that make it potentially very susceptible to the confounding influence of various forms of measurement error. It is critical to disentangle the effects of random and systematic forms of measurement error (such as method variance) from the resultant estimate of implicit bias, as doing so would improve the predictive validity of the IAT (Oswald et al., 2013).

This paper outlines a systematic examination of potential sources of error variance for the IAT, mapping the overall influence of random error and method variance on IAT scores using the Multitrait-Multimethod approach to Confirmatory Factor Analysis (CFA-MTMM).

1.1. Inferring implicit attitudes through the IAT

Blanton et al. (2006) outlined that implicit attitude assessment involves two distinct levels of measurement; a conceptual level, and an observed level. At the conceptual level, implicit attitude tasks are believed to measure the discrepancy between two implicit evaluations about a construct of interest, ostensibly revealing implicit bias towards an attitude construct like race. At the observed level, implicit attitude scores are devised by calculating the difference between two behavioural responses, in this case, the average reaction times for congruent and incongruent block trials for the IAT[4] (for an overview of the IAT procedure and scoring, see Greenwald et al., 1998; see also Greenwald et al., 2003).

---

[3] This is perhaps unsurprising, when even the construct of 'implicit attitudes' has been surrounded with so much theoretical controversy (Feest, 2020; Mitchell & Tetlock, 2017).

[4] This algorithm leads to the IAT_{RAW} score. It is noted that the IAT_D score (Greenwald, Nosek, & Banaji, 2003), which is the current suggested scoring algorithm for the IAT, has a further step of dividing the difference between reaction times for congruent and incongruent block trials, by the within-individual standard deviation (calculated across congruent *and* incongruent trials). Though mathematically different, as demonstrated by Blanton et al. (2015) the scoring methodologies yield highly correlated (r=.95) outcomes, and thus can be considered to be *conceptually equivalent.*

The observed behavioural responses (reaction times) are thus used to *infer* the magnitude of the abstract construct of implicit attitudes (see Blanton, Jaccard, Gonzales, & Christie, 2006). However, implicit attitude measures cannot perfectly capture the abstract attitude constructs they aim to assess. Attitudes are abstract in nature which increases the difficulty in measuring them. As such, there is a discrepancy between the observed scores of the test and the 'true' value of the conceptual construct, this discrepancy is measurement error (Cote & Buckley, 1987; Spector, 2006). High amounts of measurement error reduce the reliability and validity of the estimated scores, thereby limiting the effectiveness of the measurement instrument. When considering implicit attitude measures, there are numerous design aspects that could likely contribute significant amounts of error variance to the scores, resulting in confounded estimates of 'implicit attitude'.

1.2. Random and Systematic Forms of Error Variance

Classical test theory, often attributed to Spearman (1904), states that any observed score, such as an IAT effect score, is comprised of two main components; the 'true' score, or trait component (i.e. the implicit attitude construct) and an error component (that accounts for the imperfection of the measurement). Recent conceptualisations of classical test theory, have further identified error variance as being comprised of both random and systematic error components (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). In measurement, high rates of random and systematic error can be problematic as they reduce the clarity and veracity of the estimated trait variance. This typically results in poor psychometric properties, such as low reliability and difficulties in establishing estimates of construct validity, such as convergence among like measures.

IATs and other implicit attitude measures have typically demonstrated lower reliability levels than explicit attitude measures (Buchner & Wippich, 2000; Krause et al., 2010), and often below that which would be considered satisfactory for applied measurement (see Rae & Olson, 2018). Furthermore, attempts to demonstrate convergent validity for implicit attitude measures have proven unsuccessful with multiple implicit attitude measures of the same construct/domain repeatedly shown to be barely related (e.g. Bosson, Swann, & Pennebaker, 2000; Buhrmester,

Blanton, & Swann, 2011; Falk, Heine, Takemura, Zhang, & Hsu, 2013; Foroni & Bel-Bahar, 2010; Hofmann & Schmitt, 2008; Krause et al., 2010; Pieters, van der Vorst, Engels, & Wiers, 2010; Rudolph, Schröder-Abé, Schütz, Gregg, & Sedikides, 2008). Were systematic and random error to be a significant contributor to observed implicit attitude scores, it would provide a clear rationale for these poor psychometric findings. This next section explores numerous design characteristics inherent to implicit attitude measures that are likely increasing their vulnerability to both random and systematic forms of measurement error, thereby potentially reducing the veracity of IAT results.

1.2.1. Potential sources of random error variance for IATs

Random sources of error are non-trait-related forms of variance that indiscriminately influence the whole of a set of scores (Nunnally, 1978), in effect increasing the 'noise' in the data, making it more difficult to clearly estimate the construct of interest. Random sources of error are likely introduced to implicit attitude measures due to the reliance on rapid latency or 'speeded' responding paradigms. In an IAT, the aim of the task is to categorise the stimuli as quickly (and as accurately) as possible, and these reaction times form the key experimental data. Speeded responding is highly vulnerable to random error variance, due to natural variability in motor response execution, lapses in attentional focus, or differences in participant processing speed and attentional capacity (Han, Czellar, Olson, & Fazio, 2009; Poitou & Pouget, 2012). As such, speeded responding has the potential to introduce substantial non-trait or random error variance, resulting in volatile, unstable and variable results (Fiedler, Messner, & Bluemke, 2006; Lane et al., 2007). In addition, abstract constructs such as attitudes and personality are known to be less reliable than tools measuring more overt constructs, such as cognitive functions (Cote & Buckley, 1987; Spector, 2006). Implicit attitude measures are thus highly susceptible to random error variance, as they use the speeded response-time methodology to examine the abstract construct of attitudes.

Random error variance is sometimes viewed as a marginal issue for measurement, as it is theorised to be completely random, thereby having negligible overall impact on the results,

because of the assumption that error variance is spread evenly throughout the data. However, in the case of the IAT this assumption is implausible given it would require error to influence the congruent and incongruent trials of an IAT *equally* (for an overview of the IAT procedure see Greenwald et al., 1998). This is unlikely given the documented difference in cognitive effort required for the congruent and incongruent trials (Steffens, 2004; Williams & Themanson, 2011), and the finding that greater task difficulty results in greater random error variance (Brown & Heathcote, 2008). It is thus very plausible that measurement error is having a greater influence on the more cognitively challenging incongruent trials than the more intuitive congruent trials, resulting in non-random error distribution for IATs. This biasing influence could result in the production of larger IAT effect scores that ostensibly reveal greater levels of prejudice than would typically be produced, not because of any difference in implicit attitudes, but rather due to method-based systematic influences (see also Back, Schmukle, & Egloff, 2005 for further reasons that error may differentially impact congruent and incongruent trials in the IAT; Fiedler et al., 2006; Mierke & Klauer, 2003; Stülpnagel & Steffens, 2010).

