

A Practical Guide to Paper and Pencil Implicit Measures of Attitudes

Denise Sekaquaptewa

University of Michigan

Patrick Vargas

University of Illinois, Champaign-Urbana

William von Hippel

University of Queensland

In B. Gawronski & K. Payne (Eds.), *Handbook of Implicit Social Cognition*.

A Practical Guide to Paper and Pencil Implicit Measures of Attitudes

Social psychologists have long been interested in indirect or implicit measures of attitudes, for what they can tell us about the non-conscious components of attitudes and for their utility in controlling socially desirable responding and demand effects on measurement (Fazio & Olson, 2003). A variety of implicit measures have emerged and proven useful, many of which rely on computerized, reaction time (RT)-based assessments. In this chapter, we review measures that are indirect but are decidedly more "low-tech" than RT-based assessments, and describe their design, implementation, and scoring. Many of these low-tech measures are based on early theories and definitions of attitude, an important consideration given the strong influence of measurement on theorizing about the attitude construct (Ostrom, 1989).

How measurement technologies influence our concept of "attitude"

Modern definitions of attitudes are quite different from classic definitions. Consider the widely cited definition of attitude as, "a psychological tendency that is expressed by evaluating a particular entity with some degree of favor or disfavor," (Eagly & Chaiken, 1993, p. 1). The definition is relatively simple, limiting the attitude construct to positive and negative valence. When Thurstone (1928) declared that "attitudes can be measured" he was facing a much broader conceptualization of the attitude construct. He wrote, "It will be conceded at the outset that an attitude is a complex affair which cannot be measured by any single numerical index," (p. 530). Allport, for example, defined attitude as, "a mental and neural state of readiness, organized through experience, exerting a directive or dynamic influence upon the individual's response to all objects and situations with which it is related," (1935, p. 810; see also definitions by Krech & Crutchfield,

1948; Smith, Bruner, & White, 1956). When Thurstone focused on measuring the evaluative dimension of attitudes he (perhaps inadvertently) guided attitudes researchers toward a more restrictive conceptualization of attitudes.

More recently we have seen a similar influence of method on theory, with the development and increasing popularity of priming techniques (Fazio, Jackson, Dunton, & Williams, 1995) and the Implicit Association Test (IAT; Greenwald, McGee, & Schwartz, 1998) to implicitly measure attitudes. We wonder, though, whether limiting attitude measures to implicit and explicit categories, and focusing so much of the implicit attitude research on priming techniques and the IAT, might divert attention from other types of measures that are grounded in classic theorizing on attitudes but do not fall easily into the implicit/explicit dichotomy. Indeed, the implicit/explicit attitude distinction is preceded by a five-category taxonomy of attitude measures developed by Cook and Selltiz (1964). We believe this more complex classification of attitude measures might remain useful today. Just as implicit and explicit measures are thought to reflect the operation of associative and propositional processes, respectively, other classes of measures might reflect the operation of other types of cognitive processes.

Cook and Selltiz's five classes were: self-report, physiological, behavioral, partially structured, and performance on objective tasks. The first three categories are fairly consistent with contemporary conceptualizations. Self-report measures are presently known as explicit measures; examples include Thurstone, semantic differential, and Likert scales. Recent physiological measures include facial electromyographic activity (e.g., Cacioppo & Petty, 1979) and event related brain potentials (Cacioppo, Crites, Berntson, & Coles, 1993). Behavioral measures range from the lost letter technique (people are

more likely to mail “lost” letters addressed to organizations toward which they are favorably disposed; Milgram, Marm, & Harter, 1965) to the tendency for people to affiliate with similar others (Byrne, Ervin, & Lamberth, 1970). In contemporary attitude research behavior is generally considered an outcome, rather than a measure of attitudes (but see Albarracin & Wyer, 2000; Ouelette & Wood, 1998; Patterson, 2001). Cook and Selltiz’s last two categories (and an additional category that we introduce here) are of particular relevance for the present chapter.

Partially Structured Measures

Partially structured measures are those in which respondents are presented with ambiguous stimuli, and asked to interpret the stimuli in some manner. A classic example is the Thematic Apperception Test (TAT; Morgan & Murray, 1935; Murray, 1965). The TAT consists of a set of set of ambiguous images that are shown to a respondent whose task is to make up stories about the images. Trained coders review the stories and assign scores on dimensions of interest (e.g., need for power [Winter, John, Stewart, Klohnen, & Duncan, 1998]; attitudes toward union labor [Proshansky, 1943]). A variety of different partially structured measures have been developed, including some that do not rely on open-ended responses (e.g., Sokolowski, Schmalt, Langens, & Puca, 2000; Vargas, von Hippel, & Petty, 2004). Partially structured measures are implicit in that they do not require respondents to intentionally retrieve stored evaluative information about attitude objects or evaluate propositional information. Rather, partially structured measures simply allow respondents to describe, or react to, some ambiguous stimulus involving a particular attitude object. According to Cook and Selltiz (1964) these measures are based on the idea that, “. . . perception of stimuli that are not clearly structured is influenced by the

perceiver's own needs and dispositions... the expressed response corresponds directly to the individual's attitude," (pp. 47-48).

Performance on Objective Tasks

Attitude measures examining respondents' performance on objective tasks are presented as, "specific tasks to be performed... tests of information or ability, or simply as jobs that need to be done," (Cook & Selltiz, p. 50). Contemporary implicit measures relying on response times, such as the IAT and the evaluative priming task, fit into this category. Early performance measures include Hammond's information error test, in which participants were presented with a "quiz" featuring questions in which the response options were both false, but in opposite directions and equidistant to the true answer. For example, to assess attitudes toward U.S. President George W. Bush, this item might be used: "The percentage of U.S. citizens living in poverty by the end of G. W. Bush's first term was (a) 8.5% (b) 15.3%." Other performance measures examined respondents' proclivity to accept or reject logically flawed but emotionally charged arguments, or to accept attitude congruent arguments as most plausible (Saucier & Miller, 2003; Thistlethwaite, 1950; Waly & Cook, 1965). Like partially structured measures, performance measures are also implicit, in that they do not require respondents to intentionally indicate whether an attitude object is good or bad. Rather, as noted by Cook and Selltiz, these measures operate on the assumption that "performance may be influenced by attitude, and that a systematic bias in performance reflects the influence of the attitude," (p. 50).

Measures of Linguistic Biases

Although Cook and Selltiz's (1964) taxonomy of attitude measures rather neatly encompasses most contemporary attitude measures, an additional category that focuses on linguistic processing is also useful, given the growing body of work on the subtle and nuanced ways language choices reflect interpersonal processes and perceptions (Semin, 2008; von Hippel, Sekaquaptewa, & Vargas, 2009). Based on the assumption that different aspects of socially relevant information are encoded in natural language, researchers have turned to analyzing language to reveal how people interact with the world. Personality researchers have analyzed natural language to identify different dimensions of personality (John, Angleitner, & Ostendorf, 1988; Saucier & Goldberg, 1996), and – even more broadly – the measurement of meaning, itself (Osgood, Suci, & Tannenbaum, 1957; Pinker, 2007). Attitudes researchers have similarly analyzed natural language to identify markers of various types of beliefs and feelings (Pennebaker, Mehl, & Niederhoffer, 2003; Semin, 2008; von Hippel et al., 2009).

