Evaluating the Construct Validity of Implicit Association Tests using Confirmatory Factor Analytic Models

Susan Ruth Chequer

B.A. (Hons.), MPsych (Clin.)

Submitted in fulfilment of the requirements for the
Degree of Doctor of Philosophy
School of Psychology
University of Tasmania, Australia

October 2014

Declaration

This thesis contains no material which has been accepted for a degree or diploma by the university or any other institution. To the best of my knowledge and belief, this thesis contains no material previously published or written by another person except where due acknowledgement is made, nor does it contain any material that infringes copyright. This thesis may be available for loan and limited copying in accordance with the Copyright Act, 1968.

_____________________________ *Date* _____________________

Susan Chequer, MPsych. (Clin.)

Statement of Ethical Conduct

The research associated with this thesis abides by the international and Australian codes on human and animal experimentation, the guidelines by the Australian Government's Office of the Gene Technology Regulator and the rulings of the Safety, Ethics and Institutional Biosafety Committees of the University.

_____________________________ *Date* _____________________

Susan Chequer, MPsych. (Clin.)

Acknowledgements

I would like to acknowledge the guidance and encouragement of my primary supervisor Prof Rapson Gomez, particularly in the beginning years of my PhD. I would never have thought it possible to undertake such a technically orientated research project without your support. Secondly, I must sincerely thank Dr Michael Quinn my dear friend and colleague. There is no way I can repay your generosity, for you have given so much time and energy to supporting me in this endeavour. I wish to acknowledge the support of Dr Ben Schüz over the last 18 months, thank you for being the face of the Psychology Department and for stepping up when it was most needed. Further thanks go to my secondary supervisor Peter Ball, Prof Jenn Scott, and Dr Matthew Palmer for their support and/or feedback.

To my loving and long-suffering husband Nick, thank you for enduring this process and enabling me to fulfil this ambition. There is no way I could have done this without you. To my parents, thank you for all your support, encouragement and guidance. Your words have shaped who I am. To my dad and dear friends Emily, Thug and Kain, thank you for proofing portions of this thesis. Anna, massive thanks for your hundreds of supportive text messages and continual encouragement, it meant the world. Also Kat for your words of wisdom. Thanks to my running partners over the years who kept me (mostly) sane: Anna, Rhea, Jess, Tegan, Sandra, Katherine and Alex. I would like to acknowledge my boss Dr Georgina O'Donnell for being so understanding and encouraging during this process. A big thank you to all my family and friends who continue to fill my life with love and laughter. Lastly, my everlasting gratitude goes to my Lord who gives me perseverance and purpose.

Abstract

The Implicit Association Test (IAT) is the most widely used method for assessing implicit bias and prejudice. By avoiding the need for introspection, the IAT is suggested to be a more valid indicator of prejudice than explicit measures of attitudes (i.e. questionnaires). However, implicit attitudinal literature has demonstrated highly variable associations between IAT scores and various outcomes. Such inconsistencies imply IAT scores may be significantly influenced by measurement error, which could thwart efforts to accurately estimate underlying attitudes. The aim of the present thesis was to examine the construct validity of the IAT using Confirmatory Factor Analytic models (CFA) to account for the confounding influences of measurement error.

Three studies examined various aspects of the validity of IATs using data from 198 student participants of the University of Tasmania, Australia. Study One assessed the internal consistency and internal convergent validity of traditional verbal IATs, fully pictorial IATs and Affective Priming Tasks (APTs) using single-group CFA. The study revealed high amount of random error variance in the implicit attitudinal data, comprising around 55% of IAT scores and 95% of APT scores. Despite the high proportion of random error, the IATs appeared to consistently assess the trait attitude constructs, though this was not true for the APTs. The APTs were consequently deemed invalid measures of implicit attitudes.

Study Two added to the findings of Study One by further accounting for method variance in the IAT data using the CTCM CFA-MTMM analytical approach. This study indicated that method variance accounts for a further third of the IAT scores, suggesting that an average IAT score is comprised of around 80% error variance (random error and method variance). Notwithstanding this, after accounting for measurement error, strong convergence was evident between the verbal and pictorial IAT formats and two of the four IATs were found to possess good construct validity. Such findings provided some optimism for the future development of psychometrically robust implicit attitude techniques.

Study Three examined the application of IATs to assess implicit attitudes whilst using latent modelling techniques to account for the significant error component of the scores. Specialised CFA models were used to reveal anti-Arab/pro-European bias in the present sample, as well as determine the effect of certain participant characteristics, such as age, on the IAT scores.

In summary, the studies of this thesis suggest that IAT scores are likely to be confounded substantially by error variance at the individual level. However, if random error and method variance are partialled out, IAT scores can provide an adequate assessment of implicit attitudes. This suggests that future IAT applications would profit from analysing sample data using CFA or other latent modelling techniques to account for the significant error component of IAT scores.

Table of Contents

List of Tables

List of Figures

List of Equations

List of Abbreviations

Attitude Measures

APT	Affective Priming Task	Computer-based priming task designed by Fazio et al. (1986) to assess implicit attitudes
IAT	Implicit Association Test	Computer-based program designed by Greenwald et al. (1998) to measure entrenched attitudes and stereotypes
PIAT	Pictorial Implicit Association Test	Fully pictorial version of the IAT that involves categorising only pictorial stimuli (Thomas et al., 2007)
VIAT	Verbal Implicit Association Test	Traditional verbal version of the IAT that involves categorising word stimuli
MRS	Modern Racism Scale	Race-related explicit attitude questionnaire consisting of six items (McConahay et al., 1981)
TDQ	Travel Destination Questionnaire	Explicit attitude questionnaire relating to preference to travel to each of 18 countries

Statistical Methods

AVE	Average Variance Extracted	A method for determining an estimate of internal convergent validity
CFA	Confirmatory Factor Analysis	Latent modelling technique used for hypothesis-testing and construct validation (Jöreskog, 1969)
CFI	Comparative Fit Index	Non-centrality based fit index used to determine how well the estimated model fits the data in Factor Analysis
CR	Composite Reliability	An estimate of internal consistency
CT-CM	Freely Correlated Trait – Freely Correlated Method	Specification approach often applied within CFA-MTMM
EFA	Exploratory Factor Analysis	Latent modelling technique aimed at exploring the relationships between latent factors
MIMIC	Multiple Indicator Multiple Causes	Latent modelling technique whereby covariates are added to a standard CFA model to determine their effects on the latent factors
MLM	Maximum Likelihood Method	A method used to estimate factors during a Factor Analysis
MTMM	Multitrait-Multimethod matrices	Latent modelling approach often applied with CFA to estimate method effects
RMSEA	Root-mean-square Error of Approximation	Non-centrality based fit index used to determine how well the estimated model fits the data in Factor Analysis
SEM	Structural Equation Modelling	A wide range of modelling methodologies designed to test hypotheses between observed and latent variables
SRMS	Standardised Root Mean Square Residual	Absolute fit index used to determine how well the estimated model fits the data in Factor Analysis

Introduction

Psychologists have been seeking ways to indirectly infer thought processes since Freud first proposed that psychological processes could occur outside of conscious awareness (Freud, 1915 as cited in Riolo, 2010). The development of modern implicit methods of attitude assessment are claimed to provide a solution to this search by enabling deeply ingrained attitudinal biases to be inferred from differences in timed categorisation tasks. The most prominent of these implicit attitude measures are the Implicit Association Test (IAT; Greenwald, et al., 1998) and the Affective Priming Task (APT; Fazio, et al., 1986).

Implicit attitude measures have been shown to possess greater predictive validity for the assessment of socially contentious constructs than traditionally used explicit attitude measures, such as questionnaires (Greenwald, Poehlman, Uhlmann, & Banaji, 2009). Such findings have encouraged a sizeable movement towards implicit attitude research, whereby APTs and IATs have been employed in thousands of empirical studies over the last two decades[1]. Many of these studies promote implicit attitude measures as a useful means to glean insights into the unconscious biases that influence people's behaviour. The IAT is frequently used to provide individualised assessment of personal implicit biases (see the Project Implicit website; Greenwald, Banaji, & Nosek, 2011) and some have even suggested the task be used as an employment screening tool to reduce the potential for discrimination within the workplace (Ayers, 2001, pp. 424-425). However, before a task can be used in this way it is first vital to establish the validity of such measurement instruments.

[1] As evident by a search of the ProQuest database.

Inconsistencies in the implicit attitudinal literature, in conjunction with poor and highly variable psychometric support for the APT and IAT, have resulted in concerns that these tasks are heavily influenced by measurement error. Error variance can reduce the reliability of measurement instruments and provide an upper limit for construct validity estimates (Cunningham, Preacher, & Banaji, 2001). The aim of this thesis was to quantify the extent of error variance in implicit attitudinal scores, and to assess the reliability and validity of APTs and IATs using Confirmatory Factor Analysis (CFA) models. CFA uses latent variable models to separate error variance from trait variance (Byrne, 1998) and its use is novel for implicit attitudinal research.

Thesis Aims and Organisation

The overall aim of this thesis was to examine the construct validity of the APT and IAT using CFA techniques. This thesis is organised into eight chapters: four introductory chapters, three empirical chapters and a general discussion.

Introductory Chapters

Chapter One provides a conceptual and theoretical introduction to implicit and explicit attitudinal processing. It introduces the APT and IAT techniques and delivers a procedural description of these measures. The potential advantages of implicit attitude measurement over explicit attitude measurement for the assessment of socially contentious constructs are introduced.

Chapter Two demonstrates the potential utility of implicit attitude measures, by providing a critical review of APT and IAT findings across three contentious research domains, that of racial prejudice, substance use and body image. These constructs benefit from assessment by implicit attitude measures, because explicit attitude measures are susceptible to issues of self-presentation bias and manipulated responding. However, inconsistencies are noted amongst the implicit research that lead to concern for the robustness of implicit attitudinal scores.

Chapter Three discusses potential sources of error variance for implicit attitude measures, and the impacts such error would have on psychometric validation of these tasks. Error variance is explained within the framework of classical test theory. It is argued that traditional analytical approaches have been inadequate for assessing implicit attitudinal data. Rather, an approach that adequately addresses the issue of error variance is required, namely Structural Equation Modelling (SEM).

Chapter Four provides a detailed introduction to SEM analytical techniques, such as Confirmatory Factor Analysis (CFA). The common factor model is explained along with the process of model specification. A detailed description of seven applications for SEM in the assessment of the reliability and construct validity of measurement instruments is delivered. This includes an introduction to Single-group CFA, Composite Reliability (CR), Average Variance Extracted (AVE), Higher-order CFA, Multitrait-Multimethod CFA (CFA-MTMM), Multiple Indicators Multiple Causes (MIMIC) models, and Multiple-groups CFA. It is shown that SEM and CFA appear well suited to assess the psychometric properties of implicit attitude measures.

Empirical Chapters

There are three empirical studies of this thesis which are presented in Chapters Five, Six and Seven. These three chapters investigate the seven aims of this study, which relate to specific assessments of the reliability and construct validity of the APT and IAT. The empirical studies employed the traditional verbal APT and verbal IAT (VIAT) that are reliant on word stimuli, as well as a pictorial type of IAT (PIAT) that only uses pictorial stimuli.

Chapter Five outlines Study One, which covers the first four aims of this research. In Study One, single-group CFA is the primary analytical approach applied in order to separate random error variance from trait variance for the implicit attitude scores. Using this approach, the reliability of the APTs and IATs are investigated using CR and AVE (*Aim 1*). The internal construct validity of these tasks is assessed using one-factor single-group CFA (*Aim 2*). Convergent and discriminant validity for the VIATs and PIATs are then evaluated using three-factor single-group CFA (*Aim 3*), as well as higher-order CFA (*Aim 4*). Together, Study One provides a thorough examination of the influence of random error variance on the implicit attitudinal scores, and delivers an estimate of the robustness and validity of the APT and IAT.

Chapter Six outlines Study Two, which expands upon the findings of Study One by separating systematic sources of error variance, such as method effects, from the estimate of trait variance produced by single-group CFA. This is possible by using the CFA-MTMM data analytic approach that separates observed scores into trait, method and error components. CFA-MTMM thus enables a strict assessment of construct validity, which is applied to the four empirical IATs (*Aim 5*).

Chapter Seven outlines Study Three, which explores the potential use of SEM approaches to facilitate substantive enquiries for IATs. This study investigates whether it is possible to ascertain an IAT effect, indicating implicit racially-related biases, by assessing the equivalency of congruent and incongruent responses using Multiple-groups CFA (*Aim 6*). The influence of certain participant characteristics, namely sex, age and travel experience, on the IAT effect scores is explored using MIMIC modelling (*Aim 7*). Study Three demonstrates the use of latent modelling techniques to facilitate applied research using implicit attitude measures.

General Discussion

Lastly, Chapter Eight provides a general discussion regarding the implications of this research. The findings of the three studies are summarised and the overall repercussions of the results are explored. This chapter covers the limitations of this thesis and provides suggestions for future research. Chapter Eight ends with the overall conclusions of this dissertation.

CHAPTER ONE

Implicit Attitudes: Conceptual and Procedural Issues

Attitudes are favourable or unfavourable dispositions towards social objects such as people, places and policies (Stanley, Phelps, & Banaji, 2008). Attitudes enable efficient responses to simple sensory information, but also inform the more complex responses required when engaging with other individuals, groups, objects or events encountered throughout life (Stanley et al., 2008). Multiple underlying processes guide attitude development. The two main processes to attitude development are firstly a reflective and intentional explicit process, and secondly, an impulsive and automatic implicit process (Haeffel et al., 2007). Explicit processes differ from implicit processes by requiring conscious awareness and cognitive effort, whereas implicit processes occur spontaneously without any need for conscious effort or awareness. These distinct processes are well documented and can be evidenced from both behavioural data and psychophysiological data using MRI and EEG scanning equipment, with separate brain regions implicated for implicit and explicit attitude processing (Amodio, 2013; Cunningham, Johnson, Gattenby, Gore, & Banaji, 2003; Cunningham, Johnson, et al., 2004; Phelps et al., 2000; Stanley et al., 2008).

This introductory chapter begins with an overview of explicit attitude measures and their limitations, emphasising the need for implicit attitude assessment techniques to overcome these limitations. The development of implicit and explicit attitudes is then explored using the dual-process theoretical framework to conceptualise the formation and relationship between these cognitive processes. The foremost measures developed to access these implicit processes will then be introduced, namely the Affective Priming Task (APT; Fazio, Sanbonmatsu, Powell, & Kardes, 1986) and the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998). A procedural description of these implicit attitude measures will clarify how each task operates and how an estimate of implicit attitudes is devised using these techniques. Implicitly measured attitudes will briefly be reviewed in relation to explicitly measured attitudes. It will be shown that there can be substantial divergence in findings between implicit and explicit methods of attitude assessment, particularly when assessing socially sensitive constructs such as racial prejudice. In these instances, implicit attitude measures have exhibited greater predictive power than explicit measures on a wide range of behavioural, judgement and physiological outcomes. This chapter concludes that the key advantage of implicit attitude measurement techniques is in the detection of attitudes people may be unwilling to openly share.

Explicit Attitude Measures and Their Limitations

Explicit attitudes refer to consciously held evaluations of people, places and constructs (Nosek et al., 2006). Explicit views are knowingly endorsed and may be offered as opinions following a process of consideration and personal reflection. The field of attitudinal measurement has traditionally depended on explicit attitude measures such as self-report questionnaires to ascertain these considered opinions (Harrison, Mclaughlin, & Coalter, 1996). Explicit attitude measures require participants to express their views towards attitude objects, often through the use of Likert or Semantic Differential scales (Dawis, 1987). Likert scales refer to a commonly utilised five-point bipolar response format, though the scales may use greater or fewer points. Participants rate level of agreement with each statement on a scale that ranges, for instance, from Strongly Disagree through to Strongly Agree (see Figure 1.1a). The Semantic Differential technique differs in that it is used to assess feelings along a continuum. For instance, participants may mark on a continuum ranging from Bad to Good how they feel about a particular question (see Figure 1.1b). While these explicit attitude measures are invaluable tools which have been used extensively for over 80 years, they suffer from several limitations (Perugini & Banse, 2007).

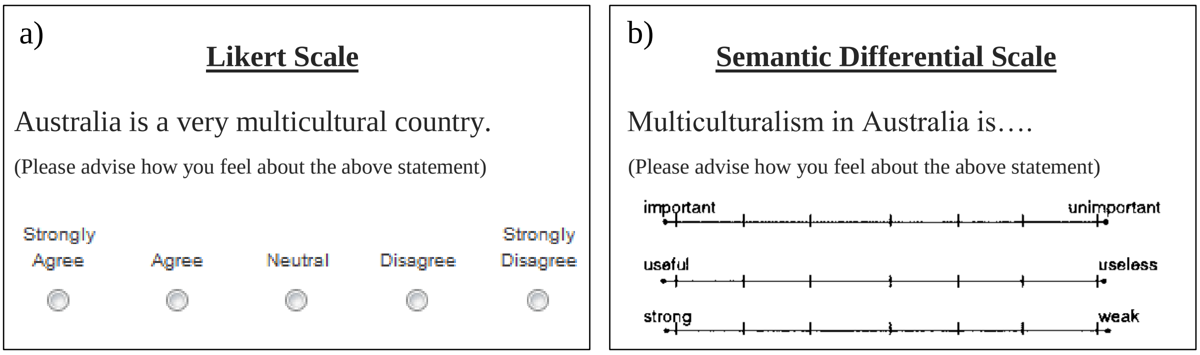

Figure 1.1. Example Likert and Semantic Differential scales for responding to explicit attitudinal questionnaires.

The primary limitation of explicit attitude measures is that there is an inherent assumption that participants have the ability and motivation to report attitudes and beliefs truthfully. However such an assumption is not always correct (Greenwald & Banaji, 1995). Self-presentation strategies, whereby a participant's true attitudes towards a topic are distorted to present more socially acceptable responses, are a commonly reported issue for explicit attitude measures. These self-presentation strategies are particularly prevalent when investigating contentious or sensitive constructs such as racial prejudice (Greenwald et al., 2002; Hofmann & Schmitt, 2008; Nosek et al., 2007; Schnabel, Asendorf, & Greenwald, 2008a; Spence, 2005). In order to overcome this potential bias, many attitudinal and personality scales contain inbuilt social desirability measures, often termed 'lie scales', which aim to assess the veracity of responses. Yet even with lie scales and other such validity checks, these instruments do not provide a way of determining a respondent's true feelings or attitudes. Rather lie scales merely alert the researcher to the possibility that the participant may not have responded with an acceptable level of bias.

Response veracity is, however, not the only issue for explicit attitude measures. There is documented evidence that people generally have very limited introspective access to the psychological processes that guide their behaviour (Nisbett & Wilson, 1977). For instance, people are often unable to explain accurately why they acted or spoke in a particular fashion, relying instead on external cues to infer a reason, such as 'I have finished all the food in my bowl so I must have been hungry' (Nisbett & Wilson, 1977). Clearly, this poses an even greater problem than self-presentational distortions for the use of self-report measures in psychological research (Gawronski, 2009) as participants cannot accurately present attitudes for which they are unaware. These issues have led psychologists to search for alternative means to assess people's 'inner minds'. In a shift from traditional explicit attitude measures, implicit attitude measures infer underlying attitudes from participants' performance on timed categorisation tasks (Gawronski, 2009). This reliance on quick reaction-times minimises issues of self-presentation distortions and limited self-awareness by reducing the participant's mental control over their responses (Stanley et al., 2008). These indirect attitude measures provide a novel approach to accessing underlying, automatic attitudes that the individual may not even be aware he or she possesses.

Conceptual Overview of Implicit Attitudes

The unintended impact of expectation on an individual's perceptions, judgements, memory and behaviour was established in some of the very first empirically-based psychological experiments (e.g. Ebbinghaus, 1885; Stroop, 1935). Since those seminal studies, research has continued to reveal that attitudes, such as stereotyping and prejudice, can also occur automatically or outside of conscious awareness (Cunningham et al., 2001; Greenwald et al., 2009; Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005; Jost et al., 2009).

Automatic attitudes are activated spontaneously like a reflex (Lane, Banaji, Nosek, & Greenwald, 2007). Automaticity is likely due to well organised semantic links within memory stores that have been developed and reinforced through personal experiences (Neely, 1977). These automatic operations are defined as implicit processes, which differ fundamentally from explicit processes that are controllable, made with awareness and require intent and cognitive resources (Nosek, 2007). As such, explicit processes require cognitive effort, whereas implicit processes occur impulsively without exertion or awareness and the attitude bearer may be completely unaware of their presence. Theories of implicit bias assert that automatic processes impact people's decisions and actions, a view which is divergent from the common perception that people are guided solely by their explicit beliefs and conscious intentions to act (Greenwald & Krieger, 2006).

In accordance with the concept of separate implicit and explicit processing routes, attitudes also consist of two related but theoretically distinct types: an explicit and an implicit type (Cunningham et al., 2001; Greenwald & Farnham, 2000; Lane et al., 2007; Nosek, 2007; Nosek & Smyth, 2007; Ranganath, Smith, & Nosek, 2008). Explicit attitudes are consciously endorsed and evident when individuals have sufficient time to provide considered responses, such as when completing a survey or questionnaire. Explicit attitudes differ from implicit attitudes because the individual is aware of the evaluative process and is thus able to assign mental effort to the development of an attitudinal appraisal. Implicit attitudes, conversely, occur spontaneously, are generally triggered by cues within the environment, and happen without any contemplation by the attitude bearer (Fazio et al., 1986). An implicit attitude has been described as “an introspectively unidentified trace of past experiences which mediates favourable or unfavourable thought, feeling or action toward an object” (Greenwald & Banaji, 1995, p. 5). In other words, aspects of past experiences automatically influence how an individual evaluates their present world. The most useful theoretical framework to conceptualise these attitude types are the dual-process theories of cognitive functioning.

Theoretical Models of Implicit Attitudinal Processes

The most prominent theories of cognitive functioning advocate a dual-process understanding of attitudes (Bargh, Chaiken, Govender, & Pratto, 1992; Fazio et al., 1986; Wilson, Lindsey, & Schooler, 2000). These theories all share the assumption that information may be processed in two ways: either automatically using simple, low-effort, readily accessible decision rules; or alternatively through a conscious, active type of process, involving effortful scrutiny of relevant information (Ranganath et al., 2008). It is assumed from application of these theories that the representation and generation of implicit attitudes occurs similarly (Bargh et al., 1992; Fazio et al., 1986; Gawronski & Bodenhausen, 2006; Wilson et al., 2000). As such, on perception of an object automatic activation of semantically connected neurons occurs in the brain (Neely, 1977). This activation pattern results in an implicit attitude that can influence further perceptions of the attitude object, the situation in which it was encountered and subsequent behaviour (Bargh et al., 1992; Fazio et al., 1986). Dual-process theories are thereby entirely consistent with the theoretical understanding of automatic implicit attitudes and more effortful explicit evaluations (Steinman, 2011).

Of the dual-process theories, the Dual-Attitude model (Wilson et al., 2000), the Motivation and Opportunity as Determinates model (MODE; Fazio et al., 1986) and the Associative-Propositional Evaluation model (APE; Gawronski & Bodenhausen, 2006) provide the most useful frameworks for understanding implicit and explicit attitudes. According to the Dual-Attitude model, people can simultaneously possess two different evaluations of the same attitudinal object: an automatically activated implicit attitude as well as a deliberate explicit evaluation that requires effortful processing to be retrieved from memory (Wilson et al., 2000). Using this approach, it is expected that people access their explicit evaluations only when they have the capacity and motivation to do so. As a result, implicit evaluation becomes the default attitude form. This theory is very similar to the MODE model which states that people will be guided by their explicit attitudes only when they have the motivation and the opportunity (such as the time or cognitive resources) to consider their views (Fazio et al., 1986). If either of these prerequisites is missing, then relatively spontaneous processing of the attitudinal object occurs, resulting in judgements and behaviour that are guided by implicit evaluations (Fazio et al., 1986).

The Associative Processing Evaluation model (APE; Gawronski & Bodenhausen, 2006) builds on these earlier dual process theories by examining the underlying processes thought to impact implicit and explicit attitudes. According to this approach, an event/object activates a pattern of stored associations in the memory that produces an automatic affective response. This associative process provides the initial source of evaluation and an implicit attitude is the result (see Figure 1.2). The particular associative pattern activated in the brain (highlighted by the bold connections in Figure 1.2) differs depending on the context in which the object was encountered. To transform an implicit attitude into an explicit evaluation, propositional reasoning is applied. For instance, a negative affective reaction, such as distaste, may be transformed within the reflective system into the reasoned proposition that "I dislike X" (Strack & Deutsch, 2004). This propositional reasoning process is a form of conscious appraisal that results in the development of an explicit attitude. During the propositional reasoning phase, the emerging attitude can be checked to ensure its consistency with already held values and with any other relevant information[2]. Following the propositional reasoning review process, the explicit attitude is formed.

[2] Theories such as cognitive consistency (Festinger, 1957) are thus solely the domain of propositional processing.

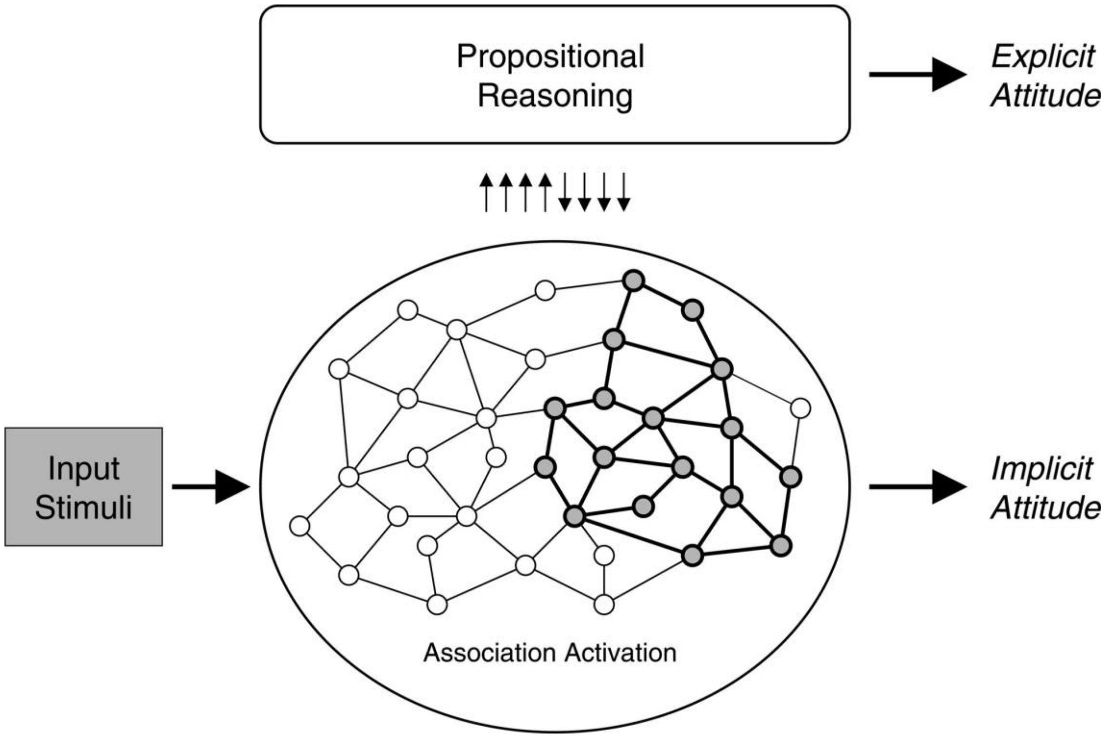

Figure 1.2. Schematic representation of the APE cognitive model depicting the interplay of associative activation and propositional reasoning in implicit and explicit attitude development (Gawronski & Bodenhausen, 2006, p. 697).

Dual-process theories provide a strong foundation for understanding the relationships between implicit and explicit attitudes. Yet the majority of implicit attitudinal research has been relatively atheoretical (See Fazio & Olson, 2003; Greenwald, 2004; Greenwald, Nosek, Banaji, & Klauer, 2005; Steinman, 2011, for further discussion). This is problematic, because even in exploratory research a theoretical framework aids decisions, including those regarding which relationships to examine further or which statistical analyses to apply. Ideally no researcher can be totally atheoretical, as preconceived notions or expectations for the research will still guide his or her behaviour and thus the research outcomes (as implicit attitude researchers should be explicitly aware). Yet Greenwald in particular has strongly questioned the popular notion that research is only valuable to the extent that it advances theory, quoting Feyerabend (1975) who stated that “any attempts to specify bounds of scientific method would be misguided… (as they) would inevitably exclude methods that are valuable in the accumulation of scientific knowledge” (e.g. Greenwald, 2004, p. 275). Proving this point, findings from purely methods-based or atheoretical empirical studies (i.e. Greenwald, et al., 1998) have consequently been incorporated into theories, such as the APE model, to further facilitate understanding of implicit attitudinal processes. The next section presents an introduction to the implicit measurement techniques that spearheaded the investigation into implicit processes and helped guide theoretical conceptualisations of these functions.

Implicit Attitude Measurement Techniques

The first insight that sparked the development of modern implicit attitude measures came from Franciscus Donders during the middle of the 20th century. Donders developed an experiment in which participants were given a small electrical shock through an electrode placed on each foot. Participants indicated which foot had received the shock by pushing a button using their hand. Donders discovered participants reacted faster when the hand responded to stimulation of the foot of the same side (i.e. right hand and right foot) compared with when the opposite foot was shocked (i.e. right hand but left foot) (Donders, 1969). Based on this finding, Donders proposed that the difference in response times for the two tasks provided an indicator of relative difficulty. The insight that underlying mental processes can be informed through differences in reaction times is the fundamental concept behind many implicit measurement techniques.

Implicit attitudinal measures specifically began to be developed in the 1980s to overcome the aforementioned limitations of explicit attitude measures and to explore the possibilities of subconscious influences on attitudes. Use of these implicit attitude measures enabled researchers to assess concepts without the participants' awareness of the target of measurement (Greenwald & Banaji, 1995). This in turn prompted a rapid expansion of interest in implicit influences on social perceptions, judgements and action (Nosek, 2007). While the distinction between implicit and explicit attitudes was theoretically driven, in practice the difference is a methodological one (Stanley et al., 2008). Explicit techniques allow time for self-reflection and the provision of a considered response, whereas implicit techniques limit the opportunity for introspection or deliberation by relying on speeded categorisation tasks (Nosek, Greenwald, & Banaji, 2006). This emphasis on fast reaction times is the primary differentiating factor between implicit and explicit attitude techniques. Implicit attitude measures typically infer underlying mental processes from differences in response latencies for the categorisation of associated stimuli (Krause, Back, Egloff, & Schmukle, 2010). Whilst there are many different variations of implicit attitude measures currently available (see De Houwer & De Bruycker, 2007; Nosek & Banaji, 2001; Steinman & Karpinski, 2008), the two most widely used techniques for measuring implicit attitudes are the Affective Priming Task (APT; Fazio et al., 1986) and the Implicit Association Test (IAT; Greenwald et al., 1998).

The Affective Priming Task (APT; Fazio et al., 1986)

The Affective Priming Task (APT; Fazio et al., 1986) was the first of the reaction-time-based tasks devised to measure implicit prejudice and has played a pivotal role in stimulating research on the assessment of implicit attitudes and stereotypes (Bargh, Chaiken, Raymond, & Hymes, 1996). The APT was designed by Fazio et al. (1986) who argued that a person's strongest and most commonly accessed attitudes were capable of being automatically activated by the mere presence of an attitude object. Further, this activation was argued to occur regardless of whether an individual was aware of their opinions towards the attitude object or not and, once activated, these automatic attitudes could influence the person's behaviour. The APT and other such priming procedures were created to measure the accessibility and strength of these automatic attitudes.

The theoretical underpinning of the priming task is that the prime (word or image) triggers an evaluative response that is either congruent or incongruent with the response required by the target (Wittenbrink, 2007). As a result, on trials where the prime and target share the same valence (i.e. they are both deemed positive or both deemed negative) the participant will take less time to make the response already triggered by the prime. When the prime and target differ in valence the response implied by the prime interferes with the target response, producing a slower response execution (Wittenbrink, 2007). The degree to which the prime facilitates a target response serves as an indicator of the strength of association between the attitude object and the target (Bargh et al., 1996).

The Implicit Association Test (IAT; Greenwald et al., 1998)

Fazio et al.'s (1986) APT paved the way for the development of the Implicit Association Test (IAT; Greenwald et al., 1998); now the most widely used method for measuring implicit attitudes or biases (Jost et al., 2009). Following a similar logic to the APT, the IAT provides an estimate of the strength of association between concepts and attributes (Lane et al., 2007). However, rather than pairing each category exemplar individually with an attribute exemplar, as is the case for the APT, the IAT uses a dual-categorisation format, whereby participants categorise attribute and category stimuli simultaneously. This enables the IAT to provide an assessment of entrenched attitudes and stereotypes in half as many trials as that required by the APT (Greenwald et al., 1998). At the time of the IAT's inception, there was some evidence of implicit social cognition, but the tools available lacked the sensitivity to detect variability among populations. Greenwald et al. (1998) hoped developing the IAT would enable valid and reliable assessment of implicit attitudes and stereotypes.

The IAT has proved popular and has been employed in well over 900 studies, far more than any other reaction-time based implicit attitude technique (Rudman & Ashmore, 2007). The IAT has been used with both clinical and community adult populations, and applied to assess many and varied constructs, including, but not limited to, political views (e.g. Greenwald et al., 1998), vegetarianism (e.g. Swanson, Rudman, & Greenwald, 2001), alcohol and other drug use (e.g. Wiers, Woerden, Smulders, & de Jong, 2002), religious differences (e.g. Rowatt, Franklin, & Cotton, 2005), racial stereotypes (e.g. Cunningham, Nezlek, & Banaji, 2004) and clinical disorders (e.g. Gschwendner, Hofmann, & Schmitt, 2008). Yet the IAT was restricted to a literate adult subject population due to its reliance on word stimuli.

Such limitation excluded numerous populations including very young children and the illiterate, which reduced the generalisability of implicit attitudinal findings. The development of a more inclusive implicit attitude measure creates greater possibility for examining attitudinal formation and development over the lifespan, which has the potential to greatly aid theoretical conceptualisation of implicit attitude development.

The Pictorial Implicit Association Test (PIAT; Thomas, et al., 2007)

Thomas, et al. (2007) overcame this previous verbal restriction by introducing a fully pictorial version of the IAT (the PIAT), which only uses pictorial stimuli for categorisation. Pictorial stimuli require less effortful mental processing than word stimuli due to stronger semantic links between pictures and the meaning they represent (Carr, McCauly, Sperber, & Parmelee, 1982; Glaser & Glaser, 1989). Because of this, the use of pictorial stimuli in the PIAT enabled significantly more efficient categorisation of the stimuli (Thomas, 2008), resulting in faster, more automatic responses. This greater automaticity, a key indicator of implicit processing, is potentially indicative of a purer measure of implicit attitudes. Furthermore, by removing the requirement of verbal fluency demanded by the traditional verbal technique, use of the PIAT opens up the possibility for assessing implicit attitudes in previously untestable populations such as younger children. It also has potential for use in cross-cultural studies as it avoids issues such as translation, salience of terms and illiteracy.

A detailed description of the procedure for each of these implicit attitudinal measures is presented in the following section. This provides a stronger conceptualisation of how each of these measures functions to deliver an estimate of implicit attitudes.

A Procedural Description of the Implicit Measurement Techniques

Implicit attitudinal measures typically require speeded categorisation of affective stimuli that are presented on a computer screen. Participants are required to respond using one of only two designated response keys on a computer keyboard. The average length of time taken to complete this action is interpreted as an indicator of the ease with which the task was executed. Although there are many implicit attitudinal tasks now available, the priming technique (APT) and the Implicit Association Test (IAT) remain the most frequently used (Lane et al., 2007). The following explanation will illustrate how implicit attitudes are inferred from each of these reaction-based techniques in turn.

The Affective Priming Task

The affective (or evaluative) priming task (APT; Fazio et al., 1986) was adapted from Neely's (1977) sequential priming paradigm and was designed to assess whether attitudes could be activated automatically. The APT is a categorisation task that requires participants quickly judge whether a word is Positive or Negative. The presentation sequence for a classic APT is as follows: a fixation point is initially displayed on-screen to focus the participants' attention, the prime word then appears briefly, followed by the target (affective) word stimulus that remains on-screen until the participant categorises it as a Positive or Negative word using two specified keys (see Figure 1.3).

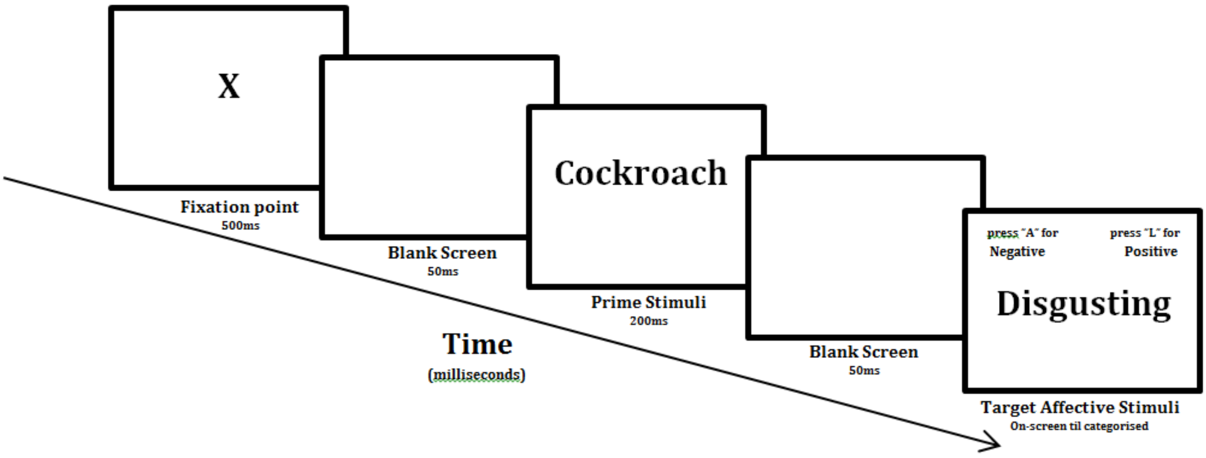

Figure 1.3. Exemplar presentation sequence for a standard APT.

The presentation of a prime is assumed to facilitate or hinder the speedy categorisation of the affective stimuli. Traditionally, the APT employs two contrasting categories of prime stimuli, for instance Insects and Flowers. Exemplars from both these categories precede the Positive and Negative word stimuli during each of the hundreds of trials required by a standard APT. Average reaction times are calculated for each of the four possible experimental conditions to develop an overall indication of implicit attitudes towards these attitude objects, which in this case are Flowers and Insects. For instance, in a particular trial a participant may witness momentarily the word “cockroach” before being asked to categorise the word “disgusting” as either Positive or Negative as quickly as possible. The classic finding is that participants make the required evaluative decisions faster when prime and target are viewed as of the same valence (both perceived to be negative, for example) than when the prime and target are of conflicting valences (Wittenbrink, 2007). The underlying assumption of the APT is that the faster the response, the stronger and thus more accessible and automatic the attitude (Bargh et al., 1996).

To ascertain an estimate of implicit evaluation, the average time in milliseconds between stimuli presentation and response is calculated for each of the four possible prime/target pairings (Insect+Positive, Insect+Negative, Flower+Positive and Flower +Negative). These averages are then compared using the following formula:

$$\text{Evaluation} = (P_Y\text{-}P_X) - (N_Y\text{-}N_X) \qquad (1.1)$$

This index captures the degree to which attitude X relative to Y yields faster responses for Positive (P) than Negative (N) targets (Wittenbrink, 2007). Thus for this example, if X related to the Flowers category and Y to the Insects category, higher scores calculated by the above formula would ostensibly indicate that the Flowers elicited more positive implicit evaluations than the Insects did. This finding is typically reported as an implicit relative preference for Flowers over Insects.

The Implicit Association Test

Similar to the APT, the IAT infers implicit evaluations on the basis of differences in reaction times for varying selections of stimuli. The principal assumption underlying the IAT is that if two concepts are highly associated, dual categorisation tasks will be easier when those concepts share the same response than when they require different responses (Greenwald et al., 1998). In other words, participants will perform more quickly when they are able to rely on well-practised associations between objects and attributes (Rudman & Ashmore, 2007). In a similar fashion to the APT, the IAT also provides a measure of *relative* implicit bias. IATs require the constructs of interest to have logical counter-constructs, like Males versus Females or Black people versus

White people. Generally, these concepts include one pair of “category” constructs (such as Flowers versus Insects) and one pair of “attribute” constructs (such as Pleasant versus Unpleasant). To assess the relative strength of associations between these pairs of concepts the IAT relies on a dual-categorisation format, which enables the IAT to provide a measure of implicit attitudes using half the number of trials required by the APT. This efficiency provides a considerable saving in terms of time and effort for the participant.

During an IAT participants are asked to rapidly classify individual stimuli representative of either a category or an attribute into one of four distinct groupings using only two responses. Participants do not deliberate about their feelings as they would if using explicit attitudinal measures, rather they categorise the items as quickly as possible (Lane et al., 2007; Nosek, 2007). By way of example, in an IAT the word “daisy” may appear on the screen and the participant’s objective is to quickly identify that stimulus as a Flower rather than an Insect by pressing a particular key on a computer keyboard. The underlying principle is that the more two concepts are congruent with a participant’s attitudes, such as Flowers are associated with Pleasant, the faster the participant will respond when these two stimuli share the same response than when incongruent stimuli, such as Flowers and Unpleasant, require a common response (Hummert, Garstka, O’Brien, Greenwald, & Mellott, 2002). In a standard IAT it is expected that participants will respond faster to the congruent stimuli pairings than the incongruent pairings (see Figure 1.4 for an example of congruent and incongruent stimuli pairs). When this expected outcome occurs it is referred to as the IAT effect.

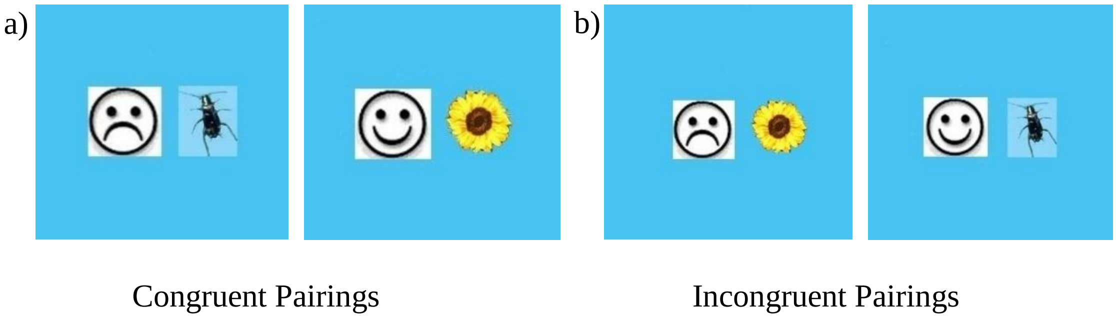

Figure 1.4. Exemplar congruent and incongruent stimuli pairings (for the Flower-Insect Pictorial IAT).

Standard IAT Procedure

A detailed description of the basic IAT procedure expands on the above example. In its original form, the IAT involves a pair of target concepts, such as Flowers and Insects, and a pair of attribute concepts, such as Pleasant and Unpleasant. There are four main components that comprise each IAT: Attribute-related stimuli categorisation, Category-related stimuli categorisation, Congruent stimuli categorisation and Incongruent stimuli categorisation. These categorisation requirements, detailed below, comprise the seven standard steps of an IAT. Steps Three, Four, Six and Seven are the empirical steps that supply the data for analysis (see Table 1.1 for an overview).

Table 1.1

The Seven Procedural Steps of a Typical Implicit Association Test

Step 1	Learn Attribute Dimension Unpleasant vs. Pleasant words
Step 2	Learn Category Dimension Insect vs. Flower words
Step 3	Congruent Dual-Categorisation Task * Unpleasant and Insect words vs. Pleasant and Flower words
Step 4	Congruent Dual-Categorisation Task * Unpleasant and Insect words vs. Pleasant and Flower words
Step 5	Learn Transposed Category Responses Flower vs. Insect words
Step 6	Incongruent Dual-Categorisation Task * Unpleasant and Flower words vs. Pleasant and Insect words
Step 7	Incongruent Dual-Categorisation Task * Unpleasant and Flower words vs. Pleasant and Insect words

* Data from these steps are used for data analytic procedures.

Attribute-related Stimuli Categorisation.

Attribute-related stimuli categorisation involves simply categorising the Pleasant and Unpleasant Attribute stimuli. This is Step One of the IAT, which facilitates learning of the attribute stimuli exemplars. Participants are asked to rapidly classify items representing two poles of an attribute dimension into their superordinate groups. For example, words such as “happy”, “freedom” and “peace” are classified as Pleasant, and words such as “filth”, “hatred” and “tragedy” are categorised as Unpleasant. Traditionally, the words are presented sequentially in the middle of a computer screen and respondents classify the words using two designated keys on a computer keyboard. Figure 1.5 presents an example stimuli presentation sequence for the attribute component involving two stimuli trials on a traditional verbal Implicit Association Test (VIAT).

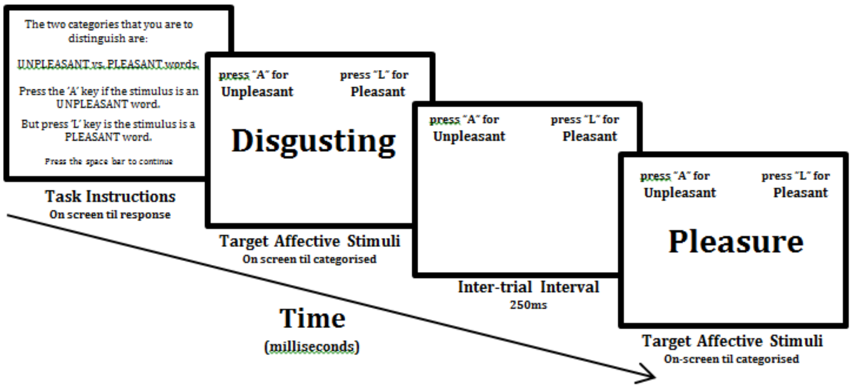

Figure 1.5. Example presentation sequence for the attribute component of a VIAT.

Category-related Stimuli Categorisation.

Step Two involves learning the category dimension and requires the similarly simple task of only categorising the category-related stimuli, such as Flower and Insect stimuli. Respondents categorise the stimuli representing Flowers and Insects using the same two keys employed for the attribute-related categorisation task.

Congruent Stimuli Categorisation.

Congruent stimuli categorisation involves dual-categorisation of attribute- and category-related stimuli simultaneously, such that the previous two tasks are combined. Step Three requires that participants sort four sets of stimuli at the same time using only two keys. Congruent stimuli categorisation requires typically more associated category and attribute stimuli be responded to using the same key. For instance, participants may be required to respond to Pleasant or Flower words using the right-hand key and to Unpleasant or Insect words using the left-hand key (see Figure 1.4a). Such stimuli pairings are referred to as congruent because it is expected that most participants will intuitively link flowers with pleasant associations and insects with unpleasant associations. Steps Three and Four are identical congruent stimuli categorisation tasks. This repetition allows for double the number of trials to be completed, whilst enabling participants a short reprieve in concentration between blocks in order to reduce fatigue.

Category-related Stimuli Categorisation.

Step Five involves learning to switch the spatial location of the categories in that the task from Step Two is repeated but in a transposed format. For example, if the appropriate response for Flower stimuli in Step Two had been the left hand key, respondents would now be required to use the right hand key, with the left hand key used to categorise the Insect stimuli. This step prepares participants for the new categorisation task that follows in Step Six.

Incongruent Stimuli Categorisation.

Incongruent stimuli categorisation is the second category-attribute pairing combination that again requires participants sort stimuli from both category and attribute groups. Step Six involves simultaneous sorting of four sets of stimuli with incongruent combinations, whereby the pairings are reversed from Steps Three and Four. As such, participants would now respond to Unpleasant and Flower words using the left hand key (in this example), and to Pleasant or Insect words using the right hand key (see Figure 1.4b). Step Seven is identical to Step Six. The pairings presented in the last two steps of this example are referred to as incongruent because for most people the stimuli are less intuitively grouped in this pairing than when in the congruent grouping.

The presentation of congruent pairings in Steps Three and Four, followed by incongruent pairings for Steps Six and Seven is switched for every second participant, so that order effects are minimised for the sample. The difference in mean response times for the two block types (the original or congruent pairings versus the reversed or incongruent pairings) provides an indication as to the direction and extent of any evaluative associations attached to the target words. This difference is calculated using the means and standard deviations of the response times in Steps Three, Four, Six and Seven (see Table 1.1), by following the scoring procedures outlined by Greenwald et al. (2003).

The IAT Effect Score (D; Greenwald, Nosek, & Banaji, 2003).

The IAT's scoring formula is, in effect, a variant of the standardised mean difference effect size measure Cohen's *d* (Cohen, 1992), and it results in an individual IAT effect score for each participant referred to as a *D* score (Greenwald et al., 2003). The IAT effect score is the key outcome of the IAT and can be interpreted using the guidelines provided by Greenwald et al. (2003). They advise that an IAT effect score greater than .60 implies strong negative implicit prejudice (i.e. strong preference for the congruent over the incongruent combination), an effect score between .35-.60 demonstrates moderate negative prejudice, .15-.35 implies slight prejudice and scores lower than .15 indicate non-existent prejudice (Greenwald et al., 2003).

To summarise, an IAT effect is evident when participants respond significantly faster to the congruent than the incongruent sets of stimuli. For the above example, if an IAT effect had occurred it would have indicated an implicit relative preference for Flowers over Insects. The IAT effect score is a measure of the size of this bias and provides a guideline as to the extent of the implicit preference.

The Pictorial Implicit Association Test

The procedure outlined above for the original verbal IAT (or VIAT) is the format for all IATs, though there is some variation in the delivery of stimuli. Although traditionally restricted to word stimuli, the introduction of the Pictorial Implicit Association Test (PIAT; Thomas et al., 2007) expanded the repertoire of stimuli usable with the IAT to include all forms of pictures (for example, see the pictorial category exemplars depicted in Figure 1.4). The methodology is exactly the same for the PIAT as the VIAT, but instead of categorising words the participants categorise pictorial representations of the attitude objects. For instance, Positive and Negative smiling icons (“smileys” or “emoticons”) replace the Positive and Negative affective stimuli, and pictures of Flowers and Insects replace the traditionally employed verbal stimuli. To provide further examples, pictures of people from different ethnic groups could replace the stereotypical name stimuli often used to ascertain implicit racial prejudices. Additionally, images of popularly branded food items could be used to assess implicit attitudes towards certain companies, or pictures of politicians’ faces might enable an estimate of implicit political preferences. Both the PIAT and the VIAT enable the assessment of countless different attitudinal constructs, but now there is also the flexibility of various visual formats.

Originally the PIAT was developed for use with very young pre-school children who were unable to read. This initial PIAT pilot group completed the task using a touch screen, which facilitated the developmental stage of the young participants (Thomas et al., 2007). In this case, the participants pressed the touchscreen on the right or left hand side to indicate their responses rather than using a standard computer keyboard.

This is not a requirement of the PIAT, but is an additional option which can be utilised depending on the capabilities of the participant group. Otherwise, all aspects of the VIAT and the PIAT have been designed to be equivalent.

Summary

Implicit attitude measures aim to infer underlying mental processes from differences in reaction times on dual-categorisation tasks. This emphasis on quick response times differs substantially from explicit methodology that allows time for reflection and consideration of the endorsed evaluations. These distinct procedural and theoretical processes, outlined in the present section, have resulted in discrepancies between an individual's implicitly and explicitly measured attitudes, especially when exploring socially contentious issues.

Advantages of Implicit over Explicit Attitude Measures in the Detection of Socially Unacceptable Attitudes

Explicitly and implicitly measured attitudes towards the same attitude object were initially expected to be reasonably similar (Greenwald et al., 1998). Yet meta-analyses that examined the strength of relationship between implicitly and explicitly measured evaluations have indicated substantial variability (Greenwald et al., 2003; Hofmann et al., 2005; Nosek, 2007). For instance, in one meta-analysis across 56 attitudinal domains, Nosek (2007) uncovered implicit-explicit correlations that ranged from strongly positive (e.g. above .70 for Pro-choice vs. Pro-life constructs) through to weakly positive (e.g. .20 for racial constructs, such as White vs. Asian). Consensus regarding these disparate results is that constructs of a more controversial nature tend to produce lower implicit-explicit correlations than less socially sensitive constructs, where conscious editing of attitude expression is presumably reduced. These results imply that the strength of the correlation between implicit and explicit attitude measures of a similar attribute is dependent on the level of stigmatisation associated with that target attitude (Swanson et al., 2001). In accordance with this theory, assessment of socially sensitive attitudes such as racial prejudice have often resulted in considerable disparity between participants' explicit and implicitly measured attitudes (e.g. Cunningham, Nezlek, et al., 2004; Greenwald et al., 1998; Hummert et al., 2002; Kawakami & Dovidio, 2001; Stanley et al., 2008).

Greenwald et al. (2009) highlighted this discrepancy when they conducted a comprehensive meta-analysis of 122 studies that employed not only the IAT, but also an explicit questionnaire and an outcome measure of relevant observable behaviour. Overall they found the attitude measures reasonably predicted outcomes on a wide range of behavioural, judgment and physiological measures, such as quality of cross-cultural interactions, consumer choices, alcohol consumption, amygdala activation, and so forth. However, one third of the studies inspected in the meta-analysis specifically investigated intergroup discrimination. Of these, the predictive validity of the IATs significantly exceeded that of the self-report measures, whereas the explicit attitude measures demonstrated superior predictive validity for domains such as political preference and consumer attitudes (Greenwald et al., 2009). This implies implicit attitude measures are stronger tools for predicting behaviours in situations where social desirability factors may be at play, whereas explicit attitude measures may be more efficacious for predicting controlled behaviours where social desirability concerns are limited (see also Dovidio, Kawakami, & Gaertner, 2002;

Perugini & Banse, 2007; Schnabel et al., 2008a; Spalding & Hardin, 1999). Several design and procedural aspects of implicit attitude measures enable a stronger estimate and greater predictive power for discriminatory or prejudiced attitudes than traditional explicit techniques (Greenwald & Krieger, 2006; Greenwald et al., 2009). Because of this, one of the primary advantages of implicit attitude measures is the detection of socially sensitive views.

Chapter Summary and Conclusion

Implicit attitude measures, such as the APT (Fazio et al., 1986) and IAT (Greenwald et al., 1998), were designed to overcome some of the limitations associated with explicit attitude measures. Over recent decades, dual-process theories have helped conceptualise implicit and explicit attitudes as being similar yet distinct processes. This has aided in the understanding of commonly reported discrepancies between implicitly and explicitly measured attitudes, particularly when assessing socially sensitive constructs. The following chapter presents a discussion of findings from the implicit attitudinal literature across the topics of racial prejudice, alcohol and other drug use, and implicit body image. These three research topics have traditionally proven difficult to assess using explicit measures and whilst not exhaustive, they are presented to illustrate the empirical utility of implicit attitude measures across a range of research domains.

CHAPTER TWO

The Application of Implicit Attitude Measures to Contentious Research Topics: A Critical Review

Implicit attitude measures have been argued to possess greater predictive validity for the assessment of socially sensitive or contentious issues (e.g. Greenwald et al., 2009). Underlying this argument is the assumption that several design and procedural aspects of implicit measurement techniques minimise opportunity for introspection and reduce a participant's control over the responses they provide (Greenwald & Krieger, 2006). As noted in the previous chapter, implicit attitude measures such as the Affective Priming Task (APT; Fazio et al., 1986) and the Implicit Association Test (IAT; Greenwald et al., 1998) differ from explicit questionnaire methods by being resistant (if not immune) to faking or deliberate manipulation of responses (Banse, Seise, & Zerbes, 2001; Nosek, 2007; Stanley et al., 2008). This poses a considerable advantage over explicit attitude measures whose validity can be substantially reduced by social desirability and self-presentational biases (Lowes & Tiggemann, 2003; Nosek et al., 2006; Perugini, 2005; Sherman, Rose, Koch, Presson, & Chassin, 2003; Spence, 2005). Such biases arise from participants distorting their true views on controversial issues to instead give more flattering or socially appropriate responses (Greenwald & Krieger, 2006). The more contentious the issue under investigation, the more likely self-presentational distortions will affect the veracity of responses (Lane et al., 2007; Nosek, 2005; Sherman et al., 2003). It is for this reason the most important applications for implicit attitude measures are in the detection of socially sensitive attitudes.

The present chapter examines findings from implicit measurement techniques across three socially contentious research topics: racial prejudice, substance use and body image. These topics have proven prime candidates for implicit attitude assessment and together provide evidence for the functional and predictive utility of implicit measurement techniques. A review of relevant APT and IAT findings are presented in turn for each of the three research domains, demonstrating that implicit attitude techniques promote an advantage over explicit attitude measures for these socially contentious topics. However, the review also reveals some inconsistencies amongst the presented research that suggest implicit attitude scores can be influenced by confounding factors, such as priming effects and task stimuli. Such extraneous factors may act to reduce the reliability and validity of implicit attitude techniques. It is argued that although there is evidence for the functional and predictive utility of implicit attitudinal measures, these psychometric concerns need to be systematically investigated.

Implicit Attitudinal Findings in the domain of Racial Prejudice

Implicit techniques can often uncover controversial or socially sensitive attitudes that may remain undetected using explicit techniques. Disparity between explicitly endorsed and implicitly held attitudes is particularly evident for racially prejudiced attitudes in Western society, where there is much social pressure to not reveal racially discriminatory attitudes due to increased acceptance of multicultural ideals (Amodio, 2013; Gilens, Sniderman, & Kuklinski, 1998). Regardless of this pressure, deeply ingrained racial prejudices can remain prevalent albeit difficult to ascertain using traditional questionnaire methods.

Racial prejudice refers to negative emotional reactions (or attitudes) toward a person or group due to race or religion (Amodio, 2013; Schweitzer, Perkoulidis, Krome, & Ludlow, 2005). Prejudice by definition involves differential treatment of one group over another (Rudman & Ashmore, 2007). As such, implicit attitude measures are well-suited to the assessment of racial prejudice because they deliver an indication of relative preference for one construct over another. Furthermore, many everyday situations involve only quick, superficial evaluations in which people are unlikely to deliberate, such as choosing whom to sit next to on a crowded bus. The speeded reaction times utilised by implicit attitude measures reduce the opportunity for deliberation, potentially producing an attitude assessment more attuned with these spontaneous decisions than would be obtained with considered reflection using explicit attitude measures. Thus the automaticity of implicit attitude measures is another advantage for the assessment of intergroup bias.

Many of the seminal studies in the field of implicit attitudes have investigated automatic racial evaluations. These key studies along with other supporting research are presented in the current section to highlight the validity and predictive utility of implicit attitude measures for the assessment of racial prejudice. How the APT and IAT can be adapted to assess inter-racial attitudes will be shown, along with what information these instruments provide and implications for the possession of negative implicit racial biases. It will be concluded that racial prejudice is well-suited to implicit attitude assessment, and is an important candidate for such research.

Changes in social norms throughout Western society commonly discourage outward expression of racial prejudice (Amodio, 2013; Devine, Plant, Amodio, Harmon-Jones, & Vance, 2002). Such a shift has been evident in a drastic reduction of prejudiced attitudes self-reported using explicit techniques (Islam & Jahjah, 2001; Schweitzer et al., 2005). However, it is difficult to determine whether decline in expressed prejudicial views is a reflection of true attitude change or merely an indication of pressure to not reveal such opinions. To avoid the possibility of appearing racially biased, participants may choose to apply self-presentation strategies to minimise any expressed prejudiced views (Devine, 1989; Devine et al., 2002; Dovidio et al., 2002; McConahay, Hardee, & Batts, 1981). This impression management reduces the potential for the attitude-bearer to be negatively evaluated by others, but it also acts to limit the validity and representativeness of traditional explicit attitude assessment measures. Such manipulated responding can result in an underestimation of racial prejudice within contemporary Western society (Amodio, 2013; Dolnicar, 2005; Dunn, Forrest, Burnley, & McDonald, 2004; Islam & Jahjah, 2001; Poynting & Mason, 2007; Schweitzer et al., 2005). By avoiding the issues of self-presentation bias, implicit attitude measures can facilitate a clearer insight into individual’s automatic evaluations and stereotypic thoughts. Deeply ingrained biases have been shown to influence behavioural outcomes, such as negative inter-racial interactions (Amodio, 2013; Dasgupta, McGhee, Greenwald, & Banaji, 2000). Because of this, the assessment of racial attitudes is arguably one of the more critical applications for implicit techniques.

Findings from the Affective Priming Task

In one of the earliest demonstrations of implicit intergroup bias, Fazio et al. (1995) applied their APT to assess systematic differences in reaction times as a function of race and valence. Participants were shown photographs of either White or Black faces before quickly categorising subsequent words as either positive or negative. The facial prime stimuli were described as distracters designed to make the task more challenging. Fazio et al. (1995) discovered that when White participants classified positively valenced words, their responses were faster when they had been exposed to White rather than Black faces. Yet when they classified negatively valenced words, their responses were faster after exposure to Black rather than White faces (Fazio et al., 1995). This pattern of findings has been widely interpreted as indicating the presence of implicit racial bias in favour of White people relative to Black people (Jost et al., 2009; Wittenbrink, Judd, & Park, 1997). Fazio et al. (1995) also found a level of relationship between priming task scores and spontaneous cross-cultural interactions. As part of the experiment, the participants interacted with an African American researcher who later rated this interaction for level of interest and friendliness (unaware of the participants score on the APT). Participants' responses on the Modern Racism Scale (an explicit questionnaire; McConahay et al., 1981) did not predict the quality of interaction with the research assistant, but their implicit APT scores did (Fazio et al., 1995). This revealed the implicit APT measure had greater predictive utility than the explicit questionnaire for the prediction of spontaneous cross-cultural interactions. This finding has been replicated (Wilson et al., 2000; Wittenbrink et al., 1997) and extended with the use of independent raters to judge the quality of cross-cultural interactions (e.g. Dovidio et al., 2002). In another study, scores on the APT were found to negatively correlate with how long participants would maintain eye contact with someone of different ethnicity (Dovidio et al., 2002). In both instances, the APT revealed greater relationship with the behavioural racially-relevant outcome than the explicit questionnaire measure did (Dovidio et al., 2002). These studies demonstrate substantial advantage for the use of APTs in the assessment of racial prejudice.

Findings from Implicit Association Tests

Fazio et al.'s APT research paved the way for the Implicit Association Test (IAT; Greenwald et al., 1998). As described in Chapter One, the IAT is based on the assumption that highly associated categories will result in faster performance than less associated categories during a forced-choice dual-categorisation task. The first IAT experiments focused on racially-relevant expected group differences between Japanese American and Korean American participants using a Japanese/Korean IAT (Greenwald et al., 1998). Due to the history of military subjugation of Korea by Japan during the first half of the 20^{th} century, this IAT was expected to reveal mutually opposed implicit attitudes. As expected, Japanese American participants responded significantly faster to Japan+Positive, Korea+Negative stimuli combinations than for the transposed conditions, while the Korean American participants responded in the opposite fashion. This finding supported the application of the IAT to the study of racial biases. Greenwald et al. (1998) then subjected White participants to a Black/White IAT in order to examine consciously disavowed evaluative differences. This study revealed the now well-replicated finding of an implicit preference for White over Black people as illustrated by significantly faster responses to the congruent (i.e. White+Positive, Black+Negative) than the incongruent (i.e. White+Negative, Black+Positive) stimuli combinations

(Cunningham, Nezlek, et al., 2004; Dasgupta et al., 2000; McConnell & Leibold, 2001; Nosek & Smyth, 2007; Rudman & Ashmore, 2007). Implicit preference for White over Black people has been well evidenced, even in children as young as six (Baron & Banaji, 2006) and four years of age (Cvencek, Greenwald, & Meltzoff, 2011). Such results imply implicit racial biases are ingrained at an early age and are thus an important area for further research.