1.2.2. Potential sources of systematic error variance for IATs

Random error variance is not the only potential issue for the IAT. Systematic error variance refers to characteristics often associated with the methodology that remain relatively consistent regardless of the construct being assessed (Coenders & Saris, 2000). Systematic forms of bias, such as method variance, can artificially inflate (or even deflate) estimates of internal consistency, test-retest reliability and inter-implicit convergent validity (Coenders & Saris, 2000), thereby contributing to misleading construct validity estimates for IATs and potentially inaccurate impressions of implicit bias. Given the abstract construct of implicit attitudes are so inextricably linked to the measurement techniques required to estimate them (such that there is confusion as to whether the term 'implicit' refers to the attitude or the measurement approach; Mitchell & Tetlock, 2017), it is pertinent to examine the influence of method variance on IAT data.

There are many potential sources of method variance for IATs which could be significantly reducing the veracity of IAT results. According to Podsakoff et al. (2003), sources of method error include common rater effects, item characteristic effects, item context effects and measurement context effects. Each of these effect types have implications for the IAT, and have been demonstrated (although not necessarily articulated as such) in previous implicit attitudinal research. In a novel application of Podsakoff et al.'s (2003) approach, these categories will now be applied to potential sources of method variance for IATs.

Common rater effects refer to artifactual covariance caused by the same respondent completing the measure. In other words, they are participant-related factors that may systematically influence the way the participant responds to the task (Podsakoff et al., 2003). For the IAT, factors such as social desirability responding, general processing speed (Blanton et al., 2006), intelligence (Stülpnagel & Steffens, 2010), task-switching ability (Back et al., 2005; Mierke & Klauer, 2003) and response style (i.e. accuracy vs. speed, Salthouse, 2000) are all factors that have been shown to influence IAT results and can contribute to common rater method effects.

Item characteristic effects refer to distinctive properties of an item/stimuli that can influence the way the participant responds, such as item demand characteristics, item ambiguity, common scale formats and positive/ negative wording of items (Podsakoff et al., 2003). Ambiguity of stimuli is known to be problematic for the IAT, often reducing the expected IAT effect (Messner & Vosgerau, 2010; Salthouse, 2000; Steffens, Kirschbaum, & Glados, 2008). There is also evidence of the participant's perception of stimuli positivity impacting IAT results (Messner & Vosgerau, 2010; Siers & Christiansen, 2013). Both of these types of method variance are item characteristic effects.

Item context effects refer to the interpretation of an item based on its relation to other items on the measure (Podsakoff et al., 2003). Interpretation of IAT stimuli and task requirements have been shown to be easily influenced by other tasks completed during a single testing session (Han et al., 2009). Block trial presentation has also been shown to systematically influence

performance, with much slower responses for incongruent trials on an IAT if completed following the congruent block of trials than vice versa (Lane et al., 2007; Williams & Themanson, 2011). Cvencek, Greenwald, & Meltzoff (2011) provide an example of this item context effect by finding significant positive attitudes towards flowers were present only when the congruent flower/good combined task was completed first, not when the incongruent flower/bad task was presented initially. These order effects are indicative of item context bias.

Lastly, measurement context effects refer to artifactual covariation produced from the context in which the responses are obtained, such as when two measures are completed at the same time, in the same location, using the same medium (Podsakoff et al., 2003). Because the IAT refers to a methodological format rather than a specific test, most validity studies require multiple versions of the IAT to be completed using the same measurement format, in the same measurement context, using similar item characteristics and a single common rater. As such measurement context effects likely also contribute to method variance for IATs, particularly in validation studies, potentially increasing the perceived overlap of the measures. Overall, there appears a high likelihood of multiple sources of method variance contributing to IAT scores which is problematic given systematic error variance can significantly influence psychometric evaluations (Mierke & Klauer, 2003).

Method variance can confound validity estimates, making it more difficult to ascertain the trait construct being assessed. This is because method variance affects the observed relationships between measurements simply due to shared method characteristics, rather than any underlying meaning (Mierke & Klauer, 2003). Method variance can thus artificially increase correlations between the absolute scores of *any* two IATs, even if they are not related by shared content (Coenders & Saris, 2000). This is problematic as it reduces the veracity of psychometric estimates for these measures, particularly when examining convergent validity. In Podsakoff et al.'s (2003) application of Cote and Buckley's (1987) research, it was demonstrated that two completely unrelated explicit attitude measures could produce an observed correlation of .23 once systematic and random error variance had been factored into the calculation (based on the

estimates derived by Cote & Buckley, 1987; of 40% systematic error and 30% random error variance). If method variance has inflated estimates of the degree of relationship between similar implicit attitudinal techniques, as Podsakoff et al.'s (2003) findings would suggest, then this implies any inter-implicit correlations smaller in magnitude than .23, for example, may potentially be regarded as solely a product of error variance. This is concerning given previously reported correlations between conceptually similar implicit attitude measures often range between zero and .30 (e.g. Krause et al., 2010; Rudolph et al., 2008). It is thus critical the influence of random and systematic error variance is clarified for the IAT.

1.3. Need for Latent Modelling Analytical Approaches to Address Measurement Error
A substantial quantity of implicit attitudinal research has employed analytical approaches based on the general linear model (e.g. ANOVA or linear regression). These analytic approaches rely on the assumption that any error incurred in the measurement of variables is completely random (Tabachnick & Fidell, 2013), which as previously demonstrated may be very unlikely for the IAT. Relying on general linear model analytic techniques would then likely result in statistical inaccuracies and biased results for implicit attitude measures (Kline, 2005). Analytical approaches reliant on the general linear model are also unable to account for systematic error or method variance.

It would thus be preferable to employ methods that are capable of identifying and partialling out systematic error or method variance, such as Structural Equation Modelling (SEM) procedures including, Confirmatory Factor Analysis (CFA; Jöreskog, 1969). These latent variable models separate error variance from the trait construct of interest, thereby accounting for non-random error distribution. However, these base methods do not allow a clear assessment of method variance.