Our organization of paper and pencil measures into three general classes (partially structured, performance, and linguistic biases) is but one way to represent the current state of research using low-tech implicit measures, and other equally valid distinctions between the various measures could be made. A more fundamental question may be to ask why it is important to consider multiple classes of attitude measures. We believe that the answer to this question lies in the fact that attitudes are complex constructs, capable of influencing thoughts and behaviors in many ways. Most researchers are wary of problems with single operationalizations of complex constructs. Repeated use of a single operation may emphasize some elements irrelevant to the central concept, and under- and over-emphasize other elements of importance. As Rosenthal and Rosnow (2008) ex-

plained, single operations tend to “underdetermine” concepts (p. 46). The current focus on implicit versus explicit, with the accompanying emphasis on priming techniques and the IAT, has clearly generated a great deal of useful information and theorizing. Nevertheless, that emphasis may now be limiting attitude research as much as advancing it. The use of multiple attitude measures that tap different underlying cognitive processes should allow a more complete understanding of the attitude construct, improve behavioral prediction, and help generate novel hypotheses about how attitudes, cognition, and behavior influence one another. Additionally, as we have argued elsewhere (Vargas, Sekaquaptewa, & von Hippel, 2007), attitude measures tapping deliberative v. spontaneous cognitive processes might be differentially related to behaviors driven by deliberative v. spontaneous processes (cf. Ajzen & Fishbein, 1977; Wilson, Lindsey, & Schooler, 2000).

In the remainder of this chapter, we review various ways that paper and pencil measures of attitudes have been used in research as implicit indicators of attitudes, and describe how such measures are designed, implemented, and scored.

Paper and Pencil Measures: What they do

Research using Partially Structured Measures

Projective measures: The classic projective measure is the Thematic Apperception Test (Morgan & Murray, 1935). Although the TAT has suffered reliability problems, its predictive validity can be impressive. For example, Winter et al. (1998) obtained explicit measures of extraversion and projective measures of the need for affiliation. These measures were weakly correlated with each other, and interacted to predict (among other things) relationship dissatisfaction and low levels of intimacy, 30 years after the projective measures had been administered. Women who were extraverted but low in im-

implicit affiliation motive, and women who were introverted but high in implicit affiliation motive (i.e., those who showed an implicit-explicit conflict), suffered more dissatisfaction and experienced more low points in intimacy.

Projective measures have been used in a variety of other ways. In an early example, Proshansky (1943) presented students known to have pro- and anti-labor union attitudes with a series of images previously judged to be neither pro- nor anti-labor and asked the students to write about each picture. Judges coded the responses as either favorable or unfavorable toward labor. Proshansky's projective technique was highly correlated with a traditional measure of attitudes toward labor unions. Such implicit/explicit consistency is no longer a hallmark of projective measures, or other implicit measures for that matter (see Hofmann, Gschwendner, Nosek, & Schmitt, 2005), but projective techniques have also been used to explore the nature of implicit/explicit discrepancies. For example, using a measure similar to that of Proshansky (1943), Thrash, Elliot, and Schultheiss (2007) demonstrated that low self-monitors show greater implicit/explicit consistency on need for achievement than high self-monitors.

More recently, measures tapping biased information processing have revealed that people with opposing but strong pre-existing attitudes can become more convinced of their original position when presented with mixed evidence due to their selective rejection of attitude-incongruent information (Ditto, Munro, Apanovich, Scepanisky, & Lockhart, 2003; Ditto, Scepanisky, Munro, Apanovich, & Lockhart, 1998; Lord, Ross, & Lepper, 1979; Saucier & Miller, 2003). This sort of biased rejection of attitude-inconsistent information has not been used as a predictor variable in many studies, but Vargas et al. (2004) have shown that contrasting a relatively neutral attitude or behavior away from

one's own can be used to predict unique variance in behavior (beyond that predicted by the IAT or explicit measures).

Name letter preferences: Most attitudes regarding the self should be accessible via implicit measurement, but the brunt of the literature has focused on assessment of implicit self-esteem. The most widely used measure of implicit self-esteem is assessment of the name-letter effect, or the tendency to prefer letters that are in one's own initials (Nuttin, 1985). Interest in this measure stems not only from its ease of use, but also from its reliability, predictive validity, and robustness even in cultures in which people do not report high levels of explicit self-esteem (Kitayama & Karasawa, 1997; Koole, Dijksterhuis, & van Knippenberg, 2001). Name letter preferences are associated with having nurturing parents (DeHart, Pelham, & Tennen, 2006), manifest themselves in occupational, geographic, and romantic partner preferences (Jones, Pelham, Carvallo, & Mirenberg, 2004; Pelham, Mirenberg, & Jones, 2002), and are sensitive to threats to the self-concept (Jones, Pelham, Mirenberg, & Hetts, 2002), negative self-thinking (Verplanken, Friborg, Wang, Trafimow, & Woolf, 2007), and daily life events (DeHart & Pelham, 2007). Perhaps most impressively, name letter preferences are associated with self-sabotaging behavior among people whose name letters are associated with negative academic and sporting outcomes (Nelson & Simmons, 2007). Name letter effects also appear to be useful in discriminating defensive individuals from others (Bosson, Brown, Zeigler-Hill, & Swann, 2003), and different varieties of name letter effects have even been documented, as narcissists find their name letters attractive but not particularly likeable (Sakellaropoulos & Baldwin, 2007).

Name letter effects have the important advantage that they can be assessed via archival records (as in Pelham et al., 2002), and thus they open up a wide variety of sources that would otherwise be unavailable for the study of implicit attitudes. But name letter effects are only one example of how people can show differential levels of preference for items associated with the self. For example, people also show a preference for their birthday numbers (Pelham et al., 2002), and they are more likely to complete word stems with positive words after answering questions about the self (Hetts, Sakuma, & Pelham, 1999). Another implicit measure of self-esteem is based on signature size (e.g., Stapel & Blanton, 2004). Measures such as signature size are also amenable to historical analysis, and one can easily imagine studies that would be possible with such measures (e.g., correlating signature size of famous historical figures with other aspects of their behavior, such as the audacity of their exploits).

A final partially structured attitude measure that has shown some success in predicting behavior is a procedure developed by von Hippel, Lakin, and Shakarchi (2005), in which participants are given success and failure feedback on two novel tasks and asked which task is more important. Participants who were self-serving by rating the task at which they succeeded as more important than the task at which they failed also tended to be self-serving by cheating on a math task, but only when they could justify their cheating as unintentional.

Research using Measures of Linguistic Bias

Linguistic Intergroup Bias (LIB): LIB is the tendency to describe attitude-consistent events with more abstract language than is used to describe attitude-inconsistent events. LIB has been documented among many social groups including Ital-

ian horse-racing clubs (Maass, Salvi, Arcuri, & Semin, 1989), Japanese baseball fans (Tanabe & Oka, 2001), and nations (Maass, Montalcini, & Biciotti, 1998), and in many types of interpersonal behaviors, including doctor-patient interactions (Watson & Gallois, 2002) and personnel selection processes (Rubini & Menegatti, 2008). LIB has even been shown to emerge in minimal groups (Moscatelli, Albarello, & Rubini, 2008) and among previously unacquainted groups who cooperate vs. compete with each other (de Montes, Semin, & Valencia, 2003). To the extent that LIB supports favorable in-group perception and maintains out-group stereotypes (Maass, 1999), LIB may be considered an implicit marker of inter-group attitudes and stereotyping. Consistent with this idea, LIB has been shown to predict evaluations of outgroup members (von Hippel, Sekaquaptewa, & Vargas, 1997).

