The Fakability and Validity of an Integrity-Based IAT

Andrew Clark & Adam W. Meade

North Carolina State University

ABSTRACT

This study concerns the validity and fakability of an integrity-based IAT (the IAT-sp); an adaptation of one developed by Fischer and Bates (2008). Data provide evidence for the assessments validity. While less predictive of deviance, the IAT-sp was found to be significantly less easily faked than an explicit integrity assessment.

PRESS PARAGRAPH

Employee theft and other counterproductive behaviors have spawned a large market of assessments intended to assess applicant integrity. Despite some promising results, these assessments are susceptible to applicant faking. This study evaluates a reaction time-based measure of integrity. This new measure correlates with engaging in counterproductive behaviors yet is much less easily faked than more traditional integrity measures.
As theft and internal fraud combine to cost organizations an estimated $400 billion dollars annually (Greenberg, 2002), there has been strong interest in the assessment of integrity. Integrity is defined as a personality trait that involves the embracing of truth, justice, and fairness and is often associated with the avoidance of deviant behaviors and compliance with social norms (Connelly, Lilienfeld, & Schmeelk, 2006). Integrity research now spans multiple fields and disciplines (Sackett & Wanek, 1996) and integrity testing has produced a number of measures that have been shown to predict outcomes such as counterproductive, extra-role, and performance behavior (Berry et al., 2007).

Integrity assessments are generally identified as being either overt or covert. Overt integrity tests, like the employee integrity index (EII; Ryan & Sackett, 1987), involve admissions of past deviance and attitudes towards deviant acts (Byle & Holtgraves, 2008). Historically such behaviors have a low base rate of admission or detection (Wimbush & Dalton, 1997). Considering the sensitive nature of such testing and the transparency of such items, there is concern that applicant faking may be responsible for these low base rates (Tett & Christiansen, 2007). One potential solution to concerns over faking is to use an implicit measure of integrity, such as the implicit association test (IAT). The purpose of the present study is to validate an IAT designed to measure integrity with respect to predictive, convergent, and discriminant validity and to compare this measure with an overt integrity measure with respect to faking.

**Self-Perceived Integrity IAT**

The IAT (Greenwald, McGhee, & Schwartz, 1998) is a computer-based assessment that utilizes response latency, or reaction time, to estimate the strength of association between stimuli categories and attributes. Greenwald and colleagues contend that the IAT can be used to uncover automatic associations of which individuals may or may not be cognizant (Greenwald &
Farnham, 2000). In its application, the IAT requires that participants pair target or attribute stimuli to target or attribute categories. The IAT has been used in over 900 studies as of 2011 (Rudman, 2011).

Early IAT investigations involved the assessment of attitudes and stereotypes, particularly the prediction of racial prejudice (Greenwald et al., 1998). Since then, IAT has been used to assess a host of constructs including self-perception and implicit personality (Schnabel, Asendorpf, & Greenwald, 2008). Recently, Fischer and Bates (2008) presented evidence suggesting that it is possible to use the IAT to assess integrity attributes and this study builds directly on that previous work. The Fischer and Bates (2008) IATi measure utilized Greenwald et al. (2002) balanced identity design (BID), consisting of three related IAT categories assessing personal (self/other), social (employee/employer) and attribute (honesty/dishonesty) contrasts.

The version of the IAT utilized in this study incorporates a number of revisions on the IATi (see Table 1). First, the social category (employee/employer) was discarded to focus on the IAT-sp’s assessment of self-concept. Second the self/other category and related stimuli were reworded from self/other to self/not-self. This change was made in response to criticism that use of a non-specific other variable may be confounding (Olson & Fazio, 2004; Pinter & Greenwald, 2005). It is possible that a respondent might make comparisons between self and a specific individual or project one’s own personality onto others, thus altering associations (Karpinski, 2004). In rewording the self/other category, the stimuli were reduced to only me and not me. The use of only two stimuli is not unprecedented and such IATs have been shown to be psychometrically sound, though they may produce mildly attenuated IAT effects (e.g., $D$ decreases from .97 to .85; Nosek et al., 2005). Lastly, this version makes use of a five block design combining matched test
and practice blocks (e.g., practice block 3 and test block 4) of the Greenwald et al.'s (2003) seven block design (Meade, 2009).

**Convergent Validity**

There are several steps in the validation of any new measure. One is to show that the new measure of a construct correlates with existing measures of the same construct. The EII (Ryan & Sackett, 1987) is a commonly used overt integrity test that involves direct admissions of past deviance and attitudes towards deviant acts.

**Hypothesis 1.** The IAT-sp will significantly correlate with the EII.

The IAT-sp should also correlate with constructs commonly found to correlate with integrity measures. A significant portion of integrity-related literature involves the identification of personality facets related to integrity (Byle & Holtgraves, 2008; Lee et al., 2008; Murphy & Lee, 1994). These studies identified significant correlations between integrity and three of the Big Five personality variables: conscientiousness, agreeableness, and neuroticism (Berry et al., 2007). The most significant correlation was found between a personality-based measure of integrity and conscientiousness ($r = .43, p<.001$; Byle & Holtgraves, 2008).