The multitrait-multimethod (MTMM; Campbell & Fiske, 1959) approach to Confirmatory Factor Analysis (CFA-MTMM) requires multiple traits (or constructs) to be assessed by multiple measures (Schumacker & Lomax, 2010). This enables a score to be partitioned into trait, random

9

error and systematic error variance components, with error covariances used to estimate the method effects (Coenders & Saris, 2000; Malhotra, Kim, & Patil, 2006). This provides a clear estimate of the influence of random and systematic error on the data and has resulted in CFA-MTMM being regarded as one of the most rigorous methods of assessing and controlling for method variance currently available (Lance, Noble, & Scullen, 2002; Marsh & Grayson, 1995). In order to evaluate the relationship between method factors the CFA-MTMM model will be specified using the correlated methods approach (Marsh & Grayson, 1995). To apply this technique, the free estimation capacity of the correlated trait-correlated method (CTCM) approach to CFA-MTMM will be used. Free estimation between latent factors allows for a direct estimate of the relationship between method factors and is well suited to validation assessment (Lance et al., 2002). This is a key advantage as it means the CT-CM CFA-MTMM approach will allow two methods, such as the verbal and pictorial IAT, to be directly compared without the confounding influence of trait or random error variance. Thereby providing a clear indication of the impact of random and systematic error variance on IAT scores, and a robust estimate of construct validity of the experimental IATs.

1.4. The Present Study

The overall aim of this study was to determine the influence of random and systematic error variance on IAT data using CFA-MTMM. The CFA-MTMM procedure requires multiple traits are assessed by multiple methods. As such, verbal (VIAT; Greenwald et al., 1998) and pictorial IAT (PIAT; Thomas, Burton-Smith, & Ball, 2007) methods will be used to assess the trait constructs of racial attitudes (Arabs vs. Europeans) and country-related attitudes (Middle East vs. Europe). Attitudes towards Arabs and the Middle East were selected as the trait constructs for this study due to strong evidence of antipathy towards persons from the Middle East that has been well documented in Australia over the past 20 years (e.g. Briskman, 2015; Dunn, Forrest, Burnley, & McDonald, 2004; Islam & Jahjah, 2001; Tufail & Poynting, 2013). It was hoped this would provide a strong IAT response for the psychometric assessments in this research.

The present study aimed to use CT-CM CFA-MTMM to estimate the proportion of trait, method and random error variance in IAT scores. It was expected that IATs would show substantive components of random and systematic error variance, but that when such error sources were accounted for, good construct validity would be demonstrated for each IAT. Construct validity of each of the IATs would be demonstrated by each task possessing greater trait than method variance (Byrne, 1998).

2. Method

All measures, manipulations, and exclusions in this study are disclosed within this methods section. This research and its methodology was approved by the Tasmanian Human Research Ethics Committee Network (HREC: Reg No. H0010891).

2.1. Participants

One hundred and ninety-eight adult students (144 female – 72%) from Tasmania, Australia, with a mean age of 26.03 years (SD=11.10 years) participated in the study. The majority of participants identified their ethnicity as "Australian" (n=174). Students received course credit for participation.

2.2. Apparatus

The Race and Country VIATs and PIATs were presented individually on a laptop PC running the Inquisit software package (Millisecond Software, 1996).

2.2.1. Verbal Implicit Association Tests

Two traditional IAT measures depicting verbal word stimuli (VIATs) were developed in accordance with Greenwald, et al. (1998). The VIATs utilised standard Pleasant and Unpleasant word stimuli, such as "love" and "hatred". The Racial VIAT employed stereotypical Middle Eastern and European first names, such as "Habib" and "Harry". The Country VIAT used Middle Eastern and European country names, such as "Iraq" and "Italy".

11

2.2.2. Pictorial Implicit Association Tests

Two Pictorial IAT (PIAT) measures were developed in accordance with Thomas et al. (2007). These tasks depicted only pictorial stimuli instead of the usual word stimuli. The PIATs in this study employed positive and negative facial icons or 'emoticons' as well as faces of Middle Eastern and European people for the Racial PIAT, and easily recognisable buildings from Middle Eastern and European countries (e.g. mosques and churches, and other distinctive landmarks) for the Country PIAT (stimuli available on request from author). All stimuli were pilot tested to ensure equal valence.

2.3. Procedure

Following consent procedures, each participant completed six empirical implicit attitude measures and two explicit attitude measures individually. All participants completed two practice IATs of a non-related attitudinal construct (i.e. Flower-Insect, as per Greenwald et al., 1998) prior to attempting the empirical attitude measures. This aimed to minimise the confounding influence of task familiarity (see Greenwald et al., 2003). Task completion was fully randomised and counterbalanced to reduce the likelihood of order effects. For the IATs, all of the practice blocks were composed of 40 trials, with the four empirical blocks consisting of 102 trials each.

2.4. Scoring of the Implicit Association Tests

Scoring of the IAT was done in accordance with the IAT_D method outlined by (Greenwald et al., 2003). According to standard scoring procedures (Greenwald et al., 2003), participant data should be excluded if $\geq 10\%$ of an IATs trials were either missing, or showed response latencies >300ms for more than were excluded. However, no participant data was excluded for these reasons. All practice trials were removed from the analysis (as per B.A. Nosek, Greenwald, & Banaji, 2006). The first two trials of each empirical block were also removed in accordance with Greenwald et al. (1998). Any individual response latencies greater than 10,000 ms was removed and replaced with the response time of the second trial to avoid missing data (Greenwald et al., 2003). For each participant, the IAT scores were divided into four equivalent data parcels consisting of 25 response latencies from each of the first congruent, second congruent, first

incongruent and second incongruent blocks of data. This enabled four equivalent IAT_D scores to be produced for each participant. The four sets of IAT scores for the combined participant pool provided the CFA input data.

2.5. Statistical Analyses

The Multitrait-Multimethod analysis (MTMM; Campbell & Fiske, 1959) was performed using a Confirmatory Factor Analytic framework within Mplus, version 6.1 (Muthén & Muthén, 2010). The CFA-MTMM analysis applied the freely correlated trait-freely correlated method (CT-CM) specification approach which allows free estimation between trait or method factors, enabling a more direct comparison of the VIAT and PIAT methods. The specified path model is presented in Figure 3.1. In this analysis the traits referred to the two constructs being measured (Racial and Country implicit attitudes) and the methods refer to the IAT formats (Verbal and Pictorial). The analysis applied the robust maximum likelihood estimation procedure (MLM), which is an extension of ML estimation however the model chi-square and standard errors of the parameter estimates are corrected for non-normality within large samples (Brown, 2006). Mplus calculates the chi-square likelihood ratio test statistic (χ^2),but this statistic is affected substantially by sample size (Brown, 2015). In view of this, model fit was assessed using the four goodness-of-fit indices derived from the Mplus program; In the incremental fit indices: comparative fit index (CFI) and the Tucker-Lewis index (TLI); and the absolute fit indices of root-mean-square error of approximation (RMSEA); and the standardised root mean square residual (SRMR). CFI and TLI $\geq$.95 are indicative of good model-data fit (Brown, 2006). Hu and Bentler (1999) suggest good fit is demonstrated with RMSEA $\leq$ 06, and SRMR $\leq$ .08.