Although some research has shown LIB to be correlated with explicit prejudice (Schanke & Ruscher, 1998), other research has failed to find a relationship between the two (von Hippel et al., 1997). The relationship between explicit prejudice and LIB may be moderated by social desirability concerns, as explicit prejudice and LIB have been found to be more highly correlated for groups that are not protected by social norms of non-discrimination, such as Islamic fundamentalists, than for groups that are normatively protected, such as Jews (Franco & Maass, 1999; for an alternative interpretation, see Gawronski, Peters, Brochu, & Strack, 2008). One interpretation of this result is that social desirability norms influence self-reported attitudes but not the LIB, but it should be noted that individuals can inhibit the LIB if the situation draws their attention to it (Douglas, Sutton, & Wilkin, 2008).

Language abstraction has been used most extensively to study inter-group and inter-personal relations, but it has also been used to study other attitudinally relevant phenomena. For example, abstract language is associated with a promotion focus, while concrete language is more typical of a prevention focus (Semin, Higgins, de Montes, Estourget, & Valencia, 2005). This pattern is thought to occur because promotion focused individuals are inclusive and broad in their emphasis on a successful outcome, whereas prevention focused individuals are more concerned with concrete details that may go awry. Abstract language use also leads to recall of earlier information than concrete language, because older memories are stored in more abstract form than more recent memories (Semin & Smith, 1999). Finally, powerful people are more likely to use abstract language than powerless people (Smith & Trope, 2006), due to the increased psychological distance felt by powerful people (Trope & Liberman, 2003). As a consequence of these various influences on the use of abstract language, care must be taken in designing manipulations and measures of language abstraction to ensure that it is tapping the construct of interest.

Breadth-based adjectives: Adjective breadth refers to the number of behaviors that could be subsumed by a particular adjective; for example, *friendly* is a broader descriptor than *punctual* (Hampson, John, & Goldberg, 1987). As with the LIB, people describe attitude-consistent behaviors with broader adjectives than they use for attitude-inconsistent behavior. This effect has been shown with in-groups and out-groups (Hamilton, Gibbons, Stroessner, & Sherman, 1992), with liked and disliked politicians (McGraw, Fischle, Stenner, & Lodge, 1996), and with the self (Karpinski, Steinberg, Versek, & Alloy, 2007). For example, in Hamilton et al. (1992) participants rated liked

and disliked nationalities on a series of broad and narrow desirable and undesirable traits. Hamilton et al. found that participants believed that liked nationalities exhibited more of the broad desirable traits than the narrow desirable traits, whereas disliked nationalities exhibited more of the narrow than broad desirable traits. Although there is not yet any research demonstrating that this measure is particularly well suited for the implicit assessment of attitudes (with the possible exception of self-attitudes; Karpinski et al., 2007), it has the advantage that differential evaluations can be measured even though people are only providing positive evaluations of groups. Thus, it seems possible that the measure may be non-reactive, and suitable for implicit measurement when social desirability is an important concern.

Stereotypic explanatory bias (SEB): SEB is the tendency to spontaneously explain stereotype-incongruent than stereotype-congruent behavior. To date, research has primarily focused on SEB as an implicit linguistic marker of gender and racial stereotyping. SEB has been shown to correlate with measures of implicit stereotyping (lexical decision making; Sekaquaptewa et al., 2003, LIB; von Hippel et al., 1997) but not a measure of implicit prejudice (the race IAT; Sekaquaptewa et al., 2003), suggesting that SEB may have convergent validity as a measure of implicit stereotyping. SEB is also increased by factors known to augment stereotyping, such as mortality salience (Schimel, Simon, et al., 1999), positive mood (Chartrand, van Baaren, & Bargh, 2007), threats to collective self-esteem (Gonsalkorale, Carlisle, & von Hippel, 2007), and suggestions that stereotyping is normative (Gonsalkorale et al., 2007). SEB is also greater in response to low than high social status targets (Sekaquaptewa & Espinoza, 2004).

Perhaps most important in establishing SEB as an indicator of implicit stereotyping are findings supporting its utility to predict the nature of inter-group interactions. SEB has been shown to predict non-verbal behaviors in inter-group interactions with African-Americans (Sekaquaptewa et al., 2003), and male vs female job applicants (Vargas, Sekaquaptewa, & von Hippel, 2004). In both studies, explicit measures of bias (the Modern Racism Scale and Attitudes toward Women Scale) did not emerge as significant predictors of these outcomes.

The predictive validity of SEB also appears to differ depending on the type of explanation provided for stereotype-inconsistency on the SEB measure. Explanations for stereotype-relevant behaviors can be categorized as internal (attributed to the person) or external (attributed to the situation). By categorizing SEB into its internal and external forms, Sekaquaptewa et al. (2003) found that engaging in external SEB (attributing stereotype-inconsistency to external or situational forces) predicted having a more negative social interaction with an African-American confederate. Conversely, engaging in internal SEB (attributing stereotype-inconsistency to internal dispositions) predicted having a more positive social interaction with the African-American confederate. This pattern may have emerged because external SEB is associated with stereotype maintenance, as stereotype-inconsistency is attributed to less controllable and perhaps fleeting situational forces. Internal SEB, on the other hand, is associated with stereotype change, as stereotype-inconsistency is attributed to stable and enduring characteristics of the actor. Therefore, an important feature of SEB is that it appears to predict positive as well as negative social interaction outcomes, depending on the locus of explanation. This use of SEB is in need of further research.

Research using Performance on Objective Tasks

In an early study using a performance-based measure, Hammond (1948) presented the information error test to respondents as a simple multiple-choice test of their knowledge. Hammond devised two different types of questions. As noted above, the first type presented respondents with two response options that were equidistant and in opposite directions from the truth. The second type involved presenting response options that are indeterminate; a current example may be "President George W. Bush is known around the world for his (a) affable (b) bumbling demeanor." Hammond found that both types of questions enabled him to differentiate between known groups of responders.

A conceptually related measure examines respondents' proclivity to accept or reject logically flawed, but emotionally charged arguments toward a socially sensitive topic. In one such study, respondents from Northern and Southern colleges were shown a series of premises and conclusions about integration, and were asked to evaluate the validity of the arguments (Thistlethwaite, 1950). Prejudiced students tended to accept flawed arguments that were emotionally charged and attitude-congruent, but reject more neutral, attitude-congruent arguments.

Paper and pencil IAT: The IAT is designed for computer-based administration, but paper-and-pencil versions have been developed for situations when computer administration is not feasible. One such paper-and-pencil IAT features on each page a column of stimuli from the target and attribute categories (e.g., a flower/insect IAT might include daffodil, rainbow, cockroach, and vomit). Respondents categorize the stimuli by marking circles on either side of the column. The critical dependent measure is the difference in the number of stimuli correctly categorized in a fixed amount of time (30 seconds in

Mast, 2004; 20 seconds in Lemm, Lane, Sattler, Khan, & Nosek, in press) between evaluatively congruent and incongruent blocks.

The Lemm et al. (in press) research has the advantage that it compared different versions of the IAT. They administered two different versions of the IAT on computer and via paper and pencil. One version featured Black and White names as stimuli, and another featured cropped pictures of Black and White faces. Lemm et al. (in press) found that the paper and pencil name version of the IAT produced a larger effect size, better test-retest reliability, and higher correlations with a computerized IAT than did the paper and pencil picture version of the IAT.