**Hypothesis 2a.** The IAT-sp will be significantly positively correlated with conscientiousness.

**Hypothesis 2b.** The IAT-sp will be significantly positively correlated with agreeableness

**Hypothesis 2c.** The IAT-sp will be significantly negatively correlated with neuroticism.

**Discriminant Validity**

Across studies, both extraversion and openness to experience have been consistently shown to have weak or non-significant relationships with integrity (Byle & Holtgraves, 2008;
Lee et al., 2005). These two variables will be used in the investigation of the IAT-sp's discriminant validity.

**Hypothesis 3a.** The IAT-sp will not be significantly correlated with extraversion.

**Hypothesis 3b.** The IAT-sp will not be significantly correlated with openness to experience.

**Criterion Validity**

Integrity is often investigated in relation to its prediction of employee deviance; specifically, counterproductive work behaviors (CWBs). CWBs are defined as intentional behaviors that harm an organization, its members, and/or purposefully impede productivity (Gruys & Sackett, 2003). As commonly defined, behaviors are only considered counterproductive if they violate some form of social or group norm (Martinko, Gundlach, & Douglas, 2002).

While there are many CWB taxonomies, the findings of meta-analytic investigations suggest that the co-occurrence of counterproductive behaviors is evidence of a personality-based general deviance construct (Berry et al., 2007; Gruys & Sackett, 2003). This deviance construct is hierarchical; the higher order general deviance construct can be broken down into target-groups, which can in turn be separated into particular facets. This CWB conceptualization is very similar to the general integrity construct, which has been shown to be nearly equally predictive of all CWB facets and components (Berry et al., 2007).

**Hypothesis 4.** Integrity, as measured through the IAT-sp, will be positively correlated with admission of counterproductive work behaviors.

**Research Question 1.** Does the IAT-sp account for more variance in CWBs than does the EII?
Faking

While overt measures of integrity prove useful in predicting performance, faking is still a concern (Dwight & Alliger, 1997). It is possible that when faking is an issue, response latency measures can be utilized to limit the extent to which applicants are able to fake (Berry et al., 2007; Dwight & Alliger, 1997). IAT effects have been shown as the least influenced by faking when compared to other forms of measurement (Kim, 2003).

The IAT’s resistance to faking is strongest when participants have little or no experience with the IAT, indicating that experience plays a part in determining an IAT score (Steffens, 2004). While generally more resistant to faking than explicit measures (Kim, 2003), it is possible to suppress scores on the IAT with and without instruction in how to do so (Fiedler & Bluemke, 2005). Fiedler and Bluemke (2005) found that participants were capable of faking by purposefully hesitating on compatible trials, making it impossible for investigators to differentiate the fake from true data. This evidence was further supported by the findings of De Houwer, Beckers, and Moors (2007) who discovered that, with appropriate instruction, IAT effects can be created from novel stimuli. Conversely, research has also shown that individuals generally have great difficulty in faking the IAT without instruction to do so (Kim, 2003). While fakable, response latency measures are still very resistant to faking when compared with other forms of measurement (Kim, 2003; Schnabel et al., 2008).

Hypothesis 5. The difference between IAT-sp scores in faking and non-faking conditions will be significantly lower than the difference between scores in EII faking and non-faking conditions.

Method

Participants
This study's participants consisted of 251 undergraduate students enrolled in an introductory Psychology course at a large southeastern university. Students were provided course credit for participation. The sample consisted primarily of Caucasian (69%) females (57%). All participants were either currently employed (43%) or had been employed within the past two years.

Measures

**Self-perception implicit association test for integrity (IAT-sp).** The newly developed IAT-sp was developed by modifying Fischer and Bates (2008) IATi as described previously. The IAT-sp was administered using the FreeIAT program (Meade, 2009).

**EII.** The employee integrity index (Ryan & Sackett, 1987) was utilized in this study as an explicit measure of workplace integrity. This index consists of 63 items measuring integrity and nine items measuring social desirability.

**Interpersonal and organizational deviance scale (IODS).** The interpersonal and organizational deviance scale devised by Bennett and Robinson (2000) was utilized in this study as a measure of deviant behavior. The scale consists of 12 items measuring deviant behaviors that are directly harmful to an organization and seven items assessing deviant behaviors that are directly harmful to individuals in that organization.

**Personality variables.** The International Personality Item Pool (IPIP) NEO Personality Inventory (NEO-PI; Goldberg, Johnson, Eber, Hogan, Ashton, Cloninger, & Gough, 2006) was utilized in this study as an explicit measure of the Big Five personality variables conscientiousness, agreeableness, neuroticism, extraversion, and openness to experience. Each scale consists of 20 items with alphas ranging from .89 to .91 (Goldberg et al., 2006).

Procedure
Participants were informed that the study consisted of three parts, two IATs and a survey consisting of 289 questions. Participants were informed that the study included questions about integrity, and CWBs and were asked to complete the two IAT procedures and a computer based survey. The survey consisted of the EII, CWB, and IPIP NEO-PI scales. For the IAT, participants were presented with instructions prior to each testing block. Following completion of the first and second sections, participants were directed to retake the IAT and EII as if they were an actual job applicant. Faking can be operationalized as the difference in the “job applicant” and