3. Results

The CT-CM CFA-MTMM analysis showed that the specified model (Figure 3.1) was a good fit to the data, χ^2 (86, N=198) =93.61, *p*=.27; CFI=.99; TLI=.98; RMSEA=.02; and SRMR=.04. The specified path model is presented in Figure 3.1 and depicts the partitioning of the IAT data

parcels into latent trait, random error and method components. It is noted that each IAT was individually assessed as per Chequer and Quinn (2020).
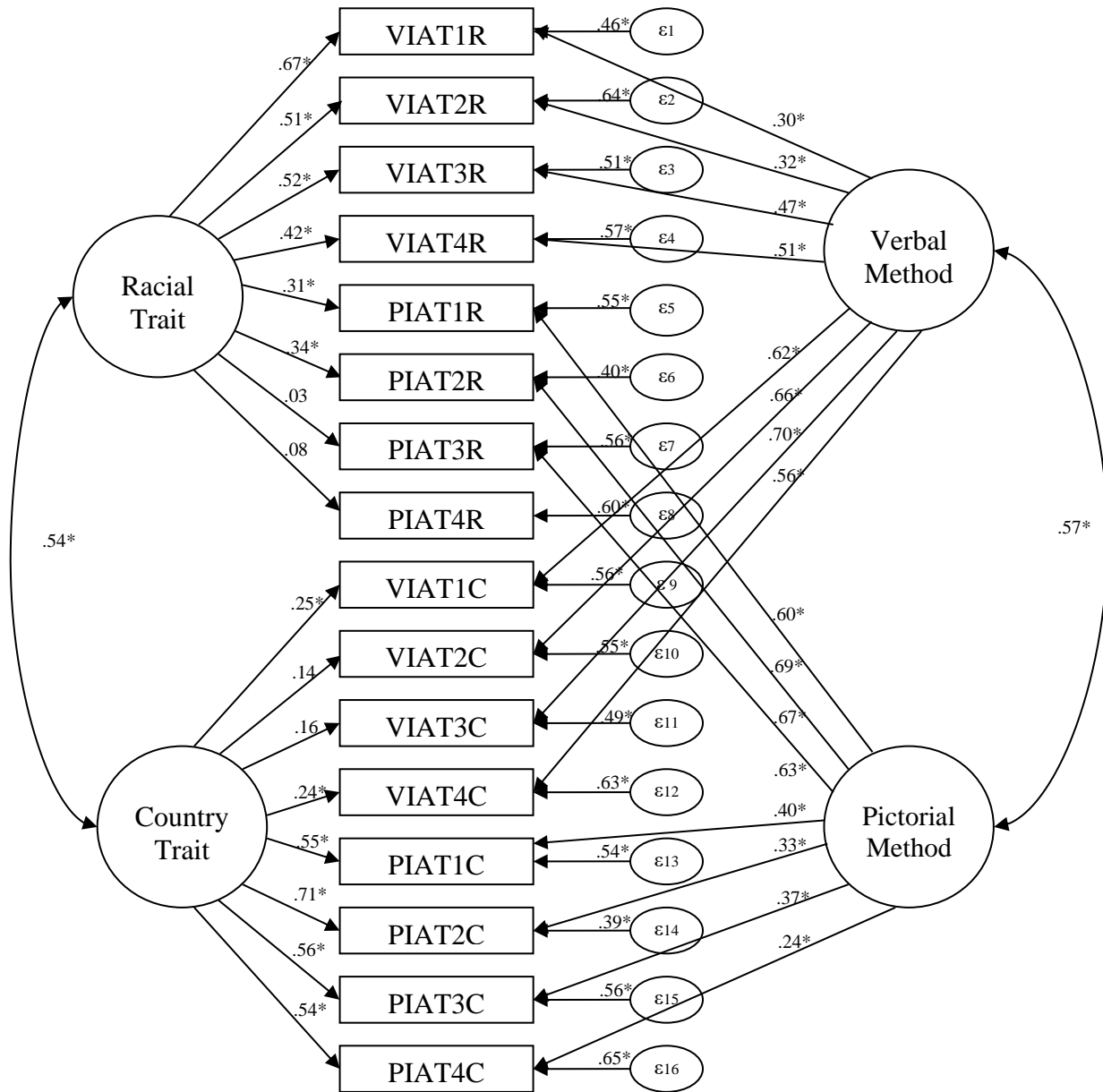
VIAT1R .46* ε1
VIAT2R .64* ε2
VIAT3R .51* ε3
VIAT4R .57* ε4
PIAT1R .55* ε5
PIAT2R .40* ε6
PIAT3R .56* ε7
PIAT4R .60* ε8
VIAT1C .56* ε9
VIAT2C .55* ε10
VIAT3C .49* ε11
VIAT4C .63* ε12
PIAT1C .54* ε13
PIAT2C .39* ε14
PIAT3C .56* ε15
PIAT4C .65* ε16

Racial Trait
Country Trait
Verbal Method
Pictorial Method

.67* .51* .52* .42* .31* .34* .03 .08
.25* .14 .16 .24* .55* .71* .56* .54*
.54*
.30* .32* .47* .51* .62* .66* .70* .56*
.60* .69* .67* .63* .40* .33* .37* .24*
.57*

Figure 3.1. CT-CM CFA-MTMM model depicting the data of four IATs that have been separated into trait, error and method components. Standardised factor loadings (STDYX) for each of the variables and residuals are displayed. *p<.001.