Word Fragment Completion: Perhaps the first paper and pencil implicit measure to be used in research on stereotyping and prejudice was word fragment completion, in a study that demonstrated that stereotype activation was not inevitable upon presentation of a stereotyped group member (Gilbert & Hixon, 1991). Since then, word fragment completion tasks have been widely used as implicit evidence that a stereotype has been activated (e.g., Sinclair & Kunda, 1999; Spencer, Fein, Wolfe, Fong, & Dunn, 1998; Steele & Aronson, 1995). But word fragment completion can also be used as a predictor variable. In one such study (Son Hing, Li, & Zanna, 2002), word fragment completion was used in conjunction with measures of explicit racism to identify aversive racists (individuals who endorse egalitarianism but nevertheless retain unconscious negative feelings toward racial outgroups; Gaertner & Dovidio, 1986). Participants who scored low on an explicit measure of prejudice (the Asian Modern Racism Scale) but high on the implicit word fragment completion measure were identified as aversive racists, and their responses were compared to individuals who scored low on both measures. Aversive racists felt

more guilty and showed less discrimination towards Asians after being reminded of previous racist transgressions against Asians (consistent with aversive racism and prejudice reduction theories; Monteith, 1996), whereas these outcomes did not emerge among individuals who scored low on both measures. Recently Son Hing and colleagues have extended this finding using the IAT as the implicit prejudice measure, suggesting that word fragment completion measures might be interchangeable with the IAT under some circumstances (Son Hing, Chung-Yan, Hamilton, & Zanna, 2008).

Memory Associations: Recent research in health psychology has revived a measure of memory associations that is similar to Freud's use of free associations to reveal people's unconscious attitudes and motivations (Ames, Franken, & Coronges, 2006; Stacy & Wiers, 2006). In an early example of this work, people were asked to provide the first word that comes to mind when they read words ambiguously related to alcohol or marijuana (Stacy, 1997). Greater frequency of responding with alcohol and marijuana-related words in this context was associated with greater subsequent alcohol and marijuana use (Kelly, Masterman, & Marlatt, 2005; Stacy, 1997). Although this measure has not been used in social psychological research, and indeed may possibly be reactive under some circumstances, its predictive ability in health psychology suggests that it would be a useful measure for attitudes researchers to explore.

How-To: Developing and Using Paper and Pencil Measures

Designing, implementing, and scoring paper and pencil measures of implicit attitudes involves many of the same issues that are important in using any type of attitude assessment; the measure should reliably assess the intended construct. Yet there are also

important differences between the three classes of paper and pencil measures and other measures.

Similar to all self-report psychological measures, responses on various types of paper and pencil measures are subject to biases regarding question wording, context, and response alternatives (Schwarz, 1999). For example, research on survey methodology has revealed that higher means are observed in data collected using rating scales ranging from -5 to 5 than when the rating scale used ranges from 1 to 10, due to people's avoidance of endorsing negative numbered response options (Schwarz, Knaeuper, Hippler, Noelle-Neumann, & Clark, 1991). To the extent that a paper and pencil measure involves properties of questionnaires known to be subject to context and wording effects (such as the use of rating scales in LIB measures), the same care should be taken to reduce the influence of such effects as when developing any self-report scale (see Sudman, Bradburn, & Schwarz, 1996).

Unlike most RT measures of attitudes, paper and pencil measures that involve simple multiple-choice responses or endorsements have the advantage of being easy to administer, because no special equipment or computer is needed, and are easy to score given their closed choice response format. However, scoring of paper and pencil measures can involve significantly more effort when open-ended responses are involved. Although this may seem at first a disadvantage of using open-ended measures, there are positives. For example, coding schemes can uncover responses or patterns in the data that may not have been evident when respondents are given distinct answer choices to choose from (Schuman & Presser, 1981). Overall, open-ended responses may provide opportunity for rich and detailed interpretations of ambiguous information for some paper

and pencil measures (e.g., partially structured measures and the TAT), but are not critical for use in all paper and pencil measures (e.g., information error test, LIB).

Developing and Using Partially Structured Measures

TAT: Many partially structured measures have the advantage of being relatively easy to administer. The main goal is to present respondents with a stimulus that is ambiguous in regard to the attitude object of interest, and let the respondents' underlying attitudes guide their interpretation of it. Several guides and coding manuals have been published on using and scoring the TAT (Jenkins, 2008; Winter, 1991). Among the issues to consider when using this technique are cultural differences indicating that certain symbols and images have clear meanings in some cultures that do not transfer across cultures (Hofer & Chasiotis, 2004). When re-administrating the TAT, test-retest reliability can be improved by instructing respondents that they should not be concerned about whether their story is the same or different from what they generated initially (Lundy, 1985).

Name letter preference: Many demonstrations of the name letter effect use archival analysis of existing public records (e.g., Pelham et al., 2002). Laboratory studies have examined name letter preferences by having respondents rate all the letters in the alphabet, or rate attitude objects of which some begin with one of the respondent's initials (e.g., Hodson & Olson, 2005). Name letter preference is indicated by comparing ratings of letters in one's name to non-name letters, or by assessing preferences for attitude objects that begin with the letters in one's initials relative to ones that do not. Name letter preferences can be examined between individuals (Does Sara like Saran Wrap more than Bob does?) or within an individual (Does Sara like Saran Wrap more than foil?).

The within-individual comparison provides an index of the extent to which an individual shows a name letter preference, which can serve as an indirect indicator of self-esteem. Ratings of objects that begin and do not begin with one of a respondent's initials, or of letters that are contained or not contained in the respondents name, are compared. A difference score is computed between ratings of the letters that appear in the attitude object or in one's name and a baseline evaluation of non-self-related letters or objects to yield a relative name preference score. Positive scores reflect the name letter effect: greater preference for objects containing one's initials than those that do not, or more positive ratings of the letters in one's own name vs. letters not appearing in one's name.

Other partially structured measures: The vignette measures used in Vargas et al. (2004) described different target characters engaged in ambiguously conflicting behaviors. For example, an item used to measure religiosity was as follows: "June didn't go to church once the whole time she was in college, but she claimed that she was still a very religious person. She said that she prayed occasionally, and that she believed in Christian ideals. Sometimes she watched religious programs on TV like the 700 Club, or the Billy Graham Crusade." Following each vignette participants were asked to rate, using 11-point scales anchored by "not at all religious" and "extremely religious," how religious they thought the main character's behaviors were, and how religious the main character was, in general. Responses to the measures are averaged to create a partially structured attitude measure score. Other sets of vignettes were designed to measure attitudes toward dishonesty and political orientations. The vignette measures were created using a typical scale development procedure (e.g., Churchill, 1979), so other sets of vignettes assessing attitudes toward a variety of different attitude objects may be developed relatively easily.

Developing and Using Measures of Linguistic Bias

LIB: Initial demonstrations of LIB used cartoon illustrations to show ingroup and outgroup members performing positive and negative behaviors. Cartoon illustrations were used because they are language-free (thus avoiding the potential of abstractions in the language of the stimulus material influencing encoding; Maass et al., 1989), although later demonstrations of LIB successfully used written passages (e.g., von Hippel et al., 1997) and videotaped footage (Gorham, 2006) to convey ingroup and outgroup behaviors. In Maass et al. (1989), participants were shown the illustrations and asked to select from among four descriptions the one they thought best described the scene.

LIB response choices are based on Semin and Fiedler's (1988) Linguistic Category Model. This model presents four linguistic categories with which individuals can describe other people. At the most concrete level are descriptive action verbs that refer to single, observable events bound to a particular physical action (such as "to hit"). Next are interpretive action verbs, which are single events but not tied to specific actions (such as "to hurt someone"). The third category, state verbs, includes psychological states without reference to a particular action but which are directed at a specific object (such as "to detest"). Finally, at the most abstract level of description is the fourth category, adjectives, which describe dispositions without reference to particular situations, targets, or behaviors (such as "violent"). LIB response alternatives were structured according to this model, and respondent's choice of more abstract descriptions for positive ingroup, and negative outgroup behaviors was evidence of LIB.