Neurobiological research also supports the application of IATs to the assessment of racial prejudice. Employing EEG and fMRI brain imaging equipment, Phelps et al. (2000) revealed IAT effect scores were correlated with amygdala activation in White participants exposed to Black (rather than White) faces (see also Amodio, 2013; Cunningham et al., 2003; Cunningham, Johnson, et al., 2004). The part of the amygdala that activated is affiliated with quick emotional responses to threat, such as fear (Amodio, 2013; Phelps et al., 2000). Evidence of a physical fear response on the viewing of Black rather than White faces for White participants during the completion of a Race IAT indicate the IAT was assessing the construct it intended to, namely automatic racial concern. No correlation between amygdala activation and the explicit questionnaire responses was evidenced (Phelps et al., 2000), illustrating the IAT was more strongly related to the body’s physical response to fear than the explicit questionnaire was. This result provides evidence for a distinction between implicit and explicit processing, and that the implicit attitude measures provided a superior assessment of racial prejudice.

Divergence between implicitly and explicitly assessed racial attitudes has often been reported (see Chapter One; Banse, 1999; Cunningham, Nezlek, et al., 2004; Cunningham et al., 2001; Greenwald et al., 1998; Hofmann et al., 2005; Nosek, 2005; Swanson et al., 2001). This incongruity between implicit and explicit attitude measures is typically interpreted within the dual-process framework, with the expectation that divergence will be greater the more contentious or socially sensitive the construct of interest is (Nosek, 2007; Sherman et al., 2003). Yet, Karpinski and Hilton (2001) notably failed to find any correlations between the IAT and explicit attitude measures, even when social desirability pressures were minimised[3]. Conversely, McConnell and Leibold (2001) reported moderately strong relationships for a socially contentious racial-related construct measured using a Racial IAT and an explicit racial prejudice questionnaire (McConnell & Leibold, 2001). These findings are inconsistent and strongly query the claim that divergence between implicit and explicit attitude measures is simply a function of stigmatisation. However, McConnell and Leibold (2001) are not the only research to show a clear link between explicitly and implicitly held negative attitudes towards racial groups.

Prejudiced Attitudes against Arab Muslims and the Middle East

Over the last decade research has revealed considerable negative affect towards Arab/Muslims by Westerners using both explicit and implicit attitude measurement techniques (e.g. Agerström & Rooth, 2009; Chopra, 2008; Dunn et al., 2004; Dunn et al., 2008; Gibson, 2008; Park, Felix, & Lee, 2007; Rooth, 2010). These views are likely the result of several well-publicised socio-political events, including the

[3] Social desirability was minimised by assessing attitudes towards apples and candy bars, which was deemed unlikely to result in participants feeling the need to monitor their responses (Karpinski & Hilton, 2001).

destruction of New York's twin towers by 19 men bearing distinctively Muslim names and the "War on Terror" predominantly instigated by the United States of America on the Middle East (Dunn et al., 2004; Rashid, 2009). Explicit questionnaire results have revealed substantial antipathy towards Arab/Muslims throughout the United States of America, England, Denmark and Australia (Chopra, 2008; Dunn et al., 2004; Dunn et al., 2008; Gibson, 2008; Islam & Jahjah, 2001; Rashid, 2009). For instance, Australian research found Arabs were perceived as the most threatening racial group (Islam & Jahjah, 2001) and concern regarding Muslims was over twice as high as concern regarding Black Africans or Indigenous Australians (Dunn et al., 2008). These explicitly stated attitudes are consistent with IAT research that reveals "Other" foreign (ostensibly unfamiliar) names are considerably preferred over Arab/Muslim names (e.g. Agerström & Rooth, 2009; Nosek et al., 2007; Park et al., 2007; Rooth, 2010; Rowatt et al., 2005). For instance, Park et al. (2007) found an implicit preference for Black names (e.g. "Jerome") over Arab names (e.g. "Muhammad") using the IAT (Park et al., 2007), a finding also supported by explicit attitudinal research (Chopra, 2008; Dunn et al., 2004). Together these studies reveal high levels of anti-Arab prejudice (both implicit and explicit) expressed by participants across the Western world.

Evidence of such extensive anti-Arab prejudice is concerning as discriminatory attitudes are often indicative of a disposition of generalised intolerance. Cunningham, Nezlek and Banaji (2004) demonstrated robust evidence for a general ethnocentric attitude underlying automatic prejudiced evaluations towards a variety of social groups. Using structural equation modelling, each of the social group factors, White versus Black, Straight versus Gay, Christian versus Jewish, Rich versus Poor, and American versus Foreign, all loaded strongly onto a single higher-order factor of implicit ethnocentrism. These results indicate that those who hold negative attitudes toward one disadvantaged group are likely to consistently have negative attitudes evoked by other culturally disadvantaged out-groups (Cunningham, Nezlek, et al., 2004). This generalisability of implicit prejudice has considerable ramifications, both for the attitude-bearer and for the out-group being discriminated against.

Implications for Negative Implicit Racial Bias

Rudman and Ashmore (2007) provided a clear demonstration of the link between possessing negative implicit racial attitudes and unambiguously harmful behaviour towards minority group members. Scores on their Racial IATs were found to significantly relate to self-reported racial discriminations, such as: verbal slurs (e.g. expressing racially or ethnically offensive comments and jokes), excluding others from social gatherings and organisations because of their ethnicity, engaging in threat, intimidation, nonverbal hostility (e.g. giving ‘the finger’) and even physically harming out-group members and/or their property (Rudman & Ashmore, 2007). Although it is unusual for participants to respond with such honesty regarding these prejudiced activities, for each item of the questionnaire participants were first asked whether they had themselves been subjected to that form of discrimination before consequently asking if they had done likewise. This process may have helped normalise the experience and encouraged honesty of responses. A relationship between implicit prejudice as measured by the IAT and such unambiguously harmful behaviours supports the validity of the IAT for prejudice-related research.

Rudman and Ashmore’s (2007) research was extended to also examine economic discrimination in the form of hypothesised budget cuts for various minority groups’ student organisations. Implicit racial biases of the predominantly White university students predicted economic discrimination against Jews, Asians and Black people. These findings support the usefulness of applying IATs to prejudice assessment. Rudman and Ashmore’s (2007) research indicate that possession of negative implicit attitudes can influence overt negative behaviours as well as more subtle decisions, both of which can profoundly impact members of the out-group. This concept of influence was further examined in a series of studies by Rooth et al. (2009, 2010) focused on employer discrimination in Sweden. Rooth et al. (2009, 2010) found implicit anti-Arab prejudice (which was detected in 94% of participating managers using a Racial IAT) was strongly and negatively correlated with the likelihood of managers providing interview opportunities for Arab job applicants (Agerström & Rooth, 2009; Rooth, 2010). These findings elucidate the possible extent of economic and social impacts that could result from the unchecked negative implicit attitudes of politicians, policy makers and managers.

The potential impacts of implicit bias are not more clearly demonstrated than in a study by Green et al. (2007) of medical doctors in the United States of America. The physicians completed measures of implicit and explicit racial bias and were randomly assigned to make a hypothetical diagnosis and expert recommendation for a 50 year old male patient who could require a thrombolysis, a standard medical treatment designed to break up blood clots, which is often used to treat heart conditions (McCaul, Lourens, & Kredo, 2012), strokes and deep vein thrombosis (Watson & Armon, 2004). The hypothetical patient happened to either be a White or a Black male. The physicians were shown to report no explicit preference for White or Black patients; however they exhibited substantial pro-White/anti-Black biases at the implicit level. The degree of the physician's pro-White implicit bias was found to be positively associated with the likelihood of recommending thrombolysis for White patients and negatively associated for similar treatment for the Black patients. This study revealed that implicit racial bias can lead to the withholding of valuable medical treatment for some patients. From this it is possible to infer that implicit bias, of which the individual may be completely unaware, can have life-or-death consequences for others (Jost et al., 2009). The research of Green et al. (2007) indicated that prejudice against Black patients was only evidenced by the IAT and not the explicit attitude measure. This underlines the integral and important role of the IAT in the assessment of automatic racial prejudice.

Summary of Implicit Attitudinal Findings for Racial Prejudice

Implicit attitude measures typically elicit a strong preference for White over Black people in the majority of Western participants, even for children as young as four years of age. This is despite explicit questionnaires that indicate prevalence of racial prejudice is on the decline. The presented review found implicit attitude scores were more strongly related to various criterion-related indices than explicit attitude measures, including: quality of cross-cultural interactions (such as perceived pleasantness and amount of eye contact), likelihood of engaging in unambiguously harmful behaviours (such as racial slurs, threats and physical harm to out-group members), economic discrimination (such as lack of interview opportunities and funding cuts), recommendations for the provision of medical treatment, and amygdala activation. Results such as these imply that implicit attitude measures deliver a more sensitive assessment tool for racial prejudice than explicit attitude measures. Given the implications of harbouring negative implicit racial biases can be extensive for out-group members, uncovering racial biases appears an important application for implicit attitude measures such as the APT and IAT.

Applications of implicit attitude techniques extend far beyond the realm of inter-racial interactions (see summaries in the meta-analyses of Greenwald et al., 2009; Hofmann et al., 2005). The following sections will present examples of two clinically-relevant research domains that are also well-suited to implicit attitudinal investigation, that of substance use and body image. The assessment of substance use has proven to be problematic using explicit techniques, predominantly due to underreporting of substance use behaviour. Whether underreporting is a function of limited insight/memory (due to effects of extensive substance use) or a hope to avoid disapproval/punishment from others, inaccurate reporting can have significant impacts for research, clinical assessment and treatment planning. Because implicit attitude techniques have proven difficult to purposefully manipulate or fake, they pose an advantage over traditional explicit questionnaire measures for substance use assessment. In the assessment of body image concerns, self-presentation distortions can impact the veracity of explicitly collected information, which is particularly problematic when examining the cognitions of clinical populations. Often there is substantial secrecy associated with eating disorders that may make the extraction of accurate information via explicit techniques very challenging. Implicit attitude tasks provide a unique way to circumvent these issues by removing the need for patients to consciously endorse their responses. Implicit attitude measures may thus enable new insights into the cognitions motivating these disordered behaviour patterns.

Implicit Attitudinal Findings relating to Substance Use

It is a well-known issue that substance users commonly under-report the quantities and frequency of their substance use (Boniface & Shelton, 2013; Brown, Kranzler, & Del Boca, 1992; Murray & Perry, 1987). This misreporting reduces the veracity of traditional explicit approaches to drug survey research. Despite being legal in Australia, use of drugs such as alcohol (generally consumed in liquid form) and nicotine (typically consumed by inhalation of cigarette smoke) have been associated with significant stigma or disapproval from others (Room, 2005; Sherman et al., 2003). This stigma reduces the likelihood that respondents will openly admit to the full extent of their substance use behaviours, either because of evaluation apprehension, a lack of insight or even self-acceptance regarding their drug consumption (Sherman et al., 2003). Significant motivating factors might also influence an individual to misrepresent their substance use, such as understating usage on first entry to a rehabilitation centre in the hope of a less restrictive treatment program, or over-reporting substance use to increase chances of diversion to drug and alcohol treatment programs rather than criminal sentencing. Inaccurate reporting of substance use behaviour can potentially result in individuals not receiving adequate support or treatment opportunities. Implicit attitudinal techniques are argued to avoid the issue of inaccurate reporting and have proved difficult to manipulate or fake (Asendorf, Banse, & Mücke, 2002; Steffens, 2004) making them well suited to substance use assessment. The present section will examine APT and IAT research related to legal drug use, illustrating the potential for indirect measurement techniques in this applied arena.

Findings from the Affective Priming Task

Ralston and Palfai (2012) provided a clinical application of the APT by adapting the task to examine the cognitive effects of alcohol consumption for students rated high in negative affect[4]. Depressive symptoms were a known contributing factor for alcohol consumption but unconscious or implicit motivations had not yet been tested. The Alcohol APT Ralston and Palfai (2012) devised used exemplar alcoholic beverages and soda drinks as the prime words, with standard positive and negative target words. Elevated depressive symptoms were found to be associated with stronger positive implicit alcohol evaluations, but only amongst students with higher coping motives (Ralston & Palfai, 2012). In other words, for students who viewed drinking alcohol as a useful coping strategy, the more depressed their affect, the more positive were their implicit attitudes towards alcohol. This unconscious motivating factor for alcohol consumption may have been more difficult to ascertain using explicit attitude measures due to potential lack of insight by the college students regarding their drinking behaviour. For instance, students may be more likely to think “I drink because that is what we do” rather than “I drink because when I am feeling low in mood I think that drinking may make me feel better able to cope with the world”. The APT thus proved a useful tool for elucidating under what condition the link between low mood and alcohol consumption would be stronges. This illustrates the priming task’s utility for informing about the underlying processes that influence engagement in substance use behaviour.

The APT has also been applied to examine contextual and motivational factors that influence attitudes towards smoking. Sherman et al. (2003) developed two APTs, one

[4] As measured by a questionnaire that paralleled the DSM-IV criteria for depression.

containing cigarette-related pictures including packaging, the other depicting cigarette-related pictures that highlighted more sensory (non-packaging-related) aspects of smoking, such as a cigarette burning in an ashtray. Using these measures, the smokers revealed positive implicit attitudes towards the sensory stimuli and negative attitudes towards the packaging stimuli. The discrepant findings were reported as showing the priming task's sensitivity to context and motivational factors associated with smoking attitudes, that social and sensory aspects of cigarettes are a stronger motivating factor rather than brand loyalty (Sherman et al., 2003). This distinction would likely have been difficult to ascertain using explicit attitude measures as smokers will often explicitly endorse or justify most things associated with their substance use behaviour (Swanson et al., 2001). Interestingly, the discrepancy between sensory and packaging stimuli was not evident in the two IATs also used in that study, which is surprising given both tasks ostensibly assess implicit attitudes towards the same constructs. Furthermore, the sensory APT was the only one of the four implicit attitude measures to correlate significantly with the explicitly reported frequency of smoking behaviour. Based on these findings the APT appeared to demonstrate better predictive validity than the IAT for smoking behaviour (see also Leventhal et al., 2008). The aforementioned research thus revealed the APT as a potentially useful technique for the assessment of substance use-related cognitions.

Findings from Implicit Association Tests

Implicit cognitions relating to substance use have also been examined using the IAT. Wiers et al. (2002) investigated attitudes towards alcohol (versus soda) using an Alcohol/Arousal IAT that differed from a traditional IAT by requiring categorisation of words associated with Arousal (e.g. "excited") and Sedation (e.g. "listless") as opposed to the standard Positive and Negative target words. Heavy drinkers revealed faster performance on the Alcohol+Arousal (Soda+Sedation) pairings than when these categories were transposed – an effect not observed for the light drinkers. However, this distinction between light and heavy drinkers vanished when the traditional evaluative categories of Positive versus Negative stimuli replaced the Arousal versus Sedation target words. Using the Positive/Negative Alcohol IAT all the participants demonstrated relatively negative implicit evaluations towards alcoholic beverages (Wiers et al., 2002). These results indicate that heavy drinkers automatically expect arousal effects from alcohol whereas light drinkers do not, however most people implicitly view alcohol in a negative light (Wiers et al., 2002). Again, these findings would have been difficult to gather using explicit attitude measures due to the likelihood that alcohol consumption would have likely been justified with a positive framework. Furthermore, the implicit association between alcohol and arousal was significantly related to reported drinking behaviour one month later (Wiers et al., 2002), which provides evidence of the relationship between IAT scores and observable behavioural outcomes.

A link between problem drinking behaviour and IAT effect scores was further elaborated by Palfai and Ostafin (2003). They used an Approach/Avoid alcohol-related IAT that differed from the traditional IAT by replacing Positive versus Negative trait stimuli with Approach (e.g. “advance”, “forward”) versus Avoid (e.g. “withdraw”, “escape”) stimuli. Higher IAT Approach scores were associated with more frequent heavy drinking episodes during the past month and higher amounts of alcohol consumed at each occasion. Higher IAT Approach scores were also linked to a number of explicitly assessed appetitive responses to alcohol, including stronger urges to drink, more positive expected outcomes and greater affective arousal responses (Palfai & Ostafin, 2003; see also Jajodia & Earleywine, 2003). It was concluded their Approach/Avoidance Alcohol IAT tapped implicit associations important for determining responses to alcohol cues (Palfai & Ostafin, 2003; see also Thush et al., 2008). The development of implicit indices of alcohol use motivation may be of value for applied treatment settings. Many individuals in treatment programs are under strong pressure to not consume alcohol and thus may not report urges or cues that indicate they are at high risk of substance consumption. The application of such a measure would mitigate this issue and could have potential utility for other substance use behaviour as well.

Sherman et al. (2003) applied the IAT to examine the addictive practice of cigarette smoking. Significant differences between the IAT effect scores of smokers and non-smokers were found, with smokers on average significantly less negative towards smoking. Although this finding appears positive for the application of IATs to the identification of substance users, it is in direct contrast with previous results from Swanson et al. (2001), whose smokers and non-smokers exhibited similarly negative implicit smoking attitudes. It is not immediately evident why there are such great inconsistencies between two ostensibly similar studies and further research is encouraged to clarify this. In Swanson et al.'s (2001) series of experiments, smokers were shown to strongly identify with a behaviour they didn't like (using a self-other IAT), even though they had high self-esteem (measured implicitly and explicitly) (Swanson et al., 2001). They characterised this pattern of implicit inconsistency for smokers as "I am good, and I identify with smoking, but smoking is bad". The smokers' explicit cognitions, however, were more in line with "I am good and I identify with smoking, and smoking is not so bad" (Swanson et al., 2001). These results indicate the presence of cognitive consistency principles at play in the explicit attitudes (see Festinger, 1957), which affirms dual-process theories such as the APE model (Gawronski & Bodenhausen, 2006; See Chapter One).

Summary of Implicit Attitudinal Findings for Substance Use

Implicit attitude measures have provided valuable information regarding motivations and cognitive effects of two legal forms of substance use. Implicit techniques have revealed unique variance in prospective alcohol use after controlling for explicit alcohol-related cognitions and background variables (Thush et al., 2008). Implicit techniques for examining attitudes can be advantageous in the field of substance use because they avoid the potential concerns of inaccurate reporting due to lack of insight/memory, potentially contributed to by substance abuse, as well as deliberate misreporting, which may be due to motivational influences such as not wanting to acknowledge the extent of substance use, hoping for an easier treatment program or wanting to gain diversionary treatment in preference to facing gaol time. For these and many other reasons, implicit measurement devices may prove useful for the assessment of substance use.

In the present section, implicit attitude techniques revealed several motivational and cognitive factors associated with alcohol consumption. Alcohol use was shown to be higher for those who associated Alcohol with arousal or positive outcomes. Future drinking behaviour was found to be significantly related to implicit attitude scores, particularly for those who perceived Alcohol use as a good coping strategy or who have depressive tendencies. Implicit attitude measures were able to differentiate smokers from non-smokers, as well as smokers that have and have not been deprived of nicotine prior to testing. Substantial cognitive inconsistencies were unveiled by the APT and the IAT for people engaging in substance use practices. For instance, smokers demonstrated implicit dislike for cigarette packaging but like for sensory aspects of smoking, or identified strongly (using implicit techniques) with a habit they didn't like despite possessing good self-esteem. These implicitly assessed cognitive inconsistencies reveal complex motivational influences driving substance use behaviour that can be separately examined using implicit attitude measures. It is unclear as to whether the APT or the IAT are the more appropriate measure for this, as very few studies have compared these two implicit techniques in this research field. The study by Swanson et al. (2001) found the APT was more highly correlated with explicitly assessed smoking behaviour than the IAT, which may potentially reveal greater predictive utility for the APT in this instance. However, there is not clear evidence to support this assertion. In general, implicit attitude measures were revealed to show several advantages over explicit attitude measures in the assessment of substance-related associations.

Implicit Attitudinal Findings in relation to Body Image

Implicit attitude techniques are lastly shown to be advantageous for the assessment of body-related attitudes. The benefit of this approach to assessment is it reduces the need for participants to reveal personal insecurities, which can ease the process of gathering important information such as cognitions that motivate eating disordered behaviours. Implicit body image refers to automatic self-evaluations specifically related to body size. Negative implicit body evaluations are often strongly associated with negative implicit self-esteem, which is in turn shown to strongly impact mental and physical health (Blechert, Ansorge, Beckmann, & Tuschen-Caffier, 2011; Spalding & Hardin, 1999; Trzesniewski et al., 2006). Implicit body image associations can thus have critical clinical implications for affect and eating behaviours, as evident in patients with eating disorders (Blechert et al., 2011; Vartanian, Herman, & Polivy, 2005). Traditional explicit assessment of body image can be problematic due to the potential reluctance of people to reveal personal insecurities (Vandromme, Hermans, & Spruyt, 2011). Concepts such as perfectionism and external validation are often associated with body image concerns, and as such the desire to be seen in a positive light can also result in greater use of self-presentation strategies (Greenwald & Banaji, 1995; Krause, Back, Egloff, & Schmukle, 2012). Such issues pose greatest challenge for the assessment of eating disorders given the nature of secrecy that often accompanies such mental illnesses. Implicit attitude measures avoid these insecurities and challenges by limiting the participant's awareness of the target of measurement. This reduced awareness removes the need for participants to be insightful about their cognitions and limits the likelihood that they will dwell on their self-evaluations. As such, there appears great potential for implicit measurement techniques to be applied to the assessment of body image concerns. The current section present findings related to body image from the APT and IAT research literature.

Findings from the Affective Priming Task

The Affective Priming Task (APT) was first adapted to examine attitudes towards body size by Bessenoff and Sherman (2000). A standard body image APT presents prime stimuli exemplars for the trait categories Thin and Fat. Typically these categories are depicted using pictures of obese and underweight women, as was the case for Bessenoff and Sherman (2000). These prime stimuli are assumed to facilitate or hinder categorisation of the Positive (e.g. “good”, “confident”) and Negative (e.g. “bad”, “ashamed”) target words. As with all priming tasks, a single prime word is presented briefly prior to each target word, which is subsequently categorised. The classic finding is that participants respond much faster when Fat primes precede Negative as opposed to Positive target stimuli, revealing an implicit prejudice against obese women (Bessenoff & Sherman, 2000).

Implicit prejudice assessed using the APT has been found to influence spontaneous behaviours. Bessenoff and Sherman (2000) convinced their participants that they would be completing a partner task with another student (who happened to be obese). Although the obese student was presently out of the room, a chair with a coat and backpack was situated in the otherwise sparse area. Participants were required to get another chair and go sit in the room to wait for their partner. It was found that participants who had shown greater anti-Fat prejudice on the APT placed their chair further away from the chair they thought the obese confederate was going to sit in than those who did not display the implicit pro-Thin/anti-Fat bias (Bessenoff & Sherman, 2000). These results reveal that implicit biases, as measured by the APT, can influence spontaneous reactions to others (see also self-esteem APT research by Krause et al., 2012; Spalding & Hardin, 1999; Vandromme et al., 2011). The implication that people can be unknowingly (and literally) distancing themselves from people they view negatively can also have substantial impacts for the target of the discrimination, such as negatively affecting the obese individual’s self-esteem

(Bosson, Swann, & Pennebaker, 2000). Overall, the aforementioned findings reveal the utility of APTs in the uncovering of sensitive attitudes such as body size.

Cognitions regarding body size are particularly important in the identification and assessment of eating disorders, such as Anorexia Nervosa and Bulimia Nervosa. The APT has previously been shown to differentiate persons with eating disorders from healthy controls. In one example, Blechert et al. (2011) elicited weight concerns using Positive and Negative prime sentences such as “when I lose weight, I feel…” (Positive) or “when I gain weight, I feel…” (Negative). They created two APTs, one with Interpersonal-related target words (e.g. “popular”, “rejected”) the other with Performance-based target words (e.g. “capable”, “weak”). The results of these APTs found that the eating disordered clients responded significantly faster when the Positive weight-related primes preceded positive Interpersonal and Performance-based target words than was found to be the case for the healthy controls (Blechert et al., 2011). These results indicated a connection between shape/weight concerns and non-appearance-related self-esteem domains such as interpersonal relationships and achievement/performance, revealing the more generalised types of self-esteem impacted by eating disorders (Blechert et al., 2011). Given clients with eating disorders often have poor insight regarding their cognitions and behaviours, this valuable insight into the self-esteem of eating disordered patients would prove very difficult to ascertain using traditional explicit attitude measures.

Findings from Implicit Association Tests

The Implicit Association Test (IAT; Greenwald et al., 1998) has also been adapted to assess pro-Thin/anti-Fat attitudes in eating disorder patients. The standard body size IAT involves the trait categories of Thin and Fat, as well as affective Positive and Negative stimuli. The typical finding is that participants will respond significantly faster to the congruent combinations (Thin+Positive, Fat+Negative) than the incongruent combinations (Thin+Negative, Fat+Positive), revealing implicit preference for the “thin ideal” and anti-Fat prejudice (Fadda, Fronza, Galimberti, & Bellodi, 2011). In one study, this IAT effect was revealed for the healthy controls as well as the eating disordered patients (including those with Anorexia Nervosa, Bulimia Nervosa and Binge Eating Disorder) (Fadda et al., 2011). In fact, there were no significant differences between the magnitudes of implicit pro-Thin/anti-Fat attitude uncovered for the participants, indicating a generalised preference for Thin over Fat (Fadda et al., 2011; Vartanian et al., 2005). Such implicit bias has been shown to differ from the explicitly reported attitudes of healthy controls (Vartanian et al., 2005), which may reflect the greater sensitivity of the indirect measurement technique over traditional questionnaire approaches for the assessment of body-related constructs.

Implicit pro-Thin/anti-Fat bias has been shown to be internalised from a very young age. Using their Pictorial IAT (PIAT), Thomas et al. (2007) found that children as young as three years of age demonstrated the ‘thin is good, fat is bad’ ideology previously identified in adult populations. Given the link between implicitly held attitudes and consequent behaviours, these findings have important implications for social development across the early years. Thomas et al.’s (2007) research was the first to reveal implicit attitudes in pre-school children to social and non-social stimuli and was thus an important advancement for implicit attitudinal research. Explicit attitude assessments for children had previously been severely limited by acquiescence effects (whereby children provide the answer they think is wanted), limited verbal capacity and lack of cognitive insight, all resulting in reduced veracity of the attitudes obtained (Spence, 2005). Implicit attitude measures, such as the IAT, have proven advantageous in this respect by avoiding these potential issues (Thomas et al., 2007). The aforementioned findings demonstrate the utility of implicit measurement techniques for the assessment of cognitions relating to body size.

Summary of Implicit Attitudinal Findings for Body Image

The reviewed findings have shown that implicit attitude measures can provide insights into the contentious issue of body image that may have otherwise been difficult to ascertain using explicit attitude measures because of issues of acquiescence, lack of insight, or self-presentational concerns. By avoiding these potential confounds, the implicit attitude measures revealed more sensitivity towards elucidating prejudiced biases than obtained using explicit attitude techniques. Implicit attitude measures were found to predict spontaneous non-verbal behaviours, such as physical proximity to a stigmatised individual and to provide valuable insights into cognitions underlying the challenging field of eating disturbances. These results demonstrate that implicit attitude measures can provide a valuable addition to traditional explicit approaches for body-size related evaluations.

Summary of Implicit Attitudinal Findings across Research Domains

The current chapter has presented substantial evidence for the utility of implicit attitude measures across the domains of racial prejudice, substance use and body image, all of which had previously proved challenging to accurately examine using self-report measures. The assembled research has illustrated evidence of predictive utility for both the APT and IAT across research topics. For instance, implicit attitude measures of racial prejudice have shown significant relationship to the quality of cross-cultural interactions and prevalence of discriminatory behaviour. Furthermore, APTs and IATs both predicted frequency and intensity of alcohol consumption, and implicit anti-Fat prejudice predicted social distancing from obese persons. The aforementioned studies thereby revealed greater predictive validity of implicit over explicit attitude measures for the outlined contentious research constructs. Together these findings provide strong evidence for the use of implicit attitude measures in the assessment of socially sensitive attitudes.

Yet implicit assessment techniques have not just been used to gain contentious information, they have also been used to provide individualised feedback regarding implicit attitudinal prejudices (see Green et al., 2007; Greenwald et al., 2009; Nosek et al., 2007). Providing individual feedback like this has been argued to increase awareness of personal unconscious biases and motivate people to apply the cognitive effort required to change such cognitions (Nosek et al., 2006). For instance, in one study examining implicit racial prejudice in physicians, following receipt of personal feedback the doctors reported higher levels of awareness regarding the influence of implicit racial prejudice, with many noting they would increase efforts to counteract this prejudice in the future (Green et al., 2007). The provision of individual feedback is standard procedure for some research groups[5]. After completion of an IAT, participants are typically informed that their "data suggests a slight/moderate/strong automatic preference for X compared to Y". This feedback has been argued to provide a positive step towards reducing negative implicit attitudes by raising awareness of automatic biases (Green et al., 2007; Rudman & Ashmore, 2007). Some have gone so far as to suggest the IAT be used as a screening tool during job interviews to reduce the potential for racial discrimination within the workplace (Ayers, 2001, pp. 424-425). The IAT could also conceivably be applied for use as a diagnostic tool with potential substance abusers. However, before implicit attitude measures can be implemented as useful measures of individual differences (and important decisions are made based on the results) it is first crucial that these tasks prove to be relatively stable measures of the construct of interest.

Concerns for the Use of Implicit Attitude Measures: Instability of Implicit Attitude Scores

The present chapter has highlighted the potential advantage implicit attitude measures provide over explicit attitude measure for the domains of racial prejudice, substance use and body image. Much support for the APT and IAT was demonstrated, with many of the findings occurring in the expected direction and relating to expected behavioural outcomes. However, this was not always the case. In some instances, implicit attitudinal research was reported that indicated divergent results for ostensibly very similar studies. Such inconsistencies raise concerns for the stability of implicit attitude techniques. For example, McConnell and Leibold (2001)

[5] For instance, the Project Implicit website (https://implicit.harvard.edu/implicit/; Greenwald et al., 2011) that enables free web-based participation in many IATs across a range of topics.

reported a moderately strong relationship between a Black/White IAT and an explicit race-related questionnaire, whereas the thorough examination by Karpinski and Hilton (2001) found almost nil correlation between similar tasks. Furthermore, Sherman et al. (2003) reported significant differences between the IAT effect scores of smokers and non-smokers on their smoking-related IAT, whereas Swanson et al. (2001) found no differences for the two participant groups, again using similar tasks. Sherman et al. (2003) also reported a significant difference between results for the sensory and packaging APTs but no difference between the sensory and packaging IATs. Likewise, Jajodia and Earleywine (2003) found their Positive Alcohol IAT correlated significantly with an explicit drinking questionnaire, whereas the Negative Alcohol IAT did not. Such inconsistencies may potentially reflect underlying differences in attitudes (and are almost always interpreted as such), but may also be indicative of instabilities in the actual implicit measurement techniques.

A growing body of evidence has revealed changes in stimuli, experimenters and previously observed material can all significantly influence implicit attitude scores. It is likely such changes are less an indication of attitude change but rather a lack of internal consistency within the measures. In this situation consistency refers to the reliability of the tasks, with high instability potentially an indicator of ‘noise’ or error in the data, which reduces the robustness of such measurement techniques. The present section describes some notable studies that illustrate the susceptibility of implicit attitude scores to influence by many confounding factors. Confounding influences can increase the inconsistency of implicit attitudinal data thereby reducing the overall robustness of implicit attitude assessment.

Evidence of the Malleability of IAT Effect Scores

Stimuli Exemplars

Dasgupta and Greenwald (2001) were one of the first to demonstrate the IAT effect score could be easily manipulated. They found that by altering the standard Black/White IAT to present famous African Americans and infamous European Americans instead, the IAT effect was reduced by more than half. This indicates that changes in the IAT stimuli can substantially impact the resulting IAT effect (see also Steffens, Kirschbaum, & Glados, 2008). It could be argued that by altering the task in that way race evaluations were confounded with familiarity, potentially producing a Famous/Infamous IAT rather than a Black/White IAT. Nevertheless, even subtle changes in stimuli have been shown to impact implicit attitude scores. Such influence has also been evident in APT research. For Sherman et al.'s (2003) research, the inclusion or not of packaging information in the smoking-related pictures determined whether a negative or a positive implicit attitude towards smoking was produced. These results indicate the significant influence of the exemplar stimuli in the creation of implicit attitude scores. Yet this susceptibility to influence is not constrained to stimuli selection as contextual cues in the environment can also result in "attitude" change as measured by the IAT effect score.

Context Effects

Lowery, Hardin and Sinclair (2001) found that the mere presence of a Black (as opposed to a White) experimenter drastically reduced the Black/White IAT effect scores for Western participants; a finding also replicated using a subliminal priming task (Lowery et al., 2001). These results indicate that stimuli and the environment can substantially affect implicit attitude scores. Context effects may also occur due to previously observed items or information. Park and colleagues (2007) clearly illustrated such context effects by requiring participants to read one page of text (either a newspaper article about the New York terrorist attack dated 12th September 2001, a health report on drinking water, or an essay on multiculturalism) before then completing an Arab/Other IAT. The IAT effects produced for the negative (newspaper article) condition were significantly greater than those in the neutral (health report) condition, suggesting exposure to negative information regarding a terrorist attack strengthened the association between Arabs and negative attributes (Park et al., 2007). Additionally, the positive (multiculturalism) condition resulted in substantially reduced IAT effects compared to those in the negative and neutral conditions (Park et al., 2007). This implies researchers potentially possess the power to manipulate the IAT effect using environmental cues or the type of stimuli chosen.

Interpretation of Category Exemplars

Han et al. (2009) demonstrated that IAT target categories can also be open to multiple interpretations based on previously observed items. Han et al.'s participants completed a questionnaire that either asked how much "people" or how much "I" like/don't like various non-race related attitude objects, before completing a standard Race IAT with Pleasant/Unpleasant category labels. As expected, the IAT effect scores revealed much lower levels of racial prejudice for the participants who had completed the "I like/dislike" questionnaire compared to the "people like/dislike" questionnaire. These results imply that attitude estimates provided by the IAT are subject to manipulation based on previous contexts (a well-recognised occurrence given this is the basic assumption underlying the priming task). In their second and third studies, Han et al. (2009) illustrated situations where the IAT effect score had changed (ostensibly showing a shift in implicit attitude) despite a high likelihood that no attitude change had occurred, as well as situations where the implementation of the IAT could obscure the detection of attitude change when change may in fact have taken place (Han et al., 2009).

Concerns for the Stability of Implicit Attitude Scores

The instability evident in the aforementioned IAT results raise some serious concerns for the ability of implicit attitude measures to accurately and reliably assess implicit attitudes. Previous research has echoed such concerns. For instance, Karpinski and Hilton (2001) queried whether implicit attitude measures, such as the IAT, assess implicit attitudes at all and Tetlock et al. (2009, p. 30) concluded there was “no scientific support” for the strong implicit-prejudice argument. Whilst there is clearly substantial evidence for the potential usefulness of implicit attitude measures, as outlined in the current chapter (see also Greenwald et al., 2009; Jost et al., 2009), these concerns regarding the validity and robustness of implicit attitude findings should not be dismissed.

One potential explanation for the inconsistencies and instabilities presented in the aforementioned findings is that the implicit attitude measures are highly influenced by error variance, which is limiting the reliability and validity of these techniques. The notion of error variance refers to any variability in scores that is not attributable to the trait attitude construct of interest. In the above discussion of factors that can influence IAT effect scores, which included interpretation of stimuli and task requirements, the testing context and previous experiences (priming effects), such factors all relate to potential sources of non-trait related variability or error variance.

Such error variance can add ‘noise’ to the data, resulting in greater inconsistency, reduced stability and less validity for measurement instruments. Previous research has documented high vulnerability to error variance for reaction-time-based procedures (Gawronski, LeBel, Banse, & Peters, 2009), which are the methodological format of most typical implicit attitude measures. Given this, it is hypothesised that the APT and IAT are susceptible to large amounts of error variance, which likely have been contributing to the inconsistencies and instability evident in implicit attitudinal research. This hypothesis will be critically examined in the coming chapters. Chapter Three explores possible sources of error variance for the APT and IAT within the context of classical test theory. It is argued that failure to adequately account for error variance in past implicit attitudinal research has contributed to the poor psychometric evidence available for the APT and IAT. Systematic analysis of error variance in implicit attitudinal data is proposed.

Chapter Summary and Conclusion

Implicit attitude measures have been shown to provide many advantages over traditionally-used explicit attitude measures for contentious research topics such as racial prejudice, substance abuse and body image. Yet despite the wide use of implicit attitude measures and their intuitive appeal for evaluating attitudes across various research domains, evidence of inconsistency and instability of implicit attitude results have led to substantial concerns regarding the robustness and validity of these techniques. The previous section revealed that implicit attitudes can be influenced by extraneous factors, such as the type of stimuli used, the specific task requirements, the testing context, and relevant previous experiences (Han et al., 2009; Karpinski & Hilton, 2001; Lowery et al., 2001; Olson & Fazio, 2004). It has been suggested that such influencing factors are examples of non-trait related variance that could be negatively impacting the stability and validity of implicit attitude measures. High amounts of error variance drastically reduce the utility of implicit attitude measures for applied research, such as the provision of personalised implicit prejudice feedback. A systematic investigation of the sources and amount of error variance in implicit attitudinal scores is required.

CHAPTER THREE

A Critical Examination of the Potential for Error Variance in Implicit Attitudinal Measurement

In Chapter Two it was demonstrated that empirical inconsistencies exist between results obtained using seemingly similar implicit techniques. This was well exemplified by findings such as that of Sherman et al. (2003) who showed that smokers could be differentiated from non-smokers using a smoking-related IAT; a finding in direct contrast to Swanson et al. (2001) who found no difference between such participant groups. Chapter Two also demonstrated that implicit attitude scores can be significantly affected by extraneous factors such as interpretation of the stimuli, the task requirements, the testing context as well as previous experiences (e.g. Han et al., 2009). Other characteristics, such as a participant’s attentional capacity, general processing speed and task-switching ability have also been shown to influence implicit attitudinal scores (Back, Schmukle, & Egloff, 2005; Fiedler, Messner, & Bluemke, 2006; Mierke & Klauer, 2003). Together these factors have the potential to introduce a substantial error component to implicit attitude data, resulting in greater inconsistency and reduced validity for these measures.

The current chapter explores the issue of ‘error variance’ and how it may occur in the course of implicit attitudinal assessment. The term error variance is used here to cover all variances that cannot be attributed to the construct of interest; in the case of this thesis, implicit attitude. Using variations on classical test theory, error variance, as defined here, is shown to include both random error and systematic error components. When left unaccounted for, these error types can greatly influence the observed research findings. Random error variance indiscriminately influences the scores, confounding validity estimation and increasing inconsistencies amongst the findings. The presence of random error might have contributed to the inconsistencies evident in past research. Systematic error variance differentially impacts upon aspects of the scores resulting in biased or misleading findings. These systematic influences are often associated with, and difficult to differentiate from, the method of measurement and may also have contributed to between study variability.

Previous psychometric evaluations of implicit attitude measures have typically failed to adequately address non-random distribution of error variance, and often revealed poor construct validity and highly inconsistent reliability estimates for implicit tasks. This review chapter will discuss these findings, and propose that such limited psychometric evidence for implicit measures are partly due to the presence of unaccounted for high error variance in the scores. It is argued traditional statistical approaches have been inadequate in addressing this issue of error variance for implicit techniques, due to (1) the assumption of random error distribution that is unlikely to hold for implicit attitude measures, and (2) the failure to control for systematic error, in particular error due to method effects. Methodologies that are capable of modelling and/or controlling for such error during analysis would be more desirable, to systematically assess the impact of error variance in implicit attitudinal data and to facilitate a more comprehensive evaluation of the construct validity of implicit attitude measures. Structural Equation Modelling (SEM) analytic approaches enable this by accounting for random error and method effects. SEM analytical techniques thus have potential to deliver a more stringent psychometric evaluation of implicit measures and a clearer view of what implicit attitude tasks actually assess.

Error Variance in Implicit Attitudinal Research

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 or prejudice towards an attitude construct like race. At the observed level, implicit attitude scores are devised by calculating the difference between two behavioural responses, such as the average reaction times for congruent and incongruent block trials for the IAT, as outlined in Chapter One (Blanton, Jaccard, Gonzales, & Christie, 2006). The observed behavioural responses (reaction times) are thus used to *infer* the magnitude of the abstract construct of implicit attitudes. However, it is highly improbable, or indeed impossible, that implicit attitude measures perfectly capture the abstract attitude constructs they aim to assess (Nunkoo & Ramkissoon, 2011). This is because the abstract nature of attitudes increases the difficulty in which they are able to be measured, leading to differences between the observed scores and the ‘true’ value of the conceptual construct (Cote & Buckley, 1987; Spector, 2006). This discrepancy relates to measurement error, which is introduced in this section within the theoretical framework of classical test theory (Spearman, 1904). Evidence for substantial error variance in implicit attitudinal data is then presented. There are numerous design aspects of implicit attitude measures, such as the APT and IAT, which likely contribute to error variance in implicit attitudinal measurement. The implications of this for the consistency and accuracy of implicit attitude assessment are discussed.

Classical Test Theory

Classical test theory, often attributed to Spearman (1904), states that any observed score (such as the IAT effect score) is comprised of two main components; the ‘true’, or trait component (i.e. implicit attitude construct) and an error component (that accounts for the imperfection of the measurement). This classical test theory framework has formed the basis of measurement theory for over one hundred years and rests on the basic classical test theory model, shown in Equation 3.1 (Eid, Lischetzke, Trierweiler, & Nussbeck, 2003).

$$y_{ijk} = t_{ijk} + e_{ijk} \tag{3.1}$$

where y_{ijk} is the observed score
t_{ijk} is the true or trait score
e_{ijk} is a residual component (error)
i is the indicator
j is the trait
k is the method

In recent conceptualisations of classical test theory, error variance has been recognised as comprised of both random and systematic components (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). This has resulted in an extension of the basic classical test theory model, such that an observed score is comprised of trait, systematic error (or method variance) and random error variance, as shown in Equation 3.2 (also Marsh & Grayson, 1995).

$$y_{ijk} = t_{ij} + m_{ik} + e_{ijk} \tag{3.2}$$

where y_{ijk} is the observed score
t_{ij} is the true or trait score
m_{ik} is a systematic or method component
e_{ijk} is a residual component (error)
i is the indicator
j is the trait
k is the method

Components of Observed Scores: Trait Variance, Systematic Error Variance and Random Error Variance

In classical test theory, trait variance refers to the portion of a score which can be directly attributed to the construct being measured (Cote & Buckley, 1988)[6]. Systematic error variance refers to characteristics (often associated with the methodology) that remain relative consistent regardless of the construct being assessed (Coenders & Saris, 2000). Random error variance refers to random variance that is not accounted for by the construct or methodology (Cote & Buckley, 1988). Significant amounts of error variance can be problematic because trait variance and error variance are inversely proportional. As such, greater error variance within an observed score reduces the amount of trait or ‘true’ variance that can possibly be present for the score (Cote & Buckley, 1988).

Both random and systematic types of measurement error can compromise estimates of reliability and construct validity, but do so in different ways. Random error variance typically adds noise to the data; resulting in greater inconsistency,

[6] In other words, the term “trait” is the mean of all true-score variables that measure the same construct (Pohl & Steyer, 2010). It is noted this meaning is distinct from the term “trait” applied in longitudinal studies, which often denotes stable and relatively consistent person-specific effects.

weakened observed correlations between similar constructs, and less clarity regarding the trait construct (Coenders & Saris, 2000). Systematic error (or method) variance can similarly weaken the observed relationships between constructs, but more typically artificially increase the observed relationships between constructs due to shared characteristics of the methodology (Cole & Maxwell, 2003; Podsakoff et al., 2003). Together both types of error variance can significantly impinge upon the consistent and accurate assessment of trait constructs by confounding validity estimation and reducing the reliability of measurement instruments.

Summary

Error variance is generated during measurement of abstract constructs, and may be either random or systematic in nature. Random error variance typically influences the observed scores in an indiscriminate manner, confounding results. Systematic types of error variance differentially impact aspects of the scores, resulting in potentially biased or misleading outcomes. In the next section, it is argued that implicit attitude measures are susceptible to both random and systematic forms of error variance by examining elements of IAT procedures most likely to result in error production.

Sources of Error Variance for Implicit Attitude Assessment

During standard implicit attitudinal assessment there are numerous factors that could introduce random and/or systematic error into the data. Such influencing factors can bias results to an unknown degree and include: differences in ability to sustain attention, fatigue, boredom, task-switching ability, cognitive functioning, general processing speed, test taking strategy, and interpretation of stimuli and task requirements. This section will predominantly focus on sources of error variance likely generated during an IAT. In this instance, the IAT is relatively representative of typical implicit attitude measures. It is argued that there is strong likelihood of random and systematic types of error variance influencing implicit attitudinal data.

Random Error in Implicit Attitude Data

Abstract constructs such as attitudes and personality are known to be more difficult and less reliably examined than tools measuring more overt constructs, such as job performance or cognitive functioning (Cote & Buckley, 1987; Kaplan & Saccuzzo, 2001; Spector, 2006). This difficulty in measurement tends to result in greater amounts of random error variance or uncertainty in the measurement of such abstract constructs. Random sources of error variance are non-trait-related forms of variance that indiscriminately influence the whole of a set of scores. Random error variance increases the erratic nature of the observed responses. This directly reduces the reliability of a measure, the ability for the task to consistently measure that which it measures (Nunnally, 1978). This is problematic, as measurement instruments can only ever be as valid as they are reliable (Cunningham et al., 2001). Great amounts of random error variance can also reduce the overall construct validity of reaction-time tasks, such as the APT and IAT, by increasing the non-trait-related ‘noise’ that is assessed. This increase in ‘noise’ makes it more difficult to clearly estimate the construct of interest, resulting in reduced construct validity (Cunningham et al., 2001; Gawronski, 2009). Further, because greater error means less trait is assessed, this limits the amount of convergence a task can have with other like tasks because the amount of overlapping trait construct is reduced. As such, random error can limit the reliability and validity of a measurement instrument. There is a high likelihood that implicit attitude tasks are compromised by a large amount of random error variance, particularly given their reliance on speeded categorisation tasks.

Error from Motor Response Execution.

Natural variability in motor response execution, which is required to perform speeded categorisation tasks, is a known source of non-trait-related inconsistency that can produce substantial proportions of random error variance (Poitou & Pouget, 2012). One of the most consistent facets of timed motor responses is the presence of large inter-trial variability (Miller & Katz, 2010). Such variability is believed to arise from the stochastic nature of neuronal firing, whereby cortical neural activity can vary quite substantially even when neuronal networks are responding to the same stimulus (Klein-Flügge, Nobbs, Pitcher, & Bestmann, 2013). Neural variability consequently results in variability in the motor response execution, which is evident for all reaction-time measures (Miller & Katz, 2010). As such, the physical motor action of executing a stimuli categorisation produces random error variance for implicit attitude measures, due in part to natural variability of neural activity (Donkin, Brown, Heathcote, & Wagenmakers, 2011; Poitou & Pouget, 2012).

Error from Attentional Lapses.

Rapid latency responding is thus a procedural format known to produces volatile, unstable and variable results (Brown & Heathcote, 2008; Buchner & Wippich, 2000; Debner & Jacoby, 1994; Lane et al., 2007). However, this volatility is further enhanced by any extraneous variables, such as blinking during stimulus presentation, momentary distractions, or lapses in attentional focus, all of which can substantially alter the accuracy and speed of responses for implicit attitudinal measures (Lane et al., 2007). Occasional attentional lapses in reaction time tasks are especially likely to be initiated by fatigue or boredom (Endler & Hunt, 1968; Nosek, Greenwald, & Banaji, 2005; Poitou & Pouget, 2012). This is problematic for implicit attitudinal measures because they involve categorisation of hundreds of stimuli in individual trials. Consistent maintenance of attention to every stimulus is thus very difficult (Salthouse, 2000), and with each and every attentional lapse greater discrepancy amongst the observed reaction times occurs (Brown & Heathcote, 2008; Debner & Jacoby, 1994). Because attentional lapses result in erratic or inconsistent responding they can be viewed as a random source of error variance.

Summary.

Random error variance poses a considerable issue for consistent and accurate implicit attitudinal assessment, and is likely a result of the tasks' reliance on rapid response latency techniques. Rapid response techniques are susceptible to erratic responding due to natural variability in motor response execution, attentional limitations, fatigue and boredom. Factors contributing to random error in implicit data potentially limit the reliability and construct validity of tasks such as the APT and IAT. However, systematic sources of error variance can pose an even greater issue as they can bias results in a way that may lead to misleading or inaccurate findings.

Systematic Influences of Error Variance on Implicit Attitude Measures

Systematic measurement error refers to relatively consistent extraneous influences that impact upon observed scores, regardless of the construct being assessed (Coenders & Saris, 2000). Systematic sources of error variance can likely result in increased or decreased observed relationships between constructs, in part because they can act differentially within a particular tool or measure leading to inaccuracies in the interpretation of findings (Cole & Maxwell, 2003; Podsakoff et al., 2003). One of the main sources of systematic influence is method variance, which is variance attributable to the methodology (Coenders & Saris, 2000). The term method refers to concrete aspects of the testing methodology, such as the content of specific stimuli, the response format and the context of the testing process (Geiser & Lockhart, 2012). Method variance can also be interpreted in a more abstract fashion that incorporates response biases due to influences such as social desirability and acquiescence (Podsakoff et al. 2003). All these influences have the potential to systematically influence responses on an attitude measure, leading to misrepresentative findings.

High Amounts of Method Variance in Attitude Measures.

Attitude measures have previously been shown to be quite susceptible to method variance, potentially leading to biased results. In one of the most comprehensive reviews of explicit questionnaire measures (see Podsakoff et al., 2003), Cote and Buckley (1987) reviewed 70 published data sets sourced from psychology, sociology, marketing, business and education literatures. They found that of all the constructs examined, explicit attitudes questionnaires possessed the greatest amount of method variance (using the Multitrait-Multimethod approach to Confirmatory Factor Analysis, CFA-MTMM, which is expanded upon later in this chapter). Specifically, Cote and Buckley (1987) revealed data from the explicit attitude questionnaires reflected about 40% method variance, 30% random error variance and only 30% trait variance related to the attitude construct of interest. These results imply that systematic forms of error variance, such as method effects, can significantly influence attitude assessments.

The findings of Cote and Buckley's (1987) research relate to explicit attitude measures; however it could be speculated that the abstract construct of attitudes would also lead to high amounts of systematic error variance when examined implicitly. Explicit and implicit attitudinal constructs have been shown to be highly related, albeit distinct constructs (Cunningham et al., 2001; Nosek, 2007). This relatedness may mean implicit attitudes are similarly susceptible to high amounts of systematic influence acquired during assessment by non-perfect measurement techniques.

Potential for Method Effects to Act as a Differentially Biasing Effect in the IAT.

Certain methodological features of the IAT make it susceptible to systematic biases that differentially influence the congruent and incongruent trials from which the IAT effect score is devised. Sources of systematic influence such as task-switching costs, task presentation order, intelligence and general processing speed can artificially bias IAT effect scores by facilitating responses on the congruent trials whilst the incongruent trials remain consistently difficult (Back et al., 2005; Fiedler et al., 2006; Mierke & Klauer, 2003; Schnabel et al., 2008a; Stülpnagel & Steffens, 2010). This biasing influence results 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.

Task-switching Ability.

The process of task-switching, which is required in an IAT when participants change between congruent and incongruent blocks of trials, provides one example of a systematic source of error. This process of transition demands re-learning the congruent/incongruent associations necessary for the correct categorisation of the stimuli. Such a cognitive shift requires significant mental exertion, cognitive flexibility and task-switching ability (Back et al., 2005; Fiedler et al., 2006; Mierke & Klauer, 2003). It has been argued that participants who are able to switch from one block type to the other more easily will produce greater IAT effect scores overall compared to those who have struggled to transition to the new task requirements (Back et al., 2005). This is because participants with greater task-switching ability tend to use this skill to progress quickly through the congruent trials, however the incongruent trials remain consistently difficult (Back et al., 2005; Fiedler et al., 2006; Mierke & Klauer, 2003). This would mean a participant with good task switching would have faster than normal congruent categorisation but equally slow incongruent categorisation, generating a larger discrepancy between the average latency of these two trial types than would otherwise be expected (Back et al., 2005). Because of the way the IAT effect score is calculated, greater discrepancy between the congruent and incongruent trials results in a larger IAT effect score, ostensibly revealing greater prejudice when in reality the aforementioned is merely a result of method variance. Similar arguments can also be presented for the systematic influence of intelligence and general processing speed, which likewise facilitate the congruent trials to a greater extent than the incongruent trials, leading to the incorrect impression of greater prejudice (Blanton et al., 2006; Fry & Hale, 2000; Jensen, 1993; Stülpnagel & Steffens, 2010). Task-switching, intelligence and general processing speed are non-trait related and participant-specific methodological influences that can artificially inflate IAT effect scores. But these are not the only sources of systematic bias for implicit attitudinal measures.

Block Presentation Order.

The presentation order of congruent and incongruent blocks delivers a further source of method effects, as the block type initially categorised can impact upon responses for the second block type completed. This is a type of context effect, whereby if the congruent block of trials is initially completed the incongruent block that follows typically produces much slower reaction times than is the case when the incongruent block is completed prior to the congruent block (Greenwald et al., 1998; Schnabel et al., 2008a). This order effect is likely because the incongruent trials are much more cognitively taxing than the congruent trials (Lane et al., 2007; Steffens, 2004; Williams & Themanson, 2011). As such, when a participant moves from the more challenging incongruent block of trials to the easier congruent trial block, their responses are facilitated because going from the context of the incongruent trials makes the congruent trials appear even easier. Likewise, the context of the easier congruent trials makes moving to the incongruent trials even more challenging. Because of this, order effects associated with the order in which the congruent and incongruent trials were completed, a purely method-induced effect, can result in an increased or reduced individual IAT effect score. Other sources of method variance for IATs include whether participant's adopt a speed-focussed or accuracy-focussed response style (Pachella, 1974; Salthouse & Hedden, 2002; Williams, Hultsch, Tannock, Strauss, & Hunter, 2005) and intepretation of stimuli and response categories (Han et al., 2009; Salthouse, 2000; Steffens et al., 2008). Such sources of potential systematic bias are expanded upon in Chapter Six.

Summary

The prior section has outlined clear evidence for the presence of random and systematic types of error variance in implicit attitudinal data. This confounding non-trait-related variance stems from many sources including motor response execution, attentional capacity, task-switching ability, processing speed, block trial order, and test taking strategy. Some of these sources of measurement error likely result in random error variance that could negatively influence the reliability and validity of the implicit attitude measure, but in an indiscriminate manner. Whereas systematic sources of error variance differentially influence certain aspects of implicit techniques (i.e. the congruent and incongruent trials of an IAT), which can significantly bias the resultant implicit attitudes that are purportedly revealed for a participant. Although it is currently unknown as to what extent these types of error variance are influencing implicit attitudinal data, the aforementioned evidence suggests the impacts could be sizeable. The next section examines previous psychometric evidence for the APT and IAT, arguing that the poor construct validity and reliability estimates evident are an indication of high amounts of random and systematic error variance negatively impacting upon the data.

Past Psychometric Evidence for Implicit Attitude Measures

Traditional correlation and regression-based analytical approaches are reliant on the assumption that error variance is randomly distributed amongst the scores, thereby not biasing results in one direction or another. As such, traditional analytical techniques do not overtly account for random or systematic error variance during analysis because error is considered inconsequential to the results. Past psychometric evidence for implicit attitudinal measures, such as the APT and IAT, has often relied on these analytical approaches. However, by neglecting to consider non-randomly distributed error, past findings may have been negatively influenced to an unknown degree by random and systematic measurement error. Were the hypothesis of substantial random and systematic error in implicit attitude measures confirmed, it would be expected that this would result in poor psychometric evidence for these measures, such as inadequate reliability and construct validity. This is because random error variance can confound estimates of construct validity and reduce the overall reliability of measurement instruments. Systematic forms of error can bias results by increasing or decreasing observed relationships between constructs, and as such the impact systematic error variance would have on psychometric evaluations is somewhat unknown (see Cole & Maxwell, 2003; Podsakoff et al., 2003). The current section presents previous psychometric evidence for the construct validity and reliability of the APT and IAT. It is argued that poor psychometric results provide further support for the hypothesis of substantial error variance in implicit data.

Construct Validity of Implicit Attitude Measures

Construct validity is the degree to which a test measures an intended hypothetical construct (Messick, 1990). Prominent types of construct validity include convergent and discriminant validity. Convergent validity assesses whether a measure is related to other measures of the same construct (Campbell & Fiske, 1959). Good convergent validity is evident when two measures of theoretically related constructs correlate highly. Discriminant validity occurs when a measure is distinct from other measures in expected ways (Campbell & Fiske, 1959)[7]. Evidence for discriminant validity occurs when supposedly unrelated constructs reveal the expected null relationship.

Discriminant Validity Evidence between Implicit and Explicit Attitude Measures

Prior empirical research has demonstrated great divergence between findings obtained using implicit and explicit attitude measurement techniques. Weak or highly variable correlations between the two measurement types are often reported, with correlations around or below r=.30 (e.g. Banse, 1999; Bosson et al., 2000; Cunningham et al., 2001; Devine et al., 2002; Gawronski, 2002; Green et al., 2007; Hofmann et al., 2005; Hummert et al., 2002; Karpinski & Hilton, 2001; Kawakami & Dovidio, 2001; Perugini, 2005; Rudolph, Schröder-Abé, Schütz, Gregg, & Sedikides, 2008; Schnabel, Asendorf, & Greenwald, 2008b; Thush et al., 2008; Vartanian et al., 2005; Wiers et al., 2002). Given the theoretical view that implicit and explicit attitudes are conceptually distinct constructs (as presented in Chapter One; see Cunningham, Nezlek, et al., 2004; Cunningham et al., 2001; Gawronski &

[7] It is noted that validity assessments for implicit attitude techniques are very different from that of evaluating self-report scales, because implicit measures represent procedural formats that can be applied to assess any number of attitudinal constructs (Lane et al., 2007). As such, two versions of the same implicit attitude measure may have little in common with each other apart from the basic task structure. This complicates psychometric evaluation of the measures, as both general (e.g. format) and specific (e.g. construct) issues of validity and reliability require examination. Nevertheless, it should still be expected that similar implicit attitude measures of the same construct would converge.

Bodenhausen, 2006; Haeffel et al., 2007; Wilson et al., 2000), such findings are to be expected and can be interpreted as evidence of discriminant validity for these measures (see also Gawronski, 2002; Greenwald et al., 2009; Lane et al., 2007; Rowatt et al., 2005; Strack & Deutsch, 2004).

Implicit and explicit attitudes, however, do not always diverge. In the field of consumer attitudes and political opinions, there are a number of instances where relatively high correlations between implicit and explicit attitude measures have been reported (e.g. Nosek & Hansen, 2008; Olson & Fazio, 2004). It has been argued that this convergence is a reflection of reduced need to hide undesirable responses (Swanson et al., 2001). As such, it is expected that stronger correlations would be revealed for consumer attitudes, for example, because there is generally little need to mask one’s food preferences. Yet low implicit-explicit convergence have also been reported for non-sensitive implicit associations, such as attitudes towards apples versus candy bars (Hofmann et al., 2005; Karpinski & Hilton, 2001; Olson & Fazio, 2004; Vandromme et al., 2011). Random error variance may have contributed to such empirical inconsistencies by increasing the variability of the scores, reducing the reliability of implicit attitudinal measures. The more error within the implicit scores, the less trait construct is able to be assessed. Because of this, random error effectively reduces the proportion of score that is comparable, for random error refers to inter-item inconsistencies and thus is not measuring anything per se. The smaller portion of trait variance that can be related to other estimates of trait construct (by other measures) is thus reduced thereby limiting the overall potential for convergent validity between tasks. The amount of random error in a measure thus provides an upper limit for observed relationships between tasks (Cunningham et al., 2001). Such conflictual findings may imply that implicit attitude measures are indeed affected by the confounding and limiting effects of random error variance.

Convergent Validity Evidence for Implicit Attitude Measures

Based on the dual-process theoretical framework of cognitive processing it is theorised that multiple implicit attitude measures of the same construct should prove strongly related, demonstrating convergent validity for these tasks (Fazio et al., 1986; Gawronski & Bodenhausen, 2006; Wilson et al., 2000). However, this has overwhelmingly not been found to be the case (Hofmann & Schmitt, 2008; Rudolph et al., 2008; Schnabel et al., 2008a; Sherman et al., 2003). In a study by Rudolph et al. (2008), numerous implicit attitude measures of self-esteem were examined, including the IAT, APT and a recently developed Single-Category IAT (SC-IAT; Steinman & Karpinski, 2008). Correlations between these implicit attitude measures revealed minimal inter-relationships between the tasks, despite efforts to improve the reliability of the implicit techniques (Rudolph et al., 2008). To exemplify this point, the non-significant correlation between the self-esteem IAT and APT was r=.07 (Rudolph et al., 2008), which is indicative of basically nil relationship between the measures (see also Jajodia & Earleywine, 2003; Krause et al., 2010; Sherman et al., 2003; Thomas, 2008). This failure to find evidence of convergent validity amongst different implicit measures of the same construct is "worrisome, both theoretically and empirically… (for if these implicit attitude measures) truly assess the same construct, then, by definition, they should overlap to a greater degree than they do" (Bosson et al., 2000, p.640; see also Hofmann & Schmitt, 2008).

Failure to find convergence among implicit attitude measures could be a direct result of high amounts of random error variance providing an upper limit for inter-implicit correlations. As discussed above, random error reduces the reliability of a measure by increasing inconsistencies in the assessment data. This reduces the amount of relevant attitude construct being measured, thus reducing possible inter-correlations with other like measures. Banse (1999) has argued the lack of convergence between implicit measures is a result of the low reliability (high error variance) of these tasks.

Reliability of Implicit Attitude Measures

Reliability refers to the dependability or consistency of measurement (Nunnally, 1978). The present section critically examines two prominent forms of reliability evidence for the IAT and APT, evidence based on cross-sectional data (i.e. internal consistency estimates) and evidence based on longitudinal data (i.e. test-retest reliability). Typically, implicit attitude measures have been found to possess considerably lower levels of reliability than explicit attitude measures (Buchner & Wippich, 2000; Fazio & Olson, 2003; Greenwald et al., 2009; Hofmann & Schmitt, 2008; Krause et al., 2010). The hypothesis of this thesis is that this discrepancy is caused by high amounts of random error limiting the reliability of implicit measures.

Internal Consistency

IATs have usually produced adequate internal consistency estimates, using Cronbach’s (1951) coefficient alpha, within the range of .80 (Asendorf et al., 2002; Bosson et al., 2000; Egloff, Schwerdtfeger, & Schmukle, 2005; Greenwald et al., 1998; Krause et al., 2010; Rudolph et al., 2008; Schmukle & Egloff, 2004). Such findings are acceptable for basic research, but higher estimates are encouraged for applied investigations (Nunnally, 1978). In contrast, internal consistency estimates for the APT typically sit around .55 (Banse, 1999; Bosson et al., 2000; Kawakami & Dovidio, 2001; Klauer & Musch, 2001; Krause et al., 2010; Rudolph et al., 2008; Spruyt, Hermans, Pandelaere, De Houwer, & Eelen, 2004; Vandromme et al., 2011), which is below satisfactory (Nunnally, 1978). Were random error variance to comprise a substantial portion of implicit attitudinal scores, it would be expected that internal consistency estimates would be less than adequate due to the erratic influence of random error reducing the consistency in which the tasks measure constructs. Given internal consistency estimates appear worse for APTs than IATs it could be hypothesised that random error variance poses a greater issue for the APT than the IAT. A systemic review would be required to investigate this claim.