3.1. Evidence of Random Error and Method Effects in the IAT

Substantial and significant error variance loadings onto all indicators were found (mean error loading .53), suggesting significant amounts of random error were present in the scores. Substantial and significant factor loadings of the indicators onto the latent method factors were also demonstrated. The average of these factor loadings was .50, above that suggested for meaningful latent effects (Gorsuch, 1983). This provides strong evidence of significant method effects within the IAT data.
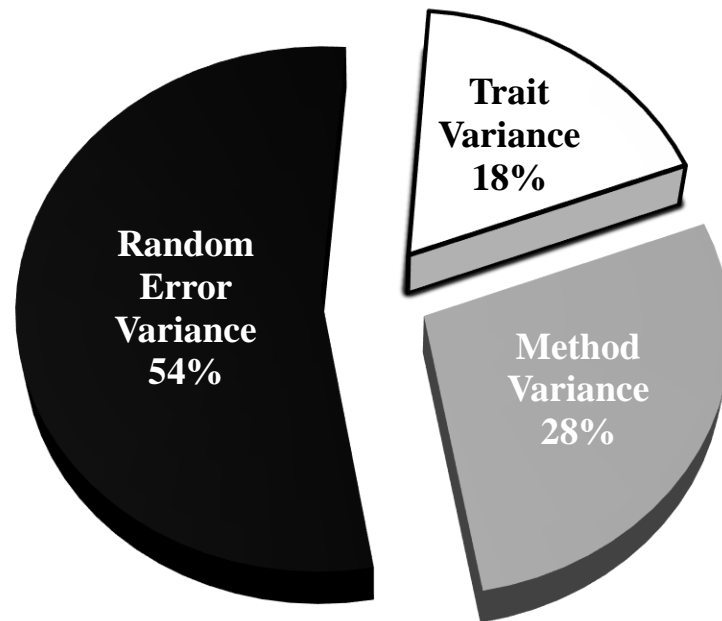
3.2. Composition of Trait, Method and Random Error Variance in IAT Data

Individual variance parameters resulting from the CT-CM CFA-MTMM analysis are presented in Table 3.2. The values for trait and method are the squared standardised loadings, and, together with the error variances, they specify the amount of variance in each IAT data parcel attributable to trait, method and random error effects respectively. Random error variance appeared relatively stable, accounting for about 54% of variance (see Table 3.2). Overall, the parameter estimates revealed 28% of the variance was attributable to method effects (see Table 3.2). This is a substantial portion, particularly when compared to the lesser 18% of variance attributable to the trait construct supposedly being tapped by the technique. These results reveal that mean IAT effect scores were comprised of 54% random error variance, 28% method variance and only 18% trait variance (see Figure 3.2). In other words, over 80% of an IAT effect score was shown to be error variance compared to less than 20% reflecting the trait construct of interest.

Table 3.2.
Squared Standardised Trait and Method Loadings, and Error Variance in IAT Effect Scores.

	Trait Variance	Method Variance	Error Variance
Race			
VIAT1R	.45	.09	.46
VIAT2R	.26	.10	.64
VIAT3R	.27	.22	.51
VIAT4R	.18	.26	.57
Mean variance	*.29*	*.17*	*.55*
PIAT1R	.10	.36	.55
PIAT2R	.12	.48	.40
PIAT3R	.01	.45	.56
PIAT4R	.01	.40	.60
Mean variance	*.05*	*.42*	*.53*
Country			
VIAT1C	.06	.38	.56
VIAT2C	.02	.44	.55
VIAT3C	.03	.49	.49
VIAT4C	.06	.31	.63
Mean variance	*.04*	*.41*	*.56*
PIAT1C	.30	.16	.54
PIAT2C	.50	.11	.39
PIAT3C	.31	.14	.56
PIAT4C	.29	.06	.65
Mean variance	*.35*	*.12*	*.54*
OVERALL			
Mean variance (SD)	*18 (.16)*	*.28 (.15)*	*.54 (.08)*

Figure 3.2. Graphical representation of the percentage variance of average IAT effect scores attributable to trait, method and random error variance.

3.3. Construct Validity Results for the IATs using CFA-MTMM

When examining the individual parameters, inconsistent construct validity evidence was revealed. Two of the four IATs, the Racial VIAT and the Country PIAT, were found to possess higher levels of trait (average 32%) than method (average 14.25%) variance. This ostensibly supports the construct validity of these two measures. Conversely, the other two IATs (the Country VIAT and the Racial PIAT) presented the opposite pattern, with little trait accounted for (average 4.5%) and a substantial proportion of method variance present (41.38%; see Table 3.2). As such, the CFA-MTMM analysis simultaneously provided strong construct validity evidence for the Racial VIAT and the Country PIAT, whilst weakening support for the Racial PIAT and the Country VIAT.

4. Discussion

4.1. Evidence of Random Error and Method Effects in IAT Scores

Random error variance was found to be the greatest contributor to IAT effects, comprising over 50% of the IAT effect scores. Random error variance is known to significantly confound validity estimates, and can provide an upper limit for observed correlations between like tasks (Cole & Maxwell, 2003; Cunningham et al., 2001). The present findings of high random error, may help explain the poor convergent validity estimates typically reported for IATs (e.g. Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005; Krause et al., 2010). Random error variance is also likely to have limited other construct validity estimates of this task.

Systematic error variance was also shown to play a substantial role in IAT effect scores, accounting for about a third of the variance in the IAT scores according to the parameter estimates depicted in Table 3.2. Findings of such a large component of systematic error imply past estimates of convergence between implicit attitudinal measures may have been inflated to an unknown degree. Such inflation could mean estimates of convergent validity for the IAT have been inaccurate, and potentially poorer than originally anticipated - a concerning prospect given the often very low inter-implicit correlations reported in the implicit attitudinal literature (e.g. Krause et al., 2010; Rudolph et al., 2008).

Given approximately 80% of the IAT effect scores are attributable to error variance (random and systematic) it would seem near impossible to deliver accurate estimates of implicit attitudes without analytically addressing error variance in the data. The current findings call to question the validity of previous implicit attitudinal findings based on non-latent analytical assessment that have failed to account for such substantial error components. It is recommended all future IAT research routinely employ latent modelling techniques to account for error variance in implicit attitudinal data.

4.2. Discrepant Construct Validity Evidence for the IAT using CFA-MTMM

An examination of the parameter estimates (presented as variances in Table 3.2) provided a discrepant overview of the construct validity of the experimental IATs in this study. The Racial VIAT and Country PIAT were revealed to possess strong construct validity, with levels of trait variance greater than method variance. However, the Country VIAT and Racial PIAT presented the opposite findings, failing to demonstrate adequate construct validity because the method variance was greater than the trait variance for the tasks. These results are difficult to interpret given each method was both deemed satisfactory and unsatisfactory depending on which construct was assessed. Likewise, each construct was either acceptable or not depending on which method was used. As such, there was not clear evidence to support the use of a particular IAT format nor the finding that one construct was more easily accessed than the other.

Seemingly nonsensical results such as these reduce the overall construct validity evidence for the IAT format and emphasise the need for each IAT to be individually validated. The direct implication is that the IAT is not able to be easily adapted to assess many and varied constructs, as it was designed to do (Greenwald et al., 2009). But rather, in trying to be so adaptable, the IAT appears to have forfeited the ability to be a consistently valid measurement technique. Highly standardised and psychometrically valid measures, such as IQ tests for instance, cannot be changed based on the focus on interest, for it is evident that any changes to the format and questions will affect the veracity of the result. The IAT being an easily adaptable measure that is also psychometrically robust is thus an improbable outcome. In order for the IAT to increase psychometric robustness, each and every IAT requires thorough psychometric validation (see also Lane et al., 2007). Future research would do well to focus on standardisation and thorough psychometric validation of a small number of IATs, rather than focusing on extensive applications for the IAT.