LIB studies using cartoon drawings to show ingroup/outgroup behaviors have the advantage of strong control over extraneous, background factors that may influence re-

spondents. Additionally, it is an easy matter to convey group membership by changing simple visual cues (such as shirt colors associated with particular *contrada*; Maass et al., 1989). LIB measures using written passages may also easily convey group membership using simple cues, such as conveying the gender of the protagonist by using gender-suggestive names, or by attaching a photograph said to be of the protagonist (e.g., von Hippel et al., 1997). When photographs are used to convey group membership (such as a Black man v. a White man), it is wise to use more than one picture of a target category (such as Black male) because idiosyncratic facial characteristics can influence social judgments (Berry & Wero, 1993). Photographs can also be pretested to ensure similarity in extraneous facial characteristics using Rhodes's (1988) facial rating scales (as in Gorham, 2006).

In creating LIB scores, different procedures have been used successfully. Maass and her colleagues (1989) simply asked respondents to choose the best description from the four alternatives, and compared the level of abstraction chosen for ingroup positive/outgroup negative behaviors to ingroup negative/outgroup positive behaviors. Von Hippel and colleagues (1997) used a subtraction method, in which participants were asked to rate each of the four descriptions on how well it described the passage; the endorsement of the most abstract description of a stereotype-inconsistent behavior was subtracted from the endorsement of the most abstract description of stereotype-consistent behavior. Positive difference scores indicated LIB.

BART: The breadth-based adjective rating task (BART; Karpinski et al., 2007) was introduced as an indirect measure of self-esteem. A list of 144 adjectives is presented, shown in previous research to vary in terms of valence and breadth (Hampson et al.,

1987), and selected such that broad and narrow traits are equated on valence (see Karpinski et al., 2007, for the complete list of BART adjectives). Respondents rate how well each adjective describes the attitude object on a scale from 1 (does not describe at all) to 9 (describes very well). To compute a BART score, ratings within each of the four BART categories (broad positive, narrow positive, broad negative, narrow negative) are averaged. The averages from the broad negative and narrow positive are summed, and this sum is subtracted from the sum of the broad positive and narrow negative categories. Higher final scores indicate more positive attitudes toward the self, person, or group.

One advantage of the BART is that potentially useful alternative scores can easily be computed, such as a total trait valence score (positive trait average minus negative trait average) and a total breadth score (broad trait ratings minus narrow trait ratings). One disadvantage of the BART is that respondents may find it tedious to rate 144 adjectives; however, meaningful results have been found using only half of the entire list of traits (retaining the traits that are at the extremes of valence and breadth; Karpinski et al., 2007, Experiment 2).

SEB: A measure of SEB typically consists of a series of sentence beginnings representing group members performing behaviors consistent or inconsistent with their group stereotype. SEB measures involve the comparison of responses to targets from two social groups; to illustrate, we use the example of male and female to assess gender SEB. Four types of SEB items are required, reflecting the crossing of target gender and stereotype-consistency: male targets performing male-stereotypic behaviors; male targets performing female-stereotypic behaviors; female targets performing female-stereotypic behaviors; female targets performing male-stereotypic behaviors (see Sekaquaptewa and

Espinoza (2004) for a useful example of how to pre-test SEB items). Previous SEB studies have used three items of each type, plus filler items designed to be stereotype-neutral. Respondents are asked to "complete the sentence" with the first response that comes to mind, using any endings they wish as long as it forms a grammatical sentence.

In computing an SEB score, it is important to use pre-determined criteria to classify SEB responses. Responses should be coded by independent raters as to whether they explain the behavior presented in the sentence beginning (e.g., Susan went to the auto show... because her dad wanted her to), or simply continue the sentence without providing an explanation (e.g., Susan went to the auto show... last weekend). A general SEB score can be computed by subtracting the number of explanations provided for stereotype-consistent SEB items from the number provided for stereotype-inconsistent items. Positive difference scores indicate SEB: greater explanation of stereotype-inconsistency over stereotype-consistency.

SEB sub-scores can also be computed according to whether the explanations reflect internal or external attributions. For example, the response "Susan went to the auto show... because her dad wanted her to" would be scored as an external explanation because her behavior was explained by a force external to her, whereas the response "because she loved cars" would be scored as an internal explanation because it places the motivation for Susan's behavior within her own disposition or preferences. Because of the implications of locus of explanation for stereotype maintenance, using these internal and external SEB subscales may provide an indicator of SEB that is more sensitive in predicting different behavioral outcomes (Sekaquaptewa et al., 2003).

Developing and Using Measures of Performance on Objective Tasks

Paper and pencil IAT: Recent reviews and methodological discussions of the IAT tend to raise issues that are probably relevant to both the computerized and paper and pencil versions of the IAT. Such issues include the categories and category labels chosen (e.g., whether to contrast science with humanities, arts, or some other non-science discipline; use of the label homosexual or gay), specific stimuli/category exemplars (for the category "USA," a picture of Vice President Dick Cheney vs. a picture of an American flag), number of trials (40 being recommended for the critical trials), and order of presentation of trials within an IAT (Greenwald et al., 1998) as well as order of IAT and other (explicit) measures (Nosek, Greenwald, & Banaji, 2005).

Information error test: The information error test requires effort in item development but is easy to score. Items should present informational questions about an attitude object with response choices that are not the true answer to the question, but that are equally "wrong," i.e., equidistant from the truth. As put by Hammond, "In constructing an item the principle requirement was to eliminate reality, the truth of the matter, as a factor and thus *force the respondent into a choice of errors*," (1948, p. 40, emphasis in original). The direction of the error reveals the respondent's underlying attitude. The original questionnaire developed by Hammond used only two response alternatives, and assigned a point for choosing the pro-attitude object response alternative and no point for choosing the anti-attitude object alternative, to produce an index in which positive scores reflect positive attitudes. However, Hammond did note that future tests could be "constructed with four error-choices to provide for 'intensity' of error." (p. 43). Thus, conceivably an information error test could be devised with response options that vary in their distance

from the truth (from being slightly off to being substantially distant from the true answer), providing a greater range of scores.

Word Fragment Completion: Generally one creates a list of target words and fashions word fragments by leaving out at least two specific letters or word stems by leaving out word endings. For example, to assess death accessibility, Schimel et al. (2007) developed target words such as "coffin" and created the word fragment "coff __" (in this case, a word stem "coff__" could also have been created). Typically word fragments or stems are designed so they can be completed with alternative completions or endings (such as "coffee" in this example). Pretesting can confirm that word fragments can be completed with non-target as well as target words (e.g., Steele & Aronson, 1995). In the implicit memory literature, however, word fragments have been used in which the fragments were designed to have only one solution, and memory was demonstrated by an increased likelihood of arriving at a solution (e.g., Tulving, Schacter, & Stark, 1982). Such a procedure could theoretically be used with word fragment or stem completion that is intended to measure attitudes or construct accessibility.

Researchers may use word frequency tables (e.g., Dahl, 1979) to compare frequency of target words to alternatives. "Coffee" may be more often used in the English language than "coffin," for example. In addition, published work on word usage may be used to develop filler items that can be completed with equally frequent non-target concept words. And norms even exist for different stem completions (Graf & Williams, 1987).

In coding word fragments, one may use a stringent criterion of assigning one point for each item completed with the target word. Alternatively, one could use more

lax criteria by assigning points for word completions that are close in meaning to the target concept, even for a filler item (e.g., completing the intended filler item "_ r _ _ e" as "grave."