Test-Retest Reliability

Test-retest reliability estimates tend to reveal even poorer estimates than internal consistency coefficients for implicit attitude measures (Egloff et al., 2005). Test-retest reliability coefficients for the IAT range from r=.27 to r=.69 with an average of approximately r=.50 (Banse, 1999; Bosson et al., 2000; Cunningham et al., 2001; Egloff et al., 2005; Greenwald & Farnham, 2000; Krause et al., 2010; Lane et al., 2007; Nosek, 2007). This is concerning as even a correlation of .50 demonstrates an alarming degree of inconsistency as 75% of the variability of scores over time can be considered unrelated to the trait construct. Even greater inconsistency is revealed in the test-retest reliabilities of the APT, with values ranging from a maximum of r=.56 (ostensibly comparable to the IAT; Kawakami & Dovidio, 2001) to r=.28 and even r=-.06, which is evidently quite unsatisfactory (Banse, 1999; Bosson et al., 2000; Krause et al., 2010). These findings of minimal test-retest reliability led Banse

(1999) to assert that the stability of priming effects was marginal at best, and Hofmann and Schmitt (2008) to note implicit attitude measures (including the IAT) are still unable to produce psychometric properties anywhere near the magnitude demonstrated by explicit attitude techniques. The highly inconsistent and generally poor reliability that is evident implies that substantial random error variance is impacting upon the implicit attitudinal data. Again, the influence of systematic error variance remains unknown.

Summary

The presented review of psychometric evidence for implicit attitude measures has uncovered much inconsistency, limited reliability and poor convergent validity amongst implicit assessment techniques. Indeed the findings have led Bosson et al. (2000) to conclude:

> "the study of implicit self-esteem (for instance) may be a boondoggle[8]. Right now, the psychometrics simply are not there" (Bosson et al., 2000, p. 641).

Such conclusions may be indicative that implicit attitude measures, such as the IAT and APT, are inherently flawed and of no empirical value because they do not validly measure trait constructs (Banse, 1999; Bosson et al., 2000; Karpinski & Hilton, 2001; Tetlock & Mitchell, 2009). This conclusion is, however, in direct contrast to the findings of the applied research presented in Chapter Two, whereby the IAT and APT were shown to possess reasonable predictive validity for a wide range of behavioural outcomes, such as quality of cross-cultural interactions, intensity and

[8] The term boondoggle refers to an unnecessary, wasteful or fraudulent project (Soanes & Stevenson, 2006).

frequency of alcohol consumption, and social distancing from prejudiced individuals. Such findings imply that implicit attitude measures may possess some functional utility in the assessment of socially sensitive attitudes, although the poor psychometric evidence may undermine such a conclusion.

Reconciling the divergent views that are presented based on the outcomes of psychometric and applied research literatures appears difficult, on the surface at least. However, the hypothesis of the present dissertation addresses this by suggesting implicit attitude measures do examine implicit attitudes, but the inconsistent findings, and sub-optimal reliability and validity amongst implicit assessment techniques is *directly a result* of the high levels of error variance inherent in the tasks. In the next section it is argued that it is unknown to what extent random and systematic error variance has influenced implicit attitudinal findings because the analytical approaches that have typically been used to examine implicit data are inadequate. Newer analytical approaches that account for random and systematic error variance are required to assess the potential utility of implicit attitude measures in applied research.

Past Psychometric Inadequacies in the Analysis of Implicit Data

The history of reported psychometric inadequacies for implicit attitude measures is arguably contributed to by past researcher’s failure to account for potentially large portions of error variance in the data. As mentioned earlier, traditional analytical approaches are based on the assumption of randomly distributed error variance. This assumption is unlikely to hold for implicit attitudinal measures. Because of this, high quantities of unaccounted for error variance may be indiscriminately or systematically influencing implicit attitude scores, impeding accurate estimation of underlying attitudes (Fazio & Olson, 2003; Gawronski, 2009; Rudolph et al., 2008). The present section argues that traditional analytical approaches are inadequate for examining implicit attitudinal data.

Inaccurate Assumption of Random Error Distribution for Implicit Measures

Research hypotheses for implicit attitude measures have traditionally been tested using correlational analyses or approaches based on the general linear model, such as regression or Analysis of Variance (ANOVA). These analytic approaches assume any error incurred in the measurement of variables is completely random (Tabachnick & Fidell, 2001). As a result of those assumptions error variance is assumed to be spread evenly throughout the data. In the case of the IAT, it would be assumed that error would be influencing the congruent and incongruent trials *equally* and thereby having little overall impact on the results. Because of this assumption, these statistical techniques test the observed variables directly and do not account for non-random distribution of error variance (Schumacker & Lomax, 2010). Yet as outlined earlier, this assumption is highly unlikely given the very plausible scenario that the incongruent trials are more cognitively taxing than the congruent trials (Steffens, 2004; Williams & Themanson, 2011) resulting in differential influence from systematic sources of error variance such as task-switching ability, intelligence, general processing speed and task presentation (Back et al., 2005; Blanton et al., 2006; Fiedler et al., 2006; Mierke & Klauer, 2003; Stülpnagel & Steffens, 2010). Random error variance in reaction-time tasks is noted to be more problematic with increased task difficulty (Brown & Heathcote, 2008), implying even random sources of error variance may pose a bigger issue for the incongruent that the congruent trials in an IAT. Because of this, there is very limited likelihood of completely random distribution of error variance for implicit attitude measures (see also Nunkoo & Ramkissoon, 2011).

Given the likelihood of uneven error distribution, analytical techniques that assume random distribution of error variance, such as traditionally used analytical approaches that examine observed scores, will likely result in statistical inaccuracies and biased results for implicit measures (Kline, 2005; Nunkoo & Ramkissoon, 2011). To conclude, correlational/regression and ANOVA-based analytical approaches are unable to account for non-random influences of error variance and are thus incapable of accurately estimating the reliability and construct validity of implicit attitude measures. Instead, a new analytical approach that is able to systematically evaluate error variance in implicit attitudinal research is required to address these past limitations. Structural Equation Modelling (SEM) offers one such approach.

The Structural Equation Modelling Approach to Addressing Error Variance in Implicit Attitude Measures

Structural Equation Modelling (SEM) provides one suitable tool to facilitate an evaluation of error variance for implicit attitude measures. SEM employs latent variable models, with multiple measures of each construct, to mathematically separate error variance from the trait construct of interest (Cunningham et al., 2001). These structural relations can clarify the relationships between variables as well as the impact of latent variables and error variance on the scores produced during testing (Byrne, 2005). SEM procedures such as Confirmatory Factor Analysis (CFA) provide one avenue for partialling random error variance from the implicit attitudinal data. Another more advanced technique, the Multitrait-Multimethod approach to Confirmatory Factor Analysis (CFA-MTMM) can expand upon this by accounting for both random error variance as well as method variance. These analytical approaches will be briefly introduced below.

Confirmatory Factor Analysis

Confirmatory Factor Analysis (CFA; Jöreskog, 1969) examines the amount of variance shared by items (or trials for an IAT/APT) in order to identify a common factor, which in this case is the underlying implicit attitude. Although CFA is based on classical test theory, in that it assumes observed scores are comprised partly of the trait construct being assessed and partly of random error variance, it differs in that it is unable to exactly measure the abstract construct of random error. As such, the concept referred to as 'random measurement error' in CFA is actually unique variance that is a combination of 'true' random error and reliable variance (or 'uniqueness') specific to the individual item (observed score) being examined (Brown, 2006). It is not possible to disentangle random error from uniqueness, however, parcelling the data (so that the observed score is comprised of several scores) can be one way to minimise its influence. CFA separates random error from the observed data, thus enabling a much more accurate and less confounded assessment of the trait construct.

Cunningham et al. (2001) applied CFA to examine the reliability and construct validity of the IAT. They found that after accounting for random error variance the

IAT data produced a reasonable test-retest reliability estimate of .68. This coefficient shows improved stability over time than traditional estimates (~.50; Lane et al., 2007), in terms of general psychometric standards (Kline, 1998). The findings of improved reliability post random error removal supports the theory that random error variance has provided an upper limit for estimates of the IAT's reliability. Cunningham et al.'s (2001) investigation compared two versions of the IAT, an APT and the Modern Racism Scale. They found consistency across the implicit attitude measures after correcting for random error variance (Cunningham et al., 2001). Other researchers have also revealed stronger convergence between implicit attitude techniques after applying CFA (Gawronski, 2002; Greenwald & Farnham, 2000). The finding of improved reliability and construct validity estimates for implicit tasks following the use of CFA implies there is a substantial random error component influencing IAT effect scores. These results provide strong evidence for the utility of CFA to account for such random error variance in implicit attitudinal data.

The Multitrait-Multimethod Approach (MTMM)

Confirmatory Factor Analysis (CFA) can facilitate accurate estimates of reliability and validity for implicit attitude techniques by accounting for random error variance (as per Cunningham et al., 2001). However, more advanced latent modelling techniques such as the Multitrait-Multimethod approach (MTMM) can deliver even greater stringency for construct validity estimates by accounting for both random and systematic forms of error variance. MTMM requires multiple traits (or constructs) to be assessed by multiple measures (Schumacker & Lomax, 2010), an estimate of method effects is then calculated based on the error covariance (Coenders & Saris, 2000; Malhotra, Kim, & Patil, 2006). In this way, MTMM can examine the systematic influences that are strongly tied to the measurement design, which are otherwise difficult to disentangle. This is particularly important for implicit attitude measures given implicit attitudes can only be assessed using specialised measurement techniques, and these techniques appear highly susceptible to systematic influences liable to bias the prejudice estimates of IAT effect scores.

Summary

Structural Equation Modelling (SEM) techniques provide one avenue to address the likely issue of error variance in implicit attitudinal measurement. Confirmatory Factor Analysis (CFA) can account for and partial out random error variance from the observed scores, enabling a more robust estimate of the reliability and construct validity of the measure. The application of CFA-MTMM then accounts not only for random measurement error, as with CFA, but further identifies variance specific to the method used in data collection (systematic error variance), delivering an even more directed estimate of the trait being assessed (Lance, Noble, & Scullen, 2002). CFA-MTMM can also provide an indication as to what magnitude of random error variance and method variance has been confounding the implicit attitudinal scores. A more detailed introduction to these various SEM analytical procedures will be covered in the following chapter, Chapter Four. It will be argued that SEM can be applied to examine many different facets of the reliability and construct validity of measurement techniques and appears well suited to facilitate a systematic review of error variance in implicit attitudinal measurement.

Chapter Summary and Conclusion

To conclude, the present chapter has argued that assessment of implicit attitudes is susceptible to substantial amounts of random and systematic error variance. Sources of such error variance have been shown to be closely intertwined with the IAT's measurement design and are difficult to differentiate from the true attitude construct aiming to be examined. Error variance is likely to have significantly confounded to an unknown degree previous estimates of implicit attitudes, as well as efforts to evaluate the psychometric properties of these tasks. Traditionally relied upon correlation and regression-based analytical techniques have been shown to be inadequate to account for the likely non-random distribution of error variance within implicit attitudinal data. As such, past psychometric evidence for implicit attitude measures was likely substantially confounded by error variance, which would have contributed to the highly variable and generally poor psychometric findings evidenced. Preliminary evidence of improved reliability and validity for the IAT following removal of error variance (using CFA), further supports the argument for significant error variance in implicit attitudinal data. These findings indicate that a systematic examination of error variance in implicit attitudinal data is required in order to sufficiently investigate the reliability and construct validity of these tasks. Such a review is possible using Structural Equation Modelling (SEM) approaches, such as CFA and CFA-MTMM, which can clarify the impacts of random and systematic error variance on implicit attitudinal research.

CHAPTER FOUR

Structural Equation Modelling: Applications for Construct Validation of Measurement Instruments

The previous chapters have argued that implicit attitude measures may have potential utility for assessing socially sensitive constructs; however at present the reliability and validity of these instruments is questionable. There is reason to suspect that existing implicit attitudinal research has been significantly confounded by random and systematic error variance. Structural Equation Modelling (SEM) procedures may provide a solution to this by accounting for random and systematic forms of 'error variance' during analysis. SEM can thus facilitate a systematic review of measurement error in implicit attitudinal data. Two prominent forms of SEM, Confirmatory Factor Analysis (CFA) and the Multitrait-Multimethod approach to CFA (CFA-MTMM), were introduced in Chapter Three. These analytical approaches can evaluate random error and method effects in data, enabling more accurate estimates of the psychometric properties of measurement techniques. The primary focus of the current chapter is to provide a non-technical conceptualisation and illustration of the main applications for SEM in the assessment of reliability and construct validity. It will also be demonstrated how SEM may be useful to assess substantive enquiries for the IAT, by examining equivalency testing and the influence of covariates on latent factors. Each of these applications has potential utility in the psychometric validation of implicit attitude measures.

The first section of this chapter presents an introduction to fundamental Structural Equation Modelling processes that guide Confirmatory Factor Analytic models, namely the common factor model, model specification and model estimation. Model specification enables CFA to be adapted to investigate different specific hypotheses. These applications for SEM are expanded upon in the second section of this chapter. Various SEM models are shown to provide a comprehensive approach to evaluating the reliability and validity of measurement instruments. It is argued that latent modelling techniques provide a suitable and thorough avenue for systematically examining the psychometric properties of implicit attitude measures.

Fundamental Structural Equation Modelling Processes

SEM is a comprehensive statistical approach used to test hypotheses about the relationships between observed and conceptual (or latent) variables (Hoyle, 1995). Latent variables are constructs that are unable to be directly observed such as implicit attitudes, which are instead inferred from directly observable variables such as reaction times (Brown, 2006). This conceptual framework is based heavily in the work of classical test theory (Spearman, 1904), as described in Chapter Three, which states that an observed score is comprised of both the trait construct being assessed (the latent variable) as well as measurement error.

SEM refers to a wide spectrum of latent modelling methodologies, such as path models, factor analyses, multiple group comparisons and multi-level modelling (Gau, 2010). The main focus of the current dissertation is on the factor analytic capabilities of SEM using Confirmatory Factor Analysis (CFA; Jöreskog, 1969). CFA is used to evaluate whether latent constructs are definable by a certain set of items (Jöreskog, 1969). CFA is thus often applied to determine the veracity of specified hypotheses regarding the relationships depicted in the observed data (Jöreskog, 1969; Schumacker & Lomax, 2010). This quality of specification makes CFA a precise yet flexible analytical tool. In the current section key theoretical and practical concepts integral to the application of SEM techniques are introduced. The common factor model will initially be described as it forms the foundation of factor analysis. This will be followed by a brief introduction to model specification and estimation. This introductory section aims to provide a conceptual overview of how SEM approaches, such as CFA, are developed to examine various research enquiries.

The Common Factor Model

The common factor model (Thurstone, 1947) aligns with classical test theory in that each observed variable is viewed as a linear function of one or more common factors (or latent variables) and one unique factor, often referred to as error variance (Brown, 2006). During CFA, the variance of each indicator (which is an observed variable) is partialled into two parts, the common variance and the unique variance. The common variance is the variance accounted for by the latent factor and is estimated using the shared variance between the indicators. For instance, the shared variance for four survey questions regarding political preferences (the indicators) can be used to determine the common latent construct of political attitudes (the common factor). A common factor or latent variable is thus an unobservable variable, hypothesised to influence more than one observed measure (Brown, 2006). Unique variance, in contrast, is the variance in the indicators not accounted for by the common factor (Coenders & Saris, 2000). Unique variance is a combination of ‘true’ random error variance, which is measurement error or unreliability in the indicator, and uniqueness, which is variance specific to the indicator but independent of the latent construct (Brown, 2006). As mentioned in Chapter Three, it is not possible to disentangle random error variance from uniqueness, however the influence of uniqueness can be minimised by using specially derived data parcels.

Data parcelling is a statistical process whereby scores from two or more observed responses are averaged and then the parcelled scores replace the item scores during CFA (Bandalos, 2002). Data parcels typically have higher communality than individual item indicators, meaning that the ratio of common variance to unique variance is larger (Meade & Kroustalis, 2006). This is because the uncorrelated sources of variance within an item and across all items in a domain are used to determine unique variance (Little, Cunningham, Shahar, & Widaman, 2002). Therefore, the more scores that have been aggregated to form the parcel the less uncorrelated sources of variance are present, reducing the overall level of ‘uniqueness’ in the indicator (Little et al., 2002). As each data parcel is then ostensibly interchangeable, there is almost no uniqueness for each indicator (Bandalos, 2002; Little et al., 2002). The data parcelling process thereby helps minimise the influence of uniqueness on the common factor model.

The Common Factor Model: Basics for CFA

The common factor model is often depicted as a pictorial representation of the relationships between common factors and indicators or test items (Thurstone, 1947). This model clearly illustrates the influence of unique variance and trait variance on the observed indicators. To aid interpretation of common factor models, the following reviews some conventions integral to CFA. Figure 4.1 presents a path diagram of a common factor model. Following the conventions of factor analysis, latent factors are depicted in ovals, whereas the indicators are represented by squares or rectangles. The common factor (η_1) and error variances ($\varepsilon_{1\text{-}4}$) are thus depicted in ovals (as it is not possible to directly observe them); whereas the observed variables or indicators (Y1-4) are represented in rectangles (see Figure 4.1)[9]. Within the path diagram, the unidirectional arrows (→) represent the factor loadings (λ, or *lambda*), which are the regression slopes (or direct effects) for predicting the indicators from the latent factor (η, or *eta*). The direct effect is a directional relationship between two variables, which is typically tested using an ANOVA or multiple regression analysis (Hoyle, 1995). These paths also relate the unique (error) variances (ε, or *epsilon*) to the indicators.

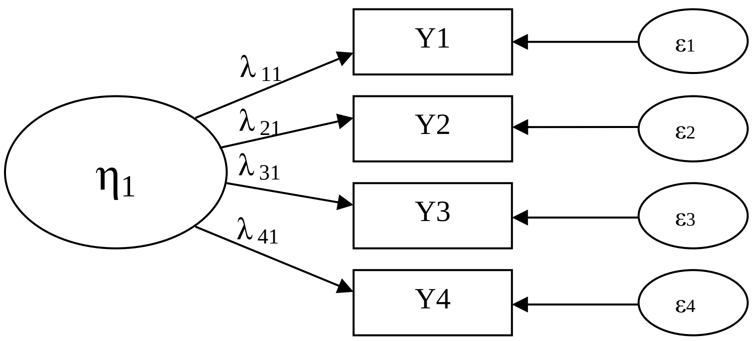

Figure 4.1. Path diagram of a one-factor CFA model.

For the simple factor solution depicted in Figure 4.1, there is a single latent factor (η_1) and four indicators (Y1-4). Typically three or four indicators are recommended per latent factor (Brown, 2006). This is because if there was only one indicator it

[9] It should be noted that SEM conforms to the conventions of matrix algebra. As such, matrices are represented by uppercase Greek letters, such as Λ (*lambda*) and Ψ (*psi*); and specific elements of these matrices are symbolised using lowercase Greek letters, e.g. λ, ψ and ε (see Brown, 2006).

would be assumed that the latent factor perfectly measured the observed variable, which is rarely the case due to the confounding effects of error variance as previously outlined (Schumacker & Lomax, 2010). In an example CFA model of a questionnaire, Y1-4 represent four questionnaire items that are comprised partly of the latent construct (e.g. political attitudes, η_1) and partly by uniqueness/error variance ($\varepsilon_{1\text{-}4}$). The factor loadings (λ) that result from such an analysis are interpreted to signify how strongly each indicator loads onto the latent factor; that is, the degree to which the variable represents the common factor. Substantive and significant factor loadings imply the indicators adequately measured the latent construct. The model presented in Figure 4.1 can also be described in a mathematical format to explain the regression functions portrayed. The regression equations can be summarised by the following four separate equations:

$$
\begin{aligned}
Y_1 &= \lambda_{11}\eta_1 + \varepsilon_1 \\
Y_2 &= \lambda_{22}\eta_1 + \varepsilon_2 \\
Y_3 &= \lambda_{33}\eta_1 + \varepsilon_3 \\
Y_4 &= \lambda_{44}\eta_1 + \varepsilon_4
\end{aligned}
\tag{4.1}
$$

The set of equations depicted above can be further condensed into a single equation which describes the relationships among the observed variables (y), latent factors (η) and unique variance (ε) as measured by the factor loadings (Λ) [10]:

$$
y = \Lambda_y \eta + \varepsilon \tag{4.2}
$$

[10] Uppercase lambda is used here as this equation refers to the full correlation matrix.

With some minor variations, these fundamental equations can be used to calculate various aspects of the sample data from the factor analysis parameter estimates, including variances, covariances and means of the input indicators (Brown, 2006).

The Process of Model Specification and Estimation in CFA

The common factor model provides the foundation for Confirmatory Factor Analysis (CFA; Jöreskog, 1969). CFA is a theoretically driven analytical process, whereby models are specified to examine particular relationships or hypotheses about the data. The first step in CFA is model specification, the process of formally setting out the model to be estimated (Hoyle, 1995). During this process, any relationships that are to be examined between the variables are clearly mapped out, such as how many common factors are needed, which indicators will load on which factors and whether the factors are correlated or not (Thompson, 2004). To do this, a statement regarding a set of parameters that depict the nature of the relationships between two variables is created (Hoyle, 1995). Parameters are typically specified to be either fixed, normally at a value of zero, or freely estimated using the variance-covariance input matrix[11]. Due to these specification requirements, a strong empirical or conceptual foundation is essential to guide model specification (Brown, 2006).

Once a theoretically-grounded CFA model is proposed it is then compared against the collected dataset to determine if the relationships among the variables in the proposed model adequately describe the input data (Saltin & Strand, 1995). In order to achieve this, a mathematical operation aimed at minimising the difference between the predicted and the observed variance-covariance matrices is applied, which entails

[11] A covariance matrix is very similar to a correlation matrix except it is unstandardised; that is, covariances are measured in the units of the original variables (Guarino, 2005).

a fitting function or estimation method. There are many methods that can be used to estimate the common factor model; including principle factors analyses, or weighted and unweighted least square analyses. However, the most widely used fitting function for CFA is the maximum likelihood (ML) procedure (Brown, 2006)[12]. The main aim of ML estimation is to find the model parameter estimates that maximise the likelihood of potentially replicating the data were the data to be collected from the same population again. This process allows for a statistical evaluation, using the goodness-of-fit indices (described in Chapter Five), of how well the pattern of fixed and free parameters specified in the factor solution reproduces the pattern of variances and covariances seen in the observed data (Hoyle, 1995). The preferred factor solutions are generally the most meaningful and yet statistically simplistic solutions (Harman, 1960).

When a CFA analytic program arrives on a set of parameter estimates that adequately reflect the observed relationships and cannot be improved upon ‘convergence’ is achieved (Brown, 2006). When convergence occurs it implies the model has been run successfully. On occasions, ML estimations can fail to converge on a final set of parameter estimates that adequately reflect the data. This results in an “improper solution”, such as “Heywood cases”. In these situations there may be an out of range estimate, such as an indicator with a loading above 1.0, or a negative error variance (Brown, 2006). The presence of improper solutions can be indicative of a poorly specified model or due to instabilities with the testing instrument.

[12] An extension of ML estimation is the maximum likelihood method (MLM) for estimation. MLM is a very similar fitting function as ML, however the model chi-square and standard errors of the parameter estimates are corrected for non-normality within large samples (Brown, 2006).

Summary

This section has introduced the common factor model, which is the underlying framework for conceptualising CFA models. It was demonstrated that CFA models can be specified to examine particular relationships between constructs or to answer detailed research enquiries. The precision and flexibility with which CFA performs such estimations establishes the utility of SEM techniques for the assessment of various applied research enquiries. In the following section, prominent applications for SEM techniques in the assessment of reliability, construct validity and substantive hypothesis testing will be described. It will be evident that SEM can provide a sophisticated and thorough approach to the psychometric investigation of measurement instruments, such as implicit attitude techniques.

Applications for SEM Techniques in Reliability Estimation, Construct Validation and Substantive Hypothesis Testing

The present section demonstrates how various research questions concerning the psychometric properties of measurement instruments can be examined using SEM analyses. The primary focus of this section is the application of CFA to construct validation, however, reliability estimation, equivalency assessments and the influences of covariates on latent constructs will also be outlined. It is argued that SEM, and particularly CFA, provide a valuable avenue through which the influence of error variance on implicit attitude scores could be examined. It is concluded CFA analytical techniques appear suitable to facilitate a thorough psychometric review of implicit attitude measures.

Application 1: Testing for Internal Construct Validity using Single-group CFA

Single-group CFA is one of the most basic applications of SEM, yet it can be used to estimate critical psychometric information for a measurement device. Single-group CFA can determine whether significant amounts of random error are in the data, as well as if the measure accurately assesses the hypothesised trait construct (Schumacker & Lomax, 2010). To examine these enquiries, a single-group CFA model is specified, as per the model depicted in Figure 4.1. Presuming the model is deemed a good fit for the data (using the fit indices described in Chapter Five) the factor loadings for the latent trait and error components are examined. Factor loadings greater than .32 are typically deemed a minimum standard for acceptability in applied psychological research because it indicates there is at least 10 percent shared loading between the variable and the factor (Gorsuch, 1983). However, from a statistical point of view the choice of threshold for a meaningful loading is often arbitrary and higher factor loadings around .5 or .6, are preferred (Gorsuch, 1983).

The factor loadings of the indicators onto the latent attitude factor (represented by $\lambda_{11\text{-}41}$ in Figure 4.1) show whether the latent trait construct is significantly and substantively assessed by the observed scores. Good internal construct validity for a measure would be demonstrated if all indicators were significant and greater than .32, although much higher factor loadings would provide stronger evidence of good construct validity (Gorsuch, 1983). Good internal construct validity for the political questionnaire depicted in Figure 4.1 would imply the items (Y1-4) were all consistently adequate measures of the same underlying construct of political preferences. Furthermore, significant and substantive factor loadings of the indicators onto the error factors (ε_{1-4}) imply that random error variance comprises a significant portion of the observed scores.

For the implicit attitude measures, single-group CFA could be applied to each IAT or APT individually to determine whether the implicit attitude scores provide an adequate and consistent estimate of the latent implicit attitude construct being investigated. If this were the case, it would reveal whether each IAT or APT possessed adequate internal construct validity, a vital prerequisite for any measure. Single-group CFA could also determine whether significant proportions of random error variance comprise the implicit attitude scores, as hypothesised. This is a key advantage of CFA for the current dissertation.

Application 2: Estimating Reliability using Composite Reliability and Average Variance Extracted

The other crucial prerequisite for a task, reliability, can also be estimated using the same one-factor single-group CFA model outlined above. Reliability is the degree to which a test consistently measures that which it measures (Nunnally, 1978). Reliability is often assessed by examining how well all the items of a test relate to each other, an estimate referred to as internal consistency. A popular internal consistency estimator is Cronbach's (1951) coefficient alpha (α), which rates greater inter-correlations as indicative of more consistency amongst test items and thus better stability/reliability for the test. Despite being widely used in behavioural and social research for more than 60 years, Cronbach's (1951) coefficient alpha has been shown to provide a sub-optimal indicator of reliability due to not accounting for error variance, as well as issues of under- and over-representation (Novick & Lewis, 1967;

Raykov, 1997; Zimmerman, 1972). Unless all the scale items have equivalent factor loadings; a model type referred to as tau-equivalent (Graham, 2006), coefficient alpha has been found to underestimate composite reliability at the population level by quite substantial amounts at times (Novick & Lewis, 1967; Raykov, 1997). Such inaccuracies have led to the conclusion that coefficient alpha cannot be considered a dependable estimator of measure reliability (Raykov, 1997). Rather, it has been argued that reliability estimation be reported using Composite Reliability (CR) and Average Variance Extracted (AVE) instead (Fornell & Larcker, 1981). CR and AVE can both be easily applied to single-group CFA. This matches the stronger reliability estimate afforded by CR and AVE with the statistical rigour of CFA to deliver a significant advantage over Cronbach's alpha.

In SEM, Composite Reliability provides an estimate of the internal consistency of a task by assessing the extent to which a set of indicators share in the measurement of a latent construct (Hair, Black, Babin, Anderson, & Tatham, 2006). CR estimates thus differ from Cronbach's alpha estimates by examining the reliability of the latent construct, after random error variance is removed, rather than examining the reliability of the individual test items. CR delivers an estimate similar to the reliability of the summated scale and will typically reveal stronger reliability estimates than Cronbach's α, unless items are tau-equivalent (Raykov, 1997). Fornell and Larcker (1981) propose CR be calculated using the formula depicted in Equation 4.3. Adequate reliability for a measure is revealed if CR estimates are greater than .60 (Tseng, Dörnyei, & Schmitt, 2006) or .70 (Hair et al., 2006).

$$\rho_\eta = \frac{(\sum \lambda_i)^2}{(\sum \lambda_i)^2 + \sum Var\,(\varepsilon_i)} \tag{4.3}$$

where ρ_η is the composite reliability,
λ_i is the factor loading *i*,
$Var(\varepsilon_i)$ is error variance for the factor loading *i*

Average Variance Extracted (AVE) compliments Composite Reliability (CR) by providing an estimate of how much variance within the indicators is explained by the common factor (Hair et al., 2006). In other words, AVE measures how much of the trait construct is accounted for by the attitude scores. This is essentially a test of internal convergent validity. AVE can be calculated using the formula depicted in Equation 4.4, similarly specified by Fornell and Larcker (1981). AVE results represent the ratio of total variance due to the latent variable and can vary between 0 and 1. AVE values greater than .50 are considered satisfactory because they indicate that at least 50% of the variance in a measure is due to the hypothesised underlying trait (Bagozzi, 1991; Dillon & Goldstein, 1984; Hair et al., 2006; Tseng et al., 2006). This score is thus particularly important as it indicates what proportion of trait versus error variance is accounted for by the observed scores. A result of greater than .50 implies good validity for both the construct and the individual variables, revealing acceptable internal convergent validity for the measure.

$$\rho_{vc\,(\eta)} = \frac{\sum \lambda_i^2}{\sum \lambda_i^2 + \sum Var\,(\varepsilon_i)} \tag{4.4}$$

where $\rho_{vc\,(\eta)}$ is the average variance extracted,
λ_i is the factor loading *i*,
$Var(\varepsilon_i)$ is error variance for the factor loading *i*

Together, CR and AVE provide a psychometrically robust estimate of the internal consistency and internal convergent validity of a task, such as the IAT or APT. Given the concerns outlined in Chapter Three regarding the reliability of these implicit attitude measures, such statistical processes provide a very crucial application for CFA in the psychometric evaluation of these tasks. CR and AVE can be applied to assess whether the IAT and APT provide a consistent and adequate measure of the implicit attitude constructs of interest. The AVE assessment can also be used to estimate what proportion of random error variance is confounding the implicit attitudinal data.

Application 3: Testing for Construct Validity using Single-group CFA

The one-factor single-group CFA models described above provide a critical foundation for reliability and validity estimation of implicit attitude measures. However, the hypothesis-testing capabilities of CFA position it well for addressing more complex construct validity estimation, such as the provision of strong convergent and discriminant validity evidence (Brown, 2006). As described in Chapter Three, there is scarce convergent validity evidence in support of implicit attitudinal measures. This may be due, at least in part, to not applying appropriate statistical processes that account for the confounding influence of error variance. CFA delivers such an estimate by enabling the simultaneous modelling of data from more than one task, using a model with multiple factors (such as the three-factor single-group CFA model presented in Figure 4.2). Using such a model, convergent validity is evident if different indicators of theoretically similar or overlapping constructs are strongly interrelated. For example, if two implicit attitude measures, such as the APT and IAT, loaded onto a single latent factor, implicit attitudes, the assertion that the two tasks were measuring a similar underlying construct would be supported. Discriminant validity, on the other hand, is demonstrated if the indicators of theoretically distinct constructs are not highly inter-correlated. For example, discriminant validity of implicit and explicit attitude types would be supported if implicitly measured attitudes and explicitly measured attitudes loaded onto separate factors, and these factors were not correlated strongly enough to imply that a broader construct had been incorrectly separated (Brown, 2006). Discriminant validity would also be evident if the inter-implicit correlation was significantly stronger than the implicit-explicit correlation.

To expand on the aforementioned example, Figure 4.2 presents a pictorial representation of a three-factor CFA model. In this case the three factors refer to two implicit attitude measures, the APT and IAT, and an explicit attitude questionnaire of the same construct. To support the convergent validity of the two implicit attitude measures it would be expected the correlation between them (‘*a*’ in Figure 4.2) would be strong and significantly larger than the correlation between either of these two tasks and the explicit attitude questionnaire (‘*b*’ in Figure 4.2). If this discrepancy was observed the discriminant validity of the implicit and explicit attitude measures would be supported, reinforcing the theoretically proposed distinction between these attitude types (Gawronski & Bodenhausen, 2006). This application has the potential to provide critical convergent and discriminant validation for the attitude measures.

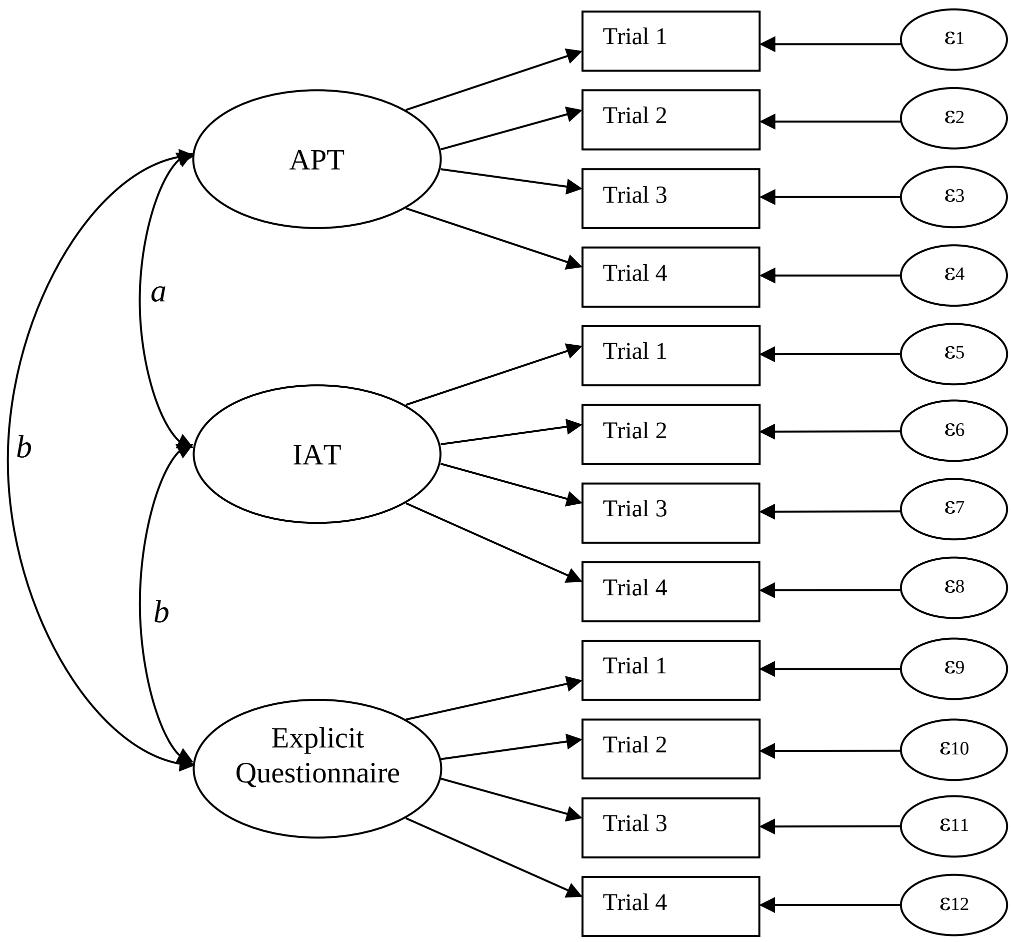

Figure 4.2. Three-factor CFA model to assess the convergent and discriminant validity of implicit and explicit attitude measures.

Application 4: Testing for Construct Validity using Higher-Order Factor Analysis

Convergent and discriminant validity evidence can be further obtained using higher-order CFA modelling. All CFA models presented thus far have been first-order models, where only one level (a first order) of latent factors was involved. However, in some instances there are reasonable theoretical grounds to assume that the first-order factors can be at least partially explained by some higher-order factor structure (Schumacker & Lomax, 2010). Probably the most well-known example of such a structure is in the field of intelligence. Individual first-order factors such as verbal comprehension, perceptual reasoning and working memory, are generally considered to be explained by a higher-order factor of general intelligence or '*g*'. In the case of implicit attitudinal measurement, the APT and IAT are both theoretically considered to be examining the same latent implicit attitudinal factor. Thus the APT and IAT could be considered first-order factors, explained by a higher-order implicit attitude factor. The process for testing such hypotheses is referred to as higher-order factor analysis.

In a standard CFA model with two or more factors, the factors are generally specified to be inter-correlated. This means it is assumed there are relations between these factors, but any specific information about the nature of these relationships is unknown. Higher-order CFA models provide a theory driven approach to examining these between-factor inter-correlations. The aim of higher-order factor analysis is to create a more parsimonious account of the correlations among the first-order factors (Brown, 2006). Because a more refined structure is applied onto the first-order model, the higher-order model results in a more theoretically comprehensive solution than is able to be produced using a standard first-order CFA. To illustrate, a higher-order CFA model suitable for assessing general intelligence is presented in Figure 4.3. In this model the first-order CFA model is comprised of three latent factors: Verbal Comprehension, Perceptual Reasoning and Working Memory. The second-order model incorporates the higher-order latent factor of General Intelligence[13].

[13] In a higher-order CFA there are additional error variances attached to the first-order latent factors that indicate the amount of variance left unexplained by the higher-order factor (see Figure 4.3).

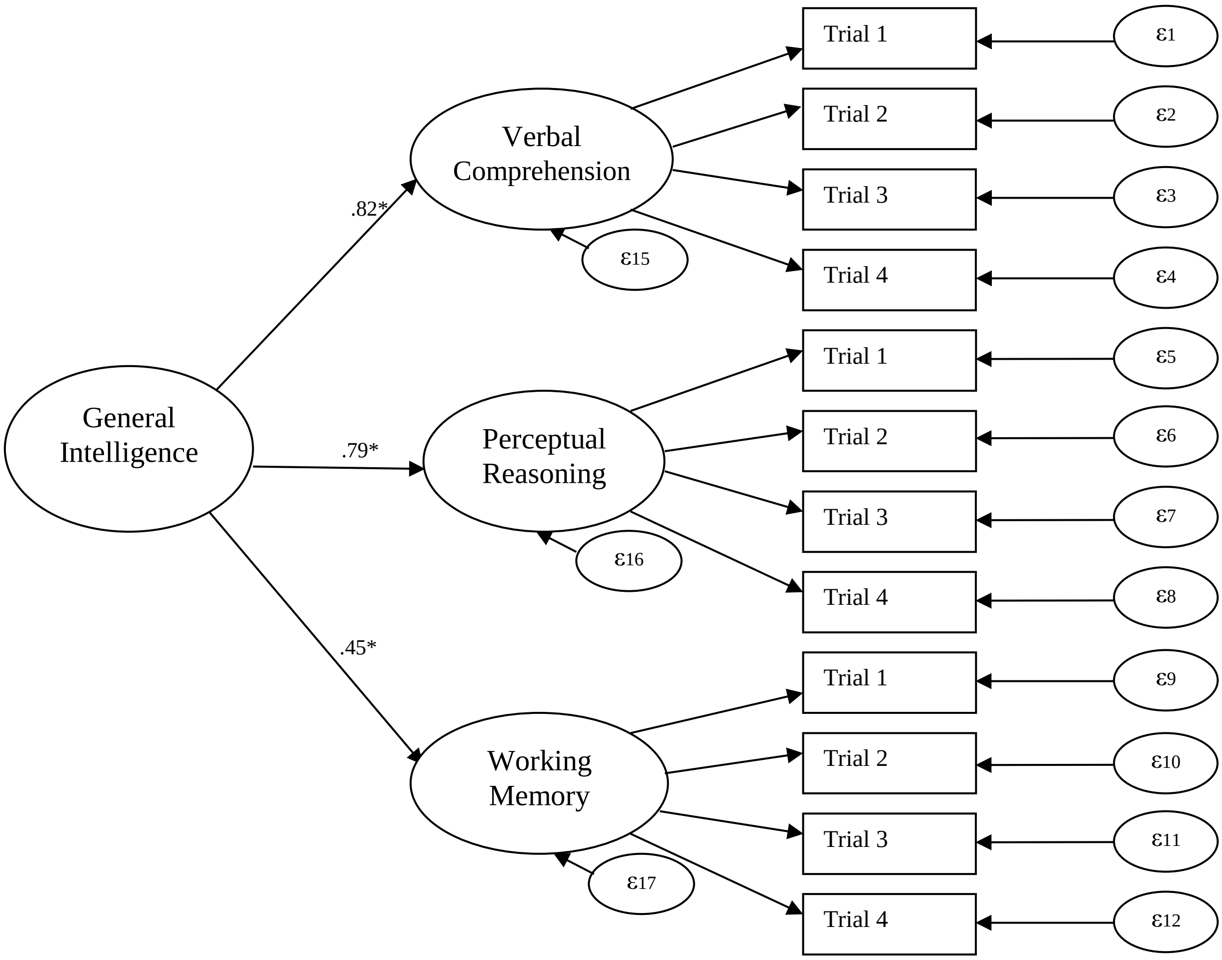

Figure 4.3. Example Higher-order CFA model of General Intelligence based on the Verbal Comprehension, Perceptual Reasoning and Working Memory sub-tests. *p<*.05.*

The structural equations that result from the analysis provide an indication of the strength of the relationship between the first- and second-order factors (Schumacker & Lomax, 2010). In this way, it is possible to tell which first-order factors provide the strongest measure of the second-order factor. For example, in Figure 4.3 Verbal Comprehension has a factor loading of .82 on the higher-order factor, which means it is the strongest measure of General Intelligence, followed by Perceptual Reasoning (.79) and then Working Memory (.45), with all three being statistically significant. In other words, all three tests assessed a substantive level of General Intelligence, but the Verbal Comprehension sub-test provided the most comprehensive measure of this construct[14].

The application of higher-order CFA to implicit attitude measures would enable a comparison of various implicit techniques to gauge which task provides the strongest assessment of the implicit attitude (higher-order factor) being examined. For instance, an APT and IAT of the same attitude construct could be compared using higher-order CFA to determine which task delivers the strongest measure of the implicit attitudinal construct, or indeed, whether the measures are comparable. Such a comparison would deliver vital information regarding the relative strengths of different attitudinal instruments for measuring implicit attitudes. The analysis could also provide convergent validity evidence for the measures, adding to the construct validity evidence of these techniques.

To summarise, higher-order CFA provides one avenue for assessing the construct validity of implicit attitude measures free from the confounding influence of random error variance, whilst enabling clear comparisons between like measures. However, further more sophisticated applications of SEM allow for the examination and quantification of not only random error variance, but also systematic types of variance associated with specific methodologies used in data collection. Such a sophisticated application of CFA delivers an even stronger assessment of construct validity than possible using single-group CFA alone.

---

[14] The factor loadings described above are created to illustrate a point and are not a true reflection on the various contributions of these sub-types on general intelligence. This example was designed purely to illustrate the utility of higher-order CFA models for increasing understanding of the relationships between first-order latent factors.

Application 5: Construct Validation of a Measuring Instrument using the Multitrait-Multimethod Approach

As introduced in Chapter Three, the Multitrait-Multimethod (MTMM) approach provides one avenue for estimating systematic forms of error variance, such as method effects, which are likely to significantly influence implicit attitudinal scores. Method effects form a systematic or consistent type of bias, regardless of what construct is being assessed, and can therefore impact upon the observed data in unknown ways (Meade, Watson, & Kroustalis, 2007; Podsakoff et al., 2003). Although standard CFA is able to partial random error variance, the method effects remain confounded with the trait construct during analysis and thereby continue to influence the results. In their classic article, Campbell and Fiske (1959) outlined a strategy for using Multitrait-Multimethod (MTMM) matrices to account for method variance in addition to random error variance when evaluating the construct validity of psychological measures. This approach enables a clear estimate of random error and systematic error variance in the data. The MTMM framework has since become a popular and critical tool for construct validation (Lance et al., 2002).

The MTMM design initially proposed by Campbell and Fiske (1959) involved directly examining elements of the correlation matrix among the trait by method measurements. This approach was cumbersome, lacked statistical significance testing and often resulted in confusion (Coenders & Saris, 2000; Gignac, 2009; Schmitt & Stults, 1986). By the early seventies, MTMM matrices began being analysed using structural equation models (Brown, 2006). These models provided more accurate reliability and validity estimates by partitioning the observed scores into trait, error and method components. The MTMM methodology can be applied using different frameworks within SEM, such as CFA, correlated uniqueness and the true score approach (Coenders & Saris, 2000). Of these, the CFA framework is particularly well suited to the application of MTMM data and there is strong support for its use in applied behavioural research (Brown, 2006; Cole & Maxwell, 2003; Marsh & Grayson, 1995). It has been argued that the enterprise of validity has reached a pinnacle with the CFA-MTMM strategy due to the way it can be applied to incorporate all forms of quantitative validity research (Gignac, 2009).

The CFA-MTMM approach requires multiple traits (or constructs) to be assessed by multiple measures (Schumacker & Lomax, 2010). This is so each observed variable loads onto one trait factor and one method factor. The method factor accounts for error covariances or systematic method effects (Coenders & Saris, 2000), which can be used to determine the amount of method variance present in the observed data. Given implicit attitude measures may be quite susceptible to systematic error, as outlined in Chapter Three, the CFA-MTMM approach could deliver a more accurate estimate of construct validity for the APT and IAT by accounting for both systematic and random error influences. As with single-group CFA, both convergent and discriminant aspects of construct validity can be assessed. In CFA-MTMM convergent validity is evidenced when measures of the same trait correlate highly even when they were assessed using different methods (Schumacker & Lomax, 2010). Discriminant validity is obtained when correlations between measures of different traits using the same method are low (Nosek & Smyth, 2007). By comparing the fit of structural models, the relative merits of different hypotheses about the structure of trait and method variance can be systematically tested (Nosek

& Smyth, 2007). Model specification is vital for facilitating this hypothesis-testing capability of CFA-MTMM.

Prominent Specification Approaches for CFA-MTMM Models

CFA-MTMM models are typically specified using one of two approaches, the correlated methods approach and the correlated uniqueness approach (Marsh & Grayson, 1995). The correlated methods approach estimates the latent method factor by examining the shared variance in the indicators for a particular method. This approach corresponds directly to the original conceptualisation of MTMM matrices by Campbell and Fiske (1959) whereby each indicator is composed of trait, method and error (or unique) variance. The correlated uniqueness approach obtains a measure of method variance by investigating the correlations between error variances for each method type rather than by examining latent method factors (Brown, 2006). The strongest advantage of the correlated uniqueness approach is that it rarely results in improper solutions. The correlated method model, while being more prone to improper solutions, allows for an evaluation of the relationship between method factors; something which is not possible using the correlated uniqueness approach (Brown, 2006). Given the current study is keenly interested in the relationship between methodologies (i.e. the IAT and APT); the correlated methods approach appears the more appropriate.

Several techniques have been proposed for applying the correlated methods approach. The correlated trait-correlated method (CTCM) enables free estimation between trait factors or method factors (Marsh & Grayson, 2005). This approach differs from the correlated trait-uncorrelated method (CTUM) that constrains correlations among method factors to be zero (Marsh & Grayson, 1995). The correlated trait-correlated method minus one model (CT-C(M-1)) uses one method less than the methods assessed, so that one method provides a 'standard' from which the others are compared (Eid, et al., 2003). Further approaches include latent difference modelling (Pohl, Steyer & Kraus, 2008) and latent means modelling strategies that examine the mean method effect rather than the impact of specific methodologies (Pohl & Steyer, 2010). Whilst all of these modelling approaches have certain advantages and disadvantages, the freely correlated trait-freely correlated method (CT-CM) has been applied most extensively to construct validation assessments (Lance et al., 2002) and may aid comparing different implicit attitude measures. Typical CT-CM CFA-MTMM analyses involve at least two methods that assess at least two traits, with correlations between trait factors or between method factors freely estimated. Free estimation between latent factors allows for a direct estimate of the relationship between these factors. This is a key advantage of the CT-CM CFA-MTMM approach as it means that two methods, such as the APT and IAT, can be directly compared without the confounding influence of trait or random error variance. Such a comparison is not possible using alternate methods such as CT-C(M-1)."

An example of how CFA-MTMM could be used to assess construct validity for the implicit attitude measures is presented in the CFA-MTMM path diagram of Figure 4.4. Two attitude constructs, racism and political preferences (the traits), were each measured by two implicit attitude measures, an APT and an IAT (the methods). Thus four measures were involved, each consisting of four sets of trials. Data from the trial sets (the indicators) loaded onto one trait and one method factor. Using the CT-CM

CFA-MTMM approach, method variance and random error variance are partialled by examining the shared variance of specific groups of indicators. To enable such an analysis, correlations among the trait and method factors were freely estimated so as to determine the relationships between these variables. Correlations between the trait and method factors were fixed to zero as there should be no relationship between those latent factors. The error variances were freely estimated but were restrained so as not to correlate with other error variances (see Figure 4.4 for the resulting model). Because the relationship between the two method factors is freely estimated it can provide an indication regarding how strongly related the APT and IAT methodologies are. CFA-MTMM could also be applied to the verbal and pictorial adaptations of the IAT to assess the convergent validity of these measures, which could greatly increase the psychometric support currently available for these tasks.

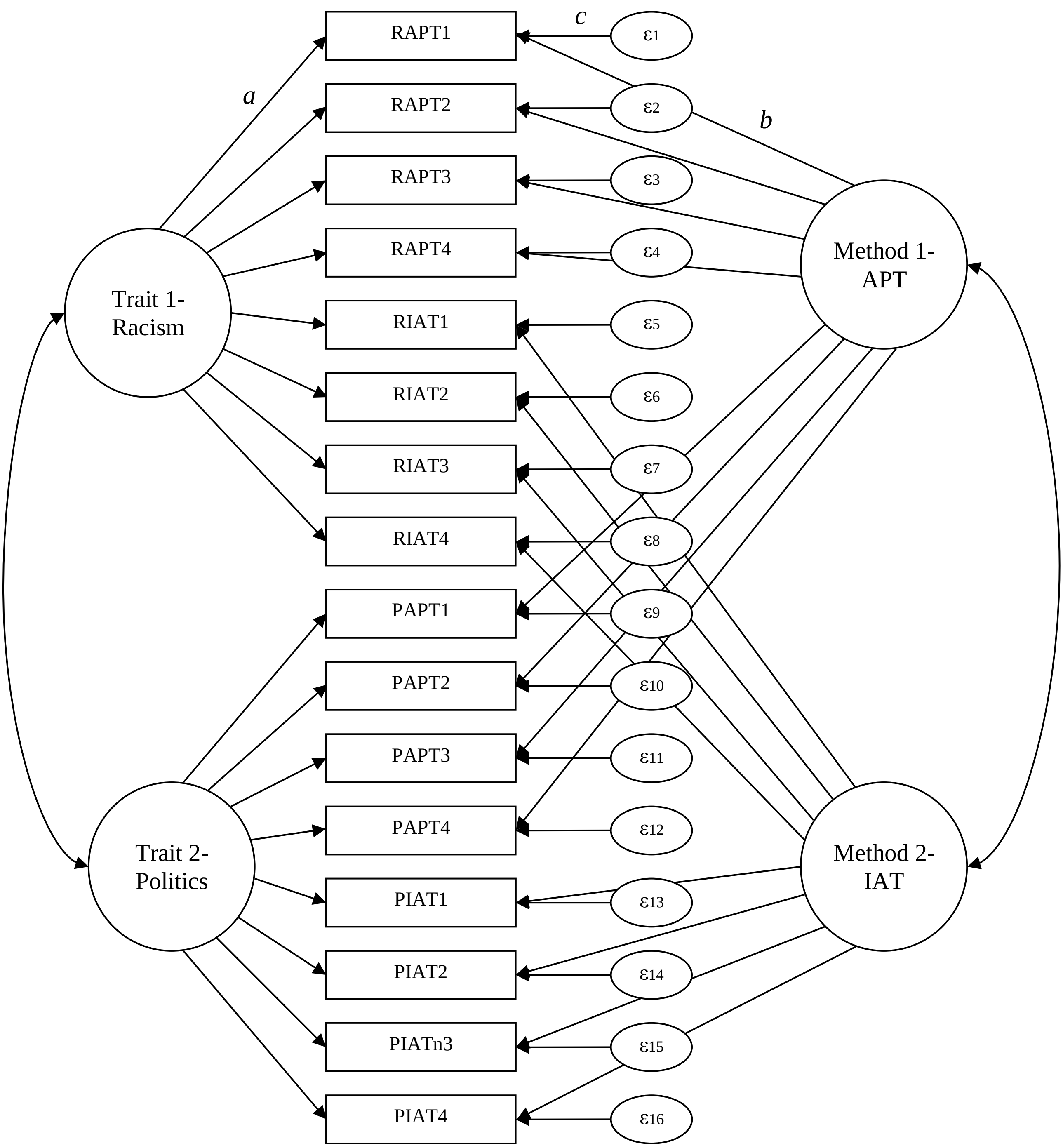

Figure 4.4. A CT-CM CFA-MTMM model depicting two traits (Race and Politics) and two methods (APT and IAT).

Estimating Construct Validity by Examining Individual Parameters

In a CFA-MTMM analysis, scrutinising the individual parameters (i.e. the factor loadings) can deliver a stringent assessment of the construct validity of measurement techniques (Byrne, 1998). In Figure 4.4, the individual parameters relating to the traits are represented by the pathways under "*a*", the parameters relating to the methods are represented by the paths under the "*b*" and the paths under the "*c*" represent the random error variances. When examining these individual parameters, adequate construct validity for a task is demonstrated if the proportion of trait variance is greater than method variance (Byrne, 1998). The significance of the variances, or the squared standardised loadings, are also taken into consideration. As such, for Figure 4.4 a^2 is the trait variance for the first data parcel RAPT1, b^2 is the method variance for that same indicator and c is the error variance. If overall $a^2 > b^2$ it would provide support for the construct validity of the measurement instrument.

This comparison of trait and method variances could provide an exacting assessment of the construct validity of implicit attitude measures that is over and above what is possible by using CFA alone. This is because the estimate of trait variance is no longer confounded by method-related variance. In addition, the individual parameters could provide an indication of what proportion of the implicit attitudinal scores is assessing the trait construct, and likewise what proportion is 'assessing' random and systematic types of error. The CFA-MTMM analysis is thus critical for confirming or otherwise the hypothesised impact of error variance on implicit attitudinal tasks. In summary, examining the construct validity of implicit attitude measures using CFA-MTMM enables random and systematic error variance to be accounted for, resulting in a more accurate estimate of the psychometric properties of the APT and IAT.

Application 6: Testing for Invariant Factorial and Latent Mean Structures using Multiple-groups CFA

SEM techniques have been shown to provide a psychometrically rigorous avenue for assessing the construct validity of measurement instruments in the case of individual groups of participants. But often researchers are interested in comparisons between more than one participant groups. Researchers may want to determine if there are significant differences between Group A and Group B on a particular measure; or whether sex, culture, age or any other variable impacted the way the measures were completed. These substantive enquiries may be very applicable to the implicit attitudinal measures, for if such measures are deemed to be psychometrically adequate then it would be important to ascertain what information about implicit attitudes have been uncovered in a way that still accounts for error variance. Multiple-groups CFA may provide one means for achieving this. Usually, multiple-groups CFA is used to simultaneously test data from two separate participant groups. However, the methodology could also be applied to determine if there was a significant difference between the latent means of the congruent and incongruent 'groups' of data, even if they are sourced from the same participant sample. Applied this way, multiple-groups CFA would provide a way of ascertaining whether an IAT effect had occurred for the sample whilst accounting for the confounding influence of error variance. However, before the average reaction times (or levels of a trait) for two different 'groups' can be compared, it is first crucial to assess whether a score of X by one group is equivalent to a score of X for the other group. If the trait scores are not comparable across groups then differences between groups in mean levels are potentially artifactual and may be substantively misleading (Reise, Widaman, & Pugh, 1993). This process of determining equivalency across groups is referred to as testing measurement invariance, which aims to ensure that a particular measure is operating in the same way for different groups of people (Burns & Haynes, 2006). For the IAT, a test of measurement invariance would be required to ascertain whether an average reaction time of X for the congruent trials is equivalent to an average reaction time of X for the incongruent trials.

Multiple-groups CFA provides a strong analytic framework for evaluating invariance across distinct groups (Reise et al., 1993). By running simultaneous CFAs for two or more groups, the measurement and structural parameters of the comparative models are tested such that any group differences between the latent models are revealed (Brown, 2006). Multiple-groups CFA requires that constraints are applied to the models, such as like parameters, in a step by step process. This process enables a detailed examination of the equivalency of the measurement (measurement invariance) and structural (population heterogeneity) solutions (Brown, 2006). By applying this thorough approach to invariance testing it is possible to ascertain whether the factor structure, the structural equivalence, the intercepts, the error variance, the variance (or standard deviation) and the latent mean scores are the same for each group (Burns, Gomez, Hafetz, & Walsh, 2006). These tests provide a direct contrast of the aforementioned aspects of the groups in order to determine comparability, or indeed, significant differences between the groups. Multiple-groups CFA require two separate input matrices and the analyses follow a specific procedural format.

To begin the assessment of measurement invariance, the factor structure of Group A and Group B's models are examined and compared. For the IAT, Group A could refer to the congruent trial data, and Group B the incongruent trial data. The initial assessment of factor structure for these ‘groups’ is referred to as equal form invariance (or configural invariance) and it means the number of factors and the pattern of indicator-factor loadings is identical across groups (Brown, 2006). This is the least restricted model which subsequent models are evaluated against using the nested chi-square (χ^2) difference test, a measure of the relative difference in chi-square value for two nested models. After establishing equal form (configural) invariance, the next series of analyses entail increasingly restrictive constraints. Factor loading equality is initially examined (metric invariance), then equality of intercepts (scalar invariance or strong factorial invariance) and finally the equality of the error variances (residual invariance or strict factorial invariance) (Brown, 2006). These four separate analyses comprise the assessment of measurement invariance.

Once measurement invariance is established, multiple-groups CFA can be applied to test the equality of latent means. This is the key aspect of multiple-groups CFA of interest for the IAT, as the analysis could reveal whether an IAT effect was present or not by showing whether there is a significant discrepancy in the latent means for the congruent and incongruent data. If the latent means for the congruent trials were found to be significantly smaller than the latent mean for the incongruent trials it would imply a positive IAT effect had occurred, that the expected attitudinal bias was revealed. The comparison of latent means is somewhat analogous to the comparison of observed group means completed using a *t*-test or ANOVA (Brown, 2006). However, unlike these traditional ways of determining whether an IAT effect has occurred, the comparison of latent means accounts for error variance, thereby providing a more accurate indication of the implicit biases evidenced.

Group comparisons of latent means are only meaningful if the factor loadings (metric) and indicator intercepts (scalar) have been shown to be invariant. As such, the comparison of latent means cannot occur until after the analyses of measurement invariance are completed. In order to compare two latent means, the latent mean of one model (the congruent trial data) is constrained to be zero, whereas the second latent mean (for the incongruent trial data) is allowed to be freely estimated. If this produces a significant latent mean for the second 'incongruent' group it indicates there is a significant difference between the means of the two groups, i.e. that an IAT effect was present for the sample. Figure 4.5 provides a conceptual model of this assessment of latent means using multiple-groups CFA. For a two group comparison, the difference between latent group means is equal to the latent mean for the second group and the sign of the second group's latent mean provides a guide as to which mean was higher (Thompson & Green, 2006). For the example IAT data in Figure 4.5, a significant and positive result indicates the average reaction time for the incongruent trials was significantly slower than the congruent trials, demonstrating the expected IAT effect.

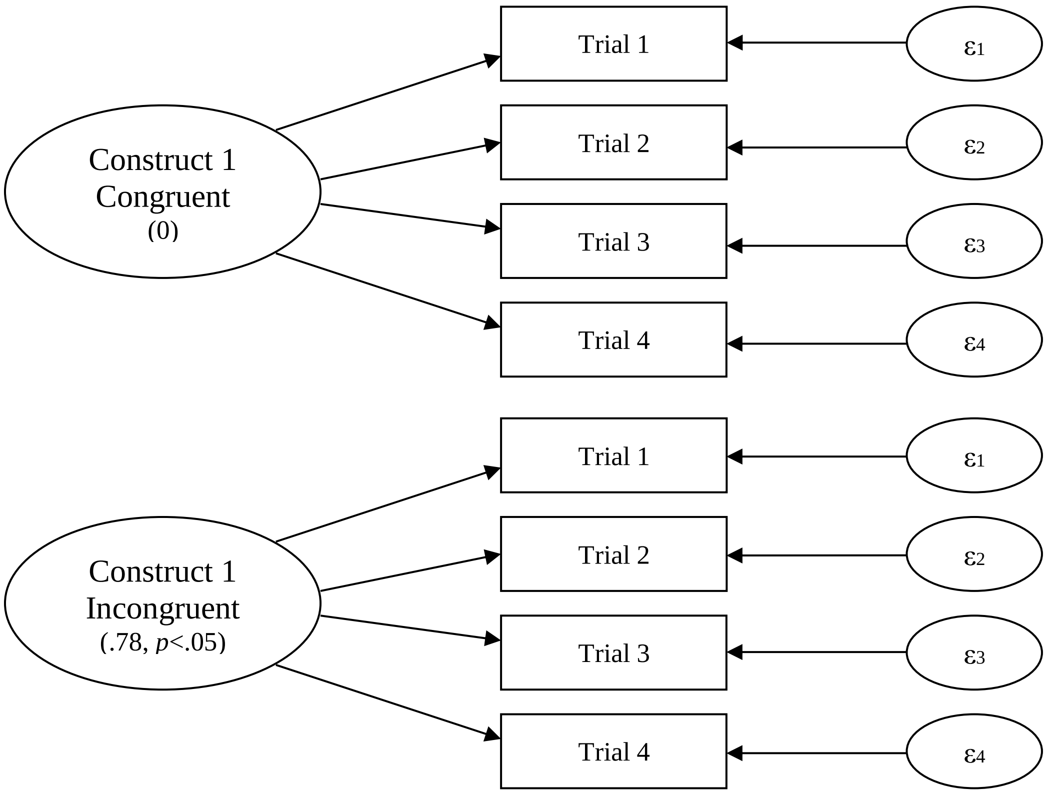

Figure 4.5. Multiple-groups CFA model assessing the IAT effect for Construct 1.

As outlined above, Multiple-groups CFA could provide one avenue to compare the congruent and incongruent IAT results to determine if an IAT effect had occurred. This approach would avoid the confounding influence of error variance on the results, potentially delivering a more accurate impression of the implicit biases of the sample. A similar approach would also be appropriate for the APT data.

Application 7: Testing for the Effects of Covariates on the Latent Factor Structure using MIMIC Models

The final application of SEM covered in this chapter provides another avenue to assess group difference, but is a considerably simpler approach than multiple-groups CFA (Thompson & Green, 2006). Multiple Indicators, Multiple Causes (MIMIC)

Models only requires a single dataset and can be used to assess the impacts of participant's characteristics such as age, sex, or political preferences, on the latent factor. As such, MIMIC models could be used to determine whether specific participant characteristics are related to greater amounts of implicit prejudice as measured by the IAT or APT.