4.3. Implications and Future Directions for Implicit Attitudinal Measurement

The key finding of this research is that error variance is the primary contributor to IAT scores. Because of this, it is important that error variance is accounted for using structural equation

modelling (SEM) analytical strategies, such as CFA, before any applied meaning can be gleaned from the data. The disadvantage of requiring SEM, is this analytical approach restricts the use of IATs as a measure of individual implicit bias.

4.3.1. Requirement for SEM to Account for High Error Variance in IAT Scores

The finding that IAT scores were comprised of over 80% error variance has significant implications for scoring and interpretation of IAT data. Indeed, it would prove almost impossible to ascertain a measure of implicit attitudes without first addressing the substantial error component inherent in the scores. In the current study, SEM analyses such as CFA & MTMM were demonstrated to be effective options for addressing the error component and enabling a clearer impression of the trait implicit attitude constructs under investigation. It is hoped routine application of SEM to all implicit attitude data would help increase the psychometric standards of laboratory techniques such as the IAT and allow for more confidence in IAT findings. Implicit attitude researchers are encouraged to focus on developing a set of psychometrically validated IATs, such as the Racial VIAT and Country PIAT of the present study. These measures could then be used for applied research, but the resulting IAT data would still require latent modelling analysis to account for error variance prior to ascertaining trait implicit attitudes. Research that fails to account for error variance using latent modelling is very unlikely to provide an accurate reflection of the participants' implicit associations and should be treated with caution.

4.3.2. IAT Effect Scores as a Measure of Individual Bias

The need for SEM analytical strategies to account for significant error variance in the IAT effect scores has substantial implications for the use of the IAT effect score as a diagnostic tool of personal implicit prejudices. SEM can only ever deliver an estimate of an individual's effect score based on the scores of a whole sample population (see Skrondal & Laake, 2001). This means that although it is theoretically possible for an individual's IAT effect score to be determined using SEM, at best it could only ever produce a quasi-individual diagnostic estimate. Even then, the practicability, or lack thereof, of such an approach would render the concept implausible. This means that in order to gain theoretically meaningful results from IATs using

SEM, IAT research must be constrained to examining implicit bias at the sample population level rather than an individual by individual basis. Because of this, it appears inappropriate to present feedback regarding personal implicit prejudices when the task is not sensitive enough to provide this accurately. Given the current findings, it is strongly suggested future IAT research provide summarised findings at a sample population level, after latent modelling analytical strategies are applied, rather than delivering personal, and potentially misleading, feedback to participants.

4.4. General Conclusions Regarding the Impact of Error Variance on the IAT
IAT scores were shown to be heavily confounded by error variance. The primary contributor to IAT effect scores was random error, accounting for more than half of the variance, a further third was attributable to method variance and trait variance (representative of 'implicit attitude') was shown to have the least influence, accounting for less than a fifth of the IAT score. These findings significantly reduce the veracity of previously reported IAT findings where error variance was not adequately accounted for. The key implication of this finding is that all future IAT research should be required to account for these substantial portions of error variance by applying SEM analytical techniques such as CFA prior to any interpretation of the results. However, it should be noted that even once error variance was accounted for there were still significant inconsistencies in the construct validity evidence produced for the four IATs of the present study. These inconsistencies are worrisome for the application of IATs to the assessment of varied attitudinal constructs. It appears critical that each IAT is individually validated using CFA procedures to ensure adequate psychometric properties are present, prior to testing with the aim of procuring implicit attitudes.

To facilitate future implicit attitudinal research, development of a few standardised and well validated IATs is encouraged. These measures could be used to examine implicit attitudes at a sample population level, following the use of latent modelling analytical strategies to partial out the significant portion of measurement error incurred during the IAT. Whilst this may appear a labour-some and inconvenient solution for implicit attitudinal research, it is surely preferable to

the alternative of providing inaccurate and potentially misleading results that have both personal and social implications. Investing in the development of psychometrically robust implicit attitude measurement will provide dividends for the accessibility and generalisability of future implicit attitudinal findings.

4. Conclusion

Nosek and Greenwald (2009, p. 375) note that "the most important considerations in appraising validity of psychological measures are those that speak to the measure's usefulness in research and application". Whilst there have been many concerns regarding the IAT's veracity and usefulness (see Blanton et al., 2009; Krause et al., 2010; Mitchell & Tetlock, 2017; Oswald et al., 2015; Rae & Olson, 2018), there has been no clear estimate for the component of error variance in IAT scores. The present study has provided clarity on this issue, demonstrating that the IAT effect scores were comprised of over 80% combined random and systematic error variance, allowing little opportunity for trait 'implicit attitudes' to be revealed through the noise, and requiring significant statistical modifications and processing to obtain even population-level 'insights into our implicit biases'. To put it simply, the IAT was shown to be inadequately honed to provide insights into our implicit biases and its 'usefulness in research and application' is questionable, if not at times, potentially misleading. The sheer magnitude of error variance has serious implications for the use and interpretation of IAT effect scores. However, as Campbell and Fiske (1959, p. 103) note "psychologists today should be concerned not with evaluating tests as if the tests were fixed and definitive, but rather with developing better tests". It is hoped this study will spur a renewed focus on addressing measurement error in social psychological research, and encourage greater psychometric validation of tasks to enable more useful and applicable insights into implicit biases into the future.

References

Back, M. D., Schmukle, S. C., & Egloff, B. (2005). Measuring task-switching ability in the Implicit Association Test. *Experimental Psychology, 52*, 167-179. doi:10.1027/1618-3169.52.3.167

Bar-Anan, Y., & Nosek, B. A. (2014). A compartive investigation of seven indirect attitude measures. *Behavioral Research Methods, 46*, 668-688. doi:10.3758/s13428-013-0410-6

Blanton, H., Jaccard, J., & Burrows, C. N. (2015). Implications of the Implicit Association Test D-transformation for psychological assessment. *Assessment, 22*, 429-440. doi:10.1177/1073191114551382

Blanton, H., Jaccard, J., Gonzales, P. M., & Christie, C. (2006). Decoding the implicit association test: Implications for criterion prediction. *Journal of Experimental Social Psychology, 42*, 192-212. doi:10.1016/j-jesp.2005.07.003

Blanton, H., Jaccard, J., Klick, J., Mellers, B., Mitchell, G., & Tetlock, P. E. (2009). *Strong claims and weak evidence: Reassessing the predictive validity of the IAT*. Faculty Scholarship. University of Pennsylvania Law School. Pennsylvania. Retrieved from http://scholarship.law.upenn.edu/faculty_scholarship/1532

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. doi:10.1037/0022-3514.79.4.631

Briskman, L. (2015). The creeping blight of Islamophobia in Australia. *International Journal for Crime, Justice and Social Democracy, 4*, 112-121. doi:10.5204/ijcjsd.v3i2.244

Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology, 57*, 153-178. doi:10.1016/j.cogpsych.2007.12.002

Brown, T. A. (2015). *Confirmatory Factor Analysis for applied research* (2nd ed.). New York: Guilford Press.