Conclusion

The development and use of paper and pencil implicit measures of attitudes, in conjunction with recently developed computerized assessments, provides an interesting example of the evolution of our concept of attitude as influenced by available technologies. Early definitions noted that defining attitudes was a "complex affair" (Allport, 1935), and early assessment techniques typically fell short of capturing this complexity. Advancements in technology led to adjustments in theory, as the use of specialized equipment and computer technologies allowed micro-assessment of responses to stimuli. Thus, today's conceptualizations of attitude represent the confluence of our available methodologies, distinguishing, for example, between associative (as measured by RT assessments) and propositional processes (as measured using self-reports; Gawronski & Bodenhausen, 2007).

Looking forward, one may predict that new technological advances (e.g., virtual reality technologies) will remain a continual influence on our evolving concept of attitude. However, it will be important to continue to investigate the utility of paper and pencil measures of attitudes, which have emerged not only as useful predictors of important outcomes, but also show promise in capturing components of the attitude construct that are untapped by measures of associative processing (Vargas et al., 2007). Although the popularity of various technologies may change over time, paper and pencil measures may continue to contribute to our understanding of attitudes by virtue of their potential to

measure aspects of attitudes not reflected in explicit self-report, physiological, or reaction-time based assessments.

References

- Ajzen, I., & Fishbein, M. (1977). Attitude-behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin*, *84*, 888-918.
- Albarracin, D., & Wyer, R.S., Jr. (2000). The cognitive impact of past behavior: Influences of beliefs, attitudes, and future behavioral decisions. *Journal of Personality and Social Psychology*, *79*, 5-22.
- Allport, G. W. (1935). Attitudes. In C. Murchison (Ed.), *Handbook of social psychology* (pp. 133-175). Worcester, MA: Clark University Press.
- Ames, S.L., Franken, I.H.A., & Coronges, K. (2006). Implicit cognition and drugs of abuse. In R.W. Wiers and A.W. Stacy (ed.), *Handbook of implicit cognition and addiction*. Thousand Oaks, CA: Sage, pp. 363-378.
- Berry, D. S. & Wero, J. L. F. (1993). Accuracy in face perception: A view from ecological psychology. *Journal of Personality*, *61*, 497-520.
- Bosson, J. K., Brown, R. P., Zeigler-Hill, V., & Swann, W. B., Jr. (2003). Self-enhancement tendencies among people with high explicit self-esteem: The moderating role of implicit self-esteem. *Self and Identity*, *2*, 169-187.
- Byrne, D., Ervin, C. R., & Lamberth, J. (1970). Continuity between the experimental study of attraction and real-life computer dating. *Journal of Personality and Social Psychology*, *16*, 157-165.
- Cacioppo, J. T., Crites, S. L., Berntson, G. G., & Coles, M. G. (1993). If attitudes affect how stimuli are processed, should they not affect the event-related brain potential? *Psychological Science*, *4*, 108-112.

- Cacioppo, J. T., & Petty, R. E. (1979). Attitudes and cognitive response: An electro-physiological approach. *Journal of Personality and Social Psychology*, 37, 2181-2199.
- Chartrand, T. L., van Baaren, R. B., & Bargh, J. A. (2007). Linking automatic evaluation to mood and information processing style: Consequences for experienced affect, impression formation, and stereotyping. *Journal of Experimental Psychology: General*.
- Churchill, G. A. (1979). A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research*, 16, 64-73.
- Cook, S. W., & Selltitz, C. (1964). A multiple-indicator approach to attitude measurement. *Psychological Bulletin*, 62, 36-55.
- Dahl, H. (1979). *Word frequency of spoken American English*. Essex, CT: Verbatim.
- DeHart, T. & Pelham, B. W. (2007). Fluctuations in state implicit self-esteem in response to daily negative events. *Journal of Experimental Social Psychology*, 43, 157-165.
- DeHart, T., Pelham, B. W., & Tennen, H. (2006). What lies beneath: Parenting style and implicit self-esteem. *Journal of Experimental Social Psychology*, 42, 1-17.
- de Montes, L. G., Semin, G. R., & Valencia, J. F. (2003). Communication patterns in interdependent relationships. *Journal of Language & Social Psychology*, 22, 259-281.
- Ditto, P. H., Munro, G. D., Apanovich, A. M., Scepansky, J. A., & Lockhart, L. K. (2003). Spontaneous skepticism: The interplay of motivation and expectation in responses to favorable and unfavorable medical diagnoses. *Personality and Social Psychology Bulletin*, 29, 1120-1132.
- Ditto, P. H., Scepansky, J. A., Munro, G. D., Apanovitch, A. M., & Lockhart, L. K. (1998). Motivated sensitivity to preference-inconsistent information. *Journal of Personality and Social Psychology*, 75, 53-69.

Douglas, K. M., Sutton, R. M., & Wilkin, K. (2008). Could you mind your language? An investigation of communicators' ability to inhibit linguistic bias. *Journal of Language and Social Psychology, 27*, 123-139.

Eagly A.E., & Chaiken S. (1993). *The psychology of attitudes*. Fort Worth, TX: Harcourt Brace Jovanovich.

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

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

Franco, F. M., & Maass, A. (1999). Intentional control over prejudice: When the choice of the measure matters. *European Journal of Social Psychology, 29*, 469-477.

Gaertner, S. L., & Dovidio, J. F. (1986). The aversive form of racism. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 61-89). San Diego: Academic Press.

Gawronski, B., & Bodenhausen, G. V. (2007). Unraveling the processes underlying evaluation: Attitudes from the perspective of the APE model. *Social Cognition, 25*, 687-717.

Gawronski, B., Peters, K. R., Brochu, P. M., Strack, F. (2008). Understanding the relations between different forms of racial prejudice: A cognitive consistency perspective. *Personality and Social Psychology Bulletin, 34*, 648-665.

Gilbert, D. T., & Hixon, J. G. (1991). The trouble of thinking: Activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology, 60*, 509-517.

Gonsalkorale, K., Carlisle, C., & von Hippel, W. (2007). Intergroup threat increases implicit stereotyping. *International Journal of Psychology and Psychological Therapy*, *7*, 189-200.

Gorham, B. W. (2006). News media's relationship with stereotyping: The Linguistic Intergroup Bias in response to crime news. *Journal of Communication*, *56*, 289-308.

Graf, P., & Williams, D. (1987). Completion norms for 40 three-letter word stems. *Behavioral Research Methods, Instruments, & Computers*, *19*, 422-445.

Greenwald, A. G., McGhee, D., & Schwartz, J.L.K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, *74*, 1364-1480.

Hamilton, D. L., Gibbons, P. A., Stroessner, S. J., & Sherman, J. W. (1992). Stereotypes and language use. In G. R. Semin & K. Fiedler (Eds), *Language, interaction and social cognition*. (pp. 102-128). Thousand Oaks, CA: Sage Publications.

Hammond, K. R. (1948). Measuring attitudes by error choice: An indirect method. *Journal of Abnormal and Social Psychology*, *43*, 38-48.

Hampson, S. E., John, O. P., & Goldberg, L. R. (1987). Category-breadth and social desirability values for 573 personality terms. *European Journal of Personality*, *1*, 37-54.

Hetts, J. J., Sakuma, M., & Pelham, B. W. (1999). Two-roads to positive regard: Implicit and explicit self-evaluation and culture. *Journal of Experimental Social Psychology*, *35*, 512-559.

Hodson, G., & Olson, J. M. (2005). Testing the Generality of the Name Letter Effect: Name Initials and Everyday Attitudes. *Personality and Social Psychology Bulletin*, *31*, 1099-1111.

Hofer, J., & Chasiotis, A. (2004). Methodological considerations of applying a TAT-type picture-story test in cross-cultural research. *Journal of Cross-Cultural Psychology*, *35*, 224-241.