MIMIC models apply the full general SEM model, which is comprised of both structural and measurement components (see Figure 4.6; see also Hoyle, 1995). The measurement model is the component of the general model concerned with the latent variables, the part utilised by CFA. The structural model depicts relationships between the latent variables and other observed variables that are not the indicators, also referred to as covariates (Hoyle, 1995). MIMIC models thus add covariates to the CFA model to examine their direct effects on the latent factors and selected indicators (Jöreskog & Goldberger, 1975). This process enables both factors and indicators to be regressed onto observed covariates representative of group membership, such as sex, age, socio-economic status or cultural grouping. Put simply, the MIMIC model is theoretically (and practically) similar to adding a regression model onto a CFA model (see Figure 4.6).

Unlike multiple-groups CFA, MIMIC models only require a single input matrix. This one dataset contains the variances and covariances of the latent variable indicators as well as covariates that denote group membership (Brown, 2006). The use of only one dataset allows for the simultaneous testing of many covariates with relative ease, which is a significant advantage for the MIMIC approach over multiple-groups CFA. The two basic steps of MIMIC modelling are to first establish a viable CFA measurement model using the full sample, then secondly add the covariates to the model to determine their direct effects on the latent factors and selected indicators (Brown, 2006). A significant direct effect of a covariate onto a latent factor signifies the factor means are different at different levels of the covariate. This is similar to the test of equal latent means in a multiple-groups CFA. Group mean differences are generally presented on the latent models as parameter estimates of the direct effects, which can provide information regarding the size of the discrepancy between the groups (Brown, 2006).

In a MIMIC model the covariate is ordinarily assumed to be free of error variance. This is a reasonable assumption given the covariate often represents known groups (such as males versus females). These covariates are typically depicted by nominal variables that denote a category level of the groups (e.g. Sex: 0 = Male, 1 = Female). However, MIMIC models are also able to accommodate continuous predictors, which pose another potential advantage over multiple groups CFA. The basic MIMIC model presented in Figure 4.6 depicts Sex as the covariate directly affecting the two latent factors. This section of the model is referred to as the structural part of the MIMIC model, where the observed variable, Sex, is being used to predict the two latent Factors (Schumacker & Lomax, 2010). The MIMIC model diagram also indicates the prediction error for the two latent factors ($\varepsilon_{9\text{-}10}$). The part of the MIMIC model showing the two factors, the four observed indicator variables and associated error variances is the measurement part of the model, which defines the latent variable (Schumacker & Lomax, 2010). Using this MIMIC model, if the direct effect between Sex and Factor 1, for example, was significant it would indicate there was a significant difference between the male and female participants for that factor.

Applying MIMIC modelling to implicit attitudinal research would allow for an investigation into the effects of sex, age or other factors onto the implicit attitudes examined using the APT or IAT. In this way it can be seen whether males possess greater implicit biases than females, or likewise for any other relevant characteristics. MIMIC models thus enable such substantive enquiries for implicit attitude measures.

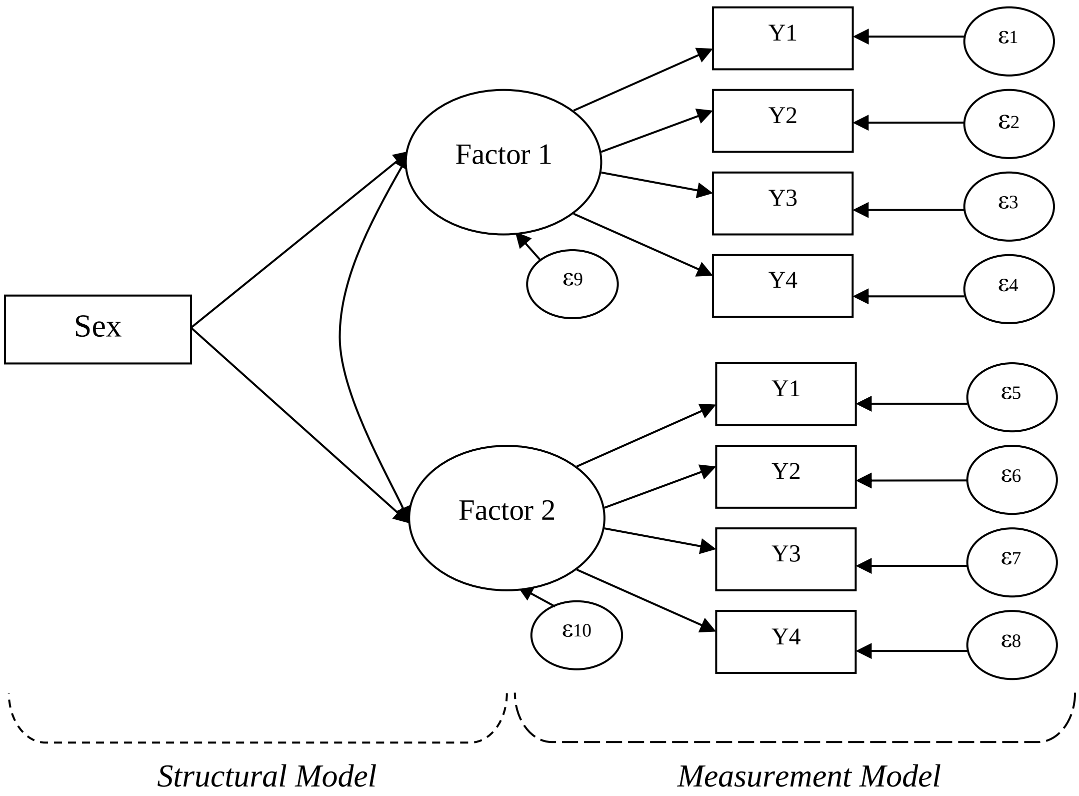

Figure 4.6. MIMIC model depicting a Sex covariate onto a two-factor model. The structural and measurement models inherent in the diagram have been highlighted.

Chapter Summary and Conclusion

SEM techniques can be applied to answer a wide range of important research questions. The prominent strength of these approaches is that they provide a sophisticated and flexible way to assess many aspects of the construct validity of measuring instruments, whilst simultaneously modelling the effects of error variance. Single-group CFA was shown to be suitable for estimating the internal construct validity of implicit attitude measures. Reliability estimations were also possible using this model when combined with CR and AVE. The construct validity of the IAT and APT may be suitably examined using single-group CFA, higher-order CFA and CFA-MTMM models. These analyses could deliver a thorough assessment of the construct validity of implicit attitude measures whilst simultaneously estimating the amount of random error variance, and in the case of CFA-MTMM, method variance that is confounding implicit data. Provided adequate psychometric support for the tasks is evidenced, multiple-groups CFA and MIMIC models can facilitate an investigation into substantive enquiries, such as whether the tasks have revealed implicit prejudice (i.e. whether an IAT effect was evident) as well as if other factors such as the age or sex of the participants influenced implicit attitudinal biases.

In conclusion, SEM analytical strategies provide a multitude of approaches for critically assessing the psychometric properties of implicit attitude measures. Because these techniques examine latent rather than observed scores, SEM approaches are likely to provide a far more psychometrically rigorous approach to examining the reliability and construct validity of implicit attitudinal techniques than has traditionally been obtained.

In the following three chapters, three empirical studies assess the psychometric properties of the IAT and APT using SEM analytical approaches. Chapter Five investigates various aspects of the reliability and construct validity of the IAT and APT whilst accounting for random error variance using single-group CFA (with CR and AVE) as well as higher-groups CFA. Chapter Six will expand on this by examining the construct validity of several IATs using CFA-MTMM to account for method effects in addition to random error variance. Chapter Seven will explore measurement invariance within the IAT using multiple-group CFA and the influence of participant characteristics on the latent implicit attitudes using MIMIC models. It will be demonstrated that SEM provides a sophisticated and thorough approach to reliability estimation and construct validation of implicit attitude measures.

CHAPTER FIVE

Study One: Reliability Estimation and Construct Validation of Implicit Attitude Measures using SEM Techniques

Implicit attitude measures, such as the Affective Priming Technique (APT; Fazio et al., 1986) and the Implicit Association Test (IAT; Greenwald et al., 1998), provide a novel and potentially useful approach for assessing socially sensitive attitudes like racial prejudice (see Chapter Two; also Nosek et al., 2005). However, as noted in Chapter Three, empirical inconsistencies and poor psychometric properties have resulted in concerns as to whether implicit attitude techniques are suitably consistent and accurate measures of implicit attitudes. High error variance has likely been a significant contributor to implicit attitudinal scores (see Chapter Three), which has not previously been taken into account when estimating the psychometric properties of implicit measures. The present study aims to rectify this by applying Structural Equation Modelling (SEM) techniques to examine various aspects of the reliability and construct validity of the IAT and APT.

This chapter outlines the first of three empirical studies, which addresses four of the seven aims of the present research. Study One assesses reliability, internal construct validity, convergent validity and discriminant validity evidence for the affective priming task (APT; Fazio et al., 1986), verbal Implicit Association Test (VIAT; Greenwald et al., 1998) and pictorial Implicit Association Test (PIAT; Thomas et al., 2007) using Confirmatory Factor Analysis (CFA) models. Mixed support for the reliability and validity of the implicit attitude measures is found, with high error variance evident for all the implicit attitude tasks. More consistent psychometric evaluation of measurement instruments is recommended so that the unique advantages of implicit attitude measures can potentially be maintained, but within the confines of more rigorous and psychometrically robust empirical research.

Implicit Attitudinal Measures; A Unique Approach to Prejudice Assessment

As discussed in Chapter One, implicit attitudinal measures provide an avenue through which new insights into our cognitions can be gleaned. Implicit attitudes measures differ from explicit attitudes measures by not requiring conscious intention or deliberation (Nosek et al., 2006). Rather, implicit attitude techniques rely on speeded categorisation tasks from which implicit attitudes are inferred. This methodology has proven quite difficult to falsely respond to due to, predominantly because of the indirect nature of the measure (making it less clear what is being assessed) and the requirement of speeded response latencies that are hard to consistently control responses on, especially over hundreds of trials (Asendorf et al., 2002; Egloff et al., 2005; Steffens, 2004). As such, implicit attitude measures have proven reasonably superior to explicit attitude measures for the assessment of sensitive attitudes that participants may be unwilling to openly share (Greenwald et al., 2009). Racial prejudice is a key example of such controversial topics.

Racial prejudice remains one of the most sensitive topics investigated in the behavioural sciences. Despite Australia’s current legislation that espouses multicultural ideals (Schweitzer et al., 2005), there is a well document history of racial intolerance within this country (Dunn et al., 2004; Islam & Jahjah, 2001). However, racial intolerance is not evenly distributed across all racial groups.

Following the well-publicised destruction of New York's twin towers by Muslim extremists in 2001 and the subsequent "War on Terror" instigated by the United States of America at the Middle East, substantial amounts of 'Islamophobia' have been reported both overseas (Agerström & Rooth, 2009; Chopra, 2008; Gibson, 2008; Nosek et al., 2007; Park et al., 2007; Rashid, 2009) as well as here in Australia (Dunn et al., 2004; Dunn et al., 2008; Islam & Jahjah, 2001). Notably, in Australia Arabs have been perceived as the most threatening racial group (Islam & Jahjah, 2001) and anti-Muslim concern is rated as over twice as high as the next most concerning out-groups, namely Black Africans and Indigenous Australians[15] (Dunn et al., 2008). Given the strong evidence of antipathy towards persons from the Middle East (see also Agerström & Rooth, 2009; Chopra, 2008; Dunn et al., 2004; Dunn et al., 2008; Gibson, 2008; Park et al., 2007; Rooth, 2010), attitudes towards Arabs and the Middle East (versus Europe and Europeans) will be investigated in the present study.

Error Variance in Implicit Attitude Measures

Implicit attitude measures, although potentially very valuable for examining racial biases, have proven to deliver inconsistent results (see Chapter Three; see also Bosson et al., 2000; Krause et al., 2010; Rudolph et al., 2008; Sherman et al., 2003). It is strongly suspected that random error variance is substantially contributing to such inconsistencies, as outlined in Chapter Three. Random error variance acts as a confounding factor, hindering the ability of any measure to accurately estimate underlying attitudes (Rudolph et al., 2008). This is because trait variance is proportional to error variance, such that the greater the error variance the less trait

[15] The irony of indigenous Australians being rated as a group not belonging to Australia is not lost on the present researchers.

variance that can be present in the measure (Cote & Buckley, 1988). High amounts of random error variance in a dataset can thus significantly confound validity estimates and may even act as an upper limit for convergent validity estimations (Cole & Maxwell, 2003; Cunningham et al., 2001; Podsakoff et al., 2003).

Podsakoff et al. (2003) demonstrated the impact of random and systematic types of error variance on the observed correlation between measures of different constructs. Podsakoff et al. (2003) applied the mathematical formula and average trait and error variances that were produced by Cote and Buckley (1987). For example, attitude-related measures were found to be comprised of 30% trait variance, 30% random error variance and 40% method variance (Cote & Buckley, 1987). When these variance estimates were inserted into the formula devised to calculate an average observed correlation, it was revealed that two perfectly correlated attitude constructs (1.00) were limited to an observed correlation of .52 following the incorporation of such error variance (Podsakoff et al., 2003). This is a Type II error, a false negative, and is concerning as it shows that even if two traits are perfectly correlated, typical levels of error variance reduce the observed correlation by half, and thus the variance explained by 70% (Podsakoff et al., 2003). These results support the assertion by Cunningham et al. (2001) that error variance has been providing an upper limit for psychometric estimates of implicit attitude measures. If it was demonstrated that high random error variance was present in the implicit attitude measurement techniques it would assist in explaining the poor inter-implicit correlations previously reported in the implicit attitude literature (see Chapter Three; also Hofmann & Schmitt, 2008; Krause et al., 2010; Rudolph et al., 2008). Thus error variance may have limited efforts to obtain satisfactory psychometric support for implicit attitude measures.

Systematic Analysis of Random Error Variance using SEM

Cunningham et al. (2001) argued for the need to systematically evaluate and account for error variance in implicit attitudinal research. Structural Equation Modelling (SEM) techniques provide one robust method to systematically account for error variance. As explained in Chapters Three and Four, CFA partials out random error variance from the implicit data thereby reducing the interference caused by unknown amounts of random measurement error. This enables more accurate assessment of the proportions of trait variance being assessed, and thus a stronger estimate of reliability and construct validity for measurement instruments. Methods such as Composite Reliability (CR) and Average Variance Extracted (AVE) can provide a measure of reliability and internal convergent validity that are not confounded by error variance, delivering a more accurate estimate of internal consistency than that afforded by traditional methods like Cronbach's Alpha (Novick & Lewis, 1967; Raykov, 1997; Zimmerman, 1972).

Confirmatory Factor Analysis (CFA; Jöreskog, 1969) is a sophisticated latent modelling analytical technique that enables construct validity estimation. Using this technique, variance associated with the construct is identified because it is shared across the items (or trials in an IAT or APT), demonstrating a commonality of variance known as the common factor. The residual variance is designated as random error variance (or unique variance). By partitioning random error variance, CFA can provide strong evidence for both convergent and discriminant validity, which are the two key sub-types of construct validity (Brown, 2006). Convergent validity is often evidenced by strong correlations (r>.50; Cohen, 1992) between conceptually similar latent constructs (Schumacker & Lomax, 2010). It is also demonstrated when two measures of the same construct (such as the IAT and the APT) are highly correlated. Discriminant validity is evidenced when indicators of theoretically distinct constructs are not highly inter-correlated (*r*<.30; Cohen, 1992), such as would be demonstrated if implicit-explicit correlations were shown to be significantly lower than inter-implicit correlations. CFA can thus provide a sophisticated assessment of construct validity for measurement instruments.

In summary, implicit attitude techniques reveal the potential for great utility in the assessment of socially contentious issues. However, psychometric evidence for the tasks has been inadequate to support their use thus far. Random error variance is believed to have been providing an upper limit for psychometric estimates of implicit attitude techniques. The present study aims to systematically examine the extent and influence of random error variance in the APT and IAT whilst assessing the socially contentious topic of anti-Arab prejudice. A thorough examination of the reliability and construct validity of these implicit attitude techniques will now be presented.

Study One

The overall aim of the present study was to apply SEM techniques, such as CFA, to examine the reliability and construct validity of Fazio et al.'s (1986) APT, Greenwald et al.'s (1998) VIAT and Thomas et al.'s (2007) PIAT. Two versions of each of these tasks were developed, one that examined implicit preferences for Europeans over Arabs (the racial attitude construct), the other that examined attitudes towards countries within certain regions of the globe, namely the Middle East and Europe (the country attitude construct). The six tasks were: a Racial APT, Country APT, Racial VIAT, Country VIAT, Racial PIAT, and Country PIAT. The series of analyses investigated the internal consistency, internal convergent validity and internal construct validity of the APT, VIAT and PIAT, as well as the convergent and discriminant validity of each of these implicit attitudinal measures.

Aim 1: Assess the Reliability of the APT, VIAT and PIAT using Composite Reliability and Average Variance Extracted

The first aim was to estimate the internal consistency and internal convergent validity of the implicit attitude measures using Composite Reliability (CR) and Average Variance Extracted (AVE). Adequate reliability was hypothesised for all the implicit attitude measures (CR>.70 and AVE>.50; Hair et al., 2006)

Aim 2: Internal Construct Validity of the APT, VIAT and PIAT using Single-group CFA

The second aim was to apply single-group CFA to individually examine the internal construct validity of each of the six implicit attitude measures. These tasks were analysed separately using the CFA model illustrated in Figure 5.1, which depicts the latent attitude construct, the observed attitude scores/ indicators (data Parcels 1-4) and the error variances ($\varepsilon_{1\text{-}4}$). It was hypothesised that all indicators would load substantively (>.32; Gorsuch, 1983) onto the latent factor, revealing good internal construct validity. Furthermore, it was expected substantial and significant error variance would also be present for all data parcels, supporting the claim of high random error in implicit attitudinal measurement.

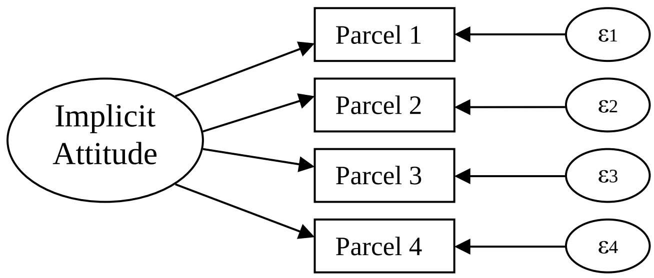

Figure 5.1. CFA path model for assessing the internal construct validity of each of the implicit attitude measures.

Aim 3: Convergent and Discriminant Validity of the Implicit Attitude Measures using Single-group CFA

The third aim was to apply single-group CFA to assess the convergent and discriminant validity of the implicit attitude measures. For each of the two attitude constructs a VIAT, PIAT, APT and explicit questionnaire were compared. The attitude constructs examined Racial preferences (towards Arabs versus Europeans) or Country preferences (towards the Middle East versus Europe). Convergent and discriminant validity evidence was assessed using the four-factor CFA model presented in Figure 5.2. A strong inter-implicit correlation was expected between the latent implicit attitude constructs (r>.50; Cohen, 1992), revealing convergence amongst the implicit tasks. A positive but weak relationship was expected between the implicit and explicit attitude measures (r<.30; Cohen, 1992), supporting the discriminant validity of implicit and explicit attitudinal constructs (see Nosek, 2007).

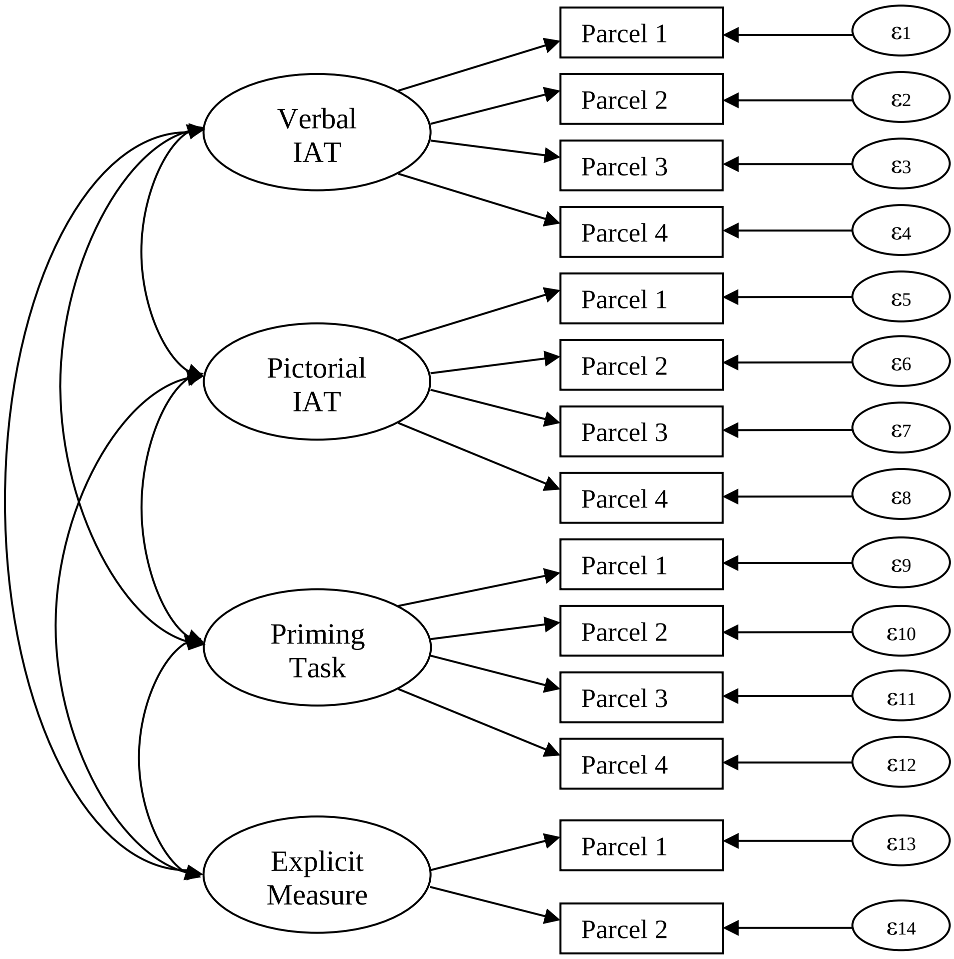

Figure 5.2. CFA path model specified to test the convergent and discriminant validity of the implicit and explicit attitude measures for a single attitude construct.

Aim 4: Convergent and Discriminant Validity for Implicit Attitude Measures using Higher-order CFA

The fourth aim examined whether the latent implicit attitude techniques all loaded significantly and substantially onto a single higher-order factor, revealing whether the tasks accessed the same underlying implicit attitudinal construct. It was expected the implicit attitude measures would load much more strongly on the implicit higher-order factor than the explicit attitude measure would, thus supporting the discriminant validity of these constructs. The specified higher-order CFA model is presented in Figure 5.3, it is an extension of the latent correlation model seen in Figure 5.2. The higher-order model enables an estimate of the relative strength of each measure in assessing the underlying construct.

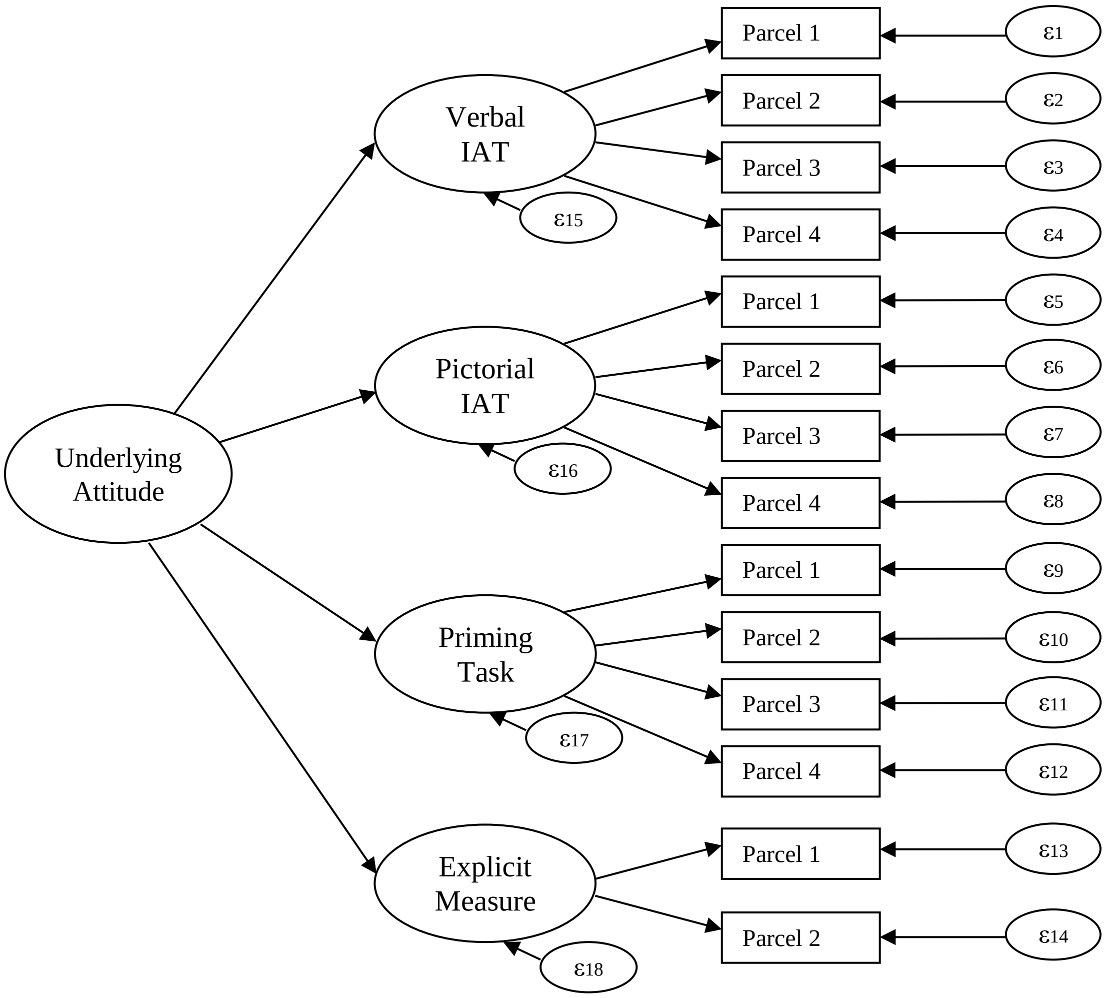

Figure 5.3. Higher-order CFA model specified to provide further convergent and discriminant validity evidence for the implicit and explicit attitude measures.

Method

Participants

Responses of 205 first year psychology students from the University of Tasmania and persons known to the investigator were collected for ten tasks. Of these participants, the data of three were used for piloting purposes. Of the remaining 202 participants, four were excluded due to missing data or because more than 10% of trials for a single IAT had latencies of less than 300ms, implying the participant did not adequately attend to the stimuli (in accordance with Greenwald, Nosek and Banaji, 2003). Consequently, the final dataset contained 198 participants. One hundred and forty-four participants were female (Mean age = 25.08 years, SD=10.86) and 54 were male (Mean age = 29.11yrs, *SD*=11.70). The overall mean age of participants was 26.03 years (*SD*=11.10). The self-identified ethnic make-up of the sample was as follows: 174 Australians, ten Europeans, eight Asians, three North Americans, two persons from the Middle East and one African.

Apparatus

Participants completed six IATs, two APTs, two explicit questionnaires and answered a few demographic questions regarding their age, ethnicity and how well travelled they believed they were. Each of these tasks is expanded upon below.

Implicit Association Tests (IAT; Greenwald et al., 1998)

IATs are computer-based programs that require attribute- and category-type stimuli be simultaneously allocated into their respective categories as quickly as possible. For the current study, each IAT was presented individually on a laptop computer running the Inquisit software package (Millisecond Software, 1996).

Stimuli categorisation required pressing the designated key (“A” for the left hand and “L” for the right hand) on the computer keyboard using the participant’s left- or right-hand forefinger. Stimuli response logos were situated to the top left and top right of the display to act as memory prompts regarding which key was designated for each category type. These prompts remained on-screen throughout each block of stimuli. The stimuli were centred on the display, remaining on-screen until the participant’s response. Stimuli were either black text on a white background (for the VIATs) or a picture 3.5cm square in size presented on a blue background (for the PIATs). The stimuli could be broadly categorised as either attribute- or category-related. The attribute-related stimuli were words or pictures that could be categorised as Pleasant or Unpleasant. The category-type stimuli were words or pictures associated with the Middle East or Europe. Stimuli were randomly presented and there were ten exemplars for each stimuli category. The inter-trial interval between stimuli response and presentation of the next stimulus was 250ms. Figure 5.4 presents an example stimuli presentation sequence for the attribute component involving two stimuli trials on a verbal IAT (VIAT).

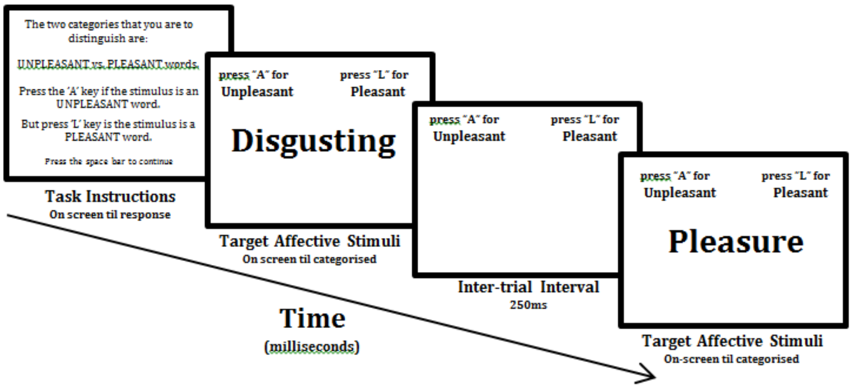

Figure 5.4. Example presentation sequence for the attribute component of a VIAT.

All IATs are comprised of four main components: attribute-related stimuli categorisation, category-related stimuli categorisation, congruent stimuli categorisation and incongruent stimuli categorisation. The first of these components, attribute-related stimuli categorisation, involved categorisation of just the attribute stimuli. If the stimuli exemplar was Pleasant it was categorised using the right-hand key, if the stimuli was Unpleasant it was categorised using the left-hand key. This step allowed participants to learn the Pleasant and Unpleasant exemplar stimuli.

The second component, category-related stimuli categorisation, was similarly simple, requiring categorisation of just the category-related stimuli. These stimuli, such as, Middle Eastern and European country names, were sorted into their respective categories. If the stimuli exemplar was European it was categorised using the right-hand key; if the stimuli was Middle Eastern it was categorised using the left-hand key. This step allowed participants to learn the categories of stimuli they need sort.

The remaining two components were more complicated as they required dual-categorisation of both attribute- and category-related stimuli simultaneously. The third component consisted of congruent stimuli categorisation involving simultaneous sorting of four sets of stimuli that were presented in congruent combinations such as European and Pleasant words/pictures versus Middle Eastern countries and Unpleasant words/pictures. An example of such stimulus organisation is shown in Figure 5.5a. If, for example the stimuli exemplar was European *or* Pleasant it was categorised using the right-hand key; if the stimuli was Middle Eastern *or* Unpleasant it was categorised using the left-hand key. The congruent component is typically deemed relatively easy as the two stimuli that share a response are more intuitively associated for most participants than is the case for the incongruent component (Hummert et al., 2002).

The fourth component consisted of incongruent stimuli categorisation involving simultaneous sorting of four sets of stimuli with incongruent combinations, such as Middle Eastern countries and Pleasant words/pictures versus European countries and Unpleasant words/pictures. An example of such stimulus organisation is shown in Figure 5.5b. If, for example the stimuli exemplar was Middle Eastern *or* Pleasant it was categorised using the right-hand key; if the stimuli was European *or* Unpleasant it was categorised using the left-hand key.

Figure 5.5. Screen set-up for the Country PIAT, with a Middle Eastern category exemplar present in the centre of the display. Panel a) illustrates the congruent response pairings; panel b) the incongruent response requirements.

These components were arranged into seven steps per IAT as shown in Table 5.1. Step five was a category-related stimuli categorisation task; however for this task the required key for categorising the Middle Eastern and European stimuli had been transposed from what was required in the first task. As such, if the European stimuli were formerly categorised using the right hand key, the left hand key was now required and vice versa for the Middle Eastern stimuli. This was a training step and prepared participants for the incongruent trials that would follow. Each of the dual-categorisation tasks were completed twice consecutively.

Table 5.1

The Seven Procedural Steps of a Typical Implicit Association Test

Step 1	Learn Attribute Dimension Unpleasant vs. Pleasant words
Step 2	Learn Category Dimension Middle East vs. Europe country names
Step 3	Congruent Dual-Categorisation Task * Unpleasant and Middle East countries vs. Pleasant and European countries
Step 4	Congruent Dual-Categorisation Task * Unpleasant and Middle East countries vs. Pleasant and European countries
Step 5	Learn Transposed Category Responses Europe vs. Middle East country names
Step 6	Incongruent Dual-Categorisation Task * Unpleasant and Europe countries vs. Pleasant and Middle East countries
Step 7	Incongruent Dual-Categorisation Task * Unpleasant and Europe countries vs. Pleasant and Middle East countries

* Data from these steps are used for data analytic procedures.

For each of the experimental IATs, Steps 1, 2 and 5 consisted of 40 trials. Steps 3, 4, 6 and 7 consisted of 102 trials. Within each component all stimuli were randomly presented. Congruent and incongruent set presentation was counter-balanced throughout the sample, such that the presentation order was changed for each alternate participant. Thus half the participants experienced the congruent trials before the incongruent (i.e. completed Steps 1-7 in order), while the other half completed the incongruent trials first (i.e. completed the steps in the order, 1, 5, 6, 7, 2, 3, 4). The procedure outlined was the format for all IATs in this study.

Six different IATs were used in the present study; one verbal (VIAT) and one pictorial (PIAT) version of each of three attitude constructs: Flowers-Insects, Racial attitudes (Europeans versus Arabs) and Country attitudes (Europe versus Middle East). These IATs are described individually below.

Flower-Insect VIAT.

The Flower-Insect VIAT (Greenwald et al., 1998) was employed as an introduction to verbal IATs, familiarising participants with the IAT task requirements. As such, results from this task were not used in further analysis. The attribute stimuli were ten Pleasant and Unpleasant word stimuli, such as “love” and “hatred”. The category stimuli were the names of Flowers and Insects (excluding butterflies and spiders), such as “daisy” and “cockroach”. Full verbal stimuli are listed in Appendix G. As this IAT was for task familiarisation, the number of trials at each step was reduced (see Table 5.1), such that steps 1, 2 and 5 consisted of 20 trials and steps 3, 4, 6 and 7 consisted of 40 trials.

Flower-Insect PIAT.

The Flower-Insect PIAT (Thomas et al., 2007) was used as an introductory task to pictorial IATs, to familiarise participants with Pictorial IAT requirements. Again, results of this task were not used in further analysis. The attribute-related stimuli depicted Positive and Negative facial icons or ‘emoticons’ as shown in Figure 5.6 (see also Appendix A). The category related stimuli were images of Flowers and Insects (excluding butterflies and spiders; see Appendix B). The number of trials at each step was as for the Flower-Insect VIAT.

Figure 5.6. Exemplar Pictorial IAT attribute stimuli. A Negative attribute ‘emoticon’ is shown on the left, a Positive attribute ‘emotion’ shown on the right.

Racial VIAT.

The Racial VIAT was first used by Greenwald et al. (1998). This task used the same attribute word stimuli as the Flower-Insect VIAT. The category stimuli were typically Middle Eastern and European first names, such as “Habib” and “Lucy”. There were five female names and five male names for both racial groups. Full verbal stimuli lists are presented in Appendix G.

Racial PIAT.

The Racial PIAT is similar to the classic verbal task outlined above, except the word stimuli were replaced by graphical stimuli. The category-related stimuli depicted faces of Arabic and European people instead of the traditional word name stimuli (see Figure 5.7). The attribute-related stimuli were the same Positive and Negative facial icons or ‘emoticons’ used in the Flower-Insect PIAT (see Figure 5.6). There were four male and four female stimuli for each category group. Full stimuli are shown in Appendix C for the Arab facial stimuli and Appendix D for the European facial stimuli.

 

Figure 5.7. Exemplar Pictorial IAT category stimuli for the Racial PIAT. The left picture shows an exemplar Arab category picture, the right shows an exemplar European category picture.

Country VIAT.

The Country VIAT utilised the same Pleasant and Unpleasant words as the Flower-Insect and Racial VIATs. The category-related stimuli were names of countries located within the Middle East and Europe, such as “Iraq” and “Italy”. Full stimuli for the VIATs are listed in Appendix G.

Country PIAT.

The Country PIAT utilised the same attribute stimuli as the Flower-Insect and Racial PIATs (see Appendix A). However, instead of depicting pictures of people’s faces as category-related stimuli, the Country PIAT presented pictures of easily recognisable buildings from Middle Eastern and European countries (such as mosques and churches). There were eight stimuli for each category. Examples of the Country PIAT stimuli are shown in Figure 5.8. (See Appendix E for full Middle Eastern landmark stimuli, Appendix F for the European landmark stimuli).

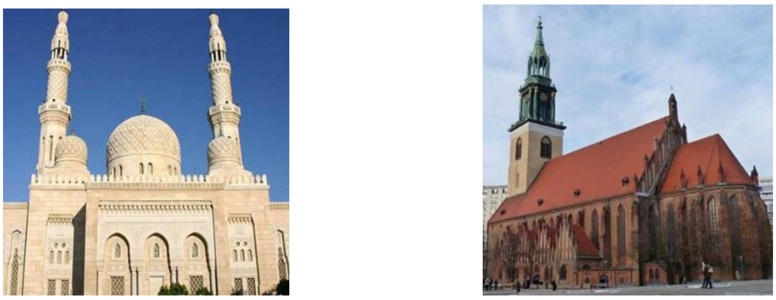

Figure 5.8. Exemplar Pictorial IAT category stimuli for the Country PIAT. The left picture shows an exemplar Middle East category picture, the right shows an exemplar European category picture.

Affective Priming Tasks (APT; Fazio et al., 1986)

APTs are computer-based programs that require speeded categorisation of Pleasant and Unpleasant stimuli following the brief presentation of a category-related prime. Two APTs were developed for the current study and were presented on a standard computer using the Inquisit software program (Millisecond Software, 1996). The target words were the Pleasant and Unpleasant attribute stimuli that required categorisation for the Racial and Country VIATs. The prime stimuli, which act as distractors, were the same as the Racial and Country VIAT category-related stimuli. As such, Arabic and European first names were the primes presented for the Racial APT, and Middle Eastern and European country names formed the prime stimuli set for the Country APT (see Appendix G).

The presentation sequence for the APTs was as follows. A blank screen appeared for 250ms, followed by a target symbol (a ‘+’) for 500ms to focus the participants’ attention. The screen would then go blank for 50ms, before the prime word (e.g. country name) appeared for 200ms. The screen then became blank for a further

50ms, before the target (affective) word stimuli appeared on screen and remained until a response was given using the appropriate key. Following a response, there was a 250ms inter-trial interval. The presentation sequence described is presented in Figure 5.9. The aim of the APTs was to categorise the second word (the attribute stimuli) as either Pleasant or Unpleasant as fast as possible without being distracted by the category-related prime. Categorisation of these target words required participants to press a key with their right hand for a Pleasant stimuli (“L” key) or with their left hand for Unpleasant stimuli (“A” key). There were 20 practice trials to train participants in the task, followed by four sets of 82 experimental trials, with a short rest break between each set of experimental trials. The time in milliseconds between stimuli presentation and response in the experimental trials provided the data for analysis.

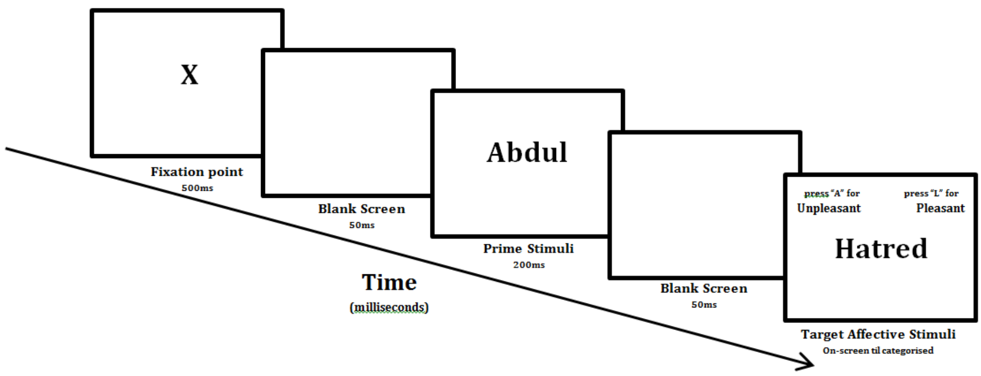

Figure 5.9. Presentation sequence for the Racial APT.

Explicit Attitude Questionnaires

Modern Racism Scale (MRS).

The explicit racial measure was the Modern Racism Scale (MRS; McConahay et al., 1981), which has been used extensively in social psychological research. The MRS consists of six items that make a statement regarding ‘Black’ people in the USA and participants rate their level of agreement on a five point scale. The MRS reports good reliability estimates (McConahay et al., 1981). However, the MRS in its original form was adjusted so as to be relevant for an Australian estimate of Anti-Arab prejudice. For instance, one question stated “Many Black people in Louisville and Jefferson County miss out on jobs or promotions because of racial discrimination.” This was altered to “Many Arabs living in Australia miss out on jobs or promotions because of racial discrimination”[16]. The six racially relevant items were located within a list of 20 questions that enquired of participants their opinion on a variety of other socially-sensitive issues: including cannabis use, suicide, homosexuality and global climate change. Questions 3, 5, 10, 12, 16, and 18 were the racially relevant items. The questionnaire provided to participants was titled ‘Student Opinions’. The précis warned participants that they were about to read some statements on a variety of issues, some of which they may agree with, others they may even find offensive. All statements were rated on a five point Likert scale (1-Strongly Disagree → 5-Strongly Agree). There was a Neutral option. The full version of this scale is provided in Appendix H.

[16] Altering the MRS in this way is certainly not unique. For instance, in one Australian study the questions were adapted to assess attitudes towards Aboriginal Australians (see Barlow, Louis, & Terry, 2010).

Travel Destination Questionnaire.

The other explicit attitude measure was a Travel Destination Questionnaire devised for the current research. This questionnaire comprised a list of 18 countries (six from Europe, six from the Middle East and six from Asia). Participants rated on a five point Likert scale how much they would like to visit each of the destinations (1-Definitely Not → 5-Definitely Would). The countries were selected from a list of tourism statistics by country (United Nations, 2009), with the top six most visited destinations for each area chosen (See Appendix I).

Subjective Measure of Travel Experience.

Participants were also asked to rate how well travelled they felt they were on a five point Likert scale (0-Not left Tasmania → 5-Well travelled; numerous countries, continents and experiences abroad). This simple measure aimed to allow for the confounding influence of travel experience on the outcomes of the other measures. The full measure is shown in Appendix J.

Procedure

Participants completed ten tasks individually at the School of Psychology, in Hobart, Launceston or North West campuses of the University of Tasmania, Australia. Participants asked questions or took short break between tests as required. Following consent procedures (see Appendix K for the Participant Information Sheet and Appendix L for the Consent Form); demographic data was obtained from the participant, including age, sex and ethnic identity, and was de-identified following the testing session. Participants also rated how well travelled they felt they were on the scale outlined above. Task procedures were explained to each participant.

Participants first completed the Flower-Insect VIAT and PIAT to gain a practical understanding and experience with the IAT procedure[17]. This data was excluded from all analyses. Following the training tasks, participants were presented with one of the following measures in a randomised and counterbalanced order: Racial VIAT, Racial PIAT, Country VIAT, Country PIAT, Racial APT, Country APT, Modern Racism Scale (MRS), or the Travel Destination Questionnaire (TDQ). Differences in stimulus presentation and response requirements were explained prior to the commencement of each task. On completion of the session the implicit associations underlying the different combinations of stimuli were fully explained to the participants and they were encouraged to ask any questions or discuss any issues that were raised during the testing procedure (see Appendix M for the Debrief Script). The ten tasks were generally completed in between one and a half and two hours.

Data Extraction and Scoring

IAT Scoring Procedure

The data produced by each IAT required substantial transformation using a procedure outlined below. Participants with any data missing or latencies less than 300ms for more than 10% of an IAT's trials were excluded (Greenwald et al., 2003). This resulted in four participants being omitted from all analyses. All practice trials were then removed (as per Nosek et al., 2006) along with any individual response latencies greater than 10,000 ms (Greenwald et al., 2003).

[17] Practice with the IAT format was important as substantial discrepancies between initial and second IAT experiences have often been reported. Whereas subsequent IATs typically produce more comparable reaction times (Greenwald et al., 2003; Greenwald et al., 2009).

In the original scoring recommendations of Greenwald et al. (1998) the first two trials of each block were also deleted, which is why the IATs had been developed with 102 trials per set (with each IAT comprised of four experimental sets). In their revised scoring protocols deleting the initial trials was no longer the suggested process (Greenwald et al., 2003). This issue required some consideration, particularly when contemplating how best to portion the data to accommodate the latent modelling analytical approach. Though the IAT scoring procedure is designed to produce one IAT effect score, an individual score is not ideal for SEM, as more than one score is required to estimate a latent factor. Although technically possible to create a score using the 204 paired congruent and incongruent reaction times for each IAT for each participant, the process was likely to be rather cumbersome and produce highly variable results. For these reasons, it was decided to parcel the IAT data into four equal parts to provide four sets of IAT scores that would produce the input data for four indicators. Data parcelling is a statistical process whereby scores from two or more items are averaged and then the parcelled scores replace the item scores in the SEM analyses (Bandalos, 2002). In order to parcel the IAT data, it was preferable for there to be only 100 data points per trial as opposed to 102 data points. As such, any deleted trials (i.e. those with response latencies above 10,000ms) were replaced with the response time of the second trial to avoid missing data. Then the first two trials for all sets were deleted.

Data parcels were created for each participant using the following procedure. Each participant generated 400 experimental response latencies for one IAT. Of these, 100 were from the first congruent block (congruent 1), 100 from the second congruent block (congruent 2), 100 from the first incongruent block (incongruent 1), and 100 from the second incongruent block (incongruent 2). Each of these blocks of 100 trials were split in four and combined to form four parcels each containing 25 response latencies from each block. Thus, each parcel contained 25 response latencies from each of the first congruent, second congruent, first incongruent and second incongruent blocks. The 25 latencies from each block created one data parcel, from which an IAT effect score was calculated (using the guidelines described in the next section). This resulted in four separate IAT effect scores for each participant (an IAT score A, IAT score B, IAT score C and IAT score D). All the participants IAT score A's were collated to provide the input data for one indicator, as were the IAT score B's for the second indicator, and so forth. These indicators are referred to as Parcels 1-4 in the models (for example, see Figure 5.1). This process was repeated for each of the IATs in the present study.

Within each data parcel, total congruent and incongruent scores were devised using the following revised scoring recommendations (Greenwald et al., 2003). Firstly, the means of correct latencies were calculated for each block separately. Then two pooled standard deviations were devised for all trials in congruent 1 and incongruent 1; and congruent 2 and incongruent 2, whereby congruent/incongruent 1 refers to the first block of congruent or incongruent trials and congruent/incongruent 2, the second block. Each incorrect latency in the dataset was replaced with the relevant block mean plus 600ms (a penalty) as per Greenwald et al. (2003). New means were then calculated for each of the blocks of trials. This process was repeated for all of the participants for each of the IATs. Table 5.2 presents the combined means and standard deviations by congruency for the four experimental IATs.

Table 5.2

Mean Latency and Standard Deviations by Construct Format and Congruency

IAT	Mean	*SD*
Racial VIAT Congruent	831.65	182.93
Racial VIAT Incongruent	1000.55	250.96
Racial PIAT Congruent	777.72	154.87
Racial PIAT Incongruent	829.88	180.36
Country VIAT Congruent	908.63	207.43
Country VIAT Incongruent	1044.21	251.17
Country PIAT Congruent	801.69	163.38
Country PIAT Incongruent	866.39	188.69

Note. *N* =198. VIAT=Verbal Implicit Association Test. PIAT=Pictorial Implicit Association Test.

After the congruent and incongruent means had been devised, the IAT effect score were produced as per Greenwald et al. (2003). The block means enabled two difference scores to be calculated: incongruent 1 – congruent 1; incongruent 2 – congruent 2. Each difference score was then divided by its pooled standard deviation. The resulting two quotients were then averaged. This produced an IAT effect score for one participant for one of their four data parcels within a single IAT. The process was then repeated for all the other data parcels and participants, for each of the four IATs. To clarify, Equation 5.1 depicts the formula used to develop an IAT effect score for a single data parcel.

$$\frac{\left[\dfrac{\mathit{MeanRTIncongruent1}-\mathit{MeanRTCongruent1}}{\mathit{SD}\ (\mathit{Incongruent1}+\mathit{Congruent1})} + \dfrac{\mathit{MeanRTIncongruent2}-\mathit{MeanRTCongruent2}}{\mathit{SD}\ (\mathit{Incongruent2}+\mathit{Congruent2})}\right]}{2} \quad (5.1)$$

The IAT scoring formula (see Equation 5.1) is very similar to Cohen's *d* (Cohen, 1992) and results in an IAT effect score referred to as a *D* score that reveals relative preference for the congruent over the incongruent stimuli pairings (Greenwald et al., 2003). Larger IAT effect scores are typically interpreted as indicative of greater negative implicit prejudice (Greenwald et al., 2003). The IAT effect scores provided the data for analysis.

Affective Priming Task Scoring

To calculate the scores for the APTs, all incorrect trials were deleted (Rydell & Gawronski, 2009). The data was then parcelled into four sets of data for further analysis in a very similar fashion to that described for the IAT. This resulted in four priming scores per participant for each APT. A priming score for an APT comparing two contrasting categories is calculated using the following formula depicted in Equation 5.2.

$$\text{Evaluation: } (P_Y\text{-}P_X) - (N_Y\text{-}N_X) \qquad (5.2)$$

This index captures the degree to which attitude X relative to Y yields faster responses for positive than negative targets (Wittenbrink, 2007). For the present study Y refers to the Middle Eastern stimuli and X refers to the European stimuli, see Equation 5.3.

$$\text{Evaluation: } (P_{\text{Middle East}}\text{-}P_{\text{Europe}}) - (N_{\text{Middle East}}\text{-}N_{\text{Europe}}) \qquad (5.3)$$

Consequently, higher scores indicate that the European stimuli yielded more positive automatic evaluations than the Middle Eastern stimuli.

Explicit Questionnaire Scoring

The relevant questions for the MRS within the “Student Opinions” questionnaire were three, five, 10, 12, 16 and 18. Of these, questions 10, 16 and 18 were reverse scored. A maximum of five points was able to be allocated to each question, with greater scores indicating stronger expressed anti-Arab sentiment. The raw scores of the relevant items provided the input data for further analysis.

To score the TDQ, the raw responses for the European, Middle Eastern and Asian countries were tallied separately. This resulted in three scores out of 30 for each participant. Higher scores indicated a higher stated desire for visiting that location.

Statistical Analyses

The factor analytic strategies applied in the current study were Exploratory Factor Analysis (EFA; Spearman, 1904) and Confirmatory Factor Analysis (CFA; Jöreskog, 1969).

Exploratory Factor Analysis (EFA)

EFA (Spearman, 1904) differs from CFA in that no specifications regarding what is expected for the model are required. Rather, EFA is an exploratory technique used to establish basic concepts or to simplify existing methods by reducing the number of items required to evaluate each construct (Saltin & Strand, 1995). EFA was applied in the current study to determine which questionnaire items provided an adequate measure of the latent construct and whether any questions should be meaningfully grouped together (i.e. parcelled). Two EFAs with maximum likelihood estimation and oblimin rotation[18] were completed in SPSS (PASW version 18). Loadings were considered substantive if greater than .37, based on the alternative formula provided by Norman and Streiner (1994, p. 139) for estimating minimum loadings in EFA with sample sizes over 100 (see Formula 5.4).

$$\text{Minimum Factor Loading} = 5.152/\ [\text{SQRT}\ (N\text{-}2)] \qquad (5.4)$$

Confirmatory Factor Analysis (CFA; Jöreskog, 1969)

The CFAs applied the maximum likelihood estimation procedure (MLM) and were performed using Mplus, version 6.1 (Muthén & Muthén, 2010). Minimum factor loadings of .32 were indicative of meaningful relationships within the CFA models, in accordance with Gorsuch (1983).

Assessing Model Fit

An evaluation of how well each model fit the input data was determined by the goodness-of-fit indices derived from the Mplus program. Mplus calculates the chi-square likelihood ratio test statistic (χ^2),but this statistic is affected substantially by sample size, with almost any model routinely rejected when the sample size is large (Brown, 2006). In view of this, the approximate fit indices provided by Mplus were used: the root-mean-square error of approximation (RMSEA), the standardised root mean square residual (SRMS), and the comparative fit index (CFI). Hu and Bentler (1999) suggest that RMSEA values close to .06 or below be taken as good fit.

[18] Factor rotation can enhance interpretability of EFA solutions without changing the underlying mathematical properties of the solution (Tabachnick & Fidell, 2001). Oblimin factor rotation is one such rotation strategy that allows for inter-correlations between factors (Brown, 2006).

However, Browne and Cudeck (1993) put forth that RMSEA values ranging between .06 and .08 could be inferred as moderate fit, and .08 to .10 as marginal fit. Hu and Bentler (1999) suggest SRMR values are close to .08 or less be taken as indication of good fit. CFI values close to .95 or above are indicative of good model-data fit (T. A. Brown, 2006), with acceptable fit determined with values above .90.

Results

Exploratory Factor Analysis for the Modern Racism Scale

An EFA with oblimin rotation was performed on the relevant questions of the MRS. The results verified good model fit: $\chi2$ (9, N=198)=47.88, p<.001. Table 5.3 depicts the factor loadings for the one-factor solution using a ML extraction.

Table 5.3

Factor Loadings for the One-Factor Solution for the Modern Racism Scale

MRS Questions	Factor Loading
MRSQ3	.63
MRSQ5	.64
MRSQ10r	-.32
MRSQ12	.77
MRSQ16r	.09
MRSQ18r	.25

Note. N=198. MRS=Modern Racism Scale.

As evident in Table 5.3, only questions three, five and 12 produced factor loadings greater than .37 (Norman & Streiner, 1994; .63, .64 and .77). As such, only data from these three questions were included in the analyses. The raw data for questions three, five and 12 formed the indicators for the latent explicit racial attitude construct.

Exploratory Factor Analysis for the Travel Destination Questionnaire

Travel Destination Questionnaire (TDQ) data were subjected to a three-factor EFA with oblimin rotation, using ML extraction. Table 5.4 shows the resulting factor loadings.

Table 5.4

Factor loadings for the Three-Factor EFA of the Travel Destination Questionnaire

	Factor 1 Middle East	Factor 2 Europe	Factor 3 Asia
TDQ1 Indonesia			-.60
TDQ2 Syria	.70		
TDQ3 Poland		.50	
TDQ4 Italy		.68	
TDQ5 Saudi Arabia	.78		
TDQ6 Thailand			-.82
TDQ7 Malaysia			-.84
TDQ8 Israel	.67		
TDQ9 France		.77	
TDQ10 China			-.33
TDQ11 UK		.54	
TDQ12 Jordan	.73		
TDQ13 Spain		.67	
TDQ14 Singapore			-.49
TDQ15 United Arab Emirates	.65		
TDQ16 Hungary		.51	
TDQ17 Lebanon	.79		
TDQ18 Japan	.40		

Note. N=198. TDQ=Travel Destination Questionnaire.

As evident in Table 5.4, all countries loaded onto their appropriate factors (with the exception of Japan that loaded highest on the Middle Eastern factor). The Asian countries were presented as distracters and were not included in any analyses. All Middle Eastern and European countries loaded strongly onto their appropriate factor (factor loadings ranging from .50-.79). Data from all twelve of these countries were thus included in subsequent analyses.

The twelve TDQ items were parcelled into two groups of European and Middle Eastern countries, roughly matched for destination popularity. A list of most popular tourist destinations (United Nations, 2009) was used to separate the countries as follows. Firstly, the most popular tourist destinations from Europe and the Middle East, (namely France and Saudi Arabia) were placed into Parcel 1, the second most popular locations (Spain and Israel) were placed into Parcel 2, the third were added to Parcel 1 and so on down the list. Parcel 1 thus consisted of the first, third and fifth most travelled to destinations in Europe and the Middle East[19], Parcel 2 contained data from the second, fourth and sixth destinations[20]. Means and standard deviations for the TDQ for each country and regional group are presented in Table 5.5. Overall, respondents stated that they 'probably would' like to visit the European countries (average rating about 4), whereas respondents appeared more diffident about visiting the Middle Eastern countries, stating they 'maybe' or 'probably would not' like to visit (average rating between 2 and 3; See Table 5.5).

[19] Namely: France, Italy, Poland, Saudi Arabia, United Arab Emirates and Syria.
[20] Namely: Spain, United Kingdom, Hungary, Israel, Jordan and Lebanon.

Table 5.5

Means and Standard Deviations of TDQ responses by Country and Group

	Mean	Standard Deviation
TDQ9 France	4.37	.89
TDQ4 Italy	4.44	.82
TDQ3 Poland	3.44	.99
TDQ5 Saudi Arabia	2.79	1.12
TDQ15 United Arab Emirates	2.94	1.20
TDQ2 Syria	2.66	.97
Europe 1,3,5	4.09	.72
Middle East 1,3,5	2.80	.94
TDQ13 Spain	4.16	.97
TDQ11 UK	4.30	1.00
TDQ16 Hungary	3.17	1.06
TDQ8 Israel	2.67	1.23
TDQ12 Jordan	2.75	1.13
TDQ17 Lebanon	2.48	1.03
Europe 2,4,6	3.88	.76
Middle East 2,4,6	2.63	.94

TDQ scores for each data parcel were determined for Parcel 1 by calculating the mean of each participants' responses for the first, third and fifth most popular Middle Eastern countries (i.e. Saudi Arabia, United Arab Emirates and Syria) and subtracting that from the mean of the responses for the first, third and fifth most popular European countries (i.e. France, Italy and Poland). The resulting score was then divided by the pooled standard deviation of all participants' results for those six countries (1.01). The formula for Parcel 1 is presented in Equation 5.5.

$$\frac{\textit{Mean}\ \text{“Euro 1, 3, 5”} - \textit{Mean}\ \text{“ME 1, 3, 5”}}{\textit{Pooled SD}\ (1.01)} \tag{5.5}$$

This formula was also applied to the countries in Parcel 2 (which were the second, fourth and sixth most popular countries in the Middle East and Europe). The pooled standard deviation of responses for these countries was 1.07. These two parcels of data were used for further analyses.

The mean and standard deviations for each of the data parcels for all of the tasks in the present study are offered in Table 5.6.

Table 5.6

Mean and Standard Deviations for all Data Parcels for each Experimental Measure

Task	Mean	SD
Racial VIAT1	.45	.33
Racial VIAT2	.42	.34
Racial VIAT3	.39	.36
Racial VIAT4	.33	.34
Racial PIAT1	.22	.38
Racial PIAT2	.15	.37
Racial PIAT3	.15	.34
Racial PIAT4	.12	.36
Country VIAT1	.32	.39
Country VIAT2	.29	.35
Country VIAT3	.28	.33
Country VIAT4	.26	.35
Country PIAT1	.24	.36
Country PIAT2	.18	.40
Country PIAT3	.16	.39
Country PIAT4	.17	.35
Racial APT1	-1.54	136.62
Racial APT2	7.02	128.41
Racial APT3	9.60	124.98
Racial APT4	8.71	141.76
Country APT1	3.69	127.96
Country APT2	25.50	142.46
Country APT3	21.77	156.34
Country APT4	9.91	144.05
Modern Racism Scale Q3	1.65	.85
Modern Racism Scale Q5	1.84	.92
Modern Racism Scale Q12	2.57	1.08
Travel Destination QnParcel1	1.28	1.01
Travel Destination Qn Parcel2	1.16	.93
Travel Experience	3.08	1.15

Internal Consistency and Internal Convergent Validity Results for the Implicit Attitude Measures

The results of the Composite reliability (CR) and Average Variance Extracted (AVE) analyses are depicted in Table 5.7. It is noted that the Heywood case was removed for the Country Priming estimate because it was confounding the data.

Table 5.7

Internal Consistency and Internal Convergent Validity of the IATs and APTs

	Composite Reliability	Average Variance Extracted
Racial VIAT	.76	.44
Country VIAT	.76	.44
Racial PIAT	.76	.45
Country PIAT	.77	.46
Racial APT	.16	.07
Country APT*	.06	.02

* Estimated following the removal of a Heywood case.

As is evident in Table 5.7, the IATs all demonstrated good internal consistency, with CR estimates well above .70 (see Hair et al., 2006). However, the internal convergent validity evidence failed to meet the required benchmark of .50 (see Hair et al., 2006) revealing a greater amount of error variance than trait variance was present in the IAT effect scores. Specifically random error variance appears to account for 55% of the IAT effect scores. The priming tasks demonstrated reliability estimates that were nowhere near adequate. The internal consistency estimates were well under .70 and the internal convergent validity estimates were almost negligible. The AVE results imply that on average around 95% of the APT scores are attributable to random error variance. Given that there was also a Heywood case in the APT data, the CR and AVE results do not support the reliability of the priming measures.

Internal Construct Validity Results for the Implicit Attitude Measures

Confirmatory Factor Analysis (CFA) was applied to assess the internal construct validity of each implicit attitude measure using the specified model illustrated in Figure 5.1. It is noted that each indicator is comprised of parcelled data.

Results for the Internal Construct Validity of the Racial VIAT

The goodness-of-fit indices for the CFA model of the Racial VIAT were: χ^2 (2, N=198)=5.12, *p*=.08; CFI=.982; RMSEA=.089; SRMR=.027. This result indicated acceptable model fit, as three of the four fit indices showed good fit and the RMSEA demonstrated marginal fit. Correlations between indicators for the CFA model are presented in Table 5.8. The standardised factor loadings (STDYX) for each of the variables and residuals for the Racial VIAT are presented in Figure 5.10.

Table 5.8

Inter-indicator Correlations for the CFA Model of the Racial VIAT

	VIAT1R	VIAT2R	VIAT3R	VIAT4R
VIAT1R	1.000			
VIAT2R	.427	1.000		
VIAT3R	.476	.463	1.000	
VIAT4R	.457	.290	.495	1.000

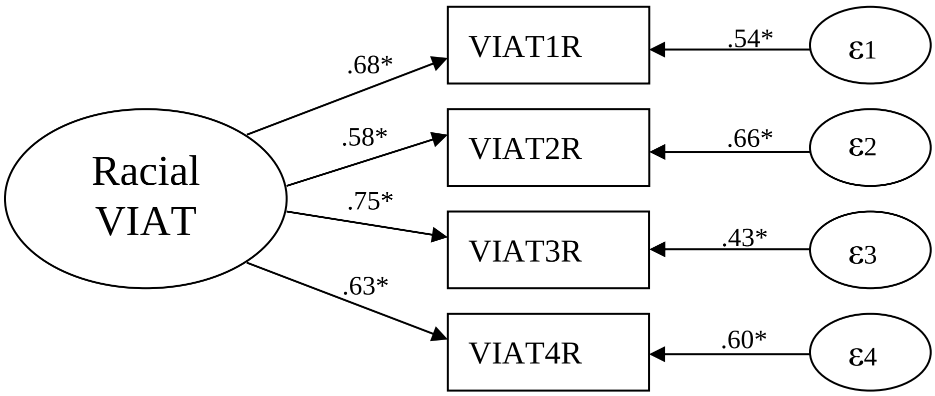

Figure 5.10. CFA model of the Racial VIAT. **p*<.001.

For the Racial VIAT all of the indicator parcels loaded significantly onto the latent factor, with high factor loadings well above .32 (factor loadings ranging from .58 to .75; see Figure 5.10). High factor loadings indicate each of the data parcels loaded substantively onto the latent implicit racial attitude factor. High and significant error variances were also evident (variances ranging from .43-.66). This provides further evidence of substantial amounts of random error variance in the IAT effect scores.