Buchner, A., & Wippich, W. (2000). On the reliability of implicit and explicit memory measures. *Cognitive Psychology, 40*, 227-259. doi:10.1006/cogp.1999.0731

Buhrmester, M. D., Blanton, H., & Swann, W. B. (2011). Implicit self-esteem: Nature, measurement, and a new way forward. *Journal of Personality and Social Psychology, 100*, 365-385. doi:10.1037/a0021341

Byrne, B. M. (1998). *Structural Equation Modelling with Lisrel, Prelis, and Simplis*. New Jersey: Lawrence Erlbaum Associates.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81-105. doi:10.1037/h0046016

Carlsson, R., & Agerstrom, J. (2016). A closer look at the discrimination outcomes in the IAT literature. *Scandinavian Journal of Psychology, 57*, 278-287. doi:10.1111/sjop.12288

Chequer, S. R., & Quinn, M. G. (2020). Construct validation of the Pictorial Implicit Association Test (PIAT): Implications for implicit attitudinal research. doi:10.31234/osf.io/z9xfu

Coenders, G., & Saris, W. E. (2000). Systematic and random method effects. Estimating method bias and method variance. *Metodoloski Zvezki, 15*, 55-74.

Cole, D. A., & Maxwell, S. E. (2003). Testing mediational models with longitudinal data: Questions and tips in the use of Structural Equation Modelling. *Journal of Abnormal Psychology, 112*, 558-577. doi:10.1037/0021-843X.112.4.558

Cote, J. A., & Buckley, M. R. (1987). Estimating trait, method, and error variance: Generalizing across 70 construct validation studies. *Journal of Marketing Research, 24*, 315-318. doi:10.2307/3151642

Cunningham, W. A., Preacher, K. J., & Banaji, M. R. (2001). Implicit attitude measures: Consistency, stability, and convergent validity. *Psychological Science, 12*, 163-170. doi:10.1111/1467-9280.00328

Cvencek, D., Greenwald, A. G., & Meltzoff, A. N. (2011). Measuring implicit attitudes of 4-year-olds: The Preschool Implicit Association Test. *Journal of Experimental Child Psychology, 109*, 187-200. doi:10.1016/j.jecp.2010.11.002

Dasgupta, N., McGhee, D. E., Greenwald, A. G., & Banaji, M. R. (2000). Automatic preference for white Americans: Eliminating the familiarity explanation. *Journal of Experimental Social Psychology, 36*, 316-328. doi:10.1006/jesp.1999.1418

Dunn, K. M., Forrest, J., Burnley, I., & McDonald, A. (2004). Constructing racism in Australia. *Australian Journal of Social Issues, 39*, 409-430.

Falk, C. F., Heine, S. J., Takemura, K., Zhang, C. X. J., & Hsu, C. W. (2013). Are implicit self-esteem measures valid for assessing individual and cultural differences? *Journal of Personality*, 1-35. doi:10.1111/jopy.12082

Feest, U. (2020). Construct validity in psychological tests - the case of implicit social cognition. *European Journal for Philosophy of Science, 10*, 1-24. doi:10.1007/s13194-019-0270-8

Fiedler, K., Messner, C., & Bluemke, M. (2006). Unresolved problems with the "I", the "A", and the "T": A logical and psychometric critique of the Implicit Association Test (IAT). *European Review of Social Psychology, 17*, 74-147. doi:10.1080/10463280600681248

Foroni, F., & Bel-Bahar, T. (2010). Picture-IAT versus Word-IAT: Level of stimulus representation influences on the IAT. *European Journal of Social Psychology, 40*, 321-337. doi:10.1002/ejsp.626

Gawronski, B., Morrison, M., Phills, C. E., & Galdi, S. (2017). Temporal stability of implicit and explicit measures: A longitudinal analysis. *Personality and Social Psychology Bulletin, 43*, 300-312. doi:10.1177/0146167216684131

Gorsuch, R. L. (1983). *Factor Analysis* (2nd ed.). New Jersey, USA: Laurence Erlbaum Associates.

Greenwald, A. G., Banaji, M., & Nosek, B. A. (2015). Statistically small effects of the Implicit Association Test can have societally large effects. *Journal of Personality and Social Psychology, 108*, 553-561. doi:10.1037/pspa0000016

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. doi:10.1037//0022-3514.74.6.1464

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. doi:10.1037/0022-3514.85.2.197

Han, H. A., Czellar, S., Olson, M. A., & Fazio, R. H. (2009). Malleability of attitudes or malleability of the IAT? *Journal of Experimental Social Psychology, 46*, 286-296. doi:10.1016/j.jesp.2009.11.011

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the Implicit Association Task and explicit self-report measures. *Personality & Social Psychology Bulletin, 31*, 1369-1385. doi:10.1177/0146167205275613

Hofmann, W., & Schmitt, M. (2008). Advances and challenges in the indirect measurement of individual differences at age 10 of the Implicit Association Test. *European Journal of Psychological Assessment, 24*, 207-209. doi:10.1027/1015-5759.24.4.207

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modelling, 6*, 1-55.

Islam, M. R., & Jahjah, M. (2001). Predictors of young Australians' attitudes toward Aboriginals, Asians and Arabs. *Social Behavior and Personality, 29*, 569-580. doi:10.2224/sbp.2001.29.6.569

Jöreskog, K. G. (1969). Structural analysis of covariance and correlation matrices. *Psychometrika, 43*, 443-477. doi:10.1007/BF02293808

Kline, T. J. B. (2005). *Psychological Testing: A practical approach to design and evaluation*. Calgary, Canada: SAGE.