Hofmann, W., Gschwendner, T., Nosek, B. A., & Schmitt, M. (2005). What moderates implicit-explicit consistency? *European Review of Social Psychology*, *16*, 335-390.

Jenkins, S. R. (2008). *A handbook of clinical scoring systems for thematic apperceptive techniques*. Mahwah, NJ: Erlbaum.

John, O. P., Angleitner, A., & Ostendorf, F. (1988). The lexical approach to personality: A historical review of trait taxonomic research. *European Journal of Personality*, *2*, 171-203.

Jones, J. T., Pelham, B. W., Carvallo, M., & Mirenberg, M. C. (2004). How do I love thee? Let me count the Js: Implicit egotism and interpersonal attraction. *Journal of Personality and Social Psychology*, *87*, 665-683.

Jones, J. T., Pelham, B. W., Mirenberg, M. C., & Hetts, J. J. (2002). Name letter preferences are not merely mere exposure: Implicit egotism as self-regulation. *Journal of Experimental Social Psychology*, *38*, 170-177.

Karpinski, A., Steinberg, J. A., Versek, B., & Alloy, L. B. (2007). The breadth-based adjective rating task (BART) as an indirect measure of self-esteem. *Social Cognition*, *25*, 778-818.

Kelly, A.B., Masterman, P.W., & Marlatt, G.A. (2005). Alcohol related associative strength and drinking behaviours: Concurrent and prospective relationships. *Drug and Alcohol Review*, *24*, 489-498.

Kitayama, S., & Karasawa, M. (1997). Implicit self-esteem in Japan: Name letters and birthday numbers. *Personality and Social Psychology Bulletin*, *23*, 736-742.

Koole, S. L., Dijksterhuis, A., & van Knippenberg, A. (2001). What's in a name: Implicit self-esteem and the automatic self. *Journal of Personality and Social Psychology*, *80*, 669-685.

Krech, D., & Crutchfield, R. (1948). *Theory & problems of social psychology*. New York: McGraw-Hill.

Lemm, K. M., Lane, K. A., Sattler, D. N., Khan, S. R., & Nosek, B. A. (in press.) Assessing Implicit Cognitions with a Paper-Format Implicit Association Test. In T.G. Morrison & M. A. Morrison (Eds.). *The psychology of modern prejudice*. Hauppauge, NY: Nova Science Publishers.

Lord, C. G., Ross, L., & Lepper, M. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, *37*, 2098-2109.

Lundy, A. (1985). The reliability of the Thematic Apperception Test. *Journal of Personality Assessment*, *49*, 141-145.

Maass, A. (1999). Linguistic intergroup bias: Stereotype perpetuation through language. *Advances in Experimental Social Psychology*, *31*, 79-121.

Maass, A., Montalcini, F., & Biciotti, E. (1998). On the (dis-)confirmability of stereotypic attributes. *European Journal of Social Psychology*, *28*, 383-402.

Maass, A., Salvi, D., Arcuri, L., & Semin, G. (1989). Language use in intergroup contexts: The linguistic intergroup bias. *Journal of Personality and Social Psychology*, *57*, 981-993.

- Mast, M.S. (2004). Men are hierarchical, women are egalitarian: An implicit gender stereotype. *Swiss Journal of Psychology, 63*, 107-111.
- McGraw, K. M., Fischle, M., Stenner, K., & Lodge, M. (1996). What's in a word? Bias in trait descriptions of political leaders. *Political Behavior, 18*, 263-287.
- Milgram, S., Marm, L., & Harter, S. (1965). The lost-letter technique: A tool of social research. *Public Opinion Quarterly, 29*, 437-8.
- Monteith, M. J. (1996). Affective reactions to prejudice-related discrepant responses: The impact of standard salience. *Personality and Social Psychology Bulletin, 22*, 48-59.
- Morgan, C.D., & Murray, H.A. (1935). A method for investigating fantasies. *Archives of Neurology and Psychiatry, 34*, 289-304.
- Moscatelli, S., Albarelo, F., & Rubini, M. (2008). Linguistic discrimination in minimal groups: The impact of status differentials. *Journal of Language and Social Psychology, 27*, 140-154.
- Murray, H.A. (1965). Uses of the thematic apperception test. In B.I. Murstein (Ed.) *Handbook of Projective Techniques*. New York: Basic Books, Inc.
- Nelson, L. D., & Simmons, J. P. (2007). Moniker maladies: When names sabotage success. *Psychological Science, 18*, 1106-1112.
- Nosek, B. A., Greenwald, A. G. & Banaji, M. R. (2005). Understanding and Using the Implicit Association Test: II. Method Variables and Construct Validity. *Personality and Social Psychology Bulletin, 31*, 166-180.
- Nuttin, J. M. (1985). Narcissism beyond Gestalt and awareness: The name letter effect. *European Journal of Social Psychology, 15*, 353-361.
- Osgood, C.E., Suci, G.J., & Tannenbaum, P.H. (1957). *The measurement of meaning*.

Urbana, IL: University of Illinois Press.

Ostrom, T. M. (1989). The interdependence of attitude theory and measurement. In A.R. Pratkanis, S.J. Breckler, & A.G. Greenwald (Eds.), *Attitude structure and function* (pp. 11-36). Hillsdale, NJ: Erlbaum.

Ouellette, J.A., & Wood, W. (1998). Habit and intention in everyday life: The processes by which past behavior predicts future behavior. *Psychological Bulletin*, *124*, 54-74.

Patterson, M.L. (2001). Toward a comprehensive model of non-verbal communication. In W.P. Robinson & H. Giles (Eds.) *The new handbook of language and social psychology* (pp. 159-176). Hoboken, NJ: John Wiley & Sons Ltd.

Pelham, B. W., Mirenberg, M. C., & Jones, J. T. (2002). Why Susie sells seashells by the seashore: Implicit egotism and major life decisions. *Journal of Personality and Social Psychology*, *82*, 469-487.

Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology*, *54*, 547-577.

Pinker, S. (2007). Language as an adaptation by natural selection. *Acta Psychologica Sinica*, *39*, 431-438.

Proshansky, H. M. (1943). A projective method for the study of attitudes. *Journal of Abnormal and Social Psychology*, *38*, 393-395.

Rhodes, G. (1988). Looking at faces: First-order and second-order features as determinants of facial appearance. *Perception*, *17*, 43-63.

Rosenthal, R., & Rosnow, R. L. (2008). *Essentials of behavioral research: Methods*

and data analysis (3rd ed.). McGraw-Hill.

Rubini, M., & Menegatti, M. (2008). Linguistic bias in personnel selection. *Journal of Language and Social Psychology, 27*, 168-181.

Sakellaropoulo, M. & Baldwin, M. W. (2007). The hidden sides of self-esteem: Two dimensions of implicit self-esteem and their relation to narcissistic reactions. *Journal of Experimental Social Psychology, 43*, 995-1001.

Saucier, G., & Goldberg, L. R. (1996). Evidence for the Big Five in analyses of familiar English personality adjectives. *European Journal of Personality, 10*, 61-77.

Saucier, D. A., & Miller, C. T. (2003). The Persuasiveness of Racial Arguments as a Subtle Measure of Racism. *Personality & Social Psychology Bulletin, 29*, 1303-1315.

Schanke, S. B., & Ruscher, J. B. (1998). Modern racism as a predictor of the linguistic intergroup bias. *Journal of Language and Social Psychology, 17*, 484-491.