Results for the Internal Construct Validity of the Racial PIAT

The goodness-of-fit indices for the CFA model of the Racial PIAT were: χ^2 (2, N=198)=5.88, *p*=.05; CFI=.980; RMSEA=.099; SRMR=.027. This result indicates acceptable model fit, because three of the four fit indices demonstrated good fit and the RMSEA indicated marginal fit. Inter-indicator correlations for the CFA model are presented in Table 5.9. Figure 5.11 presents the standardised factor loadings (STDYX) for each of the variables and residuals for the Racial PIAT.

Table 5.9

Inter-indicator Correlations for the CFA Model of the Racial PIAT

	PIAT1R	PIAT2R	PIAT3R	PIAT4R
PIAT1R	1.000			
PIAT2R	.543	1.000		
PIAT3R	.367	.449	1.000	
PIAT4R	.363	.487	.456	1.000

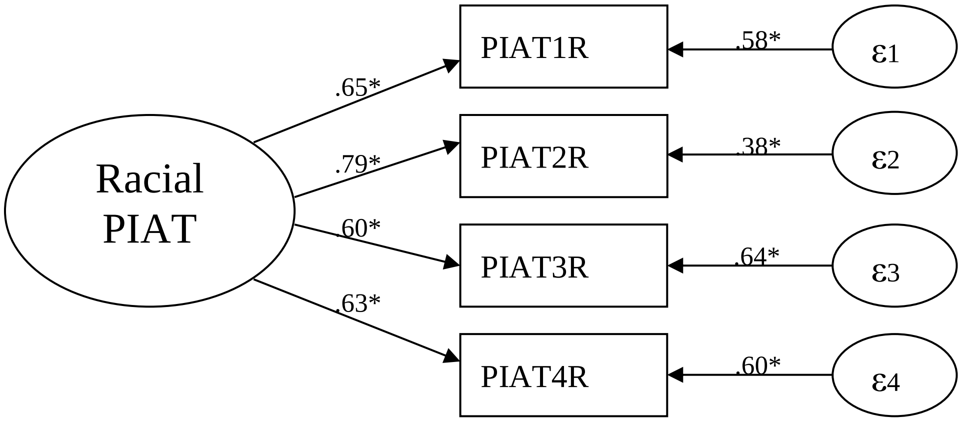

Figure 5.11. CFA model of the Racial PIAT. *p<*.001*.

For the Racial PIAT all of the indicators loaded significantly onto the latent factor and presented substantial factor loadings (again well above .32, ranging from .60 to .79; see Figure 5.11). This means each of the PIAT data parcels loaded highly on the latent implicit racial attitude factor. Error variances were also quite high for this task (ranging from .38-.64) implying significant portions of random error variance were present within the IAT effect scores.

Results for the Internal Construct Validity of the Country VIAT

The goodness-of-fit indices for the CFA model of the Country VIAT were: χ^2 (2, N=198)=3.73, *p*=.16; CFI=.989; RMSEA=.066; SRMR=.022. This result also showed acceptable model fit, because three of the four fit indices demonstrated good fit and the RMSEA indicated a moderate level of fit. Inter-indicator correlations for the CFA model are presented in Table 5.10. Figure 5.12 presents the standardised factor loadings (STDYX) for the Country VIAT.

Table 5.10

Inter-indicator Correlations for the CFA Model of the Country VIAT

	VIAT1C	VIAT2C	VIAT3C	VIAT4C
VIAT1C	1.000			
VIAT2C	.438	1.000		
VIAT3C	.444	.527	1.000	
VIAT4C	.449	.410	.381	1.000

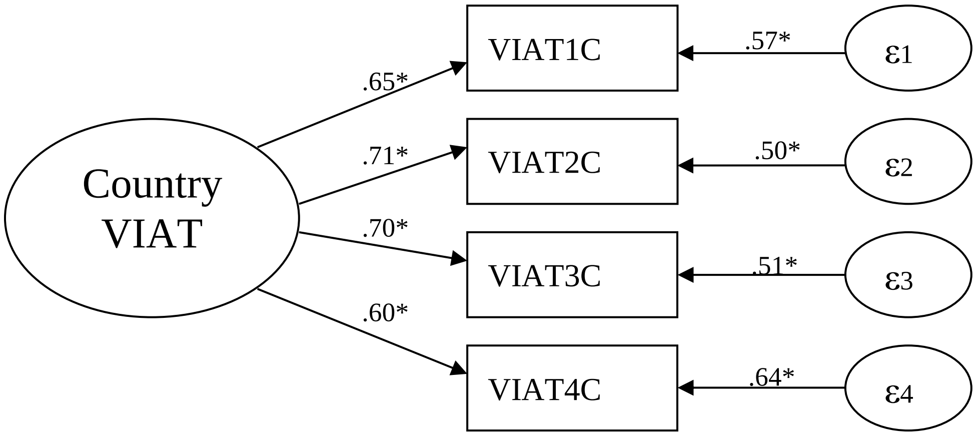

Figure 5.12. CFA model of the Country VIAT. **p*<.001.

For the Country VIAT all of the indicators loaded significantly onto the latent factor and presented substantial factor loadings (again well above .32, ranging from .60 to .71; see Figure 5.12). Again the error variances were also found to be large and significant (ranging from .50-.64).

Results for the Internal Construct Validity of the Country PIAT

The goodness-of-fit indices for the CFA model of the Country PIAT were: χ^2 (2, N=198)=.13, *p*=.94; CFI=1.000; RMSEA=.000; SRMR=.004. All four of the fit indices showed good model fit. Inter-indicator correlations for the CFA model are presented in Table 5.11. Figure 5.13 presents the standardised factor loadings (STDYX) for the Country PIAT.

Table 5.11

Inter-indicator Correlations for the CFA Model of the Country PIAT

	PIAT1C	PIAT2C	PIAT3C	PIAT4C
PIAT1C	1.000			
PIAT2C	.528	1.000		
PIAT3C	.449	.514	1.000	
PIAT4C	.381	.453	.396	1.000

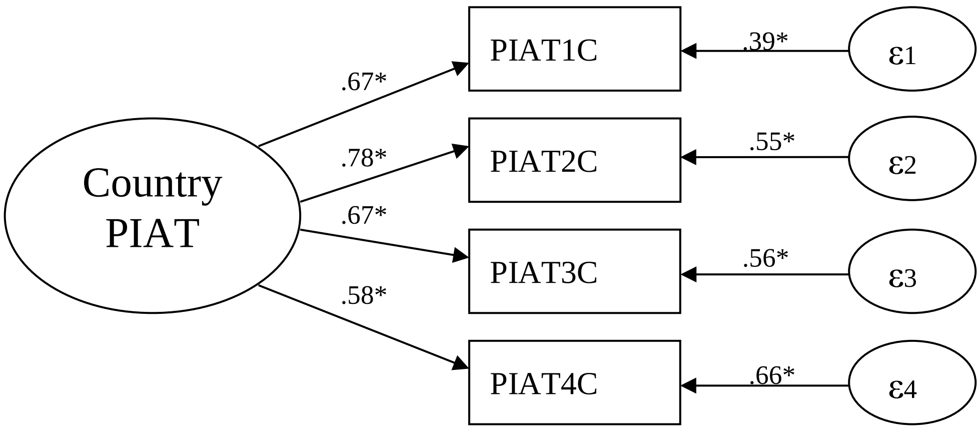

Figure 5.13. CFA model of the Country PIAT. **p*<.001.

As can be seen in Figure 5.13, all the indicators loaded significantly onto the latent factor with substantial factor loadings (factor loadings ranging from .58 to .78). The error variances again were substantial and significant (ranging from .39-.66).

Results for the Internal Construct Validity of the Racial APT

Goodness-of-fit indices for the CFA model of the racial priming task were as follows: χ^2 (2, N=198)=.55, *p*=.76; CFI=1.000; RMSEA<.001; SRMR=.021. All four of the fit indices indicated good fit. Inter-indicator correlations for the CFA model are presented in Table 5.12. In contrast to the fit indices, the standardised factor loadings (STDYX; Figure 5.14) for the Racial APT showed minimal and non-significant parameter estimates for the indicators onto the latent factor (loadings ranging between .01 and .21). The only substantial factor loading loaded in an unexpected direction (see the negative factor loading of PR4; Figure 5.14). However, the error variances are mostly significant and very substantial (ranging between .80 and 1.00). These findings demonstrate inadequate support for the internal construct validity of the Racial APT.

Table 5.12

Inter-indicator Correlations for the CFA Model of the Racial APT

	PR1	PR2	PR3	PR4
PR1	1.000			
PR2	.041	1.000		
PR3	-.067	.035	1.000	
PR4	-.006	-.094	-.074	1.000

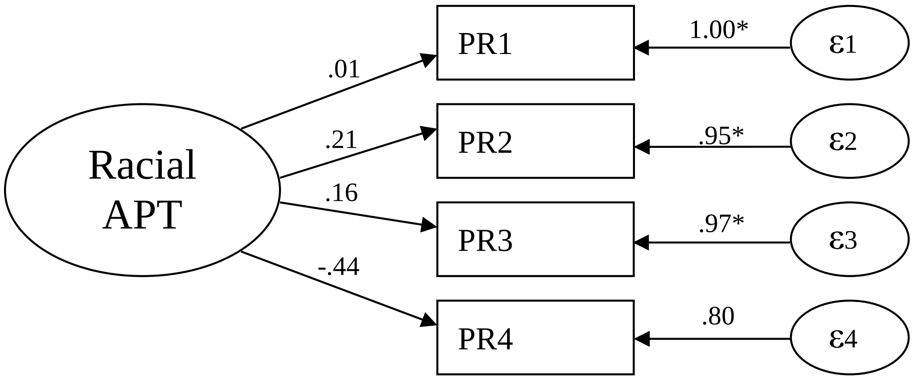

Figure 5.14. CFA of the racial priming task. **p*<.001.

Results for the Internal Construct Validity of the Country APT

The goodness-of-fit indices for the CFA model of the country priming task are as follows: χ^2 (2, N=198)=.26, *p*=.88; CFI=1.000; RMSEA=.000; SRMR=.023. Again all four of the fit indices signify good fit. Inter-indicator correlations for the CFA model are presented in Table 5.13. In Figure 5.15, it is evident the standardised factor loadings (STDYX) for the Country APT provide evidence against the internal construct validity of this measure, with minimal and non-significant parameter estimates for the indicators onto the latent factor found. This analysis resulted in an improper solution, or Heywood case. This is evident in the second data parcel whereby a parameter estimate with an out-of-range value (PC2; 1.10) and a negative indicator error variance (-.20) are visible (see Figure 5.15). Results such as these reveal the data is not fitting well with the model. For the Country APT, the error variances are again substantial and significant (.98, *p*<.001 for the non-out-of-range variances). These results reveal inadequate support for the Country APT's internal construct validity.

Table 5.13

Inter-indicator Correlations for the CFA Model of the Country APT

	PC1	PC2	PC3	PC4
PC1	1.000			
PC2	.135	1.000		
PC3	-.034	.169	1.000	
PC4	.006	-.161	-.087	1.000

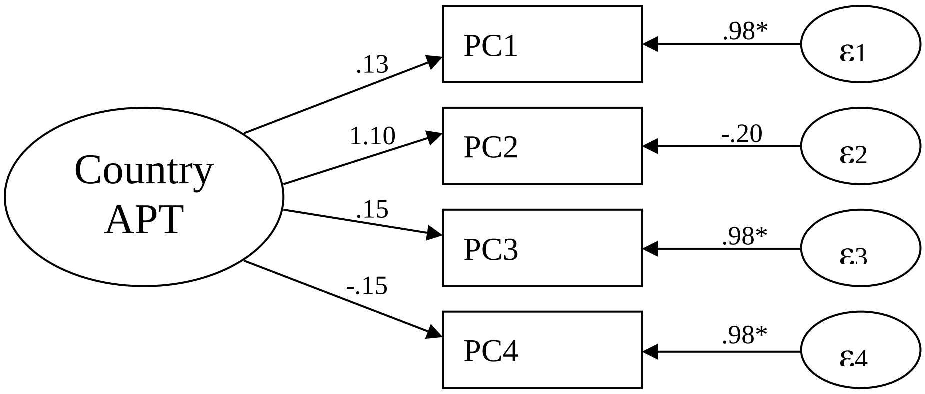

Figure 5.15. CFA of the country priming task. **p*<.001.

Retesting of APT Results with Reparcelled Data

To ensure the APT results were not caused simply by a random parcelling effect, the priming task dataset was completely reparcelled into eight new data sets. Priming task scores were then recalculated using these new data parcels and the CFA models were re-tested.

Internal Construct Validity Results for the Reparcelled Racial Priming Data.

The CFA for the reparcelled racial priming task data still showed good model fit, $\chi 2$ (20, N=198)=18.72, p=.54; CFI=1.000; RMSEA=.000; SRMR=.080. Inter-indicator correlations for the CFA model are presented in Table 5.14. Yet the highly variable factor loadings were predominantly small and non-significant (see Figure 5.16). Interestingly, the sixth data parcel (PR6) loaded strongly in the opposite direction to that expected (β=-.71, p<.05). Random error variance continued to account for the majority of the priming score.

Table 5.14

Inter-indicator Correlations for the CFA Model of the Reparcelled Racial APT Data

	PR1	PR2	PR3	PR4	PR5	PR6	PR7	PR8
PR1	1.000							
PR2	-.018	1.000						
PR3	.069	-.136	1.000					
PR4	-.037	.135	-.285	1.000				
PR5	.116	.018	-.152	.198	1.000			
PR6	-.098	-.196	-.087	.065	-.101	1.000		
PR7	.038	-.203	-.022	-.087	-.103	.171	1.000	
PR8	.093	.106	.134	-.136	.044	-.267	-.031	1.000

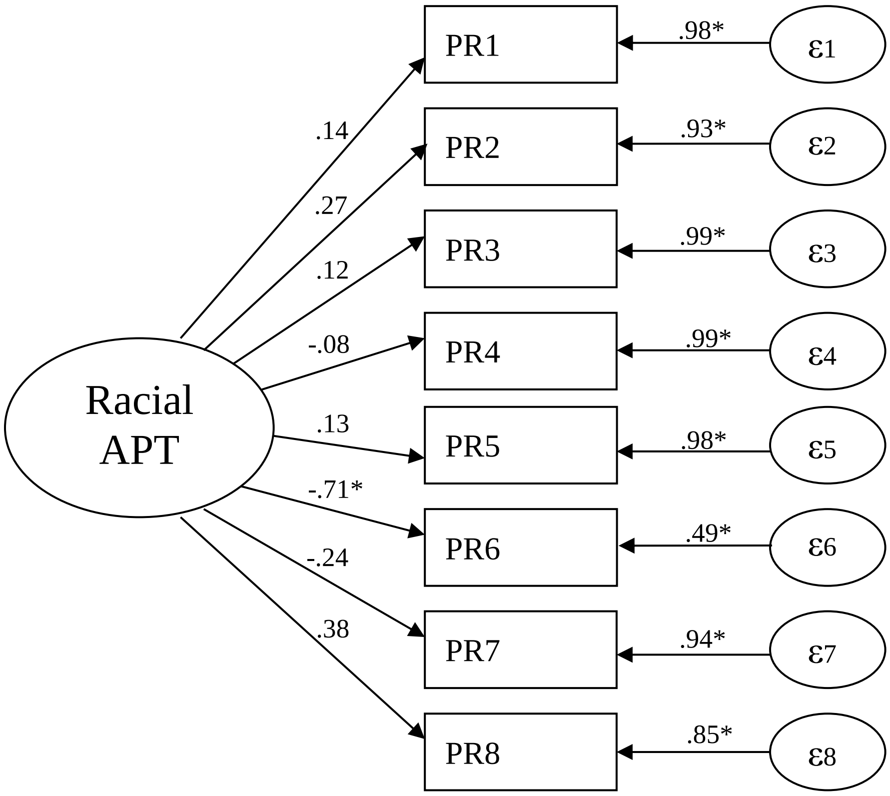

Figure 5.16. CFA of reparcelled racial priming data. *p<.001.

Internal Construct Validity Results for the Reparcelled Country Priming Data.

The reparcelled data for the Country APT would not converge. As such, no goodness-of-fit indices were able to be calculated and the estimated factor loadings cannot be presented. Non-convergence can be an indicator of unstable data (Brown, 2006). The Country APT was thus deemed to possess unsatisfactory levels of internal construct validity.

Because neither APT showed adequate internal construct validity even after reparcelling, all priming task data were excluded from further analyses.

Convergent and Discriminant Validity Results for the IAT using Three-factor Single-group CFA

The third aim of the present study was to assess the convergent and discriminant validity of the IATs using CFA. The Racial and Country attitude constructs were examined separately. Strong convergence between IATs was expected and the inter-implicit correlations were hypothesised to be significantly greater than the implicit-explicit correlations. The priming tasks were excluded from these analyses due to their lack of internal consistency and construct validity.

Convergent and Discriminant Validity Results for the Racial Attitude Construct

The goodness-of-fit indices for the racial attitude construct three-factor CFA model indicated good model fit: χ^2 (41, N=198)=42.77, *p*=.40; CFI=.997; RMSEA=.015; SRMR=.041. Inter-indicator correlations for the CFA model are depicted in Table 5.15. The standardised factor loadings (STDYX) for the three-factor CFA for the racial attitude construct are presented in Figure 5.17.

Table 5.15

Inter-indicator Correlations for the 3-Factor CFA Model of the Racial Attitude Data

	VIAT 1R	VIAT 2R	VIAT 3R	VIAT 4R	PIAT 1R	PIAT 2R	PIAT 3R	PIAT 4R	MRS Q3	MRS Q5	MRS Q12
VIAT1R	1.000										
VIAT2R	.427	1.000									
VIAT3R	.476	.463	1.000								
VIAT4R	.457	.290	.495	1.000							
PIAT1R	.310	.262	.316	.308	1.000						
PIAT2R	.346	.278	.295	.318	.543	1.000					
PIAT3R	.127	.106	.200	.284	.367	.449	1.000				
PIAT4R	.141	.141	.213	.275	.363	.487	.456	1.000			
MRSQ3	.091	.136	.063	.056	.060	.043	-.004	.031	1.000		
MRSQ5	.116	.046	.111	.131	.145	.194	.200	.142	.446	1.000	
MRSQ12	.138	.127	.124	.143	.125	.153	.141	.090	.478	.483	1.000

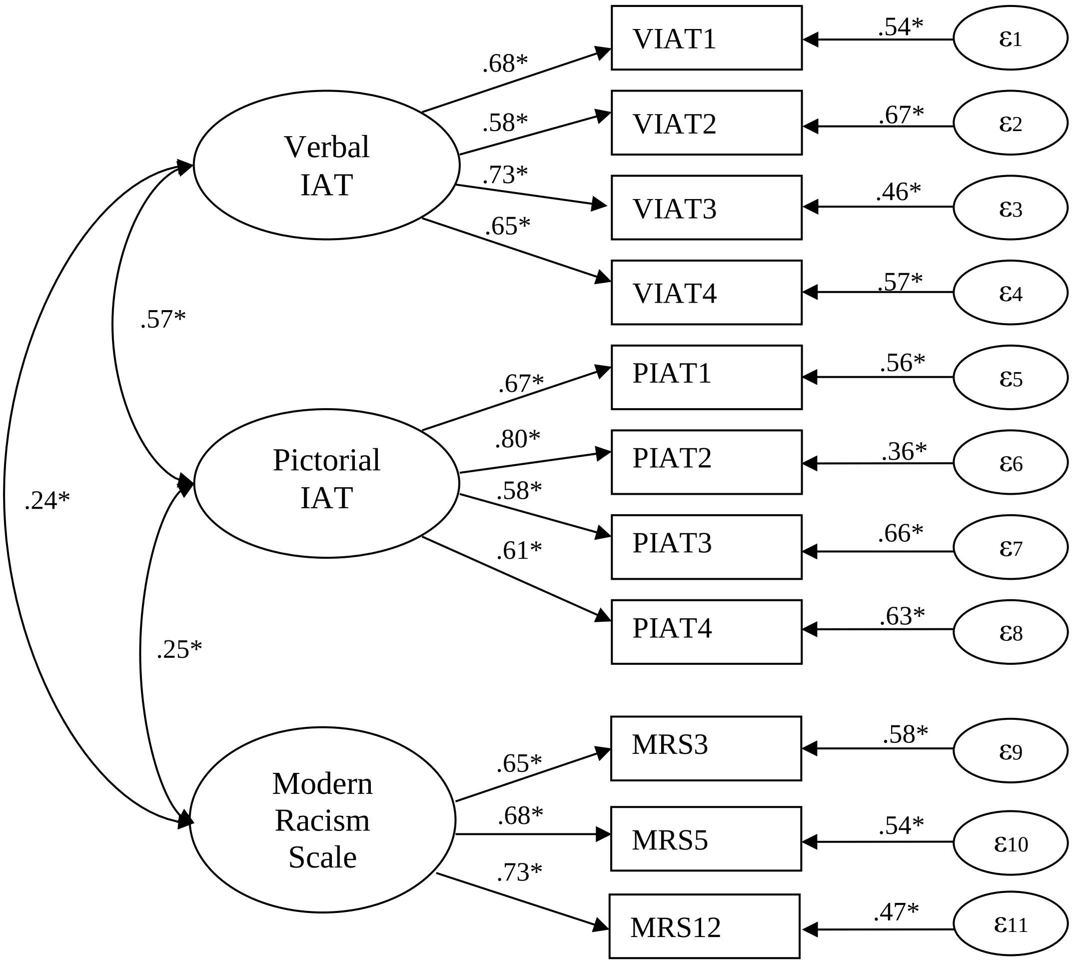

Figure 5.17. CFA of tasks assessing the racial attitude construct. *p<.001.

The model presented in Figure 5.17 supports the convergent validity of the IATs, with a strong positive inter-implicit correlation demonstrated. As hypothesised, the correlation between the VIAT and PIAT (r=.57) was significantly greater than the correlation between either of these implicit measures and the Modern Racism Scale (r=.24, z=3.98, p<.001 for the VIAT-MRS correlation; and r=.25, z=3.87, p<.001 for the PIAT-MRS correlation). These results support the discriminant validity of the IATs and questionnaire measures.

Convergent and Discriminant Validity Results for the Country Attitude Construct

For the country attitude construct, the goodness-of-fit indices for the three-factor CFA model were as follows: χ^2 (60, N=198)=56.92, *p*=.59; CFI=1.000; RMSEA >.001; SRMR=.033. All four of the fit indices indicate this model also replicated the variances and covariances of the input data very well. Inter-indicator correlations for the CFA model are depicted in Table 5.16. Figure 5.18 presents the standardised factor loadings (STDYX) for the three-factor CFA for the country attitude construct.

Table 5.16

Inter-indicator Correlations for the 3-Factor CFA Model of the Country Attitude Data

	VIAT 1C	VIAT 2C	VIAT 3C	VIAT 4C	PIAT 1C	PIAT 2C	PIAT 3C	PIAT 4C	TDQ 1	TDQ 2
VIAT1C	1.000									
VIAT2C	.438	1.000								
VIAT3C	.444	.527	1.000							
VIAT4C	.449	.410	.381	1.000						
PIAT1C	.297	.287	.201	.247	1.000					
PIAT2C	.292	.226	.248	.277	.528	1.000				
PIAT3C	.326	.187	.235	.289	.449	.514	1.000			
PIAT4C	.220	.115	.129	.238	.381	.453	.396	1.000		
TDQ1	.294	.262	.175	.227	.298	.200	.179	.190	1.000	
TDQ2	.267	.199	.088	.168	.257	.233	.207	.199	.671	1.000

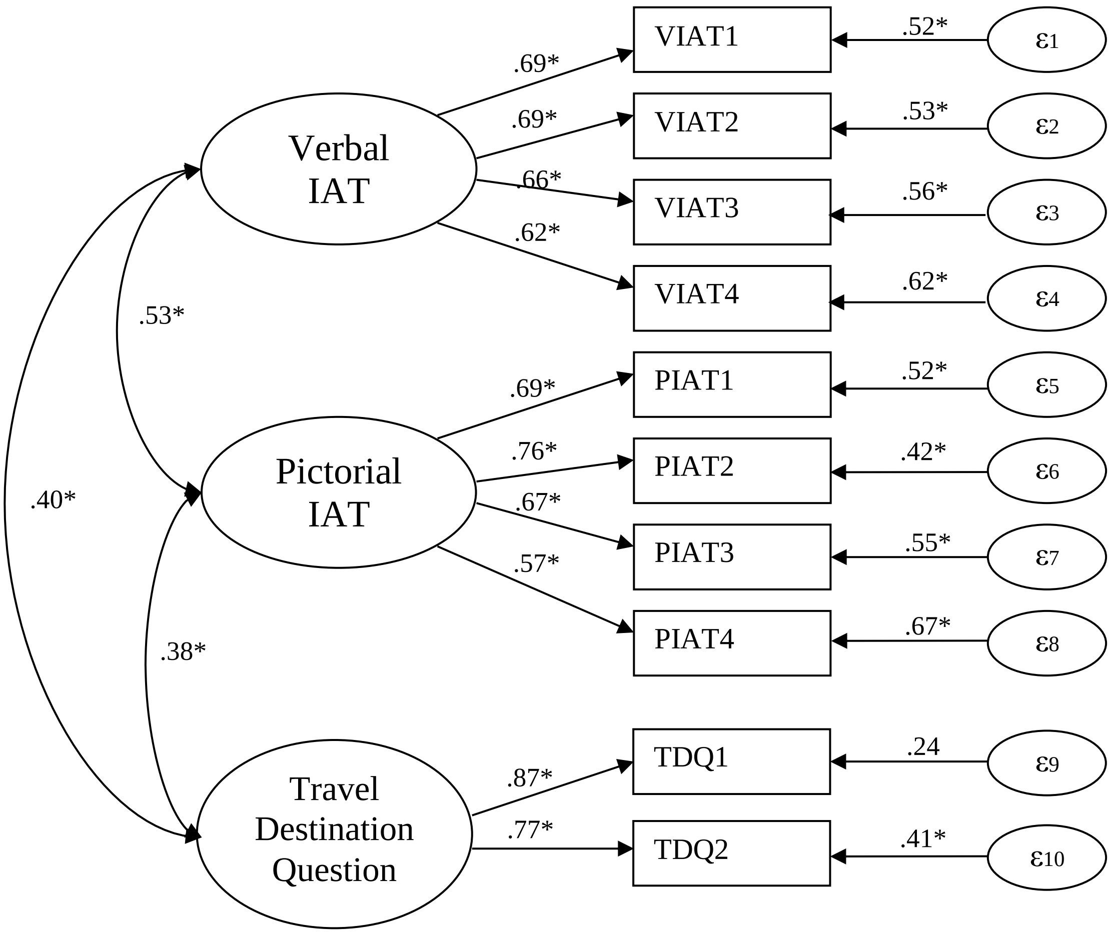

Figure 5.18. CFA of tasks assessing the country attitude construct. *p<.001.

The results depicted in Figure 5.18 show a strong positive correlation between the latent IAT measures, indicating support for the convergent validity of the VIAT and PIAT within the country attitude construct. The correlations between the VIAT and the PIAT were stronger (r=.53) than between either of these tasks and the explicit Travel Destination Questionnaire (TDQ; r=.40, r=.38 respectively). However the discrepancy between these correlations was not significant (z=1.64, p=.10, z=1.88, p=.06), due to the TDQ loading reasonably highly on the implicit latent factor. Although the TDQ loaded more highly on the implicit latent factor than the MRS, a comparison of the two explicit attitude measures revealed a non-significant discrepancy (greatest difference: z=1.77, p=.08).

Convergent and Discriminant Validity Results for the IAT using Higher-Order CFA

To assess the level of shared variance between the latent attitude factors the higher-order structure of the CFA measurement models of Figure 5.3 were examined. It was hypothesised the VIAT, PIAT and, to a lesser degree, the explicit attitude measure would all load positively onto the second-order implicit attitude factor.

Higher-order CFA Results for the Racial Attitude Construct

For the racial attitude construct, the goodness-of-fit indices for the higher-order CFA model were as follows: χ^2 (41, N=198)=42.77, *p*=.40; CFI=.997; RMSEA=.015; SRMR=.041. All four of the fit indices revealed the model shows good to excellent fit. This was expected as assessing the higher-order factor structure does not alter the fit indices of the first-order model. The correlational matrix is identical to that presented in Table 5.15. Figure 5.19 presents the standardised factor loadings (STDYX) for the higher-order CFA for the racial attitude construct.

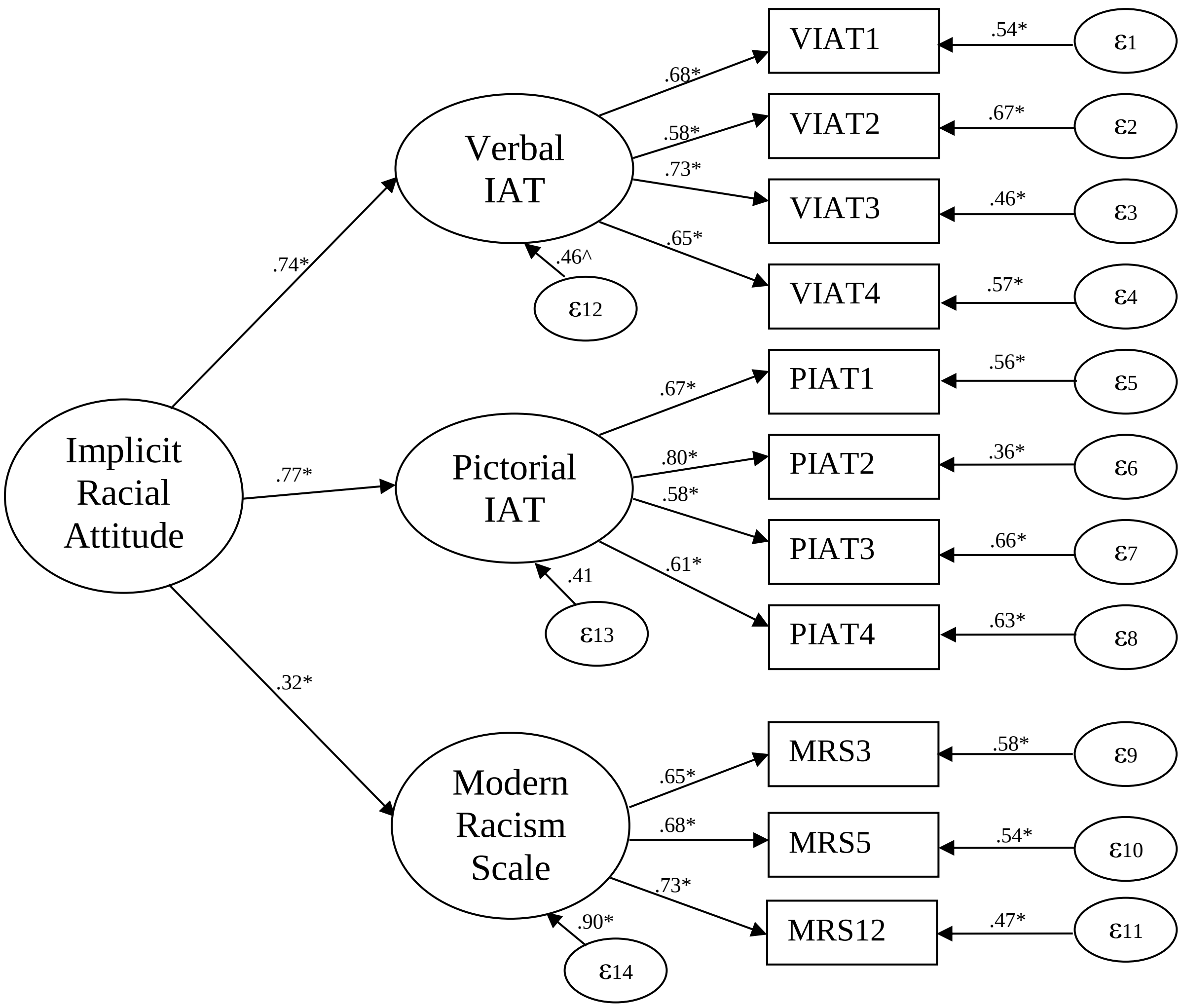

Figure 5.19. Higher-order CFA of the racial attitude construct. ^ $p<.05$, *$p<.001$.

The higher-order CFA model revealed strong support for the convergent validity of the VIAT with the PIAT for the racial attitude construct. The latent factors representing both of these tasks loaded strongly (and comparably) onto the higher-order implicit racial attitude factor (β=.74 and β=.77 respectively; see Figure 5.19). The explicit Modern Racism Scale also loaded significantly onto the same higher factor, with a moderate factor loading of β=.32. This implies that all three tasks were measuring a similar underlying attitude. Further, the average of the two parameter estimates for the IATs onto the higher-order factor were found to be significantly greater than the equivalent factor loading for the explicit questionnaire (z=6.11, p<.001), supporting the discriminant validity of these measurement types.

Higher-order CFA Results for the Country Attitude Construct

For the country attitude construct, the goodness-of-fit indices for the higher-order CFA model were again consistent with the three-factor CFA model indices: χ^2 (60, N=198) =56.92, *p*=.59; CFI=1.000; RMSEA<.001; SRMR=.033. The correlational matrix was depicted in Table 5.16. Figure 5.20 depicts the standardised factor loadings (STDYX) for the higher-order CFA for the country attitude construct.

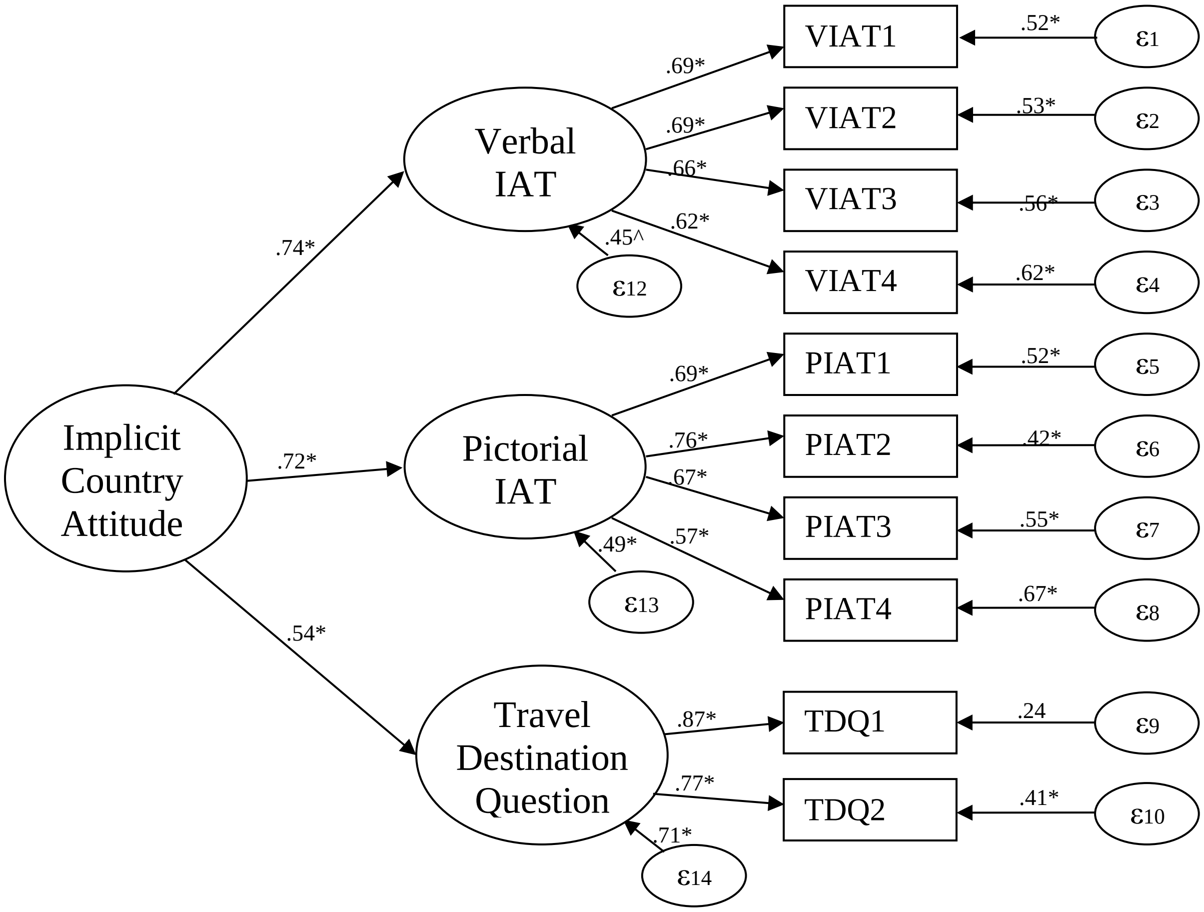

Figure 5.20. Higher-order CFA of the county attitude construct. ^*p*<.05, **p*<.001.

The higher-order CFA model provided strong support for the convergent validity of the VIAT with the PIAT for the country attitude construct. The latent factors representing both of these tasks again loaded strongly (and comparably) onto the higher-order implicit attitude factor (β=.74 and β=.72 respectively; see Figure 5.20). The explicit Travel Destination Questionnaire was also found to load significantly and quite strongly onto the higher-order factor (β=.54). This implies all three tasks were measuring a very similar underlying attitude. The average of the two factor loadings for the IATs were again found to be significantly greater than the factor loading for the explicit questionnaire on the higher-order factor (z=3.00, p=.002). This supports the discriminant validity of the implicit and explicit attitude measures.

Discussion

The overall aim of the present study was to estimate the reliability and construct validity of the IAT and APT using a structural equation modelling approach. Such an approach is novel for evaluating implicit attitude measures and it confirmed the hypothesis of significant random error variance in implicit attitudinal scores. The results of the four aims covered in the current study are discussed in this section. The somewhat mixed support for the implicit attitude measurement techniques are presented, with the IATs showing reasonable reliability and construct validity following the removal of random error, whilst the APTs fell well short of such standards. Support for the comparability of the VIAT and PIAT was also demonstrated. Implications of these results are discussed in the following sections.

Internal Consistency and Internal Convergent Validity of Implicit Attitude Measures

The first aim of the present study was to estimate the reliability of the implicit attitude measures (VIAT, PIAT and APT) using Composite Reliability (CR) and Average Variance Extracted (AVE).

Verbal and Pictorial Implicit Association Tests

The present results showed that the VIATs and PIATs all possessed satisfactory internal consistency estimates following the removal of error variance. This implies that the IATs are consistently measuring a construct. However, the AVE results showed that less than half of the IAT effect scores were attributable to trait variance, thereby indicating inadequate internal convergent validity. These results suggest random error variance comprises about 55% of the IAT effect scores, which highlights the importance of accounting for random error when analysing IAT data.

Affective Priming Tasks

Composite reliability findings for the APTs revealed minimal consistency in the APT scores. The AVE results indicated that over 90% of the APT scores were attributable to random error variance. These results provide serious concerns for the use of APTs given they indicate the priming tasks are barely measuring the construct of interest, capturing random error variance instead. The CR and AVE results thus imply that the APTs are grossly inadequate and inconsistent implicit attitudinal measures.

Summary

Whilst both types of IAT were found to be internally consistent following the removal of random error, this was not the case for the APTs. Further, the IATs were only marginally below adequate internal convergent validity, but the APTs did not come close to meeting this criterion. These results demonstrate evidence of significant portions of random error in implicit attitude measures. However, when this error is accounted for the IAT has potential for use as a stable attitudinal measure, whereas the APTs appear very unreliable.

Internal Construct Validity of the Implicit Attitude Measures

The second aim was to investigate the internal construct validity of the IAT and APT using single-group CFA.

Implicit Association Tests

The CFA latent models (presented in Figures 5.10-5.13) were found to suitably replicate the variances and covariances present in the IAT data. Factor loadings of the indicators onto the latent attitude factors were all found to be high and significant for each of the tasks. This suggests the VIATs and PIATs provided stable measurement of the attitude construct they purport to assess, delivering considerable support for the internal construct validity of the IATs following the removal of random error variance. Evidence of high and significant error variances for each IAT further supports the argument of high error variance in IAT effect scores.

Affective Priming Tasks

The CFA results for both APTs depicted minimal and non-significant factor loadings for the indicators onto the latent attitude factors. This means the APT scores were not assessing adequately (or even at all) the latent construct of implicit attitudes. In contrast, the error variances were extraordinarily high and significant throughout both APTs, revealing that around 95% of the priming score could be attributed to error variance. Results such as these do not bode well for the internal construct validity of the APT. Even re-parcelling both data sets did nothing to remedy this; however it did result in one model failing to converge. An improper solution could have been caused by a structurally misspecified model, sampling fluctuations or inconsistencies within the input data (Brown, 2006). Such inconsistencies may be indicative of an inadequate measurement technique.

Summary

Significant amounts of random error variance were evidenced for all implicit attitudinal measures. Once error variance was accounted for, all IATs were found to possess good internal construct validity using single-group CFA, whereas the APTs were found to barely tap implicit attitudes at all. Such contrasting findings imply that the VIAT and PIAT can both provide an adequate measure of implicit latent attitudes once random measurement error has been appropriately managed. In direct contrast, the APTs have been revealed to lack reliability or validity; APTs were thus deemed inadequate measures of implicit attitudes and were excluded from all further analyses.

Convergent and Discriminant Validity of IATs using Single-group CFA

The third aim of this study was to evaluate the convergent and discriminant validity of the VIAT and PIAT using CFA measurement models. The convergent validity of the two formats of the IAT was strongly supported by the specified CFA models, with a strong inter-implicit correlation between the latent factors (see Figures 5.17 and 5.18). This is the first psychometric support for the construct validity of a fully pictorial version of the IAT. Discriminant validity of the implicit and explicit attitude measures was also evidenced as the inter-implicit correlations between the VIAT and PIAT were significantly greater than the implicit-explicit correlations for both attitude constructs. Such a finding supports the theoretical view that implicit and explicit attitudes are distinct constructs (Nosek & Smyth, 2007).

Summary

The expected convergent and discriminant validity evidence for the implicit and explicit latent attitude factors were provided by the single-group CFAs.

Convergent and Discriminant Validity of IATs using Higher-order CFA

The fourth aim of the current study was to determine how well each task assessed a latent second-order implicit attitude factor, whilst gaining added convergent and discriminant validity evidence through the use of higher-order CFA. The results presented in Figures 5.19 and 5.20 showed the factor loadings between the first-order implicit latent constructs (representing the VIAT and PIAT) and the second-order latent construct (of implicit attitude) were substantial, significant and very comparable (between β=.72 and β=.77). These results imply neither the VIAT nor

PIAT provided a better estimate of the implicit attitude construct; rather the tasks were incredibly similar in their ability to tap substantial quantities of the same underlying attitudinal construct. Further, the explicit attitude measures also loaded positively onto the higher-order attitude factor for both attitude constructs. This suggests both the implicit and explicit attitude measures were accessing very similar underlying attitudes. However, the fact that the explicit attitude measures loaded much less highly than the implicit attitude measures endorsed the discriminant validity of the implicit and explicit techniques. This finding also supports the current theoretically proposed distinction between implicit and explicit attitudes (see Chapter One; Cunningham, Nezlek, et al., 2004; Cunningham et al., 2001; Gawronski & Bodenhausen, 2006; Haeffel et al., 2007; Wilson et al., 2000).

Summary

Findings from the higher-order CFA analyses revealed strong convergent validity for the VIAT and PIAT methodologies. The IATs were shown to provide a good measure of the underlying implicit attitude factor, more so than the explicit attitude measures. These findings further support the argument that implicit and explicit attitudes should be more usefully partitioned as distinct, albeit similar, constructs.

Inconsistent Psychometric Findings for Implicit Attitudinal Measures

The implicit attitudinal measures were found to possess a significant amount of random error variance that comprised more than half of the observed scores. Such a finding emphasises the importance of accounting for error variance using latent modelling analytical approaches. Following the application of CFA there was inconsistent psychometric support for the implicit attitude measures. Specifically, whilst the psychometric properties of both IAT formats improved following removal of random error variance, the APT scores were left with little substance remaining. This was an unexpected finding, the implications of which will be discussed below.

Inadequate Psychometric Evidence for the APT

The APTs of the present study were shown to possess a distinct lack of reliability and validity. This finding adds to previous research that had also flagged poor psychometric properties for APTs (Banse, 1999; Bosson et al., 2000; Kawakami & Dovidio, 2001; Krause et al., 2010; Rudolph et al., 2008). However, such previous research had not so clearly documented exactly how poor the APT's psychometric properties are. Most seriously, the current study found over 95% of the APT scores appeared attributable to random error variance and only a minuscule amount related to the construct of interest. Put simply, the APTs seemed to inconsistently be measuring very little (if any) of the implicit attitudes they were designed to examine. Priming procedures have been popular for assessing intergroup bias since their conception in the mid-1980s (Fazio et al., 1986), despite very limited reliability and validity evidence produced for the task during the last 30 years (Jost et al., 2009). This scarcity of reported psychometric support for APTs is unsurprising given the present findings, and it is concerning that APTs have been used for over two decades without thorough psychometric investigation.

The current results enable a different perspective on previous research comparing IATs and APTs. In the past, very poor convergent validity has often been reported between these tasks, with near-zero correlations not an abnormal finding (see Bosson et al., 2000; Greenwald & Farnham, 2000; Hofmann & Schmitt, 2008; Krause et al., 2010; Rudolph et al., 2008; Sherman et al., 2003). In explaining such results, it has been suggested that divergence between the IAT and APT is due to differences in stimuli categorisation requirements, involving exemplar- versus category-related associations (Gawronski, 2009). Others have argued poor psychometric properties are simply a characteristic of all implicit attitude measures (Bosson et al., 2000; Tetlock & Mitchell, 2009). However, after examining the results of the present study it appears far more likely the lack of convergence between the APT and the IAT is a reflection of the instability and lack of construct validity possessed by the priming procedure, which is divergent from the relative psychometric robustness evidenced for the IAT. It is recommended priming procedures are avoided in applied implicit attitudinal research, at least until a more valid technique can be devised. Psychometric investigation is strongly encouraged to explore this matter further.

Strong Preliminary Support for the Psychometric Stability and Validity of IATs

In direct contrast to the APT results, all the IATs revealed good internal consistency, convergent validity and construct validity following removal of the substantial and significant portions of random error variance apparent in the IAT data. This indicates that when the confounding influence of random error variance has been removed, the IAT effect scores consistently measured the implicit attitude construct they were designed to assess. This consistency, especially across IATs, is a considerable improvement on the often variable results typically reported in the past using data analytic approaches based on observed scores (Hofmann & Schmitt, 2008; Rudolph et al., 2008; Schnabel et al., 2008a; Sherman et al., 2003). The finding of relative consistency in the present study emphasise the potentially volatile influence that random error has on implicit attitudinal scores, which is problematic if not accounted for. It is encouraging the portion of trait variance measured by the IATs does appear to adequately assess the construct of implicit attitudes, despite the fact trait variance is limited by the amount of error variance in the scores (Cote & Buckley, 1987). Due to of the significant portion of random error variance in IAT effect scores, any IAT data not examined using latent modeling techniques will likely be hindered greatly in the provision of accurate or representative results. Routine estimation of error variance before interpretation of IAT scores is thus critical.

Following the use of SEM strategies, strong convergent validity evidence for the VIAT and PIAT was evidenced. This is the first psychometric support for a fully pictorial IAT. The results indicated the PIAT was comparable to the VIAT in terms of reliability and construct validity using CR, AVE, Single-group CFA and Higher-order CFA. In particular, strong convergent validity evidence was demonstrated by large effect sizes between the two IAT types. Such findings are a substantial improvement on previous research using traditional correlational or regression-based analytical techniques, which only rendered small to medium correlations between the VIAT and PIAT (Thomas, 2008). The comparable loadings of the latent VIAT and PIAT factors on to the higher-order factor provide further evidence that these tasks both assess a substantial and similar amount of trait attitude construct. Such results demonstrate that both the VIAT and PIAT are available options for use in applied behavioural research. The PIAT provides an interesting addition to the available implicit attitude techniques. Avoiding the requirement of verbal fluency allows the PIAT to expand the potential participant pool for implicit attitudinal research to include children and the illiterate, it also facilitates opportunity for cross-cultural investigations without the need for translation (Thomas, 2008; Thomas et al., 2007).

Another interesting implication of the current study is that the relationships between the IATs and explicit questionnaires were a lot stronger using latent modeling procedures than those typically reported from traditional analytic procedures (Banse, 1999; Bosson et al., 2000; Cunningham et al., 2001; Greenwald et al., 2009; Hofmann et al., 2005; Hummert et al., 2002; Nosek, 2007). The medium implicit-explicit correlations reported in the present study are very comparable to Cunningham et al.'s (2001) findings that also assessed a Racial VIAT and the MRS using CFA. Such consistency in research findings is an improvement on past inconsistencies in the research literature (see Chapter Two) and strengthens the need for SEM to routinely be applied to implicit attitudinal data.

Impacts of Error Variance on Observed IAT Effect Scores

Implicit-explicit relationships were evidenced to be stronger once random error variance was accounted for. This supports the theory that random error variance can result in attenuating reliability estimates and observed relationships between constructs, or in this case, measurement techniques assessing the same construct (see Coenders & Saris, 2000; Cunningham et al., 2001). The current study found that over half of the IAT effect scores could be attributed to random error variance, confirming previously asserted hypotheses (see Chapter Three). However, error variance is widely recognised as having both random and systematic components (Podsakoff et al., 2003). The current study has not accounted for the differential influence of systematic error variance. In CFA, systematic error remains partialled with the trait variance, confounding it to an unknown degree. In order to clarify the influence of both types of error on implicit attitudinal research, an analytical approach that separates random from systematic types of error is strongly recommended. The Multitrait-Multimethod (MTMM) procedure is such an approach, which would provide much needed clarity regarding this large error component of IAT effect scores and will be addressed in Chapter Six.

Chapter Summary and Conclusion

The current study has revealed that not all implicit attitude measures are consistent and valid assessors of implicit attitudes. The APT was shown to have very poor psychometric properties and to contribute very little to implicit attitudinal investigations by barely assessing the construct of interest. It is recommended that the APT be constrained to studies aimed at developing the stability and validity of the measure rather than any applied research. In contrast, both the verbal IAT and the pictorial IAT were shown to provide a relatively stable and valid measure of implicit attitudes following the removal of random error variance. Because of this, both the VIAT and PIAT appear potentially suitable for applied behavioural research. However, the use of latent modelling techniques is required to account for the high random error component of the IAT effect scores before substantive interpretation is possible. Research to investigate the influence of systematic error on IAT data would help further clarify the composition of IAT effect scores and deliver a more accurate psychometric evaluation of the validity of these measures.

CHAPTER SIX

Study Two: Examining the Construct Validity of Implicit Association Tests using CFA-MTMM

The hypothesis that there would be substantial error variance in IAT data was confirmed in the previous study. Error variance was found to account for over half of an average IAT effect score and appeared to limit internal consistency and construct validity estimates for the implicit attitude measures. However, the impact of systematic forms of error variance, such as method variance, on the IAT effect scores remains unknown. Systematic error variance could further confound findings and contribute to misleading reliability and construct validity estimates. However, there has only been limited consideration of method variance in the IAT. The present study aims to rectify this omission by investigating systematic sources of error variance using the Multitrait-Multimethod approach to Confirmatory Factor Analysis (CFA-MTMM). In Study One, random error was estimated using CFA, but the estimate of trait variance remained confounded by method effects. CFA-MTMM estimates method effects based on error covariances, thus differentiating trait and systematic error components. CFA-MTMM therefore delivers a more focussed and accurate assessment of construct validity, which should provide greater clarity regarding the influence of systematic and random error effects on IAT scores.

Systematic Error Variance in Implicit Attitudinal Research

Systematic error variance refers to relatively consistent extraneous influences that impact upon observed scores (Coenders & Saris, 2000). One of the main sources of systematic error variance are method effects. The term method refers to concrete aspects of the testing methodology, such as the content of specific items or stimuli, the response format and the context of the testing process (Geiser & Lockhart, 2012). Method effects can also be interpreted in a more abstract fashion to include response biases due to social desirability, acquiescence and halo effects (Malhotra et al., 2006; Podsakoff et al., 2003). These method-specific characteristics can systematically bias findings, typically inflating (but also potentially deflating) observed relationships between constructs (Cole & Maxwell, 2003; Podsakoff et al., 2003). Chapter Three clearly outlined the strong likelihood that the IAT is heavily influenced by systematic factors associated with the task's methodology. Despite this, few studies have examined method variance in IAT data[21].

Evidence for Method Variance within Attitudinal Data

The first evidence of method-specific variance in the IAT was reported ten years ago. Mierke and Klauer (2003) examined responses from a typical Flower/Insect Attitude IAT and a non-attitude-based Geometric IAT (that involved red or blue squares and circles as stimuli). Despite the fact the tasks were designed to have no convergent content at all, a moderate correlation between the Attitude and Geometric Shape IATs was reported (Mierke & Klauer, 2003), revealing evidence of method-related variance in the IAT data (see also Nosek & Smyth, 2007). In a later assessment of method variance, Siers and Christiansen (2012) examined three personality-based IATs assessing Extraversion, Conscientiousness and Emotional Stability, in contrast to a Flower/Insect VIAT and a Self/Positive VIAT. As with Mierke and Klauer (2003), Siers and Christiansen (2012) reported significant low-moderate correlations

[21] Of the 900 IAT focussed papers listed in PsycINFO, the present researchers have located fewer than ten studies that acknowledge method variance in the IAT.

between each of the conceptually unrelated Personality-related and Flower/Self-related IATs. CFA-MTMM analysis revealed that on average 22% of the variance in their trait IATs was attributable to method variance (Siers & Christiansen, 2013). These findings provide considerable evidence of method variance produced by the IAT methodology.

Types of Method Variance in IAT Data

Chapter Three presented an extended discussion of likely sources of systematic bias for the IAT. The current section summarises these sources within the framework of Podsakoff et al.'s (2003) thorough categorisation system for method biases, which has not previously been applied to implicit attitude measures. Podsakoff et al. (2003) outlines four sources of method variance – common rater effects, item characteristic effects, item context effects and measurement context effects – each of which may affect the IAT.

Common Rater Effects.

Common rater effects refer to artifactual covariance caused by the same respondent completing the measure. This category includes sources of error such as social desirability, mood state and acquiescence that can impact the way a participant consistently responds to the task (Podsakoff et al., 2003). Previous IAT research have found a respondent's general processing speed (Blanton et al., 2006), intelligence (Stülpnagel & Steffens, 2010) and task-switching ability (Back et al., 2005; Fiedler et al., 2006; Mierke & Klauer, 2003) contribute to method variance for IATs. These characteristics, outlined in Chapter Three, contribute to common rater effects during a validity study.

Item Characteristic Effects.

Item characteristic effects refer to distinctive properties of an item that can influence the respondent, such as item demand characteristics, item ambiguity, common scale formats and positive/ negative wording of items (Podsakoff et al., 2003). As discussed in Chapter Two, ambiguity of stimuli is problematic for the IAT and can produce greater error variance for a task (Messner & Vosgerau, 2010; Salthouse, 2000; Steffens et al., 2008). Also, patterns in reacting to the stimuli such as accuracy versus speed response style (see Chapter Three; Salthouse, 2000; Williams et al., 2005) or responding more positively to stimuli associated with the self (Messner & Vosgerau, 2010; Siers & Christiansen, 2013), can also contributes to this item characteristic form of method variance.

Item Context Effects.

Item context effects refer to the interpretation of an item based on its relation to other items on the measure, which includes priming effects and context induced mood (Podsakoff et al., 2003)[22]. Han et al. (2009) found interpretation of IAT stimuli and response categories are substantially influenced by other tasks completed during a testing session (refer to Chapter Two). The order in which congruent and incongruent trial blocks are presented also systematically influences performance on the subsequent block, with incongruent trials completed far slower following a congruent block of trials (see Chapter Three; Lane et al., 2007; Williams & Themanson, 2011). This is another example of systematic bias caused by item context effects.

---

[22] The impact of previously seen stimuli on later associations, referred to as the priming effect, is a well-known phenomenon in the implicit attitudinal literature (Dasgupta & Greenwald, 2001; Han et al., 2009; Park et al., 2007).

Measurement Context Effects.

Lastly, measurement context effects refer to any covariance artefacts produced from the context in which the responses are obtained. This can occur 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, which is problematic given systematic error variance can significantly influence psychometric evaluations (Mierke & Klauer, 2003).

Implications of Unaccounted For Method Variance on Validity Estimates

Method variance that is unaccounted for can artificially increase correlations between the absolute scores of any two IATs, even if they are not related by shared content (although method biases have also been known to deflate such estimates) (Coenders & Saris, 2000). For instance, in Podsakoff et al.'s (2003) application of Cote and Buckley's (1987) research, presented in Chapter Three, two completely unrelated explicit attitude measures (with expected zero correlation) produced an observed correlation of .23, providing clear evidence for the presence of systematic and random error variance. This correlation of .23 is not significantly different from the .39 inter-implicit correlation between Flower/Insect and Geometric Shape IATs reported by Mierke and Klauer (2003) that also demonstrated evidence of method variance (z=1.24, p<.108). It is thus a strong possibility that method variance has artificially elevates correlations between the absolute score of any two IATs (Siers &

Christiansen, 2013), regardless of whether they measure related constructs. Based on Podsakoff et al.’s (2003) findings, any reported correlation less than .23 may theoretically be regarded as a product solely of error variance. Given implicit attitude measures, such as the APT and IAT, have previously shown poor convergence amongst the tasks (Hofmann & Schmitt, 2008; Krause et al., 2010; Rudolph et al., 2008), it is vital that the effect of method variance is systematically assessed. Multitrait-multimethod analyses using CFA provide a comprehensive way to assess the potential effects of method variance on IAT data.

The Multitrait-Multimethod Approach to Estimating Systematic and Random Error Variance

The multitrait-multimethod (MTMM) approach requires multiple constructs (or traits) to be assessed by multiple measures (Schumacker & Lomax, 2010), with each construct assessed by a common methodology, or common methodologies in the case of more complex comparisons using more than two methods. In such analyses, the measures or methods may differ in terms of the data collection procedures, raters, or stimuli medium (see Coenders & Saris, 2000). MTMM applied within a CFA framework delivers a critical tool for construct validity estimation, and is regarded as one of the most rigorous methods for assessing and controlling for method variance (Lance et al., 2002; Marsh & Grayson, 1995; Meade et al., 2007).

The Multitrait-Multimethod Approach to Construct Validation

Assessment of construct validity using CFA-MTMM incorporates both convergent and discriminant validity evidence. Convergent validity is demonstrated when measures of the same trait are highly correlated, with a high positive correlation (*r*>.50; Cohen, 1992) between the two latent attitude traits, even though they were assessed using different methods (Schumacker & Lomax, 2010). It is also possible to assess the convergent validity of the latent method factors using the correlated methods specification approach for CFA-MTMM (see Chapter Four). This approach enables a unique opportunity for two methods (such as the verbal and pictorial IAT formats) to be directly compared without the confounding influence of trait or random error variance. There are no known papers within the implicit attitudinal literature to have utilised this capability of CFA-MTMM models. CFA-MTMM can further deliver strong discriminant validity evidence when correlations between measures of different traits using the same method are low (*r*<.30; Cohen, 1992; Nosek & Smyth, 2007). Overall, the construct validity of a specific measurement technique is supported when the trait variance of a measure is greater than its method variance (Byrne, 1998). This is a strict psychometric assessment of construct validity conducted by examining the individual parameter estimates. The strict assessment of construct validity afforded by CFA-MTMM has not previously been applied to the psychometric assessment of implicit attitudinal measures[23].

Study Two

Aim 5: Assess the Construct Validity of the IAT using CFA-MTMM

The fifth aim of the current dissertation, and the sole aim of the current study, was to apply the CFA-MTMM analytical framework to evaluate the validity of the verbal and pictorial versions of the IAT (VIAT and PIAT respectively). The model to be tested is presented in Figure 6.1. In this analysis, the traits refer to the two constructs

[23] CFA-MTMM has predominantly been constrained to establishing the psychometric properties of explicit attitude measures and has rarely been applied to laboratory techniques, such as the IAT.

being measured (Country and Racial prejudice) and the methods refer to the IAT formats (Verbal and Pictorial).

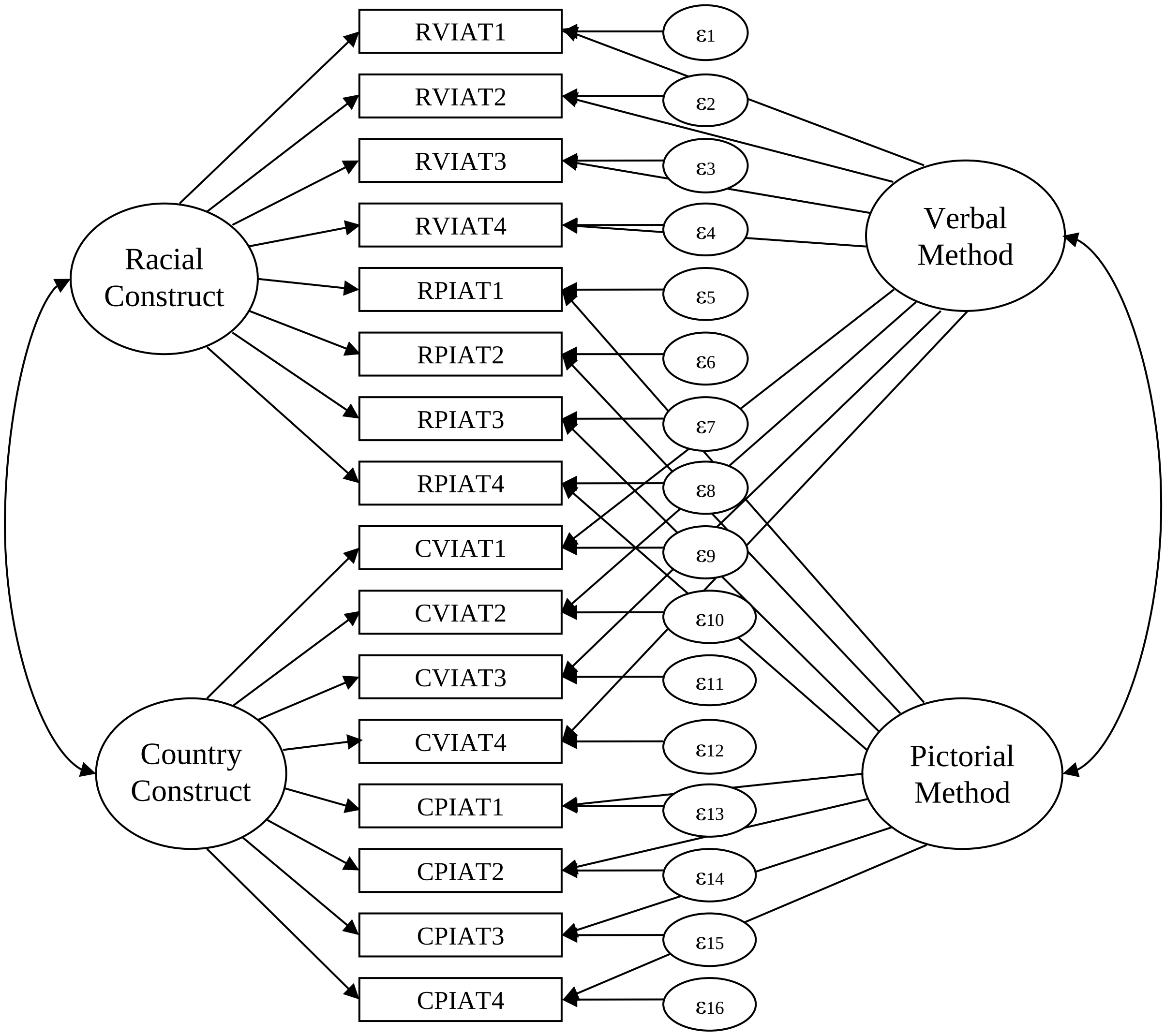

Figure 6.1. The specified path model for the CFA- MTMM analysis of the IATs, where the latent trait factors are presented on the left of the indicators and the latent method factors (and the error variances) can be seen on the right.

Within the primary objective of this study there were three specific sub-aims. The first sub-aim was to ascertain if there was any evidence of substantial method variance within the IAT data. It was hypothesised large portions of method variance would be present as evidenced by significant and substantial factor loadings of the indicators onto the latent method factors in the specified path model (see Figure 6.1). The exact proportion of variance attributable to the IAT method would also be clearly presented in the parameter estimates resulting from the analysis. The second sub-aim was to assess the comparability of the VIAT and PIAT measurement techniques. Convergent validity for the VIAT and PIAT was hypothesised, as evidenced by a strong positive correlation (r>.50; Cohen, 1992) between the latent method factors. This would provide a high level of support for the comparability of these techniques. The third sub-aim was to provide a stringent assessment of the construct validity of the IATs by scrutinising the parameter estimates. It was hypothesised that overall the trait variance would be significant and greater than the method variance for each of the four IATs; demonstrating good construct validity for the measures.

Method

Participants

Responses of the same 198 participants (Mean age: 26.03 years; *SD*=11.10) used in Study One provided the data for all analyses in the present study. For more information regarding these participants see the method section in Chapter Five.

Apparatus

The four empirical IATs described in Chapter Five: the Race VIAT, Race PIAT, Country VIAT and Country PIAT, were used in the current study.

Procedure

The procedure was as outlined in Chapter Five.

Statistical Analysis

The CFA-MTMM analysis was performed in Mplus, version 6.1 (Muthén & Muthén, 2010), using the robust maximum likelihood estimation procedure (MLM) and the freely correlated trait- freely correlated method (CT-CM) specification approach. Assessment of model fit was determined using the goodness-of-fit indices as outlined in Chapter Five.

Results

The CT-CM CFA-MTMM analysis showed that the specified model was a good fit to the data, $\chi 2$ (86, N=198)=93.61, p=.27; CFI=.99; RMSEA=.02; and SRMR=.04. The correlational matrix is depicted in Table 6.1. The specified path model is presented in Figure 6.2 and depicts the partitioning of the IAT data parcels into latent trait, random error and method components. Standardised factor loadings (STDYX) are also presented on this model.

Table 6.1

Inter-indicator Correlations for the CTCM CFA-MTMM Analysis

	VIAT 1R	VIAT 2R	VIAT 3R	VIAT 4R	PIAT 1R	PIAT 2R	PIAT 3R	PIAT 4R
VIAT1R	1.000							
VIAT2R	.427	1.000						
VIAT3R	.476	.463	1.000					
VIAT4R	.457	.290	.495	1.000				
PIAT1R	.310	.262	.316	.308	1.000			
PIAT2R	.346	.278	.295	.318	.543	1.000		
PIAT3R	.127	.106	.200	.284	.367	.449	1.000	
PIAT4R	.141	.141	.213	.275	.363	.487	.456	1.000
VIAT1C	.314	.293	.291	.359	.371	.326	.325	.253
VIAT2C	.115	.247	.336	.315	.259	.188	.215	.169
VIAT3C	.337	.341	.403	.397	.259	.241	.268	.244
VIAT4C	.241	.165	.289	.376	.329	.302	.257	.192
PIAT1C	.239	.240	.288	.288	.344	.377	.312	.230
PIAT2C	.307	.313	.237	.289	.377	.331	.254	.223
PIAT3C	.213	.215	.138	.257	.333	.369	.268	.167
PIAT4C	.237	.218	.209	.200	.159	.357	.164	.165

	VIAT 1C	VIAT 2C	VIAT 3C	VIAT 4C	PIAT 1C	PIAT 2C	PIAT 3C	PIAT 4C
VIAT1C	1.000							
VIAT2C	.438	1.000						
VIAT3C	.444	.527	1.000					
VIAT4C	.449	.410	.381	1.000				
PIAT1C	.297	.287	.201	.247	1.000			
PIAT2C	.292	.226	.248	.277	.528	1.000		
PIAT3C	.326	.187	.235	.289	.449	.514	1.000	
PIAT4C	.220	.115	.129	.238	.381	.453	.396	1.000

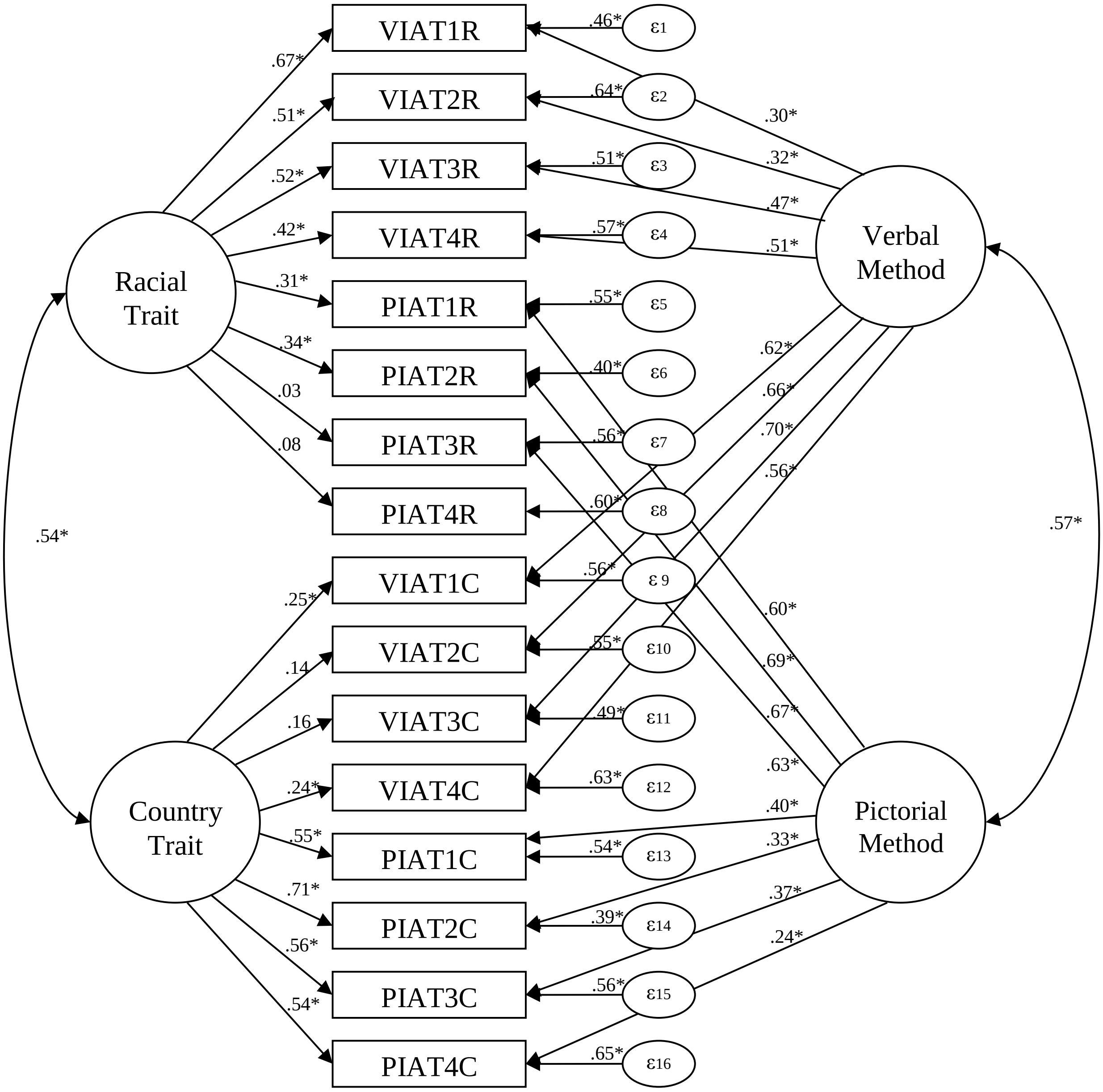

Figure 6.2. CT-CM CFA-MTMM model depicting the data of four IATs that have been separated into trait, error and method components. *p<.001.

Evidence of Method Effects in the IAT using CT-CM CFA-MTMM

As hypothesised, substantial and significant factor loadings of the indicators onto the Verbal and Pictorial latent method factors were found (see Figure 6.2). The average of these factor loadings was .50, well above the .32 cut-off for meaningful latent effects (Gorsuch, 1983). This provides strong evidence of significant method effects within IAT data.

Convergent Validity between the VIAT and PIAT using CT-CM CFA-MTMM

The second hypothesis of comparability between the verbal and pictorial IAT methodologies was substantiated with a sizeable and significant correlation of .57 between the Verbal and Pictorial latent methods, as shown in Figure 6.2. This result further enhances the convergent validity of the VIAT and PIAT.

Construct Validity Results for the IATs using CT-CM CFA-MTMM

Individual variance parameters resulting from the CT-CM CFA-MTMM analysis are presented in Table 6.2. These 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.

Table 6.2

Variance in IAT Effect Scores Accounted for by Trait, Method and Error Effects

	Trait	Method	Error
Race			
VIAT1R	.45*	.09*	.46*
VIAT2R	.26*	.10*	.64*
VIAT3R	.27*	.22*	.51*
VIAT4R	.18*	.26*	.57*
Mean (*SD*)	.29 (.11)	.17 (.08)	.55 (.08)
PIAT1R	.10*	.36*	.55*
PIAT2R	.12*	.48*	.40*
PIAT3R	.00	.45*	.56*
PIAT4R	.01	.40*	.60*
Mean (*SD*)	.05 (.06)	.42 (.05)	.53 (.09)
Country			
VIAT1C	.06*	.38*	.56*
VIAT2C	.02	.44*	.55*
VIAT3C	.03	.49*	.49*
VIAT4C	.06*	.31*	.63*
Mean (*SD*)	.04 (.02)	.41 (.08)	.56 (.06)
PIAT1C	.30*	.16*	.54*
PIAT2C	.50*	.11*	.39*
PIAT3C	.31*	.14*	.56*
PIAT4C	.29*	.06*	.65*
Mean (*SD*)	.35 (.10)	.12 (.04)	.54 (.11)
Overall			
Mean (SD)	.18 (.16)	.28 (.15)	.54 (.08)

Overall, the parameter estimates revealed an average of 28% of the variance was attributable to method effects (see Table 6.2). 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. Random error variance appeared relatively stable across the tasks, accounting for about 54% of variance (see Table 6.2). These results reveal that IAT effect scores are on average comprised of 54%

random error variance, 28% method variance and 18% trait variance (see Figure 6.3). In other words, an average of over 80% of an IAT effect score is error variance compared to less than 20% variance associated with the trait construct of interest.

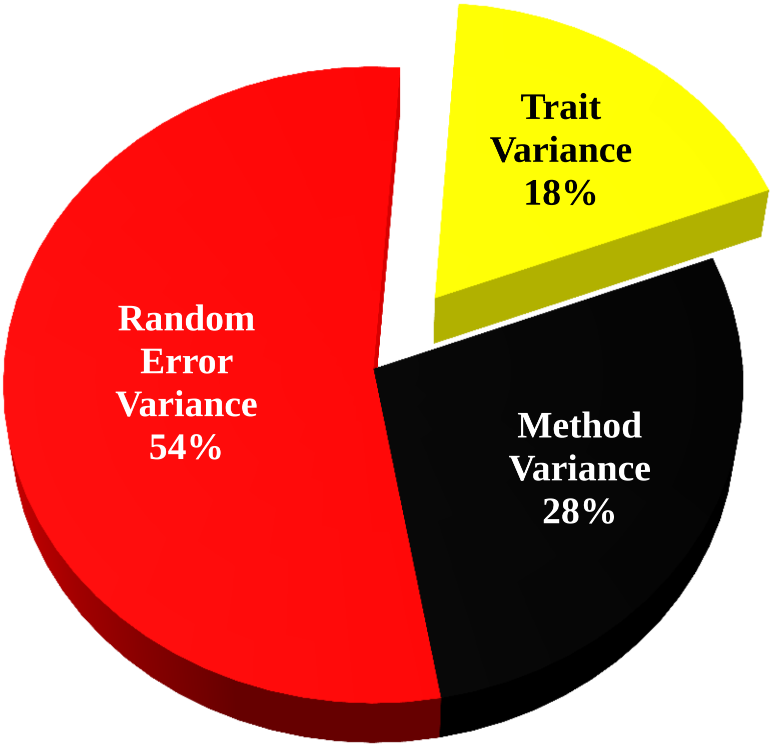

Figure 6.3. Graphical representation of the percentage variance of average IAT effect scores attributable to trait, method and random error variance.

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 (29% and 35% respectively) than method variance (17% and 12% respectively). This provides solid support for 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

(5% and 4% respectively) and a substantial portion of method variance present (42% and 41% respectively). Such findings reduce support for the construct validity of these two measures. As such, the CFA-MTMM analysis simultaneously provided strong construct validity evidence for the Racial VIAT and Country PIAT, whilst weakening support for the Racial PIAT and Country VIAT.

Discussion

There was a strong likelihood that substantial random and systematic error variance were confounding IAT effect scores, though the extent of this effect was unknown. The current study aimed to clarify this issue by applying CT-CM CFA-MTMM to investigate the influence of error variance on several aspects of the IAT's validity. The three sub-aims of this study were to determine the proportion of method variance in the scores, to compare the latent verbal and pictorial IAT methods, and to assess the construct validity of the IATs.

The results of Study Two found evidence of significant random and systematic error variance in the IAT effect scores, the sheer magnitude of which has serious implications for the use and interpretation of IAT effect scores. Furthermore, strong convergence between the verbal and pictorial IAT formats was demonstrated after the effects of random error and method effects were accounted for. Lastly, strong construct validity was shown for the Racial VIAT and Country PIAT, indicating that it may be possible to develop psychometrically solid implicit attitude measures. However, the inconsistencies evident between the validities of ostensibly similar IATs, such as the Country VIAT and Country PIAT, provide concern for the ease in which IATs can be applied to assess various constructs. Each of these findings and their implications for implicit attitudinal research will be discussed in turn.

Evidence of Systematic and Random Error Effects in the IAT using CT-CM CFA-MTMM

Method effects were revealed to play a considerable role in the IAT effect scores, as evidenced by substantial factor loadings of the indicators onto the latent method factors (see Figure 6.2) and the parameter estimates of the CFA-MTMM analysis (see Table 6.1). On average, about 30% of the IAT scores were attributable to method variance, an amount not dissimilar to the 22% method variance in IATs reported by Siers and Christiansen (2013). Such results provide reasonably consistent evidence of the significant role systematic error variance plays in IAT effect scores, which may have impacted upon the accuracy of previously reported IAT findings. The inconsistent, and often poor inter-implicit attitude correlations often reported in the implicit attitudinal literature (refer to Chapter Three; see also Krause et al., 2010; Rudolph et al., 2008; Sherman et al., 2003) may, according to present results, have actually been inflated to an unknown degree by method variance. Such inflation may imply that estimates of convergent validity for the IAT have been inaccurate, and potentially poorer than originally anticipated.

The present analyses also confirmed that random error variance comprised over 50% of the IAT effect scores, indicating it is the greatest contributor to IAT effects. Random error variance can significantly confound validity estimates, and as shown in Chapter Five, may have provided an upper limit for observed correlations between like implicit attitude measures (e.g. Hofmann et al., 2005; Krause et al., 2010;

Rudolph et al., 2008). Random error variance has likely also limited other construct validity estimates for these tasks.

It is significant that an average of 82% of the IAT effect scores was attributable to error variance, and this indicates less than one fifth (18%) of the IAT effect score measures the trait attitude construct supposedly being assessed (see Figure 6.3). Based on the current findings, it would seem near impossible to ascertain accurate estimates of implicit attitudes without analytically addressing error variance, given the IAT effect scores are confounded to such an extent. This calls into question the validity of previously reported implicit attitudinal research that were not analysed using SEM or other latent-based approaches. In summary, error variance in IAT effect scores *must* be accounted for using latent modelling techniques in order to have any hope of uncovering findings related to implicit attitudes. The implications of requiring IAT data be analysed using SEM are discussed in detail in the General Discussion, Chapter Eight.

Convergent Validity Evidence for the Verbal and Pictorial IAT Formats

The CT-CM CFA-MTMM methodology enabled a comparison of the latent method factors. This analysis revealed a substantial correlation between the latent verbal and pictorial methodologies, indicating the methods were very comparable. These results build on the findings of Study One that also found relative equivalency between the verbal and pictorial IAT formats. These findings are a substantial improvement on previous comparisons of the VIAT and PIAT that reported a much smaller correlation between these measures using traditional data analytical approaches based on observed scores (Thomas, 2008). Such an improvement in comparability implies error variance was having a limiting effect on the inter-implicit attitude correlation for related constructs, which is to be expected given random error variance is now known as the primary contributor to IAT effect scores. The present findings also provide evidence for the PIAT as a viable alternative to the traditional VIAT for attitudinal research, especially in populations where verbal stimuli are not appropriate (e.g. young children as per Thomas et al., 2007), though as indicated previously, variance in PIAT effect scores *must* be accounted for using latent modelling.

Discrepant Construct Validity Evidence for the IAT using CFA-MTMM

Despite some evidence of support for the construct validity of the IATs, this support was inconsistent. An examination of the parameter estimates (see Figure 6.2, and as variances in Table 6.1) indicated that the Racial VIAT and Country PIAT possessed strong construct validity, with levels of trait variance greater than method variance. However, the Country VIAT and Racial PIAT presented the opposite findings, thereby failing to demonstrate adequate construct validity because method variance was greater than trait variance for these tasks. Such results are very difficult to interpret given each method (verbal and pictorial) was both deemed satisfactory and unsatisfactory depending on which construct (race or country) 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 contradictory results such as the aforementioned reduce the overall construct validity evident of the IAT for, and emphasise that full psychometric investigation of each adaptation of the IAT is necessary (see also Lane et al., 2007). This is a labour-intensive requirement, as demonstrated by the current dissertation. Greater collegiality between research groups may enable the use of validated IATs, such as the Racial VIAT and Country PIAT of the current research, to be circulated (e.g. The Open Science Framework; Spies & Nosek, 2012). This could increase the overall psychometric standard of the implicit attitude measures being utilised, without the need for extensive psychometric evaluation every time research is conducted, a likely prohibitive requirement for most researchers.

Chapter Summary and Conclusion

The aim of Study Two was to investigate the influence of systematic and random error variance on the IAT effect scores using CFA-MTMM. The results revealed error variance to be the primary contributor to IAT effect scores, with more than half of variance attributable to random error, a further third attributable to method variance and trait variance shown to have the least influence on IAT results, accounting for less than a fifth of the score. These findings significantly reduce the veracity of previously reported IAT findings that failed to account for error variance. A key implication of these results is that future IAT research should account for the substantial portion of error variance by using latent modeling analytical techniques such as SEM. Once error variance was accounted for, however, 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 assessed using SEM procedures to ensure adequate psychometric properties are present, prior to testing with the aim of procuring implicit attitudes.

On the flip side, the present study revealed for the first time that two IATs, the Racial VIAT and the Country PIAT, demonstrated adequate construct validity via the stringent assessment process afforded by CFA-MTMM. Such a finding provides some hope that the IAT method can be developed to provide reasonable assessment of underlying implicit attitudes. The finding of adequate construct validity for a fully pictorial IAT (the Country PIAT) combined with strong VIAT-PIAT convergence supports the use of PIATs in implicit attitudinal research. Because of this, were IAT researchers able to consistently provide psychometrically adequate measures, there should be no barrier between the use of pictorial stimuli in comparison to verbal stimuli. However, the requirement of latent modeling techniques to analyse the data before any interpretation could occur would still be essential.

The following chapter provides an examination of how SEM techniques could be applied to examine the substantive enquiries potentially raised during implicit attitudinal research, with the confounding influence of error variance minimised. These enquiries include what the IAT scores reveal about a sample population’s implicit biases and whether sex, age or travel experience impact a person’s implicit views. It is very novel to attempt to examine such queries using SEM procedures; however to not account for error variance in this way is to likely produce inaccurate and potentially misleading conclusions.

CHAPTER SEVEN

Study Three: Examining Covariates and the IAT Effect Score using Multiple-group CFA and MIMIC models

The IAT effect score was designed to provide an indication of entrenched automatic biases. Yet the current dissertation has revealed that error variance, not trait variance, forms the majority component of these IAT effect scores. Despite this substantial limiting factor, Study Two showed that once error variance was accounted for using CFA-MTMM, the Racial VIAT and Country PIAT possessed good construct validity. This result implies that IATs have at least some potential to be developed and refined into psychometrically robust measurement instruments. Accepting this premise, the current study explored whether SEM can be used to investigate substantive enquiries for the IAT. Specifically, it was determined whether implicit biases were present for the sample, and if so, whether certain participant characteristics influenced these results. It is argued that SEM techniques such as Multiple-groups CFA and MIMIC models provide suitable means for avoiding the issue of error variance whilst obtaining substantive information from IAT data.

The Application of Implicit Association Tests to Prejudice Assessment

The purpose of the IAT is to deliver an estimate of deeply ingrained attitudinal biases. As outlined in Chapter Two, several design features of the IAT make it well suited to the assessment of racial prejudice (see also Greenwald & Krieger, 2006; Greenwald et al., 2009). Because of this, the IAT has been frequently applied to assess racial prejudice, with many studies examining White populations implicit attitudes towards other racial groups, such as Arabs, Black Africans, Hispanics, Asians and Jews (Agerström & Rooth, 2009; Baron & Banaji, 2006; Cunningham, Nezlek, et al., 2004; Greenwald et al., 1998; Greenwald et al., 2009; Nosek et al., 2005; Nosek et al., 2007; Park et al., 2007; Rooth, 2010; Rowatt et al., 2005; Rudman & Ashmore, 2007). Of these racial groups, Australians have rated Arab/Muslims as the most threatening out-group (Dunn et al., 2008; Islam & Jahjah, 2001) and it is for this reason that implicit attitudes towards this specific ethnic group were examined in the current research.

Implicit bias against Arab/Muslims has previously been investigated in large-scale, web-based research by Nosek et al. (2007). Nosek's team used a traditional verbal IAT format, which presented Pleasant/Unpleasant words along with Arab/Muslim names and Other foreign names that would be unfamiliar to a US audience. Using this task, an IAT effect was found whereby Other people's names were implicitly preferred over Arab/Muslim names (Nosek et al., 2007). Interestingly, this particular task consistently produced some of the largest group differences of all the IATs available on the Project Implicit website (Greenwald et al., 2011). Greater anti-Arab attitudes were observed among men compared to women, older compared to younger people and conservatives compared to liberals (Nosek et al., 2007).

The Use of IAT Effect Scores for Estimating Implicit Prejudice

Implicit prejudice for the IAT is typically estimated using the IAT effect score (D; Greenwald et al., 2003), which is basically calculated by examining the difference in reaction times between congruent and incongruent experimental blocks. Unlike the IAT effect, which provides a population-based overview of the sample's implicit preferences, the IAT effect score is treated more like an individual diagnostic tool

(Fiedler et al., 2006). Greenwald et al.'s (2003) interpretation guidelines state that an IAT effect score greater than .60 implies strong negative implicit prejudice, an effect score between .35-.60 demonstrates moderate negative prejudice, .15-.35 implies slight prejudice, and scores lower than .15 represent non-existent prejudice (Greenwald et al., 2003). To provide an example of such interpretation, for Nosek et al.'s (2007) study the sex difference for the Arab-Other IAT revealed male participants possessed a moderate negative prejudice (D=.48), whereas female participants showed only a slight negative prejudice (D=.24). Findings such as these imply that the IAT can be used to provide an accurate assessment of personal implicit attitudes (Fiedler et al., 2006).

The Implications of Current Findings on Previous IAT Research.

The practice of relying on IAT effect size guidelines in order to classify the strength of participants' implicit prejudice should perhaps be questioned given the findings of the previous two studies in the current dissertation. In Study Two it was revealed that over 80% of the IAT effect score was attributable to random and systematic error variance. This has substantial implications for interpretation of IAT findings, given that only one fifth of an IAT effect score may reflect the implicit attitudes of interest. Because of this, any IAT research that fails to account for error variance is likely to be significantly confounded and potentially quite misrepresentative of the implicit attitudes of the sample population.

Previous research has been severely limited by not accounting for error variance. To avoid this oversight, the IAT effect cannot be examined in the usual way by using the D score or ascertaining group differences in the average reaction times of the congruent and incongruent data via a *t*-test (or other such analysis). This is because the data remains confounded by error using these methods. Rather, more advanced latent statistical procedures are required to model and partial out error variance before determining if significant implicit bias is present. The multiple-groups approach to CFA provides one avenue to achieve this, by enabling a comparison of latent means for the congruent and incongruent IAT trial data. This technique is far more involved than a *t*-test as it requires first comparing many aspects of the two different groups to ensure comparability before a test of difference between latent means can occur. Yet given the enormity of the error variance inherent in the IAT data, the simpler traditional analytical approach is not feasible as would likely result in vast misrepresentation of the findings.

Assessing Measurement Invariance and Latent Mean Differences using Multiple-groups CFA

Multiple-groups Confirmatory Factor Analysis (CFA) enables the latent scores of a trait for two different groups to be compared. This process tests the measurement and structural parameters of two models simultaneously (one model per group) to reveal any group differences (see Chapter Four). The comparison of latent means afforded by multiple-groups CFA is typically used to determine differences between groups of participants. However, it could also be applied to examine whether there are significant differences between the latent means of the congruent and incongruent trial data, within a single participant sample. In this situation, the congruent and incongruent trial data are analysed as congruent and incongruent “groups”, even though they are not groups per se as are sourced from a single sample population. Significant discrepancy between the mean congruent and incongruent reaction times would imply a positive IAT effect, suggesting the IAT had worked as anticipated. However, for the comparison of latent means to be a useful and meaningful representation of the group differences, first several other aspects of the two models need to be shown to be comparable (Brown, 2006). This process of testing for equivalency between groups is referred to as measurement invariance and it employs a step by step process of applying constraints, such as like parameters, to both of the groups simultaneously.

Determining Measurement Invariance

Testing for measurement invariance is crucial because the latent means of two groups cannot be compared before it is first established that a score of X for one group is equivalent to a score of X for the other group. If the trait scores are not comparable across groups then any group differences could be artifactual and may be fundamentally misleading (Reise et al., 1993). The process of measurement invariance comprises numerous analyses involving increasingly restrictive constraints. The more constraints that are shown to be equivalent across the groups, the greater strength of factorial invariance revealed (see Brown, 2006). For instance, strong factorial invariance is evidenced when the factor structure (configural invariance), factor loadings (metric invariance), and intercepts (scalar invariance) are shown to be equivalent across the two groups' models. Once a reasonable level of invariance has been established it is then possible to determine if there are group differences in the latent means.

Assessing Latent Mean Differences between Groups

The test of latent mean difference is somewhat analogous to the comparison of observed group means that is traditionally completed using a *t*-test. However, the current approach compares the *latent* means of the two groups, i.e. once error variance has been accounted for. To assess the equivalence of latent means via multiple-groups CFA, the latent mean of one model is constrained to zero. If this results in a significant latent mean for the second group it indicates there is a significant difference between the two groups' means (Thompson & Green, 2006). For the IAT, such a result would reveal that an IAT effect had occurred and would ostensibly provide information regarding the population's implicit attitudes towards the constructs of interest.

Assessing the Impact of Covariates on IAT Effect Scores using MIMIC Modelling

Measurement invariance and latent mean difference analyses can deliver an appraisal of whether the IAT effect was present for each of the empirical IATs in this thesis, thereby providing insight into the overall implicit attitudes of the sample. However, there may be additional characteristics of the participants which further influence the IAT effect scores. Previous research has shown that the age, sex and political persuasion of the participants can significantly affect the strength of the IAT effect scores (Nosek et al., 2007). Specifically, males, older adults and politically conservative individuals generally produce larger IAT effect scores on tests of implicit racial bias (including against Arabs) than females, younger adults and politically progressive participants (Nosek et al., 2007). It is likely there are many other factors that could also impact IAT effect scores and the characteristics of interest would depend on the construct being assessed. For a racial IAT it could be expected that participants with greater travel experience would be more accepting of diversity, resulting in smaller IAT effect scores, than less travelled participants. Past research has found participants who associate more with minority members, or those who have completed diversity training, produce lower IAT effect scores than participant's who have not (Cashin, 2010; Rashid, 2009). As such, there is potential that a participant's sex, age, political persuasion and/or travel experience may significantly influence their IAT effect scores, with participants that are male, older, conservative and less experienced travellers potentially revealing greater implicit racial bias. The MIMIC analytical approach provides one avenue for investigating such substantive enquiries whilst accounting for the issue of error variance noted in this thesis.

MIMIC (Multiple Indicators, Multiple Causes) Models examine the influence of covariates, such as sex and age on latent factors such as implicit attitudes (see Chapter Four). MIMIC models can theoretically be conceived of as a regression model being added to a CFA model in order to examine the covariates direct effects onto the latent factors and selected indicators (Jöreskog & Goldberger, 1975). MIMIC models only require one input dataset and thus provide a very efficient analytical approach to assessing substantive enquiries (Brown, 2006). In a MIMIC model, significant direct effects between a covariate and a latent factor imply the factor means are different for different levels of the covariate. Using this approach, it is possible to determine if the covariates of sex, age and travel experience impact significantly a participant's implicit attitudes, as measured by the IAT effect score.

Study Three

The aim of Study Three was to investigate whether latent modelling techniques that account for error variance could facilitate more accurate estimates of implicit attitudinal biases for various participant populations.

Aim 6: Assess Equivalency of Congruent and Incongruent Responses using Multiple-groups CFA

The first aim of the present study, and the sixth aim of the thesis, was to determine whether an IAT effect had occurred for each of the empirical IATs using Multiple-groups CFA to test for equivalency between congruent and incongruent trial data. It was expected that a strong (or at least partial) level of factorial measurement invariance would be demonstrated between the latent congruent and incongruent models. It was hypothesised further that the assessment of latent mean difference would reveal a significant discrepancy in the latent means for the congruent and incongruent experimental blocks (“groups”). The expected positive IAT effect would be evident by a significant and positive latent mean for the incongruent group following the congruent latent group mean being constrained to zero. A conceptual path diagram for this analysis is depicted in Figure 7.1. An IAT effect for the Country VIAT, for example, would imply an implicit preference for Europe over the Middle East.

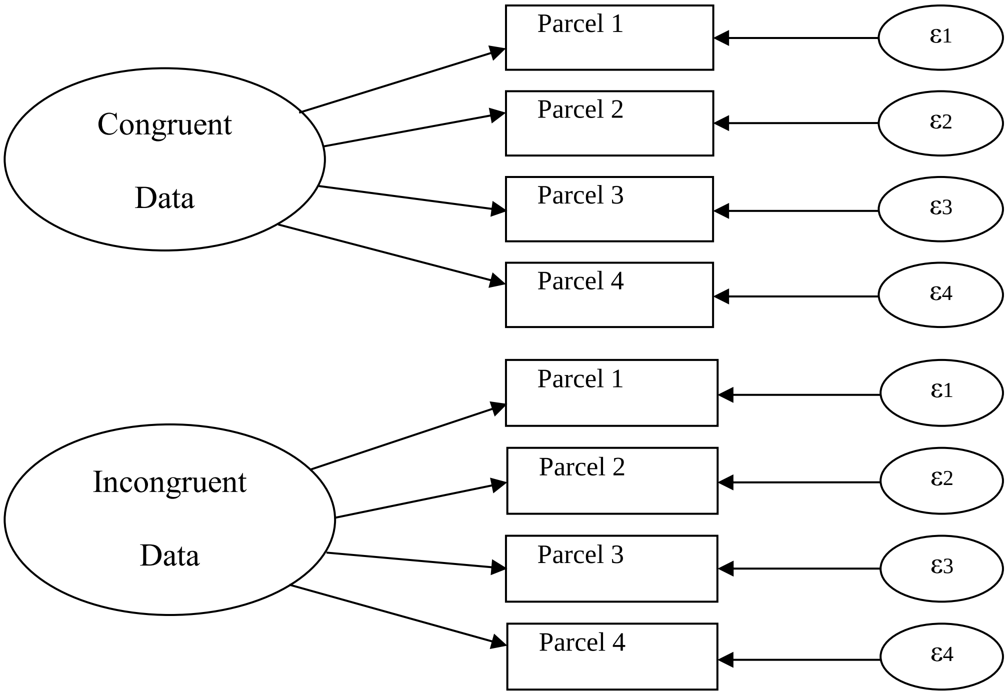

Figure 7.1. Conceptual model of the Multiple-groups CFA assessment for latent mean differences.

Aim 7: Assess the Impact of Covariates on the IAT Effect Scores using MIMIC Models

The second aim for this study, and the seventh aim of the thesis, was to determine if the participant characteristics of sex, age and travel experience significantly affected the IAT results using Multiple Indicators, Multiple Causes (MIMIC) modelling (see Figure 7.2). A significant pathway between a covariate and a latent construct would indicate significant differences in the latent IAT scores for the different categories of the participant-related covariate. Given previous research findings, male participants were hypothesised to produce higher IAT effect scores than female participants for the racially-relevant IATs (Nosek et al., 2007). It was also expected that older participants would demonstrate significantly higher IAT effect scores than younger participants (see Nosek et al., 2007). Lastly, participants with greater travel experience were expected to deliver lower IAT effect scores than less travelled participants, indicating decreased bias towards Arabs/ the Middle East.

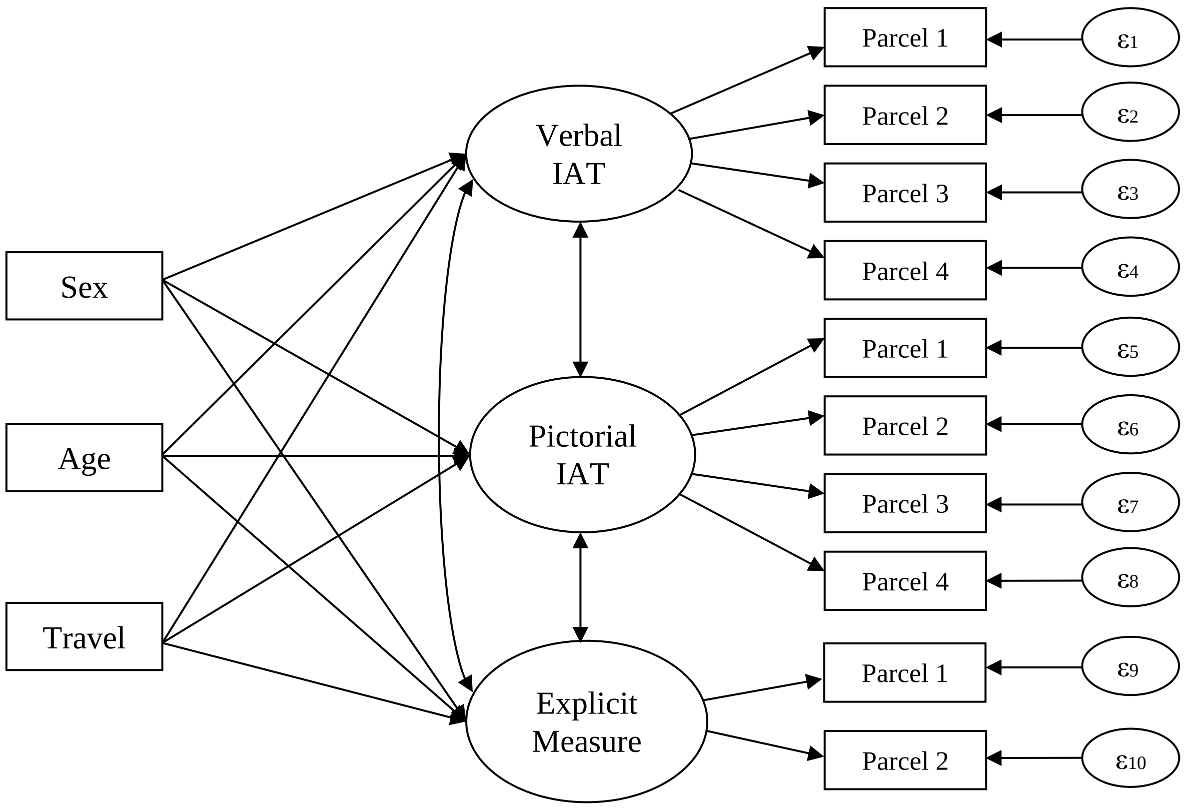

Figure 7.2. MIMIC model to evaluate the direct effects of sex, age and travel experience on the test scores.

Method

Participants

The same 198 participants (mean age: 26.03 years; *SD*=11.10 years) were used in this study as for the previous studies. For more information regarding these participants see Chapter Five.

Apparatus

The apparatus are as described in Chapter Five. Specifically, the four IATs relevant to the present analyses are the Race VIAT, Race PIAT, Country VIAT and Country PIAT. The adapted Modern Racism Scale (MRS; McConahay et al., 1981) and the Travel Destination Questionnaire (TDQ) were also used to address the current aims (see Appendix H for the Student Opinions questionnaire that includes the MRS, and Appendix I for the TDQ).

Demographic Information

Prior to testing, the participants reported their age, sex and ethnic identity. Participants were also asked how well travelled they believed they were (their Travel Experience) on a five point Likert scale, with higher scores indicative of greater travel experience (see Appendix J). The mean stated level of travel experience for the participants was 3.08 (*SD*=1.15), which implies the majority of participants had visited at least some countries overseas. The aforementioned demographic information provided the covariates for the MIMIC analysis.

Procedure

The procedure was as outlined in Chapter Five. The demographic data used for Aim Seven were obtained from the participants before they began the experimental tasks.

Statistical Analyses

Multiple-groups CFA

For the present study, multiple groups refer to data obtained during congruent (group 1) and incongruent (group 2) experimental blocks of a single IAT rather than distinct participant groups[24]. The aim for the Multiple-groups CFA was to discover if there was a significant difference between the congruent and incongruent latent means via a test of latent mean difference. However, equivalency between the two models was first required as determined by the step-by-step process of measurement invariance, which is outlined below. The outlined analytical procedure was repeated for each of the four IATs using the robust maximum likelihood estimation approach (MLM) as available in Mplus, version 6.1 (Muthén & Muthén, 2010). The Satorra-Bentler scaled chi-square calculator was used to calculate the Satorra-Bentler chi-square difference ($\Delta SB\chi 2$; Satorra & Bentler, 2001). The mean latency and standard deviations for the IAT data by construct, format and congruency are presented in Table 5.1, Chapter Five.

---

[24] It is acknowledged that because the two groups comprise data from the same participant population the test-retest CFA approach may have been appropriate. This technique was originally applied to the present study, but the models did not show convergence. Longitudinal measurement invariance has been shown to be just as validly tested using the multiple-groups approach (Vandenberg & Lance, 2000), hence the presented analyses.

The Assessment of Measurement Invariance.

Prior to the tests for equivalency, the basic CFA model was first assessed to ensure good model fit for the congruent, incongruent and total sample in three separate analyses per IAT. Once the models were found to fit the data, three types of measurement invariance were then tested: configural, metric and scalar. These models were tested sequentially following the guidelines outlined in Brown (2006), with each step assessing a different form of invariance.

The first step of the invariance procedure was an assessment of configural invariance, which refers to a comparison of the model form for the congruent and incongruent data. Configural invariance is the weakest form of invariance, whereby no parameters of the CFA model are constrained to be equal across the congruent and incongruent data “groups”. Configural invariance was achieved if the fit indices demonstrate good model fit. The second step was an assessment of metric invariance, whereby the factor loadings were constrained to be equal across the congruent and incongruent data. This provided the second model that was tested directly against the preceding less restrained configural model to ensure the second model did not provide significantly worse fit. Metric invariance was assumed if the Satorra-Bentler chi-square difference test ($\Delta SB\chi^2$) showed that the Satorra-Bentler chi-square value for the more constrained model was not significantly different to the Satorra-Bentler chi-square value for the less constrained model (see Satorra & Bentler, 2001). If a significant difference did occur, one of the loadings was released and the analysis was repeated until a non-significant difference was found. The third step involved an estimate of scalar invariance. In this model both the factor loadings and the measurement intercepts of the factors were constrained to be equal across the congruent and incongruent data "groups". If this model did not show significantly worse fit than the metric invariance model from step two then strong factorial invariance between the congruent and incongruent data was established.

The Assessment of Latent Mean Difference.

Following a strong (or at least partial) level of measurement invariance being established for the congruent and incongruent data, the test of latent mean difference could occur. This was the fourth step in this series of analyses, and involved the latent means being compared to determine equivalency. In this analysis the latent mean of the congruent group was constrained to be zero. If this constraint led to a significant and positive latent mean for the incongruent group then it provided support for the hypothesis of a positive IAT effect. If however, the resultant latent mean possessed a negative sign it would be indicative of a reverse IAT effect, implying the participants responded significantly faster to the incongruent than the congruent stimuli pairings[25]. The test of latent mean difference was expected to reveal a significant discrepancy between the congruent and incongruent data, indicating a positive IAT effect. The expected outcome of a significant difference for these analyses was in direct contrast to the previous invariance analyses that aimed to show no difference between these data groups. The assessment of latent mean difference provided the crucial indicator regarding the overall implicit attitudes of the participant group.

[25] Although a rare finding, such an effect was evident for the standard Flower/Insect VIAT when completed by a group of entomologists (Citrin & Greenwald, 1998 as cited in Lane et al., 2007).

MIMIC Models

The MIMIC approach allowed for an examination of the effects of the covariates sex, age and travel experience on the latent VIAT, PIAT and Questionnaire factors. Only one input matrix was required for the analysis, with the covariates added to the total dataset. For the present study, the covariates age and travel experience were continuous variables, whereas sex was coded categorically. The MIMIC process required a viable CFA model be tested for the complete dataset. Following this, the covariates were added and their effects on the latent factors examined. A significant direct effect of a covariate onto a latent factor implied there were significant differences in the latent means at various levels of the covariate. For the age covariate, if this path was significant and positive it would imply the latent means of the older participants were significantly higher than the latent means of the younger participants, as hypothesised. A significant and negative score on that same path would imply the younger participants were displaying more prejudiced attitudes. Unlike in multiple-groups CFA, measurement invariance was not tested for the covariates of sex, age or travel experience. The MIMIC analysis was repeated separately for the racial and country attitude constructs.

Results

Each series of analyses for the Multiple-groups CFA are presented below. They are ordered such that the racial construct is addressed first, followed by the country construct, and within each construct the verbal version of the task is presented before the pictorial. The final section of the results depicts the MIMIC analyses which assess the impacts of specific participant characteristics on the IAT effect scores.

Multiple-groups CFA – Equivalency of Congruent and Incongruent Responses

Equivalency of Responses by Congruency for the Racial VIAT

The simple one-factor CFA model was initially tested for the combined, congruent and incongruent data separately to ensure it adequately reflected the variances and covariances of the input data. The goodness-of-fit indices for the CFA model for all data for the Racial VIAT were: S-Bχ^2 (2, N=198)=2.77, *p*=.25; CFI=.999; RMSEA=.022; SRMR=.007, which revealed good model fit. The goodness-of-fit indices for the CFA model for the Racial VIAT congruent data were: S-Bχ^2 (2, N=198)=5.77, *p*=.06; CFI=.990; RMSEA=.069; SRMR=.020. A result that showed acceptable model fit, as three of the four fit indices demonstrated good fit and the RMSEA indicated a moderate level of fit (implying it was not the simplest or most parsimonious model possible). The goodness-of-fit indices for the CFA model for the Racial VIAT incongruent data showed good model fit: S-Bχ^2 (2, N=198)=5.08, *p*=.75; CFI=1.000; RMSEA>.001; SRMR=.005. As the fit indices met criterion levels for the one-factor CFA model for the combined, congruent and incongruent data, it was appropriate to continue with the test of measurement invariance.

Measurement Invariance of Congruency for the Racial VIAT.

To assess the comparability of the congruent and incongruent data three types of measurement invariance were tested: configural (equal form), metric (equal loadings) and scalar (equal intercepts). The outcomes of these tests are presented in Table 7.1.

Table 7.1

Tests of Measurement Invariance for Congruency for the Racial VIAT

Model & Invariance Level	*Df*	SBχ^2	CFI	RMSEA	SRMR	Model comparison	Δdf	ΔSBχ^2
1. Configural	4	7.43	0.996	0.047	0.014	-	-	-
2. Metric	7	15.03*	0.991	0.054	0.041	2 v 1	3	7.76
3. Scalar	10	24.46*	0.984	0.060	0.048	3 v 2	3	10.68*
3a. Scalar - RVIAT2 released	9	18.63*	0.989	0.052	0.043	3a v 2	2	3.30

Note. *N* =198. VIAT = Verbal Implicit Association Task. *dfs* = degrees of freedom. S-Bχ^2 = Satorra-Bentler Chi-square fit indices. CFI = Comparative Fit Index. RMSEA = Root Mean Square Error of Approximation. SRMR = Standardized Root Mean Square Residual. Δ = difference. *p < .05. **p < .001.

Configural invariance (Model 1 in Table 7.1) is the weakest form of invariance as no parameters of the CFA model were constrained to be equal across the groups. The fit indices for this test (see Model 1, Table 7.1) demonstrated good model fit and thus supported configural invariance. For metric invariance (Model 2 in Table 7.1) the factor loadings were constrained to be equal across congruency. The resulting model also showed good model fit, except for the significant Satorra-Bentler chi-square (SBχ^2). The chi-square is known to be inflated by sample size and routinely rejects models based on data from a large population (Brown, 2006). As such, the other fit indices were relied upon and the outcome of good model fit sustained. Model 2 was compared with Model 1, which resulted in a non-significant ΔSBχ^2, which supports the metric invariance of the congruent and incongruent data. For scalar invariance (Model 3 in Table 7.1), both the factor loadings and the measurement intercepts of the factors were constrained to be equal across congruency. The Satorra-Bentler chi-square difference test was significant, indicating non-invariance for Model 3 with the preceding model. The modification indices indicated that the RVIAT2 data parcel may have different intercepts for the congruent and incongruent data. As such, RVIAT2 was released as and the analysis was re-run (Model 3a in Table 7.1). By releasing this intercept the resulting model was found to show scalar invariance by congruency. This series of analyses provided evidence of partial measurement invariance between the congruent and incongruent data of the RVIAT. Following partial measurement invariance, it is often still possible to assess the equivalency of latent means (Brown, 2006). Given the aim was predominantly to determine if there were latent mean differences, partial measurement invariance is adequate for the current purpose.

Latent Mean Differences of Congruency for the Racial VIAT.

Latent mean invariance was examined for each IAT. This analysis involved constraining the latent mean of one model to be zero, if this resulted in a significant latent mean for the other model it showed there was a significant difference between the means of the two groups (Thompson & Green, 2006). The model resulting from the assessment of latent mean difference was compared against the scalar invariance model (Model 3a in Table 7.2), with non-invariance assumed if a significant $\Delta SB\chi^2$ was shown. Given a significant IAT effect was expected, non-invariance was the hypothesised outcome of this analysis. The results of this analysis are presented in Table 7.2.

Table 7.2

Tests of Latent Mean Difference for Congruency for the Racial VIAT

Model & Invariance Level	*df*	SBχ^2	CFI	RMSEA	SRMR	Model comparison	Δ*df*	ΔSBχ^2
3a. Scalar - RVIAT2 released	9	18.63*	0.989	0.052	0.043	-	-	-
4. Latent Means	10	101.17**	0.899	0.152	0.224	4 v 3a	1	152.63**

Note. *N* =198. VIAT = Verbal Implicit Association Task. *dfs* = degrees of freedom. S-Bχ^2 = Satorra-Bentler Chi-square fit indices. CFI = Comparative Fit Index. RMSEA = Root Mean Square Error of Approximation. SRMR = Standardised Root Mean Square Residual. Δ = difference. *$p < .05$. **$p < .001$.

The above results confirmed a significant difference (population variance) between the congruent and incongruent data for the Racial VIAT. This implies an IAT effect had occurred for this task. The output from the scalar invariance analysis provided further information regarding this difference. It indicated that when the congruent latent mean was constrained to zero, the resulting incongruent latent mean was positive (β=.81). This means the participants took longer to categorise the incongruent pairings of stimuli, as hypothesised.

Equivalency of Responses by Congruency for the Racial PIAT

The simple one-factor CFA model was tested for the combined, congruent and incongruent data of the Racial PIAT separately. The goodness-of-fit indices for the Racial PIAT combined data were: S-Bχ^2 (2, N=198)=4.89, p=.09; CFI=.996; RMSEA=.043; SRMR=.012, which revealed good model fit. The goodness-of-fit indices for the Racial PIAT congruent data were: S-Bχ^2 (2, N=198)=5.77, p=.06; CFI=.990; RMSEA=.069; SRMR=.019. This showed acceptable model fit as three of the four fit indices demonstrated good fit, and the RMSEA indicated a moderate level of fit. The goodness-of-fit indices for the CFA model for the Racial PIAT incongruent data were: S-Bχ^2 (2, N=198)=4.59, p=.10; CFI=.994; RMSEA=.057; SRMR=.018, indicating good model fit. The fit indices thus met criterion levels for the one-factor CFA model for the combined, congruent and incongruent data.

Measurement Invariance of Congruency for the Racial PIAT.

Again the congruent and incongruent data were compared using three types of measurement invariance: configural (equal form), metric (equal loadings) and scalar (equal intercepts). The outcomes of these tests are presented in Table 7.3.

Table 7.3

Tests of Measurement Invariance for Congruency for the Racial PIAT

Model & Invariance Level	*df*	SBχ^2	CFI	RMSEA	SRMR	Model comparison	Δ*df*	ΔSBχ^2
1. Configural	4	10.40*	0.992	0.064	0.018	-	-	-
2. Metric	7	21.26*	0.983	0.072	0.049	2 v 1	3	11.44*
2a. Metric – RPIAT3 released	6	12.59	0.992	0.053	0.026	2a v 1	2	1.38
3. Scalar	9	22.25*	0.984	0.061	0.034	3 v 2a	3	11.88*
3a. Scalar – RPIAT4 released	8	17.80*	0.988	0.056	0.030	3a v 2a	2	5.98

Note. *N* =198. PIAT = Pictorial Implicit Association Task. *dfs* = degrees of freedom. S-Bχ^2 = Satorra-Bentler Chi-square fit indices. CFI = Comparative Fit Index. RMSEA = Root Mean Square Error of Approximation. SRMR = Standardized Root Mean Square Residual. Δ = difference. **p* < .05. ***p* < .001.

Configural invariance (Model 1 in Table 7.3) was supported as the fit indices showed adequate model fit. Metric invariance (Model 2 in Table 7.3) was assessed by constraining the factor loadings to be equal across congruency before comparing Model 2 with Model 1. The first Satorra-Bentler chi-square difference test resulted in a significant result. As such, the modification indices were consulted and the loading for RPIAT3 was released. The resulting model (Model 2a in Table 7.3) was compared with Model 1, which produced a non-significant $\Delta SB\chi^2$. Thus the metric invariance of the congruent and incongruent data was supported for the Racial PIAT. The scalar invariance test (Model 3 in Table 7.3) also resulted in a significant Satorra-Bentler chi-square difference test. The intercept for RPIAT4 was consequently released and the resulting model (Model 3a in Table 7.3) was found to show scalar invariance. This series of analyses revealed partial measurement invariance of the congruent and incongruent data for the Racial PIAT, which as for the Racial VIAT, was adequate to proceed to the assessment of latent mean difference.

Latent Mean Difference for Congruency for the Racial PIAT.

The analysis of latent means was compared against the scalar invariance model (Model 3a in Table 7.4) using $\Delta SB\chi^2$. Table 7.4 depicts the comparison of latent mean results.

Table 7.4

Test of Latent Mean Difference for Congruency for the Racial PIAT

Model & Invariance Level	*df*	$SB\chi^2$	CFI	RMSEA	SRMR	Model comparison	Δdf	$\Delta SB\chi^2$
3a. Scalar – RPIAT4 released	8	17.80*	0.988	0.056	0.030	-	-	-
4. Latent Means	9	19.01*	0.988	0.053	0.030	4 v 3a	1	.57

Note. *N* =198. PIAT = Pictorial Implicit Association Task. *dfs* = degrees of freedom. S-Bχ^2 = Satorra-Bentler Chi-square fit indices. CFI = Comparative Fit Index. RMSEA = Root Mean Square Error of Approximation. SRMR = Standardized Root Mean Square Residual. Δ = difference. $*p < .05$. $**p < .001$.

The results of this analysis indicated a non-significant difference (or invariance) between the congruent and incongruent data for the Racial PIAT. Thus the Racial PIAT did not produce the expected IAT effect.

Equivalency of Responses by Congruency for the Country VIAT

For the Country VIAT, the simple one-factor CFA model was tested for the combined, congruent and incongruent data separately. The goodness-of-fit indices showed good model fit for the combined data: S-Bχ^2 (2, N=198)=.321, *p*=.85; CFI=1.000; RMSEA<.001; SRMR=.003. Good model fit was also revealed for the Country VIAT congruent data: S-Bχ^2 (2, N=198)=1.32, *p*=.52; CFI=1.000; RMSEA <.001; SRMR=.011. This was also the case for the Country VIAT incongruent data: S-Bχ^2 (2, N=198)=.805, *p*=.67; CFI=1.000; RMSEA< .001; SRMR=.006. The fit indices met criterion levels for the one-factor CFA model for the congruent, incongruent, and combined Country VIAT data.

Measurement Invariance of Congruency for the Country VIAT.

Configural, metric and scalar types of measurement invariance between the congruent and incongruent data were tested. The outcomes of these tests are presented in Table 7.5.

Table 7.5

Tests of Measurement Invariance for Congruency for the Country VIAT

Model & Invariance Level	*df*	SBχ^2	CFI	RMSEA	SRMR	Model comparison	Δ*df*	ΔSBχ^2
1. Configural	4	2.29	1.000	0.001	0.009	-	-	-
2. Metric	7	5.12	1.000	0.001	0.003	2 v 1	3	2.89
3. Scalar	10	7.04	1.000	0.001	0.032	3 v 2	3	1.73

Note. *N* =198. VIAT = Verbal Implicit Association Task. *dfs* = degrees of freedom. S-Bχ^2 = Satorra-Bentler Chi-square fit indices. CFI = Comparative Fit Index. RMSEA = Root Mean Square Error of Approximation. SRMR = Standardized Root Mean Square Residual. Δ = difference.

Configural invariance (Model 1 in Table 7.5) was supported as the fit indices demonstrated good model fit. The metric invariance model (Model 2 in Table 7.5) was compared with Model 1 and produced a non-significant $\Delta SB\chi^2$. This supports the metric invariance of the congruent and incongruent data. Measurement invariance was then further supported by the scalar invariance test (Model 3 in Table 7.5). This series of analyses indicated strong measurement invariance for the congruent and incongruent data of the Country VIAT.

Latent Mean Difference for Congruency for the Country VIAT.

The analysis of latent means was compared against the scalar invariance model (Model 3 in Table 7.6), with the results of the comparison of latent means presented in Table 7.6.

Table 7.6

Test of Latent Mean Difference for Congruency for the Country VIAT

Model & Invariance Level	*df*	$SB\chi^2$	CFI	RMSEA	SRMR	Model comparison	Δdf	$\Delta SB\chi^2$
3. Scalar	10	7.04	1.000	0.001	0.030	-	-	-
4. Latent Means	11	54.99**	0.928	0.100	0.178	4 v 3	1	252.52**

Note. *N* =198. VIAT = Verbal Implicit Association Task. *dfs* = degrees of freedom. $S\text{-}B\chi^2$ = Satorra-Bentler Chi-square fit indices. CFI = Comparative Fit Index. RMSEA = Root Mean Square Error of Approximation. SRMR = Standardized Root Mean Square Residual. Δ = difference. $^{**}p < .001$.

This analysis indicated a significant difference (population variance) between the congruent and incongruent data for the Country VIAT. It can thus be inferred that an IAT effect had occurred for this task. The output from the scalar invariance analysis showed that when the congruent latent mean was constrained to zero, the resulting incongruent latent mean was positive (*β*=.69). This implies that the participants took longer to categorise the incongruent pairings of stimuli than the congruent, which was as hypothesised.

Equivalency of Responses by Congruency for the Country PIAT

The simple one-factor CFA model was also tested for the Country PIAT's combined, congruent and incongruent data separately. Good model fit was evident for the combined data of the Country PIAT: S-Bχ^2 (2, N=198)=.49, *p*=.78; CFI=1.000; RMSEA< .001; SRMR=.004. Good model fit was also revealed for the Country PIAT's congruent data: S-Bχ^2 (2, N=198)=3.06, *p*=.22; CFI=.997; RMSEA=.037; SRMR=.014 and incongruent data: S-Bχ^2 (2, N=198)=.98, *p*=.61; CFI= 1.000; RMSEA< .001; SRMR=.007. The fit indices thus met criterion levels for the one-factor CFA model of the combined, congruent and incongruent Country PIAT data.

Measurement Invariance of Congruency for the Country PIAT.

Again, configural (equal form), metric (equal loadings) and scalar (equal intercepts) measurement invariance were tested. The outcomes of these tests are presented in Table 7.7.

Table 7.7

Tests of Measurement Invariance for Congruency for the Country PIAT

Model & Invariance Level	*df*	SBχ^2	CFI	RMSEA	SRMR	Model comparison	Δ*df*	ΔSBχ^2
1. Configural	4	3.98	1.000	0.001	0.011	-	-	-
2. Metric	7	5.54	1.000	0.001	0.022	2 v 1	3	1.57
3. Scalar	10	13.70	.995	.031	.032	3 v 2	3	11.90*
3a. Scalar – CPIAT4 released	9	11.58	0.996	0.027	0.030	3a v 2	2	9.25*
3b. Scalar – CPIAT3&4 released	8	9.19	0.988	0.019	0.027	3b v 2	2	6.10*

Note. *N* =198. PIAT = Pictorial Implicit Association Task. *dfs* = degrees of freedom. S-Bχ^2 = Satorra-Bentler Chi-square fit indices. CFI = Comparative Fit Index. RMSEA = Root Mean Square Error of Approximation. SRMR = Standardized Root Mean Square Residual. Δ = difference. *p < .05. **p < .001.

Configural invariance (Model 1 in Table 7.7) was supported by the fit indices showing good model fit. The metric invariance model (Model 2 in Table 7.7) was compared with Model 1 resulting in a non-significant $\Delta SB\chi^2$. This supported the metric invariance of the congruent and incongruent data. The scalar invariance test (Model 3 in Table 7.7) produced a significant Satorra-Bentler chi-square difference test. As such, the intercept for CPIAT4 was released and the resulting model analysed (Model 3a in Table 7.7). Releasing this intercept still led to a significant $\Delta SB\chi^2$. The intercept for CPIAT3 was then released. This model (Model 3b in Table 7.7) still demonstrated a significant $\Delta SB\chi^2$ when compared with Model 2. If a further intercept were released the resulting model would have been exactly the same as Model 2. This implies there is variance between the congruent and incongruent data's intercepts and thus, scalar invariance was not supported. Because of this, the Country PIAT only demonstrated weak or partial measurement invariance between the congruent and incongruent data. Although not ideal, weak or partial invariance might suffice for analyses assessing difference in latent means. For instance, Byrne, Shavelson and Muthén (1989) demonstrated how to assess differences in latent mean structures with only partially invariant measuring instruments. Using self-concept data from high school students they found that invariance evaluations could precede in the context of partial measurement invariance, however the potential for Type I or Type II errors must be considered in interpretation of results (Byrne, et al., 1989).

Latent Mean Difference for Congruency for the Country PIAT.

The analysis of latent means was compared against the metric invariance model (Model 2 in Table 7.8) due to the lack of scalar invariance. The results of the comparison of latent means are presented in Table 7.8.

Table 7.8

Test of Latent Mean Difference for Congruency for the Country PIAT

Model & Invariance Level	*df*	$SB\chi^2$	CFI	RMSEA	SRMR	Model comparison	Δdf	$\Delta SB\chi^2$
2. Metric	7	5.54	1.000	0.001	0.022	-	-	-
4. Latent Means	8	31.89**	0.967	0.087	0.103	5 v 2	1	46.37**

Note. *N* =198. PIAT = Pictorial Implicit Association Task. *dfs* = degrees of freedom. $S\text{-}B\chi^2$ = Satorra-Bentler Chi-square fit indices. CFI = Comparative Fit Index. RMSEA = Root Mean Square Error of Approximation. SRMR = Standardized Root Mean Square Residual. Δ = difference. ***p* < .001.

This analysis indicated a significant difference (population variance) between the congruent and incongruent data for the Country PIAT, which implies an IAT effect had occurred for the task. The output from the metric invariance analysis showed that when the congruent latent mean was constrained to zero the resulting incongruent latent mean was positive (β=.61). This implies the participants took longer to categorise the incongruent pairings of stimuli than the congruent, as hypothesised.

MIMIC Models – The Impact of Covariates on the IAT Effect Scores

MIMIC methodology was applied to assess whether sex, age or travel experience had significantly influenced the IAT and questionnaire results. The average IAT and questionnaire scores by sex, age and travel experience groups are presented in Table 7.9. It is noted that for ease of interpretation the age and travel experience data were separated into two groups before average IAT *D* scores were calculated.

Table 7.9

Average Scores by Covariates for Each of the Tasks.

Covariate	Total IAT Score (*D*)	VIATR (*D*)	PIATR (*D*)	VIATC (*D*)	PIATC (*D*)	MRS	TDQ
Females	0.26	0.41*	0.15	0.29	0.18	1.97	1.33
Males	0.25	0.36*	0.17	0.28	0.20	2.15	0.93
Under 40s	0.24	0.37*	0.13	0.27	0.18	1.98	1.22
Over 40s	0.35*	0.51*	0.30*	0.37*	0.22	2.23	1.19
Less Travelled	0.24	0.40*	0.14	0.26	0.16	2.01	1.34
More Travelled	0.27	0.39*	0.17	0.30	0.21	2.02	1.13

Note. N = 198. IAT effect scores (*D*) <.15 represent non-existent prejudice, .15-.35 implies slight prejudice, .35-.60 is moderate, and >.60 implies strong negative implicit prejudice (Greenwald et al., 2003). * IAT effect scores that show a moderate level of prejudice. Higher MRS scores imply greater prejudice, and higher TDQ scores demonstrate stronger preference for Europe over the Middle East.

An examination of the average IAT effect scores (presented in Table 7.9) indicated a slight to moderate level of prejudice displayed by the participants, according to Greenwald et al.’s (2003) guidelines. Overall, there were minimal sex difference for the IAT effect scores; however, some discrepancy in the scores by age was present. Older participants appeared to exhibit higher IAT effect scores than the younger participants (see Table 7.9). A more accurate estimate of these findings was delivered by the MIMIC analytical technique, which was applied separately to assess the racial attitude and country attitude data.

MIMIC Results for the Racial Attitude Construct

The goodness-of-fit indices for the MIMIC model of the race attitude construct were: χ^2 (65, N=198)=70.84, *p*=.29; CFI=.990; RMSEA=.021; SRMR=.040. All four of the fit indices indicated that this model replicated the variances and covariances of the input data very well. The correlational matrix for this analysis is depicted in Table 7.10. Figure 7.3 presents the standardised factor loadings (STDYX) for the racial attitude MIMIC model.

Table 7.10

Inter-indicator Correlations for the Racial Attitude MIMIC Model

	VIAT 1R	VIAT 2R	VIAT 3R	VIAT 4R	PIAT 1R	PIAT 2R	PIAT 3R	PIAT 4R
VIAT1R	1.000							
VIAT2R	.427	1.000						
VIAT3R	.476	.463	1.000					
VIAT4R	.457	.290	.495	1.000				
PIAT1R	.310	.262	.316	.308	1.000			
PIAT2R	.346	.278	.295	.318	.543	1.000		
PIAT3R	.127	.106	.200	.284	.367	.449	1.000	
PIAT4R	.141	.141	.213	.275	.363	.487	.456	1.000
MRSQ3	.091	.136	.063	.056	.060	.043	-.004	.031
MRSQ5	.116	.046	.111	.131	.145	.194	.200	.142
MRSQ12	.138	.127	.124	.143	.125	.153	.141	.090
SEX	-.058	-.086	-.083	-.031	.023	.056	.063	-.026
AGE	.139	.091	.146	.197	.172	.165	-.024	.122
TRAVEL	-.052	-.013	.045	.083	.092	.006	.082	.100

	MRS Q3	MRS Q5	MRS Q12	SEX	AGE	TRAVEL
MRSQ3	1.000					
MRSQ5	.446	1.000				
MRSQ12	.478	.483	1.000			
SEX	.078	.120	.065	1.000		
AGE	.044	.055	.115	.161	1.000	
TRAVEL	-.056	-.027	-.088	-.040	.062	1.000

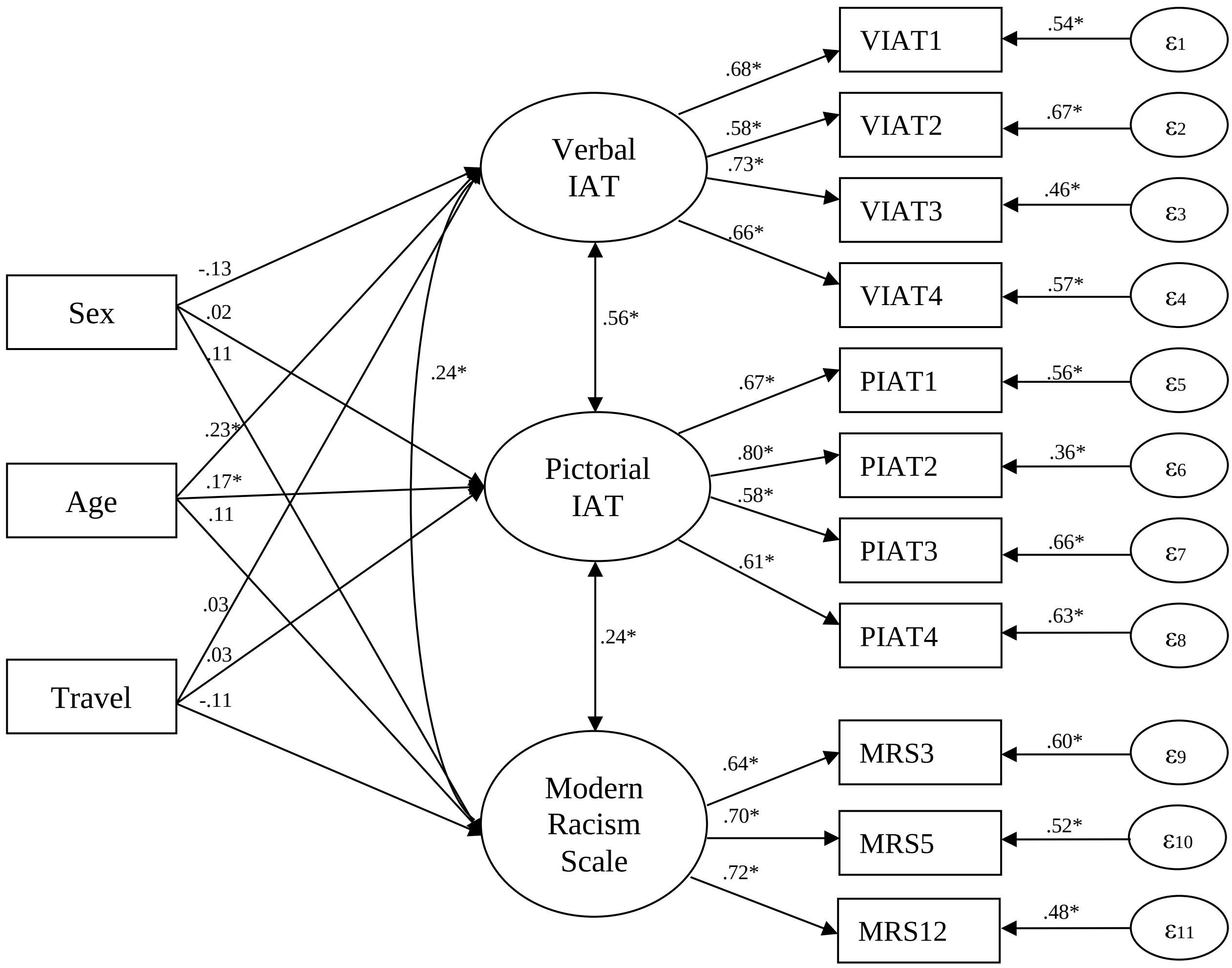

Figure 7.3. MIMIC Model of the effects of Sex, Age and Travel Experience on the tasks measuring the racial attitude construct. *p<.001.

For the racial attitude construct, the paths between Age and the VIAT and PIAT were both statistically significant (β=.23, p=.005; β=.17, p=.01 respectively). This indicates older participants produced larger IAT scores, implying that overall they posessed a stronger anti-Arab/pro-European bias than the younger participants. There were no significant differences in responses based on Sex or Travel Experience for this construct.

MIMIC Results for the Country Attitude Construct

The goodness-of-fit indices for the country attitude MIMIC model were: χ^2 (53, N=198)=51.79, *p*=.52; CFI=1.000; RMSEA< .001; SRMR=.035. All four of the fit indices indicated this model also replicated the variances and covariances of the input data very well. The correlational matrix for this analysis is depicted in Table 7.11. The standardised factor loadings (STDYX) for the country attitude MIMIC model are presented in Figure 7.4.

Table 7.11

Inter-indicator Correlations for the Country Attitude MIMIC Model

	VIAT 1C	VIAT 2C	VIAT 3C	VIAT 4C	PIAT 1C	PIAT 2C	PIAT 3C	PIAT 4C
VIAT1C	1.000							
VIAT2C	.438	1.000						
VIAT3C	.444	.527	1.000					
VIAT4C	.449	.410	.381	1.000				
PIAT1C	.297	.287	.201	.247	1.000			
PIAT2C	.292	.226	.248	.277	.528	1.000		
PIAT3C	.326	.187	.235	.289	.449	.514	1.000	
PIAT4C	.220	.115	.129	.238	.381	.453	.396	1.000
TDQ1	.294	.262	.175	.227	.298	.200	.179	.190
TDQ2	.267	.199	.088	.168	.257	.233	.207	.199
SEX	-.005	.011	-.085	.002	.055	.027	-.020	.030
AGE	.077	.071	.129	.005	.079	.037	.025	.017
TRAVEL	.044	.041	.144	.062	.005	-.102	-.099	-.131

	TDQ1	TDQ2	SEX	AGE	TRAVEL
TDQ1	1.000				
TDQ2	.671	1.000			
SEX	-.184	-.187	1.000		
AGE	-.083	-.041	.161	1.000	
TRAVEL	-.084	-.035	.009	.194	1.000

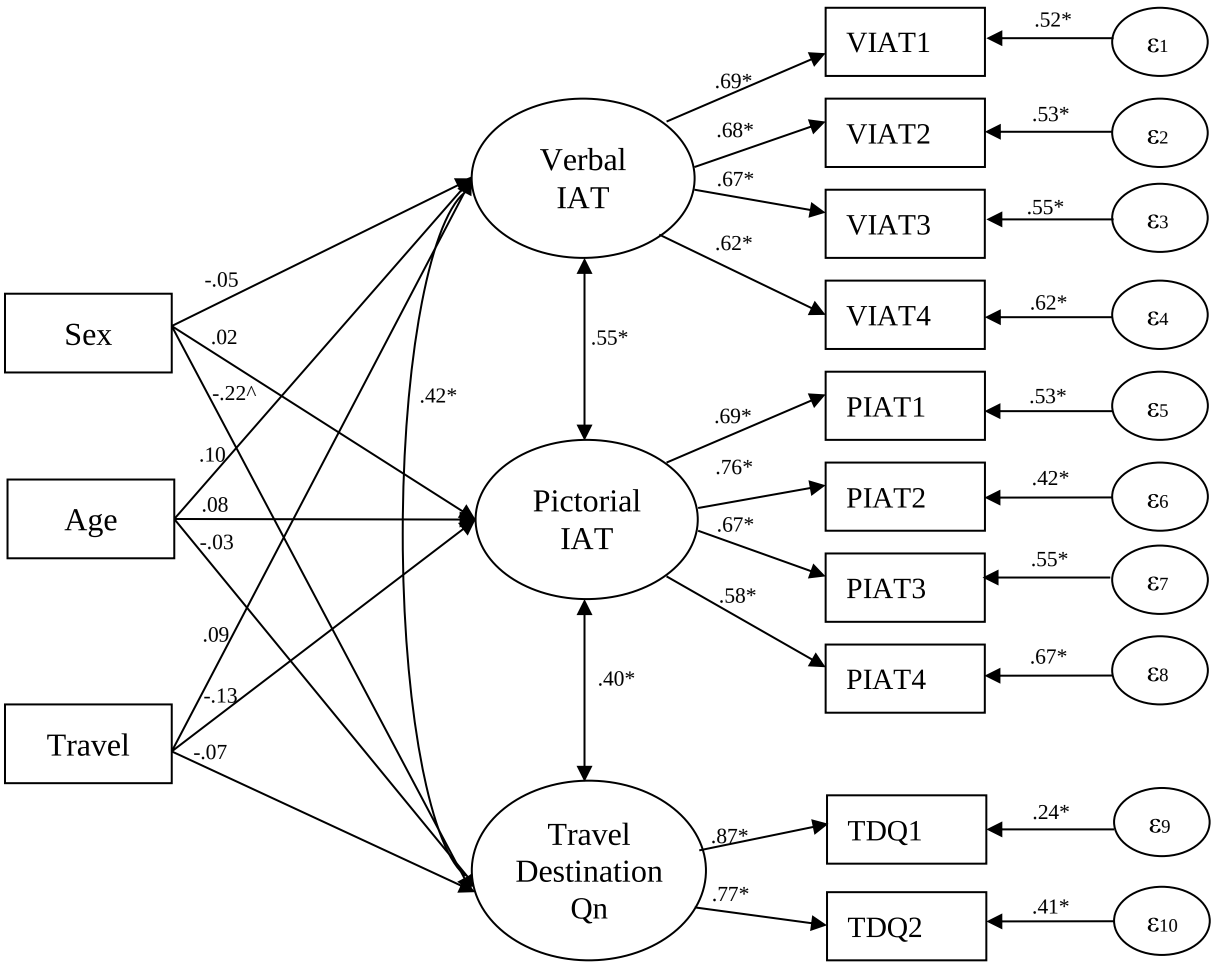

Figure 7.4. MIMIC Model of the effects of Sex, Age and Travel Experience on the tasks measuring the country attitude construct. ^p<.05, *p<.001.

For the country attitude construct, the path from Sex to the Travel Destination Questionnaire was found to be statistically significant (β=-.22, p=.005). Given the coding of the sex covariate (1=Females, 2=Males) and the negative sign of this parameter estimate, males explicitly reported a smaller bias against travel to Middle Eastern countries than females did. Or specifically, males had a mean Travel Destination Score .42 units higher than the mean of the females. This is a small to medium effect size. No other significant pathways were present.

Discussion

Study Three examined the utility of addressing IAT-related substantive enquiries using SEM analytical approaches that account for error variance. The first aim of Study Three was to investigate whether an IAT effect was present for each of the experimental IATs using the Multiple-groups CFA approach. This strategy revealed the IAT effect for three of the four IATs, the Racial VIAT, Country VIAT and Country PIAT. However, this bias was not evident for the Racial PIAT. Such findings may imply a generalised implicit preference for Europe/ Europeans over the Middle East/Arabs. The second aim of Study Three was to determine whether IAT effects were significantly influenced by the age, sex or travel experience of the participant using MIMIC modeling. Evidence of age bias in the IAT scores for the racial attitude construct and gender bias in expressed attitudes towards visiting the Middle East for the country attitude construct were found. The implications of these results will be discussed, along with the impact of harbouring anti-Arab/pro-European implicit biases, and the generalisability of the present results. It is argued that SEM has great potential for examining substantive investigations in implicit attitudinal research.

Evidence of IAT Effects as Revealed by Multiple-groups CFA

The hypothesis that participants would demonstrate an IAT effect indicating anti-Arab/pro-European implicit bias was supported for the Racial VIAT, Country VIAT and Country PIAT, but not the Racial PIAT. A bias against the Middle East in favour of Europe was expected given the plethora of anti-Middle Eastern information that has been prevalent in Western media over the last decade (see Cashin, 2010; Dunn et al., 2004). The current findings reinforce previous IAT research that reported evidence of anti-Arab/Muslim bias when compared with Whites, Christians and Black Africans (Nosek et al., 2007; Park et al., 2007; Rowatt et al., 2005). These findings support the assertion that a level of 'Islamophobia' is present throughout the Western world (Dunn et al., 2004; Dunn et al., 2008; Islam & Jahjah, 2001; Poynting & Mason, 2007).

Evidence of Unbiased Attitudes on the Racial PIAT

In contrast to the other tasks, anti-Arab/pro-European bias was not demonstrated for the Racial PIAT. Rather, latent mean invariance was observed, which suggests the participants responded comparably to the congruent and incongruent stimuli pairings. Such unbiased responding is incongruent with the results of the other three IATs as well as being contradictory to previous implicit attitude research findings (e.g. Nosek et al., 2007; Park et al., 2007) and explicit attitudinal research in Tasmanian where a strong negative association with Arabs was reported (Dunn et al., 2008). The current findings could be argued to demonstrate that the student participants of the University of Tasmania were unbiased in the face of real photos of Middle Eastern and European peoples. Previous research has indeed found university populations tend to provide more egalitarian views than those espoused by the general population (Nosek, 2007). However, given the hypothesised implicit prejudice was evident for the other three IATs, further theories for this inconsistent result should be considered.

It is possible that the apparent lack of bias in the racial PIAT results was due to the confounding influence of the attractiveness of the pictorial Arab stimuli. The present sample were predominantly female (~73%) with a mean age of 25 years. The stimuli chosen for the Racial PIAT were pictures of Middle Eastern and European people that were selected to present a positive and unbiased image for each racial group. IAT stimuli should ideally avoid all possible idiosyncratic connotations that might be unrelated to the concepts intended for measurement (Steffens et al., 2008), for it is well known stimuli can play a large role in determining IAT effects (Lane et al., 2007). Although such issues were considered during stimuli selection, during the testing process there were many instances where female participants made comment regarding the physical attractiveness of the male Middle Eastern stimuli. Two stimuli, shown in Figure 7.5, were particularly commented upon. Physical attractiveness is known to be a powerful basis for decision-making (Thorndike, 1920). It is thus a definite possibility that these stimuli were categorised primarily on attractiveness rather than racial belonging. This means the Racial PIAT results were likely confounded substantially by these stimuli.

Figure 7.5. Examples of the Arab pictorial stimuli that may have led to confounded categorisation.

The Influence of Covariates on IAT Effect Scores via MIMIC Modelling

The impact of various participant characteristics on their attitudinal scores was examined in the final aim of this thesis using MIMIC modelling. Information regarding which participant characteristics were likely to result in higher or lower levels of implicit prejudice could be beneficial given the IAT is often used in applied research settings and as a diagnostic instrument. As such, this information may help guide which population groups to target for further research or educational campaigns aimed at acceptance of diversity. The present research examined the racial-related and country-related attitudes separately.

The Impact of Age on Responses for the Implicit Racial Attitude Construct

For the racial attitude construct, significant paths between the Age covariate and the VIAT and PIAT latent factors were evident in the MIMIC model. This implies older participants produced larger IAT effect scores, suggesting they produced a stronger bias towards Europeans over Arabs than did their younger counterparts. This finding supports previous research by Nosek et al. (2007) who also found older adults produced stronger negativity towards Arab/Muslim people (compared to White people) than was the case for their younger participants. These findings imply older adults are more racially prejudiced than younger adults, which is reasonable given the relative socio-cultural environment of Australia half a century ago versus today.

The older participants of the current sample were likely to have been raised in an Australia characterised by different social norms than that of the younger participants. Half a centrury ago it was reasonably acceptable to possess negative attitudes towards minority racial groups (Cashin, 2010; Dunn et al., 2004;

McConahay et al., 1981; Poynting & Mason, 2007). For instance, the White Australia policy[26] prevailed until the 1970s in Australia (Dunn et al., 2004), which is around the time some of the 'older' participants were being born. As such, it is not unexpected that these older participants would possess a stronger anti-Arab/pro-European bias than the younger participants.

The average IAT effect score for the older participants was indicative of a moderate level of prejudice, whereas the younger participants revealed a slight prejudice according to Greenwald et al.'s (2003) guidelines. Such findings reveal lower levels of prejudice than have previously been reported in North American samples (see Nosek et al., 2007). This discrepancy may imply Australians have greater racial tolerance than Americans, given Nosek et al. (2007) quote an average *D* score of .77, which is very sizeable compared to *D*=.28 for the current study. It is noted this discrepancy is likely exacerbated by the inclusion of the much faster PIAT responses with the VIAT scores for the racial attitude construct in Study Three. Combining the data in this way produces an overall lower average IAT effect score, which is reflective of a methodological issue rather than a difference in attitude [27]. Nevertheless, even the Australian VIAT scores (*D*=.40) were almost half the magnitude of the reported American results, implying there may be cultural differences worth exploring in future research.

---

[26] An immigration policy that favoured immigrants from certain European countries, such as the UK.
[27] Specifically, the PIAT scores (*M*=.16, *SD*=.04) when combined with the VIAT scores (*M*= .40, *SD*=.05) for the racial attitude construct, substantially lower the average IAT effect score (*M*=.28, *SD*=.13). These findings highlight the difficulty in generalising IAT effect score cut-offs for level of implicit bias from the VIAT to the PIAT, as it is very unlikely an IAT effect score of X on a VIAT would imply an equivalent attitude as the same score of X produced by a PIAT. Such issues associated with the *D* scoring strategy will be discussed further in Chapter Eight.

Such cultural or age discrepancies, as noted above, may imply there is value in future cross-cultural or cross-sectional IAT research to assess global and developmental automatic racial biases. SEM approaches such as Multiple-groups CFA could clearly be applied to facilitate such investigations by determining whether an IAT effect score of X in one country or age bracket was comparable to a score of X for the relevant other group. Use of SEM in this way would enable more accurate comparisons of IAT group data.

Findings of the MIMIC Analysis for the Country Attitude Construct

The age discrepancy evident for the racial attitude construct did not generalise to the country attitude construct. Rather, the only significant pathway for the country attitude construct was between the Sex covariate and the Travel Destination Questionnaire, which indicated that males explicitly reported a greater preference for travel to Middle Eastern countries than females did. This is perhaps not surprising, given that the Middle East is viewed as a very risky location to visit by many potential travellers in the West (Sonmez & Graefe, 1998) and some have argued that males are generally less adverse to physical risks than females are (Lepp & Gibson, 2003). Furthermore, the constraints on female travellers are arguably much greater than males when travelling within Islamic countries. For instance, women are often required to dress differently (covering head, shoulders, arms and legs at times), are unable to drive in countries such as Saudi Arabia, may be segregated from males on transport and at restaurants, and women are typically expected to be accompanied by a male (Wilson & Little, 2008). As such, the finding that men expressed a greater likelihood of visiting a Middle Eastern country than women did is understandable.

Contrary to expectations, the sex difference on the TDQ was the only significant effect found for the country attitude construct, and did not generalise to the IAT data. The lack of sex differences evident for the present IAT data was contrary to previous research by Nosek et al. (2007). Their research found men demonstrated a moderate level of prejudice against Arab/Muslim names in preference of Other ethnic names, a finding substantially greater than the slight prejudice revealed by the female participants (Nosek et al., 2007). For the present study, both males and females produced average IAT effect scores within the slight prejudice range for the racial and country attitude constructs. Perhaps a larger study with a broader spectrum of participants, like the enormous web-based datasets utilised by Nosek et al. (2007)[28], would provide greater discrepancies between the travel experiences of the participants and enough power to find sex differences if they were present.

Limitations of the MIMIC Approach to Assessing Group Differences

Unlike in multiple-groups CFA, MIMIC models do not assess for measurement invariance between the groups. As such, in the present study the covariates of sex, age and travel experience were not examined for between-group equivalency. Furthermore, the multiple-group CFAs of Study Three were limited to examining equivalency of the IAT congruent and incongruent data and did not examine the covariates of interest; sex, age and travel experience. Future research could use the multiple-groups CFA approach to examine the measurement invariance of relevant covariates, before then applying the MIMIC methodology to examine the effects of these participant characteristics on the implicit attitude scores. The use of multiple-group CFA and MIMIC analytical strategies in conjunction would thus enable a stronger estimate of the impact of covariates on IAT scores.

[28] Nosek et al. (2007) report reviewing more than 22,000 Racial IATs.

Summary

Study Three produced the expected IAT effect for three of the four empirical IATs. This reveals a relative bias against the Middle East and its people in preference of Europe and its people. This bias was found to be greater for older than younger participants for the racial attiude construct. In the following sections, the implications of possessing such a bias are presented, along with the generalisability of these results. Implications arising from the unexpected results of the Racial PIAT are also examined particularly in relation to the development and use of IATs in applied behavioural research. It will be argued that stimuli should be extensively piloted prior to use, and further discussion regarding the construct actually being assessed by IATs is warranted.

Evidence of Anti-Arab/Pro-European Bias in the Present Sample

Evidence of anti-Arab/pro-European bias was found in the current sample using Middle Eastern and European first names, country names and images of buildings such as churches and mosques. A finding of relative bias against the Middle East and Arabic names in comparison to Europe and European names can be difficult to appropriately interpret, for it is confounded as to whether such a result predominantly indicates anti-Arab bias or pro-European views. SEM research by Blanton et al. (2006) revealed that Racial IAT effect scores were predominantly influenced by a tendency to associate negative attributes with Black people rather than a tendency to associate negative or positive attributes with White people. This result is consistent with substantial evidence that negative attitudes, stereotypes, emotions and events are more easily formed, more powerful and more resistant to change than positive ones (Baumeister, Finkenauer, & Vohs, 2001). Such research suggests that the findings of the current study are indicative of anti-Arab sentiment rather than anti-Arab/pro-European views.

The present findings of anti-Middle East/Arab prejudice are supportive of considerable evidence indicating negative affect towards Arab/Muslims by Westerners using both explicit and implicit measurement techniques (e.g. Agerström & Rooth, 2009; Chopra, 2008; Dunn et al., 2004; Dunn et al., 2008; Gibson, 2008; Park et al., 2007; Rooth, 2010). Anti-Arab sentiment has likely resulted from several well-publicised socio-political events, including the destruction of New York's twin towers by men with distinctively Muslim names and the resultant "War on Terror" (Dunn et al., 2004; Rashid, 2009). Because of such events, Middle Eastern names are often presented in the media in association with terrorism, extreme religious behaviour and other negative connotations (Poynting & Mason, 2007). In addition, countries such as Afghanistan, Iraq and Lebanon are often linked with war, terrorism, religious dogma and political instability (Lepp & Gibson, 2003). Likely as a result of these well-publicised associations, Arabs have been named explicitly as the most threatening racial group for Australian respondents (Dunn et al., 2008; Islam & Jahjah, 2001) and the Middle East has been rated as one of the riskiest places to visit in the world (Kozak, Crotts, & Law, 2007; Sonmez & Graefe, 1998). The present findings support such prejudices, evidencing implicit bias against Middle Eastern names as well as the whole Middle East geographic region.

The Impacts of Harbouring High Levels of Implicit Racial Prejudice

Generalisation of negative attitudes towards not only Middle Eastern people but also the entire geographic region are consistent with previous research that has revealed people can easily generalise negative affect towards whole groups of people or locations based on limited and sometimes incorrect information (Dunn et al., 2004; Lepp & Gibson, 2003; Sonmez & Graefe, 1998). For instance, Muslims have reportedly suffered quite dramatically from negative stereotypes associated with Islamic practices (Dwyer, 1993). Further, entire continents have been generalised as perilous or safe based on limited and sometimes incorrect information[29] (Lepp & Gibson, 2003). Thus, implicit prejudices can be easily formed with limited rationale, and are also easily generalisable. Cunningham, Nezlek and Banaji (2004) found that individuals who possessed implicit racial prejudice against a particular out-group, such as Arab-Muslims, were likely to also experience consistently negative attitudes towards other culturally disadvantaged out-groups, such as Black Africans, homosexuals and the poor (Cunningham, Nezlek, et al., 2004). As such, the anti-Arab prejudice revealed in the current study could have implications for interactions with many disadvantaged populations.

As previously mentioned in Chapter Two, high levels of implicit prejudice are related to the probability of engaging in unambiguously harmful actions towards members of minority groups (see Rudman & Ashmore, 2007). Although no research has yet specifically qualified what level of implicit prejudice is required for certain behaviours to occur (see Blanton & Jaccard, 2006), in general it has been found that higher IAT effect scores reflect higher propensity to react negatively towards, or

---

[29] In one example, Zambia's tourism industry was substantially affected by the USA releasing a safety warning for Zimbabwe (Lepp & Gibson, 2003).

make decisions that would negatively impact, minority group members (see Agerström & Rooth, 2009; Green et al., 2007; Greenwald et al., 2009; McConnell & Leibold, 2001; Rooth, 2010; Rudman & Ashmore, 2007). Such findings (summarised in Chapter Two) reveal that negative implicit racial biases, of which the individual may be completely unaware, can have considerable consequences for others. In a country reportedly committed to the concept of multiculturalism, such as Australia (Schweitzer et al., 2005), it is worrisome to observe such negative automatic racial prejudices.

Generalisability of the Present Findings

The present sample revealed the presence of anti-Arab/pro-European bias despite the use of a population predominantly comprised of university students. As previously mentioned, university populations are known for having less biased attitudes than the general population (Nosek, 2007). The finding of such prejudices, despite a student sample, imply that were the research to be conducted again with a community-based sample the amount of implicit relative bias against Arabs would likely be comparable or even greater than that reported by the current sample. Furthermore, according to the results of the present research, and Nosek et al. (2007), these racially biased attitudes are potentially likely to be greater for older generations of adults rather than younger people. Greater bias by older persons is consistent with culturally accepted norms as evidence by Australia’s history of racially discriminatory legislation (see Poynting & Mason, 2007). Given Australia has an aging population (Anderson & Hussey, 2000) the current findings of relative bias against Arabs and the Middle East are likely to also be evident throughout the greater Australian population.

Methodological Implications for the IAT Stemming from the Racial PIAT's Unexpected Results

In direct contrast to the results of the other IATs, the Racial PIAT did not produce the expected IAT effect, potentially due to the confounding influence of the attractiveness of the Arab pictorial stimuli. It is concerning for the robustness of IAT findings that stimuli selection impacted the IAT effect to this extent. However, such effects have been documented previously (refer to Chapter Two; see also Dasgupta & Greenwald, 2001; Gawronski, LeBel, & Peters, 2007; Han et al., 2009; Karpinski & Hilton, 2001). These studies show that the IAT is not as robust to momentary, irrelevant contextual considerations as was once believed (see Han et al., 2009; Karpinski & Hilton, 2001; Olson & Fazio, 2003).

There are two prominent implications for stimuli having such powerful influence over the IAT effect, and thus the perceived implicit attitudes these scores are meant to reflect. Firstly, these results imply substantial effort should be focused on piloting the potential stimuli exemplars used to represent constructs in an IAT. Currently, detailed piloting of stimuli rarely occurs, or at least, is rarely reported to have occurred (Fiedler et al., 2006). This is problematic given the current findings that suggest implicit attitude researchers would do well to pilot stimuli very carefully during IAT development in case the stimuli are not representative of the constructs aiming to be measured.

Secondly, if changes in the stimuli can drastically alter the resulting IAT effect score (see also Dasgupta & Greenwald, 2001) then what exactly is the IAT measuring. Do IAT's assess a person's attitude towards the concept of Europeans versus Arabs, the meaning of the stimuli (Iraq, United Kingdom, Poland, Saudi Arabia), the specific names chosen to be generic stimuli (Abdul, Charles, Habib, Penny), or other features of the stimuli not accounted for by the researcher (such as “my best friend is called Penny” or “that guy is really attractive”)? All of these options are in principle distinct from the general concept of racial prejudice (Fiedler et al., 2006), implying that the efficacy of the IAT is questionable. At very least, the thorough piloting of stimuli and the application of SEM strategies to analyse IAT data can help uncover and potentially address such issues for future research. The aforementioned methodological and theoretical implications for future IAT research are further discussed in the next and final chapter, the General Discussion (Chapter Eight).

Chapter Summary and Conclusion

SEM was found to be suitable for avoiding the issues of error variance whilst assessing substantive enquiries for the IAT. Using Multiple-groups CFA, an implicit relative bias against the Middle East and its people in preference for Europe and its people was found for the majority of the IATs. Evidence of prejudice against the Middle East/Arabs has implications for race relations in Australia and is likely reflective of similar attitudes amongst the general Australian population. This preference against the Middle East appeared stronger for the older than the younger participants of the current study for the racial attitude construct. The unexpected results of the Racial PIAT have several implications for the use of IATs in applied behavioural research, which will be discussed further in Chapter Eight.

It is noted that the MIMIC analyses of the current study provided limited findings regarding the influences of sex, age and travel experience on the IAT. However, the potential usefulness of MIMIC models in the assessment of IAT-related substantive issues has been illustrated. In a larger investigation, these same analytical processes could easily be adapted to assess the impacts of educational level, socio-economic status, geographical location, and religious or political identification on explicit and implicit attitudes. In this way, SEM analytical techniques have been shown to not simply be a means for providing stringent psychometric validation of measurement instruments, but also as a useful tool for clarifying theoretical understanding of constructs and to understand further the broader spectrum of factors that influence both explicit and implicit attitudes. SEM analytical techniques appear well situated to facilitate the ongoing use of implicit attitude measures, by providing a much needed avenue to address the significant issue of error variance for the IAT.

CHAPTER EIGHT

General Discussion

The overall aim of this dissertation was to examine the construct validity of implicit attitude measures using confirmatory factor analysis (CFA). CFA is a novel approach in implicit attitudinal research, which has typically failed to adequately account for non-randomly distributed error variance. The tasks examined were the traditional verbal Implicit Association Test (VIAT; Greenwald et al., 1998), a fully pictorial Implicit Association Test (PIAT; Thomas et al., 2006) and the Affective Priming Task (APT; Fazio et al., 1986). It was demonstrated that implicit attitudinal scores are comprised of a significant portion of error variance. However, the results reported in this thesis also demonstrate that once error variance was accounted for, the IATs had reasonable construct validity. In contrast, even if error variance is accounted for, the APTs did not show sufficient construct validity.

The first part of this chapter provides a summary of the results of the three studies presented in Chapters Five, Six and Seven of this dissertation. Following the summary, a general discussion of the implications of the thesis findings is presented. Limitations of the current research and directions for future research are also addressed. It is concluded that the IAT requires continued development and the application of latent modelling procedures to enable its utility for population-based research.

Summary of Results

This thesis is organised in three studies that examine the reliability and construct validity evidence of three commonly used implicit attitude measures: the verbal Implicit Association Test (VIAT; Greenwald et al., 1998), the pictorial Implicit Association Test (PIAT; Thomas et al., 2006) and the Affective Priming Task (APT; Fazio et al., 1986). Each of these tasks was adapted to assess racial and country-related attitude constructs pertaining to Arabs/Middle East versus Europeans/Europe. This means that the psychometric properties of six implicit attitude measures were investigated, using a sample population of 198 students of the University of Tasmania, Australia.

The first study examined the reliability and construct validity of six measures using single-group Confirmatory Factor Analysis to separate random error from the trait variance of the implicit attitudinal data. Following on from this, the second study examined systematic forms of error variance, such as method effects, in addition to random error using the Multitrait-Multimethod (MTMM) approach to CFA. The third study examined implicit attitudinal bias using Multiple-groups CFA and whether certain participant characteristics had influenced the implicit attitude scores using Multiple Indicators, Multiple Causes (MIMIC) modelling. A summary of the results of each of the studies, organised by aim, along with the main findings are displayed in Table 8.1.

Table 8.1

Summary of Main Results for the Thesis

Research Question	*Aim*	*Findings*	*Conclusion*
<u>*Study 1 (Chapter 5)*</u>			
(1) Are the Implicit Association Test (IAT) and the Affective Priming Task (APT) reliable measures of implicit attitudes?	Assess reliability of IATs and APTs using *Composite Reliability* and *Average Variance Extracted*.	The IATs showed good internal consistency (CR>.70) but inadequate internal convergent validity (AVE~.45), indicating trait variance accounts for < 50% of IAT scores. The APTs failed both assessments (CR~.1, AVE~.05).	Good internal consistency for the IATs, but more random error than trait variance. The APT is not a reliable measure of implicit attitudes.
(2) Are implicit attitude measures psychometrically robust?	Assess internal construct validity of IATs and APTs using *single-group CFA.*	The IATs had good internal construct validity (β~.60), but significant error variance was also evident (β~.45). APT scores were almost entirely comprised of random error (β~.95).	The IAT is internally robust once random error variance is accounted for. The APT is not a valid measure of implicit attitudes.
(3) Do implicit attitude measures possess sufficient convergent and discriminant validity compared to explicit attitude measures?	Assess correlations between the IATs and questionnaires using s*ingle-group CFA.*	The verbal and pictorial versions of the IAT were strongly related (r=.55). The inter-implicit correlations were significantly stronger than any implicit-explicit correlation.	The new Pictorial IAT (PIAT) appeared comparable to the traditional Verbal IAT (VIAT). Implicit and explicit attitudes seem to be similar yet distinct constructs.
(4) Are implicit and explicit tasks assessing a substantial portion of the trait attitude construct?	Assess convergent and discriminant validity of the IATs using *Higher-order CFA.*	Both versions of the IAT loaded strongly and significantly onto a higher-order implicit attitude factor (β>.70). Explicit attitude measures loaded onto this same higher-order factor but to a lesser degree (β~.40).	Strong support for the comparability of the VIAT and PIAT. The explicit attitude measures assessed a similar construct to that of the implicit attitude measures.

Research Question	*Aim*	*Findings*	*Conclusion*
Study 2 (Chapter 6)			
(5) How much of the IAT score is implicit attitude and how much is method effects or random error? If error variance is accounted for, is the IAT a good measure of implicit attitudes?	Assess the construct validity of the four empirical IATs using *CT-CM CFA-MTMM.*	The IATs were comprised of 18% trait, 28% method and 54% random error variance. The latent method factors were strongly related (r=.57). The Racial VIAT and Country PIAT showed good construct validity (trait>method variance, 32%>14%). The Country VIAT and Racial PIAT did not (trait<method variance, 5%<41%).	Error must be accounted for before providing any feedback about implicit biases. Individual psychometric validation of implicit attitude measures is required. Development of psychometrically robust IATs appears possible.
Study 3 (Chapter 7)			
(6) What implicit attitudes were revealed by the empirical IATs?	Assess equivalency of congruent/incongruent IAT responses using *multiple-group CFA.*	Test of latent mean difference revealed a positive IAT effect for the Racial VIAT ($\Delta\chi^2$=152.63, p<.001), Country VIAT ($\Delta\chi^2$=252.52, p<.001) and Country PIAT ($\Delta\chi^2$=46.37, p<.001). No IAT effect for the Racial PIAT ($\Delta\chi^2$=.57, *n.s.*).	Overall implicit preference for Europe over the Middle East revealed in the present sample. IAT stimuli can significantly impact upon the IAT effect.
(7) Do participant characteristics, such as sex, age or travel experience, impact IAT effect scores?	Examine the impact of covariates on the IAT effect scores using *MIMIC models.*	Older participants showed larger IAT effect scores than younger respondents on the Race IATs (β~.20, p<.001). Sex and travel experience did not significantly impact IAT results.	Older participants revealed a greater level of implicit anti-Arab/ pro-European bias than younger participants. Demonstration of how to assess substantive issues for the IAT using SEM.

Study One

In the first study the reliability and internal construct validity of the APTs, VIATs and PIATs was estimated using single-group Confirmatory Factory Analysis (CFA; Jöreskog, 1969). CFA models latent variables by separating the observed scores into trait and random error components, and are typically used to assess the internal construct validity of measurement instruments. However, this model was also applied to estimate internal consistency using Composite Reliability (CR) and internal convergent validity using Average Variance Extracted (AVE). The results of these analyses revealed the implicit attitudinal scores to be mostly composed of random error variance. Figure 8.1, shows that random error accounts for about 55% of the IAT effect scores and 95% of the APT scores. The large amount of random error contained in the observed scores emphasises that implicit attitudinal data should be analysed using CFA.

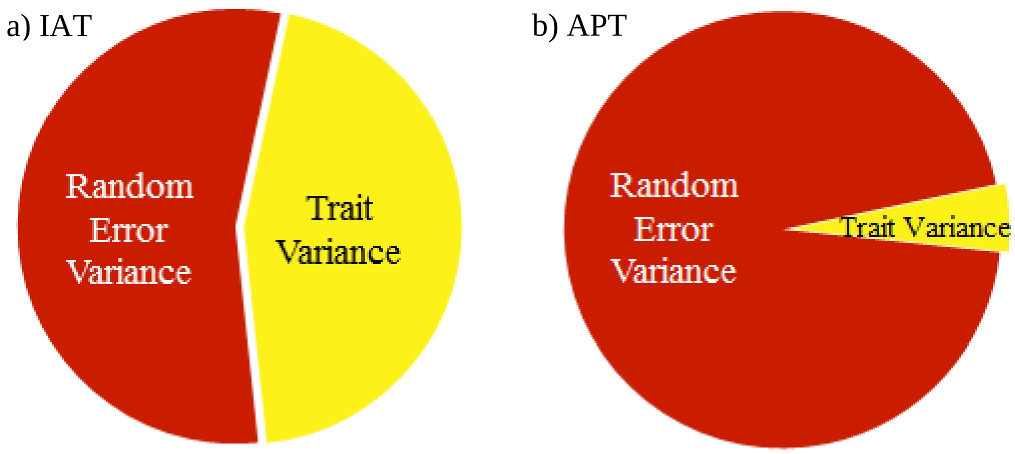

Figure 8.1. Estimated composition of random error variance and trait variance for the IAT (a) and APT (b) scores.

Once random error was accounted for in the CFA models, both the verbal and pictorial IAT formats had adequate internal consistency (CR>.70) and internal construct validity (β~.60). This suggests that the trait variance component of the IAT scores (see Figure 8.1a) provided a relatively consistent estimate of the country-related or racial-related implicit biases. This means that if random error was routinely accounted for, the trait component tapped by IATs could provide more adequate estimates of trait implicit attitudes. In contrast, the APT scores were comprised almost entirely of random error variance (see Figure 8.1b), with minimal trait variance measured (CR~.1; β~.15). This suggests that the priming tasks are invalid measures of implicit attitudes.

Convergent and discriminant validity for the IATs and explicit attitude measures were examined using three-factor CFA and higher-order CFA, which investigated the relationships between first- and second-order latent Race and Country Attitude factors. These analyses revealed the VIATs and PIATs to be comparable, with substantial covariance between the latent IAT factors (r=.55). In addition, both tasks loaded strongly and comparably onto a single higher-order implicit attitude factor (average β=.74 for the VIAT, β=.75 for the PIAT). Strong VIAT-PIAT convergence is a substantial improvement over previous estimates that had analysed observed scores, which had only showed small to medium relationships between these tasks (Thomas, 2008). Discriminant validity between the IATs and attitude questionnaires was indicated by latent inter-implicit correlations that were significantly stronger than the implicit-explicit correlations (racial construct: z=6.11, p<.001; country construct: z=3.00, p=.002). This supports the claim that implicit and explicit attitudes are distinct, albeit similar, constructs (Nosek & Smyth, 2007).

Study Two

In the second study, method effects were additionally accounted for using the multitrait-multimethod approach to CFA (CFA-MTMM). The CFA-MTMM analysis indicated that on average over half of an IAT effect score was attributable to random error variance, and method variance comprised almost a further third of the score. This meant that less than a fifth of an average IAT effect score was trait variance (see Figure 6.3, Chapter Six). The finding that over 80% of an IAT effect score was attributable to error variance has significant implications for the use of IATs in applied behavioural research.

CFA-MTMM also delivers a robust estimate of construct validity, whereby good construct validity is evident when trait variance is greater than method variance (Byrne, 1998). For the current study, good construct validity was demonstrated for the Racial VIAT and Country PIAT (trait= 32% > method= 14%). However, the Country VIAT and Racial PIAT indicated poor construct validity (trait= 5% < method= 41%), despite being ostensibly very similar tasks. This inconsistency is problematic and highlights the need for robust psychometric validation of each individual IAT. However, evidence of adequate construct validity for two of the IATs suggests that psychometrically robust verbal and pictorial IATs can potentially be developed.

Study Three

In the third study, it was determined whether IAT effects were significantly influenced by participant characteristics. According to IATs, implicit prejudice is evident when prejudice-congruent (e.g. Arab+Negative, European+Positive) stimuli pairings are responded to significantly faster than incongruent pairings (e.g. Arab+Positive, European+Negative). In a novel approach, multiple-group CFA was used to simultaneously perform CFAs for the congruent and incongruent ("groups") to determine whether various aspects of the models were equivalent. In addition, latent mean differences were tested to determine IAT effects. Implicit bias was detected in the Racial VIAT, Country VIAT and Country PIAT data, revealing implicit preference for Europe/European names over the Middle East/Arab names. No IAT effect was found using the Racial PIAT, with the attractiveness of the Arab male pictorial stimuli potentially having confounded results for this task.

Lastly, the influence of participant characteristics on IAT scores was investigated using Multiple Indicators Multiple Causes (MIMIC) modelling. MIMIC models add covariates indicative of group membership onto the standard CFA model to examine their direct effects (Jöreskog & Goldberger, 1975). The MIMIC analysis for the racial attitude construct showed that older participants had stronger implicit prejudice (*D*=.40) than the younger participants (*D*=.25). Contrary to previous research (see Nosek et al., 2007), no significant effects of sex or travel experience were found to impact the IAT effect.

Summary

The overarching results of this thesis have demonstrated the utility of CFA approaches in estimating the validity of reaction-time based assessment of implicit attitudes. CFA can be used as a systematic framework to examine psychometric properties of implicit attitude measures. In particular, CFA allowed for the separation of trait from error variance, and appears to be a useful approach to analyse implicit attitudinal data.

The present results demonstrated that implicit attitude scores are characterised by high levels of error variance. Potential sources of measurement error for implicit attitude measures were outlined in Chapter Three. They include random sources of error, such as natural variability in motor execution and attentional lapses, as well as systematic forms of error, namely block presentation order, task-switching ability and processing speed. The cumulative effect of this measurement error is that it significantly confounds the estimate of trait implicit attitude produced by these tasks. This thesis found the APT to be an invalid measure of implicit attitudes because it almost solely assessed error variance; as such it is inadequate for use in applied research settings. In contrast, both verbal and pictorial versions of the IAT had adequate construct validity after error variance was controlled for. This means that when error is addressed using latent modelling approaches, the trait variance that *is* tapped into by the IAT has the potential to provide a valid assessment of implicit attitudes. This implies that the IAT has potential utility for applied research, but that analytical approaches such as CFA are required to facilitate this use.

Implications of the Current Research

The finding that significant error variance confounds implicit attitudinal data has important implications for the use of implicit attitude measures in applied research. In particular, the poor psychometric properties of the APT suggest that it should not be used if valid assessments of implicit attitudes are aimed for. In contrast, the VIATs and PIATs have the potential to deliver reliable and valid assessment of implicit race-related attitudes. However, latent variable models are required to facilitate this use by addressing the significant amount of error variance in the IAT scores. This requirement has implication for the valid use of individualised feedback.

Implications for the Valid Use of Implicit Attitude Measures

Implications for the Validity of using Affective Priming Tasks to Assess Attitudes

In Study One, the priming measures (APTs) were found to barely assess the attitude constructs and to do so inconsistently. The poor reliability and validity evident for the APT suggests limits to its use in applied research, if these results can be generalised beyond the current studies. The standard APT script was applied in this thesis, as available through the Inquisit program (Millisecond Software, 1996) and as recommended by Fazio (personal communication, 14 May, 2009). The data produced using these standard scripts delivered consistently poor results for both the Race-related and Country-related constructs. This suggests poor reliability and validity might not be limited to the current study, especially since previous research also suggests poor psychometrics for the APT. For instance, Krause et al. (2010) examined the reliability of several implicit self-esteem measures and found sub-optimal reliabilities for the APT (r=.29). Further, Falk et al. (2013) found poor validity evidence for the APT when measuring implicit self-esteem in Canadian and

Japanese participants, concluding the APT is not suitable for individual or cross-cultural research. These findings support the conclusions of the current thesis - that the APT is neither a valid nor stable measure of implicit attitudes.

The results of this thesis corroborate previous evidence of poor validity for the APT, implying researchers should not rely on the APT to assess implicit attitudes accurately. It is of great concern that priming techniques have been used in applied psychological research for almost thirty years without sufficient psychometric validation. The results of this thesis imply that implicit biases are better assessed using the IAT, though as discussed presently; this test also has inherent limitations.

Implications for the Validity of using Implicit Association Tests to Assess Attitudes

Verbal Implicit Association Test (VIAT).

The verbal form of the IAT (VIAT) is the most widely used and researched implicit attitudinal technique (Spence, 2005), and has previously been subjected to a plethora of reliability and validity investigations. However, previous investigations of the IAT typically examined observed scores (e.g. Greenwald et al., 2009; Rudolph et al., 2008), which were demonstrated in the current thesis to be highly confounded by error variance (see also Siers & Christiansen, 2013). The present research applied a novel approach for IAT studies by addressing error variance using CFA, thereby providing a critical extension of previous research.

Study One provided psychometric support for the racial and country VIATs, with good internal consistency (CR=.76 for both), internal convergent validity (AVE=.44

for both) and internal construct validity demonstrated by the CFA models (latent construct factor loadings >.58 for the Racial VIAT, >.60 for the Country VIAT). Both VIATs also loaded substantively onto their respective second-order latent attitude factors (β=.74 for both). The VIATs thus provided a relatively consistent measure of the trait constructs once error variance was accounted for. In addition, Study Two demonstrated strong construct validity for the Racial VIAT using CFA-MTMM, as trait variance was greater than method variance (29% versus 17%). In Study Three, this same task provided evidence of anti-Arab prejudice in the sample, illustrating the functional utility of this measure.

However, the Country VIAT was not found to possess good construct validity in Study Two, as the trait variance was substantially less than the method variance (4% versus 41%). This was despite the fact that both the Country and Racial VIATs were identical with only the attribute name stimuli differing (personal names versus country names). Such inconsistencies make it difficult to ascertain whether VIATs in general have good construct validity. Previous research by Siers and Christiansen (2013) reported greater method variance than trait variance for three different personality VIATs, which means they all would have failed the strict assessment of construct validity applied in Study Three. Such findings suggest caution in using VIATs as measures of implicit attitudes. However, the fact that the VIATs have demonstrated some construct validity, once error variance was accounted for, implies that it may be possible to develop VIATs that could produce valid estimates of attitudinal constructs. Accordingly validated IATs would need to be highly refined and standardised measures that routinely demonstrate greater trait than method variance.

In summary, VIATs appear to have potential for use in applied research settings with sample populations as a measure of implicit bias. However, significant psychometric validation, involving CFA and CFA-MTMM, as well as the use of latent modelling techniques for the analysis of IAT data would be required to facilitate this potential.

Pictorial Implicit Association Test (PIAT).

The PIAT performed very comparably to the traditional VIAT in all analyses. As such, Study One demonstrated adequate support for the PIAT's reliability (CR=.76 for Racial PIAT, CR=.77 for Country PIAT), internal convergent validity (AVE=.45 for Racial PIAT, AVE=.46 for Country PIAT), and internal construct validity (latent construct factor loadings >.60 for the Racial PIAT, >.58 for the Country PIAT). Furthermore, both VIATs and PIATs loaded comparably and strongly onto the second-order latent attitude factors (Racial attitude: β=.74 VIAT, β=.77 PIAT; Country attitude: β=.74 VIAT, β=.72 PIAT), which strengthens the construct validity of the measures and the comparability of the IAT formats. In Study Two, the VIAT and PIAT latent method factors were found to be substantially related (r=.57), further reinforcing the equivalency of these tasks. Such findings provide the first psychometric support for the construct validity of a fully pictorial IAT.

PIATs offer two notable advantages over the traditional VIAT methodology. PIATs produce significantly faster reaction times than VIATs, increasing the automaticity of the task and allowing for greater numbers of trials to be run (see also Baron & Banaji, 2006; Dasgupta et al., 2000; Thomas, 2008). This effect was corroborated in this thesis. Increased automaticity is likely due to pictorial stimuli requiring less effortful processing than word stimuli (Carr et al., 1982; Glaser & Glaser, 1989). Higher automaticity in the quicker processing of pictorial over verbal stimuli could theoretically result in an even more automated (or unconscious) impression of the participants' biases. Furthermore, faster reaction times allow participants to complete the same number of trials in a shorter period of time, which may reduce fatigue or difficulties sustaining attention for the participants. These faster reaction times also enable researchers to run more trials without fatigue effects than has previously been possible. Increasing the number of trial has been suggested to increase the reliability of the IAT (Siers & Christiansen, 2013). Using the Spearman-Brown Prophecy formula, these authors suggested that the number of trials in their IATs would need to be increased from 96 to 485 trials to gain an adequate reliability level. This estimate is not dissimilar from the 408 trials used in the current study, which did reveal adequate reliability estimates for the IATs (see Study One). The routine use of PIATs would substantially reduce the time taken to complete over 400 trials and thus could be encouraged as one avenue for increasing the general reliability of IATs.

The second methodological advantage of the PIAT is that it avoids the need for verbal fluency, thereby expanding the possible participant pool for the task. For example, young children were previously excluded from IAT investigations due to a lack of verbal fluency, but are now able to participate using the PIAT (Thomas et al., 2007). The study of implicit attitudes in very young children has potential to provide rich information regarding the development of implicit stereotypes (e.g. Cvencek et al., 2011; Dohnt & Tiggemann, 2005; Thomas et al., 2007). The PIAT could also be used for cross-cultural research, as pictorial stimuli do not require specific language comprehension[30]. Such research may enhance theoretical understanding of implicit biases across sociological and geographical divides.

Summary

The APT was shown to have limited applied use due to a lack of reliability and validity, whereas both forms of the IAT appeared to have the capacity to deliver valid estimates of implicit attitudes. Yet the high amount of error variance generated by IATs and difficulties associated with the IAT effect score provide some significant limitations that require consideration for future research.

Future Use of Implicit Attitude Measures

The key finding of this dissertation is that substantial amounts of error variance confound implicit attitudinal scores. Because of this, the main implications for future research are that 1) latent modelling analytical strategies such as CFA are required to account for the error variance in implicit attitudinal data and 2) each implicit attitudinal measure requires thorough psychometric validation to support its use.

The Use of Structural Equation Modelling in Implicit Attitudinal Research

SEM strategies have previously been mainly applied to explicit attitude measures (e.g. Burns et al., 2006; Hendrick, Fischer, Tobi, & Frewer, 2013; Nelson, Benson, & Jensen, 2010) and have only sporadically been used to assess laboratory measures such as the IAT (e.g. Blanton, et al., 2006; Cunningham et al., 2001; Nosek & Smyth, 2007; Siers & Christiansen, 2012). This thesis has shown that latent variable approaches are extremely useful for analysing and interpreting implicit attitude data.

[30] It is noted that substantial effort would be required to develop stimuli that are interpreted reasonably equally by all participants. Multiple-groups CFA is one approach to ensure equivalency of stimuli.

The CFA-MTMM analysis in Study Two demonstrated that over 80% of an IAT effect score is composed of error variance. In other words, random and systematic errors are the major contributors to the IAT effect score, and the implicit attitude construct comprises a relatively small portion of the observed scores. These influences are demonstrated in the conceptual model presented in Figure 8.2. This model illustrates that observed IAT scores do contain trait variance, but are highly confounded by random and systematic error. The use of latent modelling procedures enabled greater clarity regarding the extent of this influence, with random error shown to contribute over 50% of the IAT effect scores, systematic error about 30% and trait variance less than 20%.

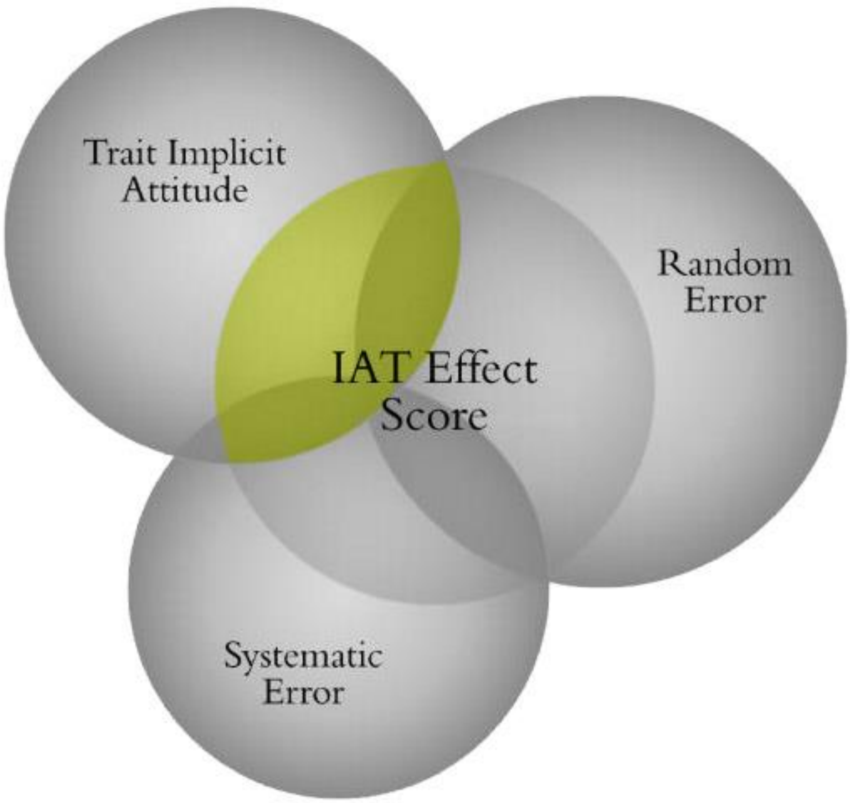

Figure 8.2. Conceptual diagram of contributing influences for IAT effect scores, with implicit attitudes tapped by the measure highlighted.

The large proportion of error variance in IAT scores makes it virtually impossible to gain a clear idea of the supposed underlying implicit attitudes using traditional analytical approaches that rely on observed scores. To adequately examine attitudes of interest, highlighted in yellow in Figure 8.2, latent modelling techniques like SEM are crucial to partial out the unwanted non-trait-related variance. The present research suggests that all future implicit attitudinal research would do well to use latent variable approaches to analyse IAT data. Research that fails to account for error variance in this way is very unlikely to provide an accurate reflection of the participants' implicit associations and should be treated with caution.

The routine application of SEM to all implicit attitude data would likely result in two substantial benefits. Firstly, SEM will greatly enhance the validity and efficacy of implicit attitude measures by addressing the significant issue of error variance in IAT effect scores. Secondly, applying SEM in this way should help increase the psychometric standards of laboratory techniques such as the IAT. This will allow more confidence in IAT findings.

Limitations of the IAT Effect Score as Currently Calculated

The need for SEM has significant implications for the use of the IAT effect score as a diagnostic tool of personal implicit prejudices. Current practice is that on completion of an IAT, the participant is provided with an indication of how strong their implicit bias is for X over Y construct based on the strength of their IAT effect score. IAT effect scores are interpreted using Greenwald et al.'s (2003) guidelines, whereby moderate levels of prejudice are indicated by *D* effect scores greater than

.35. Although arbitrary in nature (Blanton & Jaccard, 2006), these cut-off scores ostensibly provide some indication of the general level of prejudice revealed by the IAT. However, the present research has indicated limitations to implementing the IAT effect score in this way. In future, IAT effect scores would need to be standardised across IAT type and must account for error variance.

IAT Effect Score Comparability between IAT Types.

IAT effect score cut-offs, as described by Greenwald et al. (2003), are not easily transferred to findings produced using the PIAT. On average, the PIAT delivers much smaller IAT effect scores than the VIAT (Thomas, 2008). This difference reflects the fact that pictorial stimuli are easier to categorise than verbal stimuli (see Carr et al., 1982), rather than indicating a difference in attitudes. Similar issues for the IAT effect score have previously been noted for age and intelligence (Blanton et al., 2006; Hummert et al., 2002). Faster overall reaction times for the PIAT reduce the total discrepancy between the average congruent and incongruent reaction times, thereby resulting in a smaller IAT effect score. The greater efficiency of the PIAT thus makes it very difficult to compare ‘prejudice levels’ between verbal and pictorial-based IAT research. Development of comparison scores for the VIAT and PIAT could be a useful avenue for future research that would enable meaningful research discussion across method types.

IAT Effect Scores as a Measure of Individual Bias.

The use of SEM to analyse IAT data poses an even greater problem than comparability for IAT estimates of implicit bias. This thesis has emphasised the need for SEM analytical strategies to account for significant error variance in the IAT effect scores. However, SEM approaches 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. Nevertheless, individual feedback continues to routinely be presented, e.g. Project Implicit (https://implicit.harvard.edu/implicit/; Greenwald et al., 2011). Given the current findings, it is strongly suggested future IAT research provide summarised findings at a sample population level, after SEM is applied, rather than providing personal, and potentially misleading, feedback to participants.

Individual Validation of Each Implicit Attitude Task

Inconsistencies evident between the psychometric properties of the implicit attitude measures suggest that individual psychometric validation is required for each and every adaptation of the IAT.

Psychometric Validation of IATs.

The IAT was designed to be easily adapted to assess many and varied constructs (Greenwald et al., 2009; Spence, 2005). However, in trying to be so adaptable, the IAT appears to have forfeited the ability to be a consistently valid measurement technique. In Study Two it was demonstrated that two of the four empirical IATs did not pass the stringent CFA-MTMM assessment of construct validity, despite being ostensibly very comparable techniques. For instance, the only difference between the Racial VIAT (which had good construct validity) and the Country VIAT (which did not) was the category-based stimuli for one task depicted Middle Eastern and European first names (e.g. ‘Habib’ or ‘Penny’), the other Middle Eastern and European country names (e.g. ‘Iraq’ or ‘Ireland’). These discrepant results suggest IATs cannot easily and validly be applied to assess many and varied constructs, but rather each and every IAT requires thorough psychometric validation, as per the lengths outlined in this thesis (see also Lane et al., 2007).

Thorough Piloting of Stimuli Exemplars.

All components of an IAT will require thorough validation, including the stimuli and category exemplars. This is critical given the stimuli, for instance, can significantly influence the outcome of an IAT, as was evident for the Racial PIAT in Study Three (see also Dasgupta & Greenwald, 2001; Lane et al., 2007). Stimuli development is a key component of a successful and valid IAT. However, there are currently no standardised rules to aid stimuli set construction. Stimuli exemplars are typically suggested to be as specific as possible (Han et al., 2009) and to not be confused by valence (Fiedler et al., 2006). As such, reasonably neutral stimuli such as “Afghanistan”, “Habib” or “Mosque” would be viewed as more appropriate stimuli for a VIAT assessing attitudes towards the Middle East than "Terrorist" or "Torture". However, it has been argued these latter depreciative terms, although unlikely to represent the experimenter's notion of the Middle East, may be very representative of the associative structure of a prejudiced individual (Fiedler et al., 2006). In this way, neutral stimuli may actually act to conceal negative attitudes expressed by a highly prejudiced person. Development of one ideal set of stimuli considered representative of all attitudinal positions is thus likely to be very difficult.

Valid Use of the Implicit Association Test for Future Research

The Implicit Association Test was designed to assess many and varied constructs (Greenwald et al., 1998; Lane et al., 2007), yet there are no standardised rules for the construction of viable IATs (Fiedler et al., 2006). This is highly problematic for the development of a reliable and valid measurement instrument, as changes in stimuli (Study Three; see also Dasgupta et al., 2000), category exemplars (Han et al., 2009), and testing context (Lowery et al., 2001) can all highly influence the IAT results and the psychometric properties of these measures (see results of this thesis, also Lane et al., 2007). An easily adaptable measure that is also psychometrically robust seems improbable. Highly standardised and psychometrically valid measures, such as IQ tests for instance, cannot be changed on the whim of a researcher for it is evident that any changes to the format and questions will affect the veracity of the result. In order for the IAT to increase psychometric robustness, emphasis on standardisation and validation of a small number of IATs, rather than extensive applications, should be the primary focus.

Valid use of the IAT in applied research requires the tasks are first well validated, like the Racial VIAT and Country PIAT of the present research. Such standardised IATs would need to demonstrate greater trait than method variance, and use validated and representative stimuli and category exemplars. When these IATs are applied to investigate the implicit attitudes of a sample population, the data would require analysis using latent modelling approaches to account for the significant error component of the scores. In order to provide individual assessments of implicit bias, development of an IAT effect score that incorporates a correction for error would be necessary. A standardised set of comparison latent-based scores that could be used to estimate level of individual prejudice from a participant's observed IAT scores may also be an option. These endeavours would require extensive research, but are theoretically possible avenues for future assessments of implicit bias.

Summary

Significant implications for the future use of IATs and APTs were realised through the present research. The APT was shown to be an inadequate measure of implicit attitudes and should be restricted to task development aimed at improving its reliability and validity. In contrast, both the VIAT and PIAT appear to have potential for use in applied research even though significant efforts to manage the large error component of the IAT effect scores are required. Future IAT research should concentrate on psychometric development using latent modelling techniques and standardisation of a small number of measures in order to increase consistency and validity of implicit prejudice assessment. With continued development and psychometric investigation it is possible the IAT may yet prove suitable for providing valid insights into unconscious biases.

Limitations of Current Research

Limitations of the CTCM CFA-MTMM Model

The CTCM CFA-MTMM analysis provided critical evidence for the high proportion of error variance within IAT scores, a central argument for this dissertation. However, it must be noted that the CTCM CFA-MTMM approach is known to suffer from convergence and admissibility problems, often due to empirical underidentification (Lance, Noble, & Scullen, 2002). Underidentified models have less known parameters than unknown parameters. This is problematic as there are infinite number of parameter estimates that may result in perfect model fit (Brown, 2006). CTCM models are thus prone to producing improper solutions with out-of-range parameter estimates (Marsh & Grayson, 1995). It has been argued that CTCM models require an excessive amount of parameters, causing the MTMM structure to become ‘overparameterized’ (Geiser, Eid, & Nussbeck, 2008). Because of this, it has been suggested that model parameter estimates may be underestimated in CTCM modelling, presenting unrepresentatively low or even non-significant factor loadings (Geiser, et al., 2008). Given these limitations, there is a possibility that the estimates of trait variance found in these studies may be an underestimate of the true trait variance. The impact of different CFA-MTMM model specifications on trait, method and error variance estimates for IAT data is worthy of future research investigation. However, given the nature of the low trait validity estimates obtained, it is still likely that assessment of the trait variance using other models (e.g. CTCM-1) would produce low values.

Advantages of Including Additional Implicit Attitude Measures

The findings reported in this thesis are limited to the APT and IAT, and as such it is not possible to ascertain whether high error variance is a problem shared by all implicit attitude measures. There are currently numerous implicit attitudinal techniques available, including the Go/No-Go Association Task (GNAT; Nosek & Banaji, 2001), the East Affective Simon Task (EAST; De Houwer & De Bruycker, 2007), the Affect Misattribution Procedure (AMP; Payne, Cheng, Govorun, & Stewart, 2005), and the Name-Letter Test (NLT; Nuttin, 1985), as well as more recent variants of the IAT and APT, such as the Single-Category IAT (SC-IAT; Steinman & Karpinski, 2008) and the Response-Window APT (RW-APT; Krause et al., 2012). Additional implicit attitude measures in the current research would have furthered the generalisability of the findings and increased opportunity for assessing convergent validity amongst the implicit attitude measures. It is expected that high error variance also affects the newer implicit attitude measures, given recent findings of questionable validity for these tasks (Buhrmester, Blanton, & Swann, 2011; Falk et al., 2013). For instance, Falk et al. (2013) stated there has been “no improvement in the validity of these new (implicit attitude) measures” (p.21). Such findings indicate the need for thorough psychometric investigation of all implicit attitude measures, which is a critical avenue for future research.

Advantages of Including Additional Explicit Attitude Measures

Additional explicit attitude measures in this research could have provided a more conclusive estimate of the discriminant validity between implicit and explicit attitude measures. This is because the factor loadings for each measurement type could have been compared to ensure consistency. For the racial attitude construct, the Symbolic

Racism 2000 Scale (Henry & Sears, 2002) would have been a useful comparison for the Modern Racism Scale (MRS; McConahay et al., 1981), as it was devised specifically to accommodate the MRS' limitations. For the country attitude construct, further research would have been required to devise another explicit country attitude questionnaire. The additional measures could have enabled a comparison of the relationships between latent implicit and explicit attitudes using higher-order CFA, with a comparison of one-factor versus two-factor second-order models. Such an examination would enable greater clarity as to whether implicit and explicit attitude measures are better conceptualised as similar or distinct constructs (see p.280).

Representativeness of the Sample Population

A convenience sample predominantly drawn from the University of Tasmania was used for the current research. Although participants were sourced from campuses in South, North and North-Western Tasmania, such a sample is still University-based and thus unlikely to be representative of the Tasmanian or Australian populations. Further demographic information, including the occupations of the participant or their parents, educational level of their parents, and suburb of residence, would have helped facilitate an estimate of the representativeness of this sample. Future applied research may choose to employ a community-based participant sample to overcome this limitation.

Limitations of IAT Stimuli

Representativeness is also important in the selection of IAT stimuli. Given what is now known, the selection of stimuli may have required greater consideration during task development. The finding that the Racial PIAT did not produce the expected IAT effect highlights the importance of extensive piloting of suitable and representative stimuli exemplars. The development of a standardised set of validated stimuli for use with implicit attitude measures would be beneficial for applied research. Use of open science forums that encourage collaboration amongst researchers may be one way of facilitating the development and dissemination of such validated items (e.g. The Open Science Framework; Spies & Nosek, 2012).

Limitations of the Explicit Travel Assessment

A more precise estimate of travel experience may have proved beneficial. Specific questions relating to travel experience within Islamic or Middle Eastern countries may have aided the MIMIC analysis of Study Three[31]. Furthermore, enquiring whether the participants knew or were friends with anyone from the Muslim faith or Middle Eastern background may have enabled the friendship hypothesis to be evaluated. The friendship hypothesis states that the prejudice levels of people who are friends with/or in regular contact with members of the out-group, such as Arab/Muslims, are significantly lower than those who do not have contact with out-group members (Cashin, 2010; Pettigrew, 1998; Saad, 2006). Were participants to have rated level of interaction with Arab/Muslims, as opposed to travel experience, it may have been possible to ascertain whether such relationships do have a mediating influence on implicit prejudice. Future research may investigate this matter further.

[31] It is noted that given the predominance of university students in the sample, the likelihood of many of them having the financial means to explore the Middle East by this stage of life is perhaps limited.

Directions for Future Research

Based on the findings of this dissertation, the following outline some potential avenues for future research.

Psychometric Validation of Implicit Attitude Measures

Thorough psychometric validation using SEM is strongly encouraged for all implicit attitude measures being used in applied research. This includes all IAT adaptations, as well as more recent implicit attitudinal techniques, such as the EAST or AMP. Recent validity assessments have not yet provided promising evidence for the robustness or veracity of these techniques (Buhrmester et al., 2011; Falk et al., 2013; Krause et al., 2010; Rudolph et al., 2008). However, the Racial VIAT and Country PIAT of the present research were found to possess good construct validity using CFA-MTMM. Further SEM-based research would aid psychometric validation of these instruments and help clarify the potential functional utility, or not, of these tasks.

Application of SEM to Further Theoretical and Psychometric Investigation

SEM strategies proved a useful analytical approach to facilitate task validation in this thesis. Further SEM-based research opportunities are outlined, which show SEM can be used to obtain greater theoretical understanding of implicit and explicit attitude concepts, as well as further psychometric investigation.

Higher-order CFA

The theorised relationship between implicit and explicit forms of attitude assessment has been contentious (Cunningham, Nezlek, et al., 2004; Krause et al., 2010; Nosek & Smyth, 2007). Higher-order CFA can assist in determining whether implicit and explicit attitudes are more appropriately viewed as distinct constructs or different expressions of the same underlying construct. The experimental design involves a comparison of model fit for two second-order CFA models. Data from two implicit attitude measures and two explicit attitude measures is required. The two-factor second-order CFA model represents distinct second-order implicit and explicit attitude factors; the one-factor second-order CFA model represents a general attitude factor. If the two-factor second-order model provided better model fit than the one-factor solution it would imply that implicit and explicit attitudes are most suitably conceived as distinct constructs. Alternatively, if the one factor model had superior fit then a singular attitudinal construct would be implied. In this way, greater theoretical clarity regarding implicit and explicit attitudes would be facilitated.

Reliability Estimation

Higher-order CFA models can also be used to examine omega-hierarchical reliability, which can be conceptualised as an extension of composite reliability (CR) estimation (applied in Study One). The hierarchical coefficient omega examines the extent to which all of the items in a test measure the same latent variable (Zinbarg, Revelle, Yovel, & Li, 2005). However in contrast to CR, the hierarchical coefficient omega determines how much of the indicators are accounted for by the second-order factor, rather than the first. Hierarchical coefficient omega may thus prove a useful tool for further validation of the IATs by analysing how consistently and how well the latent implicit attitude construct is being measured. Although beyond the scope of the current research, longitudinal research examining the stability of IATs over time using a test-retest approach, would also be critical for establishing the reliability of implicit attitudinal techniques. Here, the use of multiple implicit and other attitude measures would provide more robust estimates of the stability of implicit attitude measurement over time. Multiple-method longitudinal designs would also allow the convergent and discriminant validity of implicit attitude measures to be assessed at both trait (stable) and occasion-specific (momentary) levels. Using a SEM framework, a future longitudinal approach might provide more comprehensive reliability and validity estimation for IAT data.

Further Identification of Systematic Error Influences for IATs

Significant method effects were found to influence IAT effect scores in Study Two. Several potential causes of systematic error for IATs were outlined in Chapter Three. Future research could investigate the specific impacts of particular sources of method effects for the IAT using the CFA-MTMM analytical framework. For instance, two IATs could be designed to capture different method effects and then compared using the CFA-MTMM methodology. Additionally, systematic factors such as intelligence, general processing speed, or task-switching ability could be determined using an intelligence test or a non-attitude based reaction-time measure, and then the results could be added as covariates using the MIMIC methodology to ascertain whether such systematic influences significantly impact the IAT effect scores. These analyses have the potential to clarify the relative influence of systematic forms of error variance on IAT effect scores.

Establishment of IAT Scoring Procedures that Account for Error Variance

The endeavour of accounting for error variance in the scoring process of IATs is an important area for future research. Were it to be possible, a new scoring process that accounted for error would deliver a much more robust estimate of implicit prejudice than currently produced by Greenwald et al.'s (2003) *D* score. In the meantime, participants can be provided with summarised feed-back based on the pooled participant's data that has been analysed with CFA. Future research may invest in the development of a standardised set of latent scores that can be reliably related to observed IAT scores in order to deliver an estimate of individual implicit bias. Such an endeavour would require extensive research and validation. In the interim, IAT research should be constrained to the assessment of sample populations that can be analysed with CFA.

Population-based Research: Cross-sectional and Cross-cultural Designs using the PIAT

Research of sample populations can provide useful insights into group attitudes, as well as shifts in attitude over time and location. A cross-sectional study examining implicit racial attitudes in children (using the PIAT) through to the elderly would provide an array of information regarding the development of attitudes across the lifetime[32]. It may also elucidate additional information regarding method effects such as the impact of age on general processing speed. The PIAT may also prove useful for examining attitudinal formation and development for specific attitudinal constructs. For example, Thomas et al. (2007) found a gender difference in preschool

[32] Multiple-methods of attitude assessment are advantageous both in longitudinal and cross-sectional research, as this strategy might provide a more robust overview of the participant's attitudes. Given the high proportion of error variance in implicit attitude measures, all analyses would need to be conducted using SEM procedures.

respondents' attitudes towards body size using a Thin/Fat PIAT, with girls showing the 'thin is good, fat is bad' ideology at a much younger age than the boys. Such a finding could have important clinical implications for development of body dissatisfaction and disturbance in girls (Hendy, Gustitus, & Lietzel-Scwalm, 2001; Lowes & Tiggemann, 2003). Whether this effect is consistent across constructs could be investigated to determine if it is specifically body size that young girls are attuned to earlier than boys, or if boys in general internalise implicit attitude constructs at a later age than girls. There is also great potential for utilising the non-language dependent qualities of the PIAT for cross-cultural investigations, as previously discussed. In summary, were well validated VIATs and PIATs to become available, there are many applications for future research using these measures to address research enquiries of theoretical and clinical value.

Conclusion

Implicit attitude measures such as the Affective Priming Task (APT) and Implicit Association Test (IAT) produce substantial quantities of random error and method variance that heavily confound the findings. This thesis demonstrated the utility of Structural Equation Modelling analytical techniques, such as Confirmatory Factor Analysis, to systematically evaluate and account for such error variance in order to assess the psychometric properties of these tasks. Implicit attitude measures were shown to not routinely deliver valid or reliable estimates of implicit attitudes.

The APT was found to be an invalid measure of implicit attitudes, as the APT scores were comprised almost entirely of error variance, with minimal trait construct assessed. In contrast, the verbal and pictorial formats of the IAT were shown to possess adequate reliability and construct validity following the removal of error variance using SEM. High convergence between the verbal and pictorial IAT formats was consistently demonstrated, which provided the first psychometric support for a fully pictorial IAT. Good construct validity was also evident for the Racial VIAT and Country PIAT using Multitrait-Multimethod matrices. These findings imply IATs can provide an estimate of implicit bias provided error in the IAT scores is accounted for using CFA.

However, the IATs do not consistently deliver adequate assessments of implicit bias. Extensive psychometric validation is required for each adaptation of the IAT. Further, all IAT data must be analysed using latent modelling procedures to partial out the significant error component of the scores. As such, future IAT research is limited to investigations of sample populations, and individual diagnostic feedback should be avoided. This is especially pertinent when investigating socially sensitive attitudes such as racial prejudice. In conclusion, the VIAT and PIAT have some potential to provide valid estimates of implicit biases in applied population-based research, but only once the significant error component of the scores is addressed.

References

Agerström, J., & Rooth, D.O. (2009). Implicit prejudice and ethnic minorities: Arab-Muslims in Sweden. *International Journal of Manpower, 30*, 43-55. doi: 10.1108/01437720910948384

Amodio, D. M. (2013). Implicit prejudice and the regulation of intergroup responses. In B. Derks, D. Scheepers & N. Ellemers (Eds.), *The Neuroscience of Prejudice* (pp. 167-189). New York: Psychology Press.

Anderson, G. F., & Hussey, P. S. (2000). Population aging: A comparison among industrialized countries. *Health Affairs, 19*, 191-203. doi: 10.1377/hlthaff.19.3.191

Asendorf, J. B., Banse, R., & Mücke, D. (2002). Double dissociation between implicit and explicit personality self-concept: The case of shy behaviour. *Journal of Personality and Social Psychology, 83,* 380-393. doi: 10.1037//0023-3514.83.2.380

Ayers, I. (2001). *Pervasive Prejudice? Unconventional Evidence of Race and Gender Discrimination*. Chicago, USA: University of Chicago Press.

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

Bagozzi, R. P. (1991). Further thoughts on the validity of measures of elation, gladness, and joy. *Journal of Personality and Social Psychology, 61*, 98-104. doi: 10.1037/0022-3514.61.1.98

Bandalos, D. L. (2002). The effects of item parceling on goodness-of-fit and parameter estimate bias in Structural Equation Modeling. *Structural Equation Modelling, 9,* 78-102. doi: 10.1207/S15328007SEM0901_5

Banse, R. (1999). Automatic evaluation of self and significant others: Affective priming in close relationships. *Journal of Social and Personal Relationships, 16*, 803-821. doi: 10.1177/0265407599166007

Banse, R., Seise, J., & Zerbes, N. (2001). Implicit attitudes towards homosexuality: Reliability, validity, and controllability of the IAT. *Zeitschrift fuer Experimentelle Psychologie, 48*, 145-160. doi: 10.1026//0949-3946.48.2.145

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. doi: 10.1037/0022-3514.62.6.893

Bargh, J. A., Chaiken, S., Raymond, P., & Hymes, C. (1996). The automatic evaluation effect: Unconditional automatic activation with a pronounciation task. *Journal of Experimental Social Psychology, 32,* 104-128. doi: 10.10.0022-1031/96

Barlow, F. K., Louis, W. R., & Terry, D. J. (2010). Minority report: Social identity, cognitions of rejection and intergroup anxiety predicting prejudice from one racially margnialized group towards another. *European Journal of Social Psychology, 40,* 805-818. doi: 10.1002/ejsp.651

Baron, A. S., & Banaji, M. R. (2006). The development of implicit attitudes: Evidence of race evaluations from ages 6 and 10 and adulthood. *Psychological Science, 17,* 53-58. doi: 10.1111/j.1467-9280.2005.01664.x

Baumeister, R. F., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology, 5,* 323-370. doi: 10.1037//1089-2680.5.4.323

Bessenoff, G. R., & Sherman, J. W. (2000). Automatic and controlled components of prejudice toward fat people: Evaluation versus stereotype activation. *Social Cognition, 18,* 329-353. doi: 10.1521/soco.2000.18.4.329

Blanton, H., & Jaccard, J. (2006). Arbitrary metrics in psychology. *American Psychologist, 61*, 27-41. doi: 10.1037/0003-066X.61.1.27

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

Blechert, J., Ansorge, U., Beckmann, S., & Tuschen-Caffier, B. (2011). The undue influence of shape and weight on self-evaluation in anorexia nervosa, bulimia nervosa and restrained eaters: A combined ERP and behavioral study. *Psychological Medicine, 41*, 185-194. doi: 10.1017/S003329170000395

Boniface, S., & Shelton, N. (2013). How is alcohol consumption affected if we account for under-reporting? A hypothetical scenario. *The European Journal of Public Health, 23*, 30-46. doi: 10.1093/eurpub/ckt016

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

Brown, J., Kranzler, H. R., & Del Boca, K. F. (1992). Self-reports by alcohol and drug abuse inpatients: Factors affecting reliability and validity. *British Journal of Addiction, 87*, 1012-1024. doi: 10.1111/j.1360-0443.1992.tb03118.x

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. (2006). *Confirmatory Factor Analysis for applied research*. New York: Guilford Press.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing Structural Equation Models* (pp. 136-162). Beverly Hills, CA: Sage.

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

Burns, G. L., Gomez, R., Hafetz, N., & Walsh, J. A. (2006). Measurement and structural invariance of parent ratings of ADHD and ODD symptoms across gender for American and Malaysian children. *Psychological Assessment*, *18*, 452-457. doi: 10.1037/1040-3590.18.4.452

Burns, G. L., & Haynes, S. N. (2006). Clinical psychology: Construct validation with multiple sources of information and multiple settings. In M. Eid & E. Diener (Eds.), *Handbook of multimethod measurement in psychology* (pp. 401-418). Washington, DC: American Psychological Association.

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

Byrne, B. M. (2005). Factor analytic models: Viewing the structure of an assessment instrument from three perspectives. *Journal of Personality Assessment, 85*, 17-32. doi: 10.1207/s15327752jpa8501_02

Byrne, B. M., Shavelson, R. J., & Muthén, B. O. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105*, 456-466. doi: 10.1037/0033-2909.105.3.456

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

Carr, T. H., McCauly, C., Sperber, R. D., & Parmelee, C. M. (1982). Words, pictures, and priming: On semantic activation, conscious identification, and the automaticity of information processing. *Journal of Experimental Psychology, Human Perception and Performance, 8*, 757-777. doi: 10.1037/0096-1523.8.6.757

Cashin, S. C. (2010). To be "Muslim-looking" in America: A comparative exploration of racial and religious prejudice in the 21st century. *Duke Forum for Law and Social Change, 2*, 125-139.

Chopra, S. (2008). *Anti-Arab and anti-Muslim sentiment amongst potential jurors: Underlying psychological constructs and tools for identification*. Paper presented at the American Society of Trial Consultants, Chicago.

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

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159. doi: 10.1037/0033-2909.112.1.155

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

Cote, J. A., & Buckley, M. R. (1988). Measurement error and theory testing in consumer research: An illustration of the importance of construct validation. *Journal of Consumer Research, 14*, 579-582.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297-334. doi: 10.1007/BF02310555

Cunningham, W. A., Johnson, M. K., Gattenby, J. C., Gore, J. C., & Banaji, M. R. (2003). Neural components of social evaluation. *Journal of Personality and Social Psychology, 85*, 639-649. doi: 10.1037/0022-3514.85.4.639

Cunningham, W. A., Johnson, M. K., Raye, C. L., Gattenby, J. C., Gore, J. C., & Banaji, M. R. (2004). Separable neural components in the processing of black and white faces. *Psychological Science, 15*, 806-813. doi: 10.1111/j.0956-7976.2004.00760.x

Cunningham, W. A., Nezlek, J. B., & Banaji, M. R. (2004). Implicit and explicit ethnocentrism: Revisiting the ideologies of prejudice. *Personality and Social Psychology Bulletin, 30*, 1332-1346. doi: 10.1177/014616720-426-4654

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., & Greenwald, A. G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality and Social Psychology, 81*, 800-814. doi: 10.1037/0022-3514.81.5.800

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

Dawis, R. V. (1987). Scale construction. *Journal of Counseling Psychology, 34*, 481-489. doi: 00122-0167/87

De Houwer, J., & De Bruycker, E. (2007). The identification-EAST as a valid measure of implicit attitudes toward alcohol-related stimuli. *Journal of Behavior Therapy and Experimental Psychiatry, 38*, 95-104. doi: 10.1016/j.jbtep.2006.10.004

Debner, J. A., & Jacoby, L. L. (1994). Unconscious perception: Attention, awareness and control. *Journal of Experimental Psychology, 20*, 304-317. doi: 10.1037/h0043350

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56*, 5-18. doi: 10.1037//0022-3514.56.1.5

Devine, P. G., Plant, A. E., Amodio, D. M., Harmon-Jones, E., & Vance, S. L. (2002). The regulation of explicit and implicit race bias: The role of motivations to respond without prejudice. *Journal of Personality and Social Psychology, 82*, 835-848. doi: 10.1037//0022-3514.82.5.835

Dillon, W., & Goldstein, M. (1984). *Multivariate analysis: Methods and applications*. New York: Wiley.

Dohnt, H. K., & Tiggemann, M. (2005). Peer influences on body dissatisfaction and dieting awareness in young girls. *British Journal of Developmental Psychology, 23*, 103-116. doi: 10.1348/026151004X20658

Dolnicar, S. (2005). Understanding barriers to leisure travel: Tourist fears as a marketing basis. *Journal of Vacation Marketing, 11*, 197-208. doi: 10.1177/1356766705055706

Donders, F. C. (1969). On the speed of mental processes. *Acta Psychologica, 30,* 412-431. doi: 10.1016/0001-6918(69)90065-1

Donkin, C., Brown, S. D., Heathcote, A., & Wagenmakers, E. (2011). Diffusion versus linear ballistic accumulation: Different models but the same conclusions about psychological processes? *Psychonomic Bulletin and Review, 18*, 61-69. doi: 10.3758/s13423-010-0022-4

Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology, 82*, 62-68. doi: 10.1037/0022-3514.82.1.62

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

Dunn, K. M., Forrest, J., Ip, D., Babacan, H., Paradies, Y., & Pedersen, A. (2008). *Challenging Racism: The anti-racism research project*. Paper presented at the 4Rs Conference: Rights, reconciliation, respect, responsibility, Sydney, Australia.

Dwyer, C. (1993). Constructions of Muslim identity and the contesting of power: The debate over Muslim schools in the UK. In P. Jackson & J. Penrose

(Eds.), *Constructions of Race, Place and Nation* (pp. 143-159). London: University College London Press.

Ebbinghaus, H. (1885). Memory: A contribution to experimental psychology. In C. D. Green (Series Ed.) Classics in the History of Psychology, Retrieved from http://psy.ed.asu.edu/~classics/Ebbinghaus/index.htm

Eid, M., Lischetzke, T., Trierweiler, L. I., & Nussbeck, F. W. (2003). Separating trait effects from trait-specific method effects in multitrait-multimethod models: A multiple-indicator CT-C(M-1) model. *Psychological Methods, 8*, 38-60. doi: 10.1037/1082-989X.8.1.38

Egloff, B., Schwerdtfeger, A., & Schmukle, S. C. (2005). Temporal stability of the Implicit Association Test - Anxiety. *Journal of Personality Assessment, 84*, 82-88. doi: 10.1207/s15327752jpa8401_14

Endler, N. S., & Hunt, J. (1968). Triple-interaction variance in the S-R Inventory of Anxiousness. *Perceptual and Motor Skills, 27*, 1098.

Fadda, E., Fronza, S., Galimberti, E., & Bellodi, L. (2011). The implicit "Thin Ideal" in eating disorders. *European Psychiatry, 26*, 720. doi: 10.1016/S0924-9338(11)72425-7

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

Fazio, R. H. (2009, 14 May). [RE: Evaluative Priming enquiry from Tasmania].

Fazio, R. H., 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. doi: 10.1037//0022-3514.69.6.1013

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology, 54*, 297-327. doi: 10.1146/annurev.psych.54.101601.145225

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. doi: 10.1037//0022-3514.50.2.229

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford, California: Stanford University Press.

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

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*, 39-50. doi: 10.2307/3151312

Fry, A. F., & Hale, S. (2000). Relationships among processing speed, working memory, and fluid intelligence in children. *Biological Psychology, 54*, 1-34. doi: 10.1016/j.intell.2010.07.003

Gau, J. M. (2010). Basic principles and practices of structural equation modeling in criminal justice and criminology research. *Journal of Criminal Justice Education, 21*, 136-151. doi: 10.1080/10511251003693660

Gawronski, B. (2002). What does the Implicit Association Test measure? A test of the convergent and discriminant validity of prejudice-related IATs. *Experimental Psychology, 49*, 171-180. doi: 10.1027//1618-3169.49.3.171

Gawronski, B. (2009). Ten frequently asked questions about implicit measures and their frequently supposed, but not entirely correct answers. *Canadian Psychology*, *50*, 141-150. doi: 10.1037/m7013848

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin*, *132*, 692-731. doi: 10.1037/0033-2909.132.5.692

Gawronski, B., LeBel, E. P., Banse, R., & Peters, K. R. (2009). Methodological issues in the validation of implicit measures: Comment on De Houwer, Teige-Mocigembra, Spruyt, and Moors (2009). *Psychological Bulletin*, *135*, 369-372. doi: 10.1037/a00014820

Gawronski, B., LeBel, E. P., & Peters, K. R. (2007). What do implicit measures tell us? Scrutinizing the validity of three common assumptions. *Perspectives on Psychological Science*, *2*, 181-193. doi: 10.1111/j.1745-6916.2007.00036.x

Geisser, C., Eid, M., & Nussbeck, F.W. (2008). On the meaning of the latent variables in the CT-C(M-1) model: A comment on Maydeu-Olivares and Coffman (2006). *Psychological Methods*, *13*, 49-57. doi: 10.1037/1082-989X.13.1.49

Geiser, C., & Lockhart, G. (2012). A comparison of four approaches to account for method effects in latent state-trait analyses. *Psychological Methods*, *17*, 255-283. doi: 10.1037/a0026977

Gibson, J. L. (2008). Intolerance and political repression in the United States: A half century after McCarthyism. *American Journal of Political Science*, *52*, 96-108. doi: 10.1111/j.1540-5907.2007.00301.x

Gignac, G. E. (2009). Psychometrics and the measurement of emotional intelligence. In C. Stough, D. H. Saklofske & J. D. C. Parker (Eds.), Assessing Emotional Intelligence; Theory, research and applications (pp. 9-40). New York: Springer Science. doi: 10.1007/978-0-387-88370-0_2

Gilens, M., Sniderman, P. M., & Kuklinski, J. H. (1998). Affirmative action and the politics of realignment. *British Journal of Political Science, 28*, 159-183. doi: 10.1017/S0007123498000143

Glaser, M. O., & Glaser, W. R. (1989). Time course analysis of the Stroop phenomenon. *Journal of Experimental Psychology: Human Perception and Performance, 8*, 875-894. doi: 10.1037//0096-1523.8.6.875

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

Graham, J. G. (2006). Congeneric and (essentially) tau-equivalent estimates of score reliability: What they are and how to use them. *Educational and Psychological Measurement, 66*, 930-944. doi: 10.1177/0013164406288165

Green, A. R., Carney, D. R., Pallin, D. J., Ngo, L. H., Raymond, K. L., Iezzoni, L. I., & Banaji, M. R. (2007). Implicit bias among physicians and its prediction of thrombolysis decisions for black and white patients. *Journal of General Internal Medicine, 22*, 1231-1238. doi: 10.1007/s11606-007-0258-5

Greenwald, A. G. (2004). The resting parrot, the dessert stomach, and other perfectly defensible theories. In J. T. Jost, M. R. Banaji & D. A. Prentice (Eds.), *The ying and yang of scientific progress: Perspectivism in social psychology* (pp. 275-285). Washington, DC: American Psychological Association.

Greenwald, A. G., Banaji, M., & Nosek, B. A. (2011). Project Implicit Retrieved 14th August, 2013, from https://www.projectimplicit.net

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review, 102*, 4-27. doi: 10.1037/0033-295X.102.1.4

Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., & Mellott, D. S. (2002). A unified theory of implicit attitudes, sterotypes, self-esteem, and self-concept. *Psychological Review, 109*, 3-25. doi: 10.1037//0033-295X.109.1.3

Greenwald, A. G., & Farnham, S. D. (2000). Using the Implicit Association Test to measure self-esteem and self-concept. *Journal of Personality and Social Psychology, 79*, 1022-1038. doi: 10.1037//0022-3514.79.6.1022

Greenwald, A. G., & Krieger, L. H. (2006). Implicit bias: Scientific foundations. *California Law Review, 94*, 945-972. doi: 10.2307/20439056

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

Greenwald, A. G., Nosek, B. A., Banaji, M. R., & Klauer, K. C. (2005). Validity of the salience asymmetry interpretation of the Implicit Association Test: Comment on Rothermund and Wentura (2004). *Journal of Experimental Psychology, 134*, 420-425. doi: 10.1037/0096-3445.134.3.420

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test; III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology, 97*, 17-41. doi: 10.1037/a0015575

Gschwendner, T., Hofmann, W., & Schmitt, M. (2008). Convergent and predictive validity of implicit and explicit anxiety measures as a function of specificity similarity and content similarity. *European Journal of Psychological Assessment, 24*, 254-262. doi: 10.1027/1015-5759.24.4.254

Guarino, M. H. (2005). A comprehensive and comprehensible guide to Structural Equation Modelling. *PsycCRITIQUES, 50*, np. doi: 10.1037/040947

Haeffel, G. J., Abramson, L. Y., Brazy, P. C., Shah, J. Y., Teachman, B. A., & Nosek, B. A. (2007). Explicit and implicit cognition: A preliminary test of a dual-process theory of cognitive vulnerability to depression. *Behaviour Research and Therapy, 45*, 1155-1167. doi: 10.1016/j.brat.2006.09.003

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate Data Analysis* (6th ed.). New Jersey: Prentice Hall.

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

Harman, H. H. (1960). *Modern factor analysis*. Chicago: The University of Chicago Press.

Harrison, D. A., Mclaughlin, M. E., & Coalter, T. M. (1996). Context, cognition, and common method variance: Psychometric and verbal protocol evidence. *Organizational Behavior and Human Decision Processes, 68,* 246-261. doi: 10.1006/obhd.1996.0103

Hendrick, T. A. M., Fischer, A. R. H., Tobi, H., & Frewer, L. J. (2013). Self-reported attitude scales: Current practice in adequate assessment of reliability, validity, and dimensionality. *Journal of Applied Social Psychology, 43*, 1538-1552. doi: 10.1111/jasp.12147

Hendy, H. M., Gustitus, C., & Lietzel-Scwalm, J. (2001). Social cognitive predictors of body image in preschool children. *Sex Roles*, *44*, 557-570.

Henry, P. J., & Sears, D. O. (2002). The Symbolic Racism 2000 Scale. *Political Psychology*, *23*, 253-283. doi: 10.1111/0162-895X.00281

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

Hoyle, R. H. (Ed.). (1995). *Structural Equation Modeling; Concepts, issues and applications*. Thousand Oaks, California: SAGE.

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.

Hummert, M. L., Garstka, T. A., O'Brien, L. T., Greenwald, A. G., & Mellott, D. S. (2002). Using the Implicit Association Test to measure age differences in implicit social cognitions. *Psychology and Aging, 17*, 482-495. doi: 10.1037//0882-7974.17.3.482

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

Jajodia, A., & Earleywine, M. (2003). Measuring alcohol expectancies with the Implicit Association Test. *Psychology of Addictive Behaviors, 17*, 126-133. doi: 10.1037/0893-164X.17.2.126

Jensen, A. R. (1993). Why is reaction time correlated with psychometric *g*? *Current Directions in Psychological Science, 2*, 53-56. doi: 10.1111/1467-8721.ep10770697

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

Jöreskog, K. G., & Goldberger, A. S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association, 70,* 631-639. doi: 10.1080/01621459.1975.10482485

Jost, J. T., Rudman, L. A., Blair, I. V., Carney, D. R., Dasgupta, N., Glaser, J., & Hardin, C. D. (2009). The existence of implicit bias is beyond reasonable doubt: A refutation of ideological and methodological objections and executive summary of ten studies that no manager should ignore. *Research in Organizational Behavior, 29,* 36-69. doi: 10.1016/j.riob.2009.10.001

Kaplan, R. M., & Saccuzzo, D. P. (2001). *Psychological testing; Principles, applications and issues* (Fifth ed.). Belmont: Wadsworth/ Thomas Learning.

Karpinski, A., & Hilton, J. L. (2001). Attitudes and the Implicit Association Test. *Journal of Personality and Social Psychology, 81,* 774-788. doi: 10.1037//0022.3514.81.5.774

Kawakami, K., & Dovidio, J. F. (2001). The reliability of implicit stereotyping. *Personality and Social Psychology Bulletin, 27*, 212-225. doi: 10.1177/0146167201272007

Klauer, K. C., & Musch, J. (2001). Does sunshine prime loyal? Affective priming in the naming task. *Journal of Experimental Psychology, 54*, 727-751. doi: 10.1080/713755986

Klein-Flügge, M. C., Nobbs, D., Pitcher, J. B., & Bestmann, S. (2013). Variability of Human Corticospinal Excitability Tracks the State of Action Preparation. *The Journal of Neuroscience, 33*, 5564 –5572. doi: 10.1523/JNEUROSCI.2448-12.2013

Kline, R. B. (1998). *Principles and practice of structural equation modelling*. New York, USA: Guilford Press.

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

Kozak, M., Crotts, J. C., & Law, R. (2007). The impact of the perception of risk on international travellers. *International Journal of Tourism Research, 9*, 233-242. doi: 10.1002/jtr.607

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

Krause, S., Back, M. D., Egloff, B., & Schmukle, S. C. (2012). A new reliable and valid tool for measuring implicit self-esteem: The Response-Window Affective Priming Task. *European Journal of Psychological Assessment, 28*, 87-94. doi: 10.1027/1015-5759/a000095

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.

Lepp, A., & Gibson, H. (2003). Tourist roles, perceived risk and international tourism. *Annals of Tourism Research, 30,* 606-624. doi: 10.1016/j.tourman.2007.08.002

Leventhal, A. M., Waters, A. J., Breitmeyer, B. G., Miller, E. K., Tapia, E., & Li, Y. (2008). Subliminal processing of smoking-related and affective stimuli in tobacco addiction. *Experimental and Clinical Psychopharmacology, 16*, 301-312. doi: 10.1037/a0012640

Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling: A Multidisciplinary Journal, 9*, 151-173. doi: 10.1207/S15328007SEM0902_1

Lowery, B. S., Hardin, C. D., & Sinclair, S. (2001). Social influence effects on automatic racial prejudice. *Journal of Personality and Social Psychology, 81*, 842-855. doi: 10.1037//0022-3514.81.5.842

Lowes, J., & Tiggemann, M. (2003). Body dissatisfaction, dieting awareness and the impact of parental influence in young children. *British Journal of Health Psychology, 8,* 135-147. doi: 10.1348/135910703321649123

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.

McCaul, M., Lourens, A., & Kredo, T. (2012). Pre-hospital versus in-hospital thrombolysis for acute myocardial infarction. *Cochrane Database of Systematic Reviews, 11*(CD010191). doi: 10.1002/14651858.CD010191

McConahay, J. B., Hardee, B. B., & Batts, V. (1981). Has racism declined in America? It depends on who is asking and what is asked? *Journal of Conflict Resolution, 25,* 563-579. doi: 10.1177/002200278102500401

McConnell, A. R., & Leibold, J. M. (2001). Relations among the Implicit Association Test, discriminatory behavior, and explicit measures of racial attitudes. *Journal of Experimental Social Psychology, 37,* 435-442. doi: doi:10.1006/jesp.2000.1470

Meade, A. W., & Kroustalis, C. M. (2006). Problems with item parceling for Confirmatory Factor Analytic tests of measurement invariance. *Organizational Research Methods, 9,* 369-403. doi: 10.1177/1094428105283384

Meade, A. W., Watson, A. M., & Kroustalis, C. M. (2007). *Assessing common method bias in organizational research*. Paper presented at the 22nd Annual Meeting of the Society for Industrial and Organizational Psychology, New York.

Messick, S. (1990). *Validity of Test Interpretation and Use*. Princeton, N.J.: Educational Testing Service.

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

Miller, P., & Katz, D. B. (2010). Stochastic transitions between neural states in taste processing and decision-making. *Journal of Neuroscience, 30*, 2559-2570. doi: 10.1523/JNEUROSCI.3047-09.2010

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

Murray, D. M., & Perry, C. L. (1987). The measurement of substance use among adolescents: When is the ‘bogus pipeline’ method needed? *Addictive Behaviors, 12*, 225-233. doi: 10.1016/0306-4603(87)90032-3

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

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. doi: 10.1037//0096-3445.106.3.226

Nelson, T. D., Benson, E. R., & Jensen, C. D. (2010). Negative attitudes toward physical activity: Measurement and role in predicting physical activity levels among preadolescents. *Journal of Pediatric Psychology, 35*, 89-98. doi: 10.1093/jpepsy/jsp040

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*, 231-259. doi: 10.1037/0033-295X.84.3.231

Norman, G. R., & Streiner, D. L. (1994). *Biostatistics: The bare essentials*. St. Louis: Mosby.

Nosek, B. A. (2005). Moderators of the relationship between implicit and explicit evaluation. *Journal of Experimental Psychology, 134*, 565-584. doi: 10.1037/0096-3445.134.4.565

Nosek, B. A. (2007). Implicit-explicit relations. *Current Directions in Psychological Science, 16*, 65-69. doi: 10.1111/j.1467-8721.2007.00477.x

Nosek, B. A., & Banaji, M. R. (2001). The go/no-go association task. *Social Cognition, 19*, 625-666. doi: 10.1521/soco.19.6.625.20886

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2005). Understanding and Using the Implicit Association Test: II. Method Variables and Construct Validity. *Personality & Social Psychology Bulletin, 31*, 166-180. doi: 10.1177/0146167204271418

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.

Nosek, B. A., & Hansen, J. J. (2008). The associations in our heads belong to us: Searching for attitudes and knowledge in implicit evaluation. *Cognition and Emotion, 22*, 553-594. doi: 10.1080/02699930701438186

Nosek, B. A., & Smyth, F. L. (2007). A multitrait-mutimethod validation of the Implicit Association Test: Implicit and explicit attitudes are related by distinct constructs. *Experimental Psychology, 54*, 14-29. doi: 10.1027/1618-3169.54.1.14

Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ranganath, K. A., Smith, C. T., Olson, K. R., Chugh, D., Greenwald, A. G., & Banaji, M. R. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology, 18*, 36-88. doi: 10.1080/10463280701489053

Novick, M. R., & Lewis, C. (1967). Coefficient alpha and the reliability of composite measurement. *Psychometrika, 32*, 1-13. doi: 10.1007/BF02289400

Nunkoo, R., & Ramkissoon, H. (2011). Structural equation modelling and regression analysis in tourism research. *Current Issues in Tourism, iFirst article*, 1-26. doi: 10.1080/13683500.2011.641947

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

Nuttin, J. M. (1985). Narcissism beyond Gestalt and awareness: The name letter effect. *European Journal of Social Psychology, 15*, 353-361. doi: 10.1002/ejsp.2420150309

Olson, K. R., & Fazio, R. H. (2004). Reducing the influence of extrapersonal associations on the Implicit Association Test: Personalizing the IAT. *Journal of Personality and Social Psychology, 86*, 653-667. doi: 10.1037/0022-3514.86.5.653

Olson, M. A., & Fazio, R. H. (2003). Relations between implicit measures of prejudice: What are we measuring? *Psychological Science, 14*, 636-639. doi: 10.1046/j.0956-7976.2003.psci_1477.x

Pachella, R. G. (1974). The Interpretation of Reaction Time in Information Processing Research. In B. Kantowitz (Ed.), *Human Information Processing: Tutorials in performance and cognition*. New York: Lawrence Earlbaum Associates.

Palfai, T. P., & Ostafin, B. D. (2003). Alcohol-related motivational tendencies in hazadous drinkers: Assessing implicit respons tendencies using the modified-IAT. *Behaviour Research and Therapy, 41*, 1149-1162. doi: 10.1016/S0005-7967(03)00018-4

Park, J., Felix, K., & Lee, G. (2007). Implicit attitudes toward Arab-Muslims and the moderating effects of social information. *Basic and Applied Social Psychology*, *29*, 35-45. doi: 10.1080/01973530701330942

Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology*, *89*, 277-293. doi: 10.1037/0022-3514.89.3.277

Perugini, M. (2005). A commentary on Spence (2005): Using implicit tasks in attitude research. A review and a guide. *Social Psychological Review, 7*, 21-24.

Perugini, M., & Banse, R. (2007). Personality, implicit self-concept and automaticity. *European Journal of Personality, 21*, 257-261. doi: 10.1002/per.637

Pettigrew, T. F. (1998). Intergroup contact theory. *Annual Review of Psychology, 49,* 65-85. doi: 10.1146/annurev.psych.49.1.65

Phelps, E. A., O'Connor, K. J., Cunningham, W. A., Funayama, E. S., Gatenby, J. C., Gore, J. C., & Banaji, M. R. (2000). Performance on indirect measures of race evaluation predicts amygdala activation. *Journal of Cognitive Neuroscience, 12,* 729-738. doi: 10.1162/089892900562552

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

Pohl, S., & Steyer, R. (2010). Modeling common traits and method effects in Multitrait-Multimethod analysis. *Multivariate Behavioral Research, 45*, 45-72. doi: 10.1080/00273170903504729

Pohl, S., Steyer, R., & Kraus, K. (2008). Modelling method effects as individual causal effects. *Journal of the Royal Statistical Society, 171*, 41-63. doi: 10.1111/j.1467-985X.2007.00517.x

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): InTech. Retrieved from http://www.intechopen.com/books/contemporary-trends-in-adhd-research/variability-noise-and-predictability-in-motor-response-times-adaptation-or-misadaptation. doi: 10.5772/28166

Poynting, S., & Mason, V. (2007). The resistable rise of Islamophobia; Anti-Muslim racism in the UK and Australia before 11 September 2001. *Journal of Sociology, 43*, 61-86. doi: 10.1177/1440783307073935

Ralston, T. E., & Palfai, T. P. (2012). Depressive symptoms and the implicit evaluation of alcohol: The moderating role of coping motives. *Drug and Alcohol Dependence, 122*, 149-151. doi: 10.1016/j.drugalcdep.2011.09.011

Ranganath, K. A., Smith, C. T., & Nosek, B. A. (2008). Distinguishing automatic and controlled components of attitudes from direct and indirect measurement methods. *Journal of Experimental Social Psychology, 44*, 386-396. doi: 10.1016/j.jesp.2006.12.008

Rashid, M. (2009). *Testing our tolerance: American attitudes toward limitations on the first amendment rights of Arab and Muslim Americans*. Department of Political Science. Barnard College, Columbia University. New York

Raykov, T. (1997). Estimation of composite reliability for congeneric measures. *Applied Psychological Measurement, 21*, 173-184. doi: 10.1177/01466216970212006

Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory Factor Analysis and Item Response Theory: Two approaches for exploring measurement invariance. *Psychological Bulletin, 114*, 552-566. doi: 10.1037/0033-2909.114.3.552

Riolo, F. (2010). The psychoanalytic status of the term "unconscious": A present-day perspective. In F. Riolo (Ed.), *The Italian Psychoanalytic Annual 2010: The unconscious and repression* (pp. 7-22). Rome, Italy: Borla Edizioni SRL.

Room, R. (2005). Stigma, social inequality and alcohol and drug use. *Drug and Alcohol Review, 24*, 143-155. doi: 10.1080/09595230500102434

Rooth, D. (2010). Automatic associations and discriminations in hiring: Real world evidence. *Labour Economics, 17*, 523-534. doi: 10.1016/j.labeco.2009.04.005

Rowatt, W. C., Franklin, L. M., & Cotton, M. (2005). Patterns and personality correlates of implicit and explicit attitudes towards Christians and Muslims. *Journal for the Scientific Study of Religion, 44*, 29-43. doi: 10.1111/j.1468-5906.2005.00263.x

Rudman, L. A., & Ashmore, R. D. (2007). Discrimination and the Implicit Association Test. *Group Processes and Intergroup Relations, 10*, 350-372. doi: 10.1177/1368430207078696

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

Rydell, R. J., & Gawronski, B. (2009). I like you, I like you not: Understanding the formation of context-dependent automatic attitudes. *Cognition and Emotion, 23*, 1118-1152. doi: 10.1080/02699930802355255

Saad, L. (2006). Anti-Muslim sentiments fairly commonplace: Four in ten Americans admit feeling prejudiced against Muslims. *Gallup News*. Retrieved from http://www.gallup.com/poll/24073/AntiMuslim-Sentiments-Fairly-Commonplace.aspx

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

Salthouse, T. A., & Hedden, T. (2002). Interpreting reaction time measures in between-group comparisons. *Journal of Clinical and Experimental Neuropsychology, 24*, 858-872. doi: 10.1076/jcen.24.7.858.8392

Saltin, J. F., & Strand, B. C. (1995). Analysis and control of newsprint quality and paper machine operation using integrated factor networks. *Pulp & Paper Canada*, *96*, 48-56.

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika, 66*, 507-514. doi: 10.1007/BF02296192

Schmitt, N., & Stults, D. M. (1986). Methodology review: Analysis of multitrait-multimethod matrices. *Applied Psychological Measurement, 10,* 1-22. doi: 10.1177/014662168601000101

Schmukle, S. C., & Egloff, B. (2004). Does the Implicit Association Test for assessing anxiety measure trait and state variance? *European Journal of Personality, 18,* 483-494. doi: 10.1002/per.525

Schnabel, K., Asendorf, J. B., & Greenwald, A. G. (2008a). Assessment of individual differences in implicit cognition: A review of IAT measures. *European Journal of Psychological Assessment, 24,* 210-217. doi: 10.1027/1015-5759.24.4.210

Schnabel, K., Asendorf, J. B., & Greenwald, A. G. (2008b). Using Implicit Association Tests for the assessment of implicit personality self-concept. In G. J. Boyle, G. Matthews & D. H. Saklofske (Eds.), *Handbook of Personality Theory and Testing* (pp. 508-528). London: Sage.

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

Schweitzer, R., Perkoulidis, S. A., Krome, S. L., & Ludlow, C. N. (2005). Attitudes towards refugees: The dark side of prejudice in Australia. *Australian Journal of Psychology, 57,* 170-179. doi: 10.1080/00049530500125199

Sherman, S. J., Rose, J. S., Koch, K., Presson, C. C., & Chassin, L. (2003). Implicit and explicit attitudes toward cigarette smoking: The effects of context and motivation. *Journal of Social and Clinical Psychology, 22,* 13-39. doi: 10.1521/jscp.22.1.13.22766

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

Skrondal, A., & Laake, P. (2001). Regression among factor scores. *Psychometrika, 66*, 563-576. doi: 10.1007/BF02296196

Soanes, C., & Stevenson, A. (Eds.). (2006). *Concise Oxford English Dictionary* (11th revised ed.). Oxford, UK: Oxford University Press.

Sonmez, S., & Graefe, A. R. (1998). Influence of terrorism risk on foreign tourism decisions. *Annals of Tourism Research, 25*, 112-144. doi: 10.1016/S0160-7383(97)00072-8

Spalding, L. R., & Hardin, C. D. (1999). Unconscious unease and self-handicapping: Behavioral consequences of individual differences in implicit and explicit self-esteem. *Psychological Science, 10*, 535-539. doi: 10.1111/1467-9280.00202

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

Spence, A. (2005). Using implicit tasks in attitude research: A review and a guide. *Social Psychological Review, 7*, 2-17.

Spies, J. R., & Nosek, B. A. (2012). Open Science Framework. *Open Science Collaboration* Retrieved 7th January, 2014, from https://osf.io/4znzp/wiki/home/

Spruyt, A., Hermans, D., Pandelaere, M., De Houwer, J., & Eelen, P. (2004). On the replicability of the Affective Priming Effect in the pronounciation task. *Experimental Psychology, 51,* 109-115. doi: 10.1027/1618-3169.51.2.109

Stanley, D., Phelps, E., & Banaji, M. R. (2008). The neural basis of implicit attitudes. *Current Directions in Psychological Science, 17*, 164-170. doi: 10.1111/j.1467-8721.2008.00568.x

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

Steinman, R. B. (2011). Advances in indirect questioning techniques in marketing research. In M. A. Goralski, H. P. LeBlanc & M. G. Adams (Eds.), *Business Research Yearbook* (Vol. 18, pp. 448-454): International Academy of Business Disciplines.

Steinman, R. B., & Karpinski, A. (2008). The Single Category Implicit Association Test (SC-IAT) as a measure of implicit consumer attitudes. *European Journal of Social Sciences, 7*, 32-42.

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behaviour. *Personality and Social Psychology Review, 8*, 220-247. doi: 10.1207/s15327957pspr0803_1

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18,* 643-662. doi: 10.1037//0096-3445.121.1.15

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

Swanson, J. E., Rudman, L. A., & Greenwald, A. G. (2001). Using the Implicit Association Test to investigate attitude-behaviour consistency for stigmatised behaviour. *Cognition and Emotion, 15*, 207-230. doi: 10.1080/0269993004200060

Tabachnick, B. G., & Fidell, L. S. (2001). *Using Multivariate Statistics* (4th ed.). Boston: Allyn and Bacon.

Tetlock, P. E., & Mitchell, G. (2009). Implicit bias and accountability systems: What must organizations do to prevent discrimination? *Research in Organizational Behavior, 29,* 3-38. doi: 10.1016/j.riob.2009.10.002

Thomas, S. R. (2008). *Investigating the Reliability and Validity of a Non-Verbal Adaptation of the IAT.* (Masters of Clinical Psychology Masters Thesis), University of Tasmania, Hobart.

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.

Thompson, B. (2004). *Exploratory and Confirmatory Factor Analysis: Understanding concepts and applications*. Washington, DC: American Psychological Association.

Thompson, M. S., & Green, S. B. (2006). Evaluating between-group differences in latent variable means. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modelling: A second course* (pp. 119-170). Greenwich, Connecticut: Information Age.

Thorndike, E. L. (1920). A constant error in psychological ratings. *Journal of Applied Psychology, 4*, 25-29. doi: 10.1037/h0071663

Thurstone, L. L. (1947). *Multiple-factor analysis*. Chicago: University of Chicago Press.

Thush, C., Wiers, R. W., Ames, S. L., Grenard, J. L., Sussman, S., & Stacy, A. W. (2008). Interactions between implicit and explicit cognition and working memory capacity in the prediction of alcohol use in at-risk adolescents. *Drug and Alcohol Dependence, 94*, 116-124. doi: 10.1016/j.drugalcdep.2007.10.019

Trzesniewski, K. H., Donnellan, M. B., Moffitt, T. E., Robins, R. W., Poulton, R., & Caspi, A. (2006). Low self-esteem during adolescence predicts poor health, criminal behavior, and limited economic prospects during adulthood. *Developmental Psychology, 42*, 381-390. doi: 10.1037/0012-1649.42.2.381

Tseng, W. T., Dörnyei, Z., & Schmitt, N. (2006). A New Approach to Assessing Strategic Learning: The case of self-regulation in vocabulary acquisition. *Applied Linguistics*, *27*, 78-102. doi: 10.1093/applin/ami046

United Nations. (2009). *World Statistics Pocketbook; Economic and social affairs*. New York: USA: United Nations Publications.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literatures: Suggestions, practices and recommendations for Organizational research. *Organizational Research Methods*, *3*, 4-70. doi: 10.1177/109442810031002

Vandromme, H., Hermans, D., & Spruyt, A. (2011). Indirectly measured self-esteem predicts gaze avoidance. *Self and Identity, 10*, 32-43. doi: 10.1080/15298860903512149

Vartanian, L. R., Herman, C. P., & Polivy, J. (2005). Implicit and explicit attitudes toward fatness and thinness: The role of the internalization of societal standards. *Body Image, 2,* 373-381. doi: 10.1016/j.bodyim.2005.08.002

Watson, L., & Armon, M. P. (2004). Thrombolysis for acute deep vein thrombosis. *Cochrane Database of Systematic Reviews, 4*(CD002783). doi: 10.1002/14651858.CD002783

Wiers, R. W., Woerden, N., Smulders, F., & de Jong, P. J. (2002). Implicit and explicit alcohol-related cognitions in heavy and light drinkers. *Journal of Abnormal Psychology, 111*, 648-658. doi: 10.1037//0021-843X.111.4.648

Williams, B. R., Hultsch, D. F., Tannock, R., Strauss, E. H., & Hunter, M. A. (2005). Inconsistency in reaction time across the life span. *Neuropsychology, 19*, 88-96. doi: 10.1037/0894-4105.19.1.88

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

Wilson, E., & Little, D. E. (2008). The solo female travel experience: Exploring the 'geography of women's fear'. *Current Issues in Tourism, 11*, 167-186. doi: 10.2167/cit342.0

Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review, 107*, 101-126. doi: 10.1037//0033-295X.107.1.101

Wittenbrink, B. (2007). Measuring attitudes through priming. In B. Wittenbrink & N. Schwarz (Eds.), *Implicit measures of attitudes* (pp. 17-59). New York, USA: Guilford Press.

Wittenbrink, B., Judd, C. M., & Park, B. (1997). Evidence for racial prejudice at the implicit level and its relationship with questionnaire measures. *Journal of*

Personality and Social Psychology, 72, 262-274. doi: 10.1037/0022-3514.72.2.262

Zimmerman, D. W. (1972). Test reliability and the Kuder-Richardson formulas: Derivation from probability theory. *Educational and Psychological Measurement, 32*, 939-954. doi: 10.1177/001316447203200408

Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α, Revelle's β, and McDonald's ωH:Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika, 70*, 123-133. doi: 10.1007/s11336-003-0974-7

Appendices

Appendix A – Positive and Negative Facial Icons and Logos


NS1


NS2


NS3


NS4


NS5


Negative Facial Logo


PS1


PS2


PS3


PS4


PS5


Positive Facial Logo

Appendix B – Flower and Insect Stimuli and Logo


Fl1


Fl2


Fl3


Fl4


Fl5


Flower Logo


I1


I2


I3


I4


I5


Insect Logo

Appendix C –Arab Facial Stimuli

Arab1

Arab2

Arab3

Arab4

Arab5

Arab6

Arab7

Arab8

Arab Logo

Appendix D –European Facial Stimuli


Euro1


Euro2


Euro3


Euro4


Euro5


Euro6


Euro7


Euro8


Euro Logo

Appendix E –Middle Eastern Landmark Stimuli


ME1


ME2


ME3


ME4


ME5


ME6


ME7


ME5

Arab World Logo

Appendix F –European Landmark Stimuli


E1


E2


E3


E4


E5


E6


E7


E5

Europe Logo

Appendix G – Verbal IAT Stimuli

Flower Words
- Hyacinth
- Marigold
- Orchid
- Rose
- Bluebell
- Daffodil
- Buttercup
- Daisy
- Violet
- Lily

Insect Words
- Ant
- Caterpillar
- Locust
- Fly
- Maggot
- Bee
- Cockroach
- Mosquito
- Wasp
- Dragonfly

Unpleasant Words
- Filth
- Grief
- Stink
- Assault
- Disaster
- Hatred
- Pollute
- Tragedy
- Ugly
- Rotten

European Names
- Sean
- Andrew
- Harry
- Charles
- James
- Lily
- Ingrid
- Suzanna
- Penny
- Mary

Arabic Names
- Abdul
- Habib
- Jamal
- Mohammed
- Rahman
- Khalidah
- Hana’
- Laylali
- Nashita
- Basha’ir

Pleasant Words
- Freedom
- Love
- Peace
- Friend
- Loyal
- Pleasure
- Honest
- Family
- Happy
- Laughter

European Countries
- France
- Italy
- Switzerland
- United Kingdom
- Sweden
- Netherlands
- Ireland
- Hungary
- Czech Republic
- Finland

Middle Eastern Countries
- Saudi Arabia
- Egypt
- Libya
- Iraq
- Bahrain
- Palestine
- Algeria
- Morocco
- Syria
- Lebanon

Appendix H – Student Options Questionnaire incorporating the adapted Modern Racism Scale (MRS)

Student Opinions

I'd like you to read some statements on a variety of issues. Some of these you might agree with; others you might even find offensive. These statements do not necessarily reflect the opinions of the researchers. For each one please indicate if you disagree strongly (1), disagree somewhat (2), have no opinion (3), agree somewhat (4), or agree strongly (5) by circling the appropriate number.

	Strongly Disagree	Somewhat Disagree	No Opinion	Somewhat Agree	Strongly Agree
1. People who are overweight tend to be a little untrustworthy	1	2	3	4	5
2. Marijuana use should be legalised in Australia	1	2	3	4	5
3. Over the past few years the government and news media have shown more respect to Arabs than they deserve	1	2	3	4	5
4. Economic concerns should always take precedence over environmental issues	1	2	3	4	5
5. Over the past few years Arabs have got more economic power than they deserve	1	2	3	4	5
6. Sometimes suicide is the only escape from life's problems	1	2	3	4	5
7. Men care for children equally as well as women	1	2	3	4	5
8. The legal drinking age in Australia should be raised to 21 years	1	2	3	4	5
9. Support services are geared towards women	1	2	3	4	5
10. It is easy to understand the resentment of Arabs living in Australia	1	2	3	4	5

	Strongly Disagree	Somewhat Disagree	No Opinion	Somewhat Agree	Strongly Agree
11. Sex education should be taught in grade one in all government primary schools	1	2	3	4	5
12. Arabs are getting too demanding in their push for acceptance of their cultural and religious practices by Western Society	1	2	3	4	5
13. It's better to be on your own than in a relationship	1	2	3	4	5
14. Being gay or bisexual is as natural as being straight	1	2	3	4	5
15. Becoming depressed is part of being old	1	2	3	4	5
16. Many Arabs living in Australia miss out on good housing because Australian owners won't rent or sell to them	1	2	3	4	5
17. You should only go to the doctor when there is something seriously wrong	1	2	3	4	5
18. Many Arabs living in Australia miss out on jobs or promotions because of racial discrimination	1	2	3	4	5
19. People with poor stamina deal with life problems by becoming depressed	1	2	3	4	5
20. Global climate change gains more media coverage than it deserves	1	2	3	4	5

Items 3, 5, 10, 12, 16 and 18 are the racially relevant items adapted from the Modern Racism Scale (MRS; McConahay, Hardee, & Batts, 1981).

McConahay, J. B., Hardee, B. B., & Batts, V. (1981). Has racism declined in America? It depends on who is asking and what is asked? *Journal of Conflict Resolution, 25*, 563-579.

Appendix I – Travel Destination Questionnaire (TDQ)

Travel Destination Questionnaire

For the following countries, I'd like you to rate how much you would like to visit them. If you have visited a country before, then think about how much you would like to go visit the country again. For each country please indicate if you definitely not, probably not, maybe, probably would, or definitely would like to visit the location by circling the appropriate number.

	Definitely not	Probably not	Maybe	Probably would	Definitely would
1. Indonesia	1	2	3	4	5
2. Syria	1	2	3	4	5
3. Poland	1	2	3	4	5
4. Italy	1	2	3	4	5
5. Saudi Arabia	1	2	3	4	5
6. Thailand	1	2	3	4	5
7. Malaysia	1	2	3	4	5
8. Israel	1	2	3	4	5
9. France	1	2	3	4	5
10. China	1	2	3	4	5
11. United Kingdom	1	2	3	4	5
12. Jordan	1	2	3	4	5
13. Spain	1	2	3	4	5
14. Singapore	1	2	3	4	5
15. United Arab Emirates	1	2	3	4	5
16. Hungary	1	2	3	4	5
17. Lebanon	1	2	3	4	5
18. Japan	1	2	3	4	5

Appendix J – Subjective Measure of Travel Experience

How well travelled would you say you are?

0 – does going across the river count?
1 – I've been to the mainland once
2 – I've been to the mainland several times/ I've been out of the country once
3 – I've been to more than one country overseas
4 – I've been to several countries overseas/ I've been to more than one continent (out of Australia)
5 – I'd say I'm pretty well travelled (ie. Numerous countries/ continents and experiences abroad)

Appendix K – Participant Information Sheet

UNIVERSITY OF TASMANIA

Evaluating the Construct Validity of Implicit Association Tasks using Confirmatory Factor Analysis Models

Dear Participant,

My name is Susan Thomas and I am completing a PhD at the School of Psychology, University of Tasmania. My supervisors are Prof. Rapson Gomez and Mr Peter Ball. Together we aim to determine the construct validity of the Implicit Association Task (IAT), a computer task designed to measure attitudes. I am looking for participants aged 18-60 years to complete ten tasks in total, which should not take more than 2 hours.

The tasks include 6 IATs that involve categorising words or pictures on a computer. The words or pictures may depict flowers and insects, names or pictures of people from different countries, names of countries or pictures of buildings from different parts of the world. Two implicit priming tasks, which are similar to the IAT and also involve categorising words, will also be presented. There will be two questionnaires: one enquires after your attitudes towards Arabs in an Australian context, the other measures travel preferences. The tasks are relatively simple to do and generally people will complete all ten tasks within 2 hours. Most people do not find the tasks stressful or disturbing. However, if you are sensitive about viewing realistic pictures of insects, photos of people, or answering questions regarding your attitudes toward Arabs, you should not take part in this study. Further, it you find you are distressed after data collection and debriefing, you may be referred to the University Psychology Clinic for extra assistance. To participate in this research it is important that you are fluent in English. Data will be gathered at the School of Psychology, University of Tasmania.

If at any stage you do not wish to continue you may withdraw at any time without prejudice. Your responses will be anonymous, identified by code numbers only, and will remain confidential. All data will be kept in secure storage at the School of Psychology for at least five years after completion of the study, after that time data will be destroyed either by shredding hard copies of documents and/or by deleting all electronic files. If you are a first-year Psychology student, course credit will be provided for your participation.

This study has received approval from the Human Research Ethics Committee (Tasmania). If you have any concerns of an ethical nature or complaints about the manner in which the project is conducted, you may contact the Human Research Ethics Committee (Tasmania) Network (Tel. 6226 7479; E-mail: human.ethics@utas.edu.au). If you have any queries regarding this research, please contact either Rapson Gomez (Tel. 6226 2887; E-mail: Rapson.Gomez@utas.edu.au), Peter Ball (Tel. 6226 7462; E-mail: P.Ball@utas.edu.au) or Susan Thomas (E-mail: susant@utas.edu.au).

A summary of the results and brief conclusions of the study will be made available towards the end of 2011 via the School of Psychology website (http://www.research.utas.edu.au).

Please keep this information letter for future reference. Thank you for considering being involved.

______________________	______________________	______________________
Rapson Gomez	Peter Ball	Susan Thomas

Appendix L – Statement of Consent

UNIVERSITY OF TASMANIA

STATEMENT OF INFORMED CONSENT

Evaluating the Construct Validity of Implicit Association Tasks using Confirmatory Factor Analysis Models

I have read and understood the "Information Sheet" for this study. Its nature has been explained and any questions that I have asked have been answered to my satisfaction. I understand that the study requires I complete some questionnaires and some computer administered tasks during an experimental session that should take about two hours, involving classifying pictures and words on a computer screen. I understand there is minimal risk for participants during the experiment.

I have been informed that all research data will be securely stored on University of Tasmania premises for a period of at least five years and will be destroyed after this time. I understand the data provided will be treated as strictly confidential and will be used only for the purposes of the research. I agree that research data gathered from me for this study may be published.

I agree to participate in this investigation, on the understanding that I may withdraw at any time during the testing process without being disadvantaged in any way, and that at that time I may request any data I had supplied be withdrawn from the research.

Name of participant: ____

Signature of participant: ______________________________ Date: __________

Statement by investigator:

I have explained this project and the implications of participation in it to this volunteer and I believe that consent is informed and that he/she understands the implications of participation.

Name of investigator: ____

Signature of investigator: ______________________________ Date: __________

Appendix M – Participant Debrief Script

Thankyou very much for your participation. I'm just going to take a few moments to explain how the Implicit Association Task (IAT) actually works.

The IAT is a measure used to assess attitudes without needing to ask you directly about your views. Remember the trials that required you to categorise four sets of pictures, say *Middle Eastern* and *European* countries as well as the *Pleasant* and *Unpleasant* words? Did you find one combination of those stimuli were easier to categorise than when they were swapped around the other way? Researchers believe the difficulty of this task is dependent on how closely associated are the two categories that share the one response. Associations between concepts can be strong or weak. For instance, many people strongly associate *Christmas* with *Presents*, but probably only weakly associate it with *Tennis*. Making decisions that involve strong associations tend to be quick and easy, whereas decisions involving weak associations tend to be slower and more difficult. So the more the concepts of *Europe* is linked with *Pleasant* and the *Middle East* with *Unpleasant*, the faster we will categorise these stimuli when they are in this combination. Conversely when *Europe* is categorised using the same key as *Unpleasant* and the *Middle East* with *Pleasant*, the task may become a tad more difficult.

By measuring your reaction time we can assess how quickly and easily each categorisation decision is made. By comparing the average reaction times for the two different arrangements of the keys we can get an indication of the strength of the associations between the two pairs of ideas (*Middle Eastern* versus *European* and

Pleasant versus *Unpleasant*). In other words, it gives us an idea of your attitudes towards Middle Eastern nations in comparison with European ones. Just to allay any fears, these scores do not provide a measure of racism. Besides, my research is actually examining how good the tasks are at measuring attitudes. So I am less interested in your scores per se, but more in how much overlap there is between everyone's scores over all the tasks. Also, remember that after you leave this room the data you have created today cannot be linked back to you in any way. Do you have any questions or concerns that have arisen from today's experiment?

Any participants with further concerns will be offered the University Psychology Clinic's phone number: 6226 2805, and are encouraged to gain free counselling there.