Krause, S., Back, M. D., Egloff, B., & Schmukle, S. C. (2010). Reliability of implicit self-esteem measures revisited. *European Journal of Personality*. doi:10.1002/per.792

Lance, C. E., Noble, C. L., & Scullen, S. E. (2002). A critique of the correlated trait-correlated method and correlated uniqueness models for multitrait-mutimethod data. *Psychological Methods, 7*, 228-244. doi:10.1037//1082-989X.7.2.228

Lane, K. A., Banaji, M. R., Nosek, B. A., & Greenwald, A. G. (2007). Understanding and Using the Implicit Association Test: IV; What we know (so far) about the method. In B. Wittenbrink & N. Schwartz (Eds.), *Implicit measures of attitudes* (pp. 59-102). New York: The Guilford Press.

Malhotra, N. K., Kim, S. S., & Patil, A. (2006). Common method variance in IS research: A comparison of alternative approaches and a reanalysis of past research. *Management Science, 52*, 1865-1883. doi:10.1287/mnsc.1060.0597

Marsh, H. W., & Grayson, D. (1995). Latent variable models of multitrait-multimethod data. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 177-198). California: Sage.

Messner, C., & Vosgerau, J. (2010). Cognitive inertia and the Implicit Association Test. *American Marketing Association, 47*, 374-386. doi:10.1509/jmkr.47.2.374

Mierke, J., & Klauer, K. C. (2003). Method-specific variance in the Implicit Association Test. *Journal of Personality and Social Psychology, 85*, 1180-1192. doi:10.1037/0022-3514.85.6.1180

Millisecond Software. (1996). Inquisit. Seattle, WA: Millisecond Software. Retrieved from http://www.millisecond.com/

Mitchell, G., & Tetlock, P. E. (2017). Popularity as a poor proxy for utility; The case of implicit prejudice. In S. O. Lilienfeld & I. D. Waldman (Eds.), *Psychological Science Under*

Scrutiny; Recent challenges and proposed solutions (pp. 164-195). West Sussex, UK: John Wiley & Sons.

Muthén, L. K., & Muthén, B. O. (2010). *Mplus user's guide* (6th ed.). Los Angeles, California: Muthén & Muthén.

Nosek, B. A., & Greenwald, A. G. (2009). (Part of) the case for a pragmatic approach to validity: Comment on De Houwer, Teige-Mocigemba, Spruyt, and Moors (2009). *Psychological Bulletin, 135*, 373-376. doi:10.1037/a0015047

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2006). The Implicit Association Test at age 7: A methodological and conceptual review. In J. A. Bargh (Ed.), *Social psychology and the unconscious: The automaticity of higher mental processes* (pp. 265-292). New York: Psychology Press.

Nunnally, J. C. (1978). *Psychometric Theory*. New York: McGraw-Hill.

Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2013). Predicting ethnic and racial discrimination: A meta-analysis of IAT criterion studies. *Journal of Personality and Social Psychology, 105*, 171-192. doi:10.1037/a0032734

Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2015). Using the IAT to predict ethnic and racial discrimination: Small effect sizes of unknown societal significance. *Journal of Personality and Social Psychology, 108*, 562-571. doi:10.1037/pspa0000023

Pieters, S., van der Vorst, H., Engels, R. C., & Wiers, R. W. (2010). Implicit and explicit cognitions related to alcohol use in children. *Addictive Behaviors, 35*, 471-478. doi:10.1016/j.addbeh.2009.12.022

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*, 879-903. doi:10.1037/0021-9010.88.5.879

Poitou, T., & Pouget, P. (2012). Variability, noise and predictability in motor response times: Adaptation or misadaptation? In J. M. Norvilitis (Ed.), *Contemporary Trends in ADHD Research* (pp. 15-28). doi:10.5772/28166

Rae, J. R., & Olson, K. R. (2018). Test-retest reliability and predictive validity of the Implicit Association Test in Children. *Developmental Psychology, 54*, 308-330. doi:10.1037/dev0000437

Rudolph, A., Schröder-Abé, M., Schütz, A., Gregg, A. P., & Sedikides, C. (2008). Through a glass, less darkly? Reassessing convergent and discriminant validity in measures of implicit self-esteem. *European Journal of Psychological Assessment, 24*, 273-281. doi:10.1027/1015-5759.24.4.273

Salthouse, T. A. (2000). Aging and measures of processing speed. *Biological Psychology, 54*, 35-54. doi:10.1016/S0301-0511(00)00052-1

Schumacker, R. E., & Lomax, R. G. (2010). *A beginner's guide to Structural Equation Modeling* (3rd ed.). New York: Routledge.

Siers, B. P., & Christiansen, N. D. (2013). On the validity of implicit association measures of personality traits. *Personality and Individual Differences, 54*, 361-366. doi:10.1016/j.paid.2012.10.004

Singal, J. (2017). Psychology's favorite tool for measuring racism isn't up to the job. *The Cut*. Retrieved from https://www.thecut.com/2017/01/psychologys-racism-measuring-tool-isn't-up-to-the-job.html

Spearman, C. (1904). "General intelligence", objectively determined and measured. *American Journal of Psychology, 15*, 201-293. doi:10.2307/1412107

Spector, P. E. (2006). Method variance in organizational research: Truth or urban legend? *Organizational Research Methods, 9*, 221-232. doi:10.1177/1094428105284955

Steffens, M. C. (2004). Is the Implicit Association Test immune to faking? *Experimental Psychology, 51*, 165-179. doi:10.1027/1618-3169.51.3.165

Steffens, M. C., Kirschbaum, M., & Glados, P. (2008). Avoiding stimulus confounds in Implicit Association Tests by using the concepts as stimuli. *British Journal of Social Psychology, 47*, 212-243. doi:10.1348/014466607X226998

Stülpnagel, R. v., & Steffens, M. C. (2010). Prejudiced or just smart? Intelligence as a confounding factor in the IAT effect. *Zeitschrift für Psychologie, 218*, 51-53. doi:10.1027/0044-3409/a000008

Tabachnick, B. G., & Fidell, L. S. (2013). *Using Multivariate Statistics* (6th ed.). Boston: Pearson Education.

Thomas, S. R., Burton-Smith, R., & Ball, P. (2007). Implicit Attitudes in Very Young Children: An adaptation of the IAT. *Current Research in Social Psychology, 13*, 75-85.

Tufail, W., & Poynting, S. (2013). A common 'outlawness': Criminalisation of Muslim minorities in the UK and Australia. *International Journal for Crime, Justice and Social Democracy, 2*, 43-54.

Williams, J. K., & Themanson, J. R. (2011). Neural correlates of the Implicit Association Test: Evidence for semantic and emotional processing. *Social Cognitive and Affective Neuroscience, 6*, 468-476. doi:10.1093/scan/nsq065