Schimmel, J., Simon, L., Greenberg, J., Pyszczynski, T., Solomon, S., Waxmonsky, J., & Arndt, J. (1999). Stereotypes and terror management: Evidence that mortality salience enhances stereotypic thinking and preferences. *Journal of Personality and Social Psychology, 77*, 905-926.

Schuman, H., & Presser, S. (1981). Questions and answers in attitude surveys. New York: Academic Press.

Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist, 54*, 93-105.

Schwarz, N., Knäuper, B., Hippler, H-J., Noelle-Neumann, E., & Clark, L. (1991). Rating scales: Numeric values may change the meaning of scale labels. *Public Opinion Quarterly, 55*, 570-582.

Sekaquaptewa, D., & Espinoza, P. (2004). Biased processing of stereotype-incongruity is greater for low than high status groups. *Journal of Experimental Social Psychology, 40*, 128-135.

Sekaquaptewa, D., Espinoza, P., Thompson, M., Vargas, P., & von Hippel, W. (2003). Stereotypic explanatory bias: Implicit stereotyping as a predictor of discrimination. *Journal of Experimental Social Psychology, 39*, 75-82.

Semin, G. R. (2008). Language puzzles: A prospective retrospective on the linguistic category model. *Journal of Language and Social Psychology, 27*, 197-209.

Semin, G. R., & Fiedler, K. (1988). The cognitive functions of linguistic categories in describing persons: Social cognition and language. *Journal of Personality and Social Psychology, 54*, 558-568.

Semin, G. R., Higgins, E. T., de Montes, L. G., Estourget, Y., & Valencia, J. F. (2005). Linguistic Signatures of Regulatory Focus: How Abstraction Fits Promotion More Than Prevention. *Journal of Personality and Social Psychology, 89*, 36-45.

Semin, G. R., & Smith, E. R. (1999). Revisiting the past and back to the future: Memory systems and the linguistic representation of social events. *Journal of Personality and Social Psychology, 76*, 877-892.

Sinclair, L., & Kunda, Z. (1999). Reactions to a Black professional: Motivated inhibition and activation of conflicting stereotypes. *Journal of Personality and Social Psychology, 77*, 885-904.

Smith, M.B., Bruner, J.S., & White, R.W. (1956). *Opinions and personality*. New York: Wiley.

Smith, P. K., & Trope, Y. (2006). You focus on the forest when you're in charge of the trees: Power priming and abstract information processing. *Journal of Personality and Social Psychology, 90*, 578-596.

Sokolowski, K., Schmalt, H., Langens, T. A., & Puca, R. M. (2000). Assessing achievement, affiliation, and power motives all at once: The Multi-Motive Grid (MMG). *Journal of Personality Assessment, 74*, 126-145.

Son Hing, L. S., Chung-Yan, G. A., Hamilton, L. K., & Zanna, M. P. (2008). A two-dimensional model that employs explicit and implicit attitudes to characterize prejudice. *Journal of Personality and Social Psychology, 94*, 971-987.

Son Hing, L. S., Li, W., & Zanna, M. P. (2002). Inducing hypocrisy to reduce prejudicial responses among aversive racists. *Journal of Experimental Social Psychology, 38*, 71-78.

Spencer, S. J., Fein, S., Wolfe, C., Fong, C., & Dunn, M (1998). Stereotype activation under cognitive load: The moderating role of self-image threat. *Personality and Social Psychology Bulletin, 24*, 1139-1152.

Stacy, A. W. (1997). Memory activation and expectancy as prospective predictors of alcohol and marijuana use. *Journal of Abnormal Psychology, 106*, 61-73.

Stacy, A. W. & Wiers, R. W. (2006). An implicit cognition, associative memory framework for addiction. In: M. R. Munafò & I. P. Albero (Eds.). *Cognition and Addiction*. Pp. 31-71. NY: Oxford University Press.

Stapel, D. A., & Blanton, H. (2004). From seeing to being: Subliminal social comparisons affect implicit and explicit self-evaluations. *Journal of Personality and Social Psychology, 87*, 468-481.

Steele, C. M. & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African-Americans. *Journal of Personality and Social Psychology*, 69, 797-811.

Sudman, S., Bradburn, N. M., & Schwarz, N. (1996). *Thinking about answers: The application of cognitive processes to survey methodology*. San Francisco: Jossey-Bass.

Tanabe, Y., & Oka, T. (2001). Linguistic Intergroup Bias in Japan. *Japanese Psychological Research*, 43, 104-111.

Thistlethwaite, D. L. (1950). Attitude and structure as factors in the distortion in reasoning. *Journal of Abnormal and Social Psychology*, 45, 442-458.

Thrash, T. M., Elliot, A. J., & Schultheiss, O. C. (2007). Methodological and dispositional predictors of congruence between implicit and explicit need for achievement. *Personality and Social Psychology Bulletin*, 33, 961-974.

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

Trope, Y., & Liberman, N. (2003). Temporal construal. *Psychological Review*, 110, 403-421.

Tulving, E., Schacter, D. L., & Stark, H. A. (1982). Priming effects in word-fragment completion are independent of recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 8, 336-342.

Vargas, P. T., Sekaquaptewa, D., & von Hippel, W. (2004). It's not just what you think, it's also how you think: Prejudice as biased information processing. In W. N. Lee & J. Williams (Eds.), *Diversity in Advertising* (pp. 93-119). New Jersey: Earlbaum.

Vargas, P. T., Sekaquaptewa, D., & von Hippel, W. (2007). Armed only with paper and pencil: “Low-tech” implicit measures of attitudes, prejudice, and self-esteem. Chapter to appear in B. Wittenbrink & N. Schwarz (Eds.), *Implicit Measures of Attitudes*. Guilford Press.

Vargas, P. T., von Hippel, W., & Petty, R. E. (2004). Using “partially structured” attitude measures to enhance the attitude-behavior relationship. *Personality and Social Psychology Bulletin*, *30*, 197-211.

Verplanken, B., Friborg, O., Wang, C. E., Trafimow, D., & Woolf, K. (2007). Mental Habits: Metacognitive Reflection on Negative Self-Thinking. *Journal of Personality and Social Psychology*, *92*, 526-51.

von Hippel, W., Lakin, J. L., & Shakarchi, R. J. (2005). Individual differences in motivated social cognition: The case of self-serving information processing. *Personality and Social Psychology Bulletin*, *31*, 1347-1357.

von Hippel, W., Sekaquaptewa, D., & Vargas, P. (1997). The Linguistic Intergroup Bias as an implicit indicator of prejudice. *Journal of Experimental Social Psychology*, *33*, 490-509.

von Hippel, W., Sekaquaptewa, D., & Vargas, P. (2009). Linguistic markers of implicit attitudes. In R. E. Petty, R. H. Fazio, & P. Brinol (Eds.), *Attitudes: Insights from the new implicit measures*. New York: Psychology Press.

Waly, P., & Cook, S. W. (1965). Effect of attitude on judgments of plausibility. *Journal of Personality and Social Psychology*, *2*, 745-749.

Watson, B., & Gallois, C. (2002). Patients’ interactions with health providers: A Linguistic Category Model approach. *Journal of Language & Social Psychology*, *21*, 32-52.

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

Winter, D. G. (1991). Measuring personality at a distance: Development of an integrated system for scoring motives in running text. In A. J. Stewart, J. M. Healy, Jr., & D. Ozer (Eds.), *Perspectives in personality: Approaches to understanding lives* (pp. 59–89). London: Jessica Kingsley.

Winter, D.G., John, O.P., Stewart, A.J., Klohnen, E.C., & Duncan, L.E. (1998). Traits and motives: Toward an integration of two traditions in personality research. *Psychological Review*, *105*, 230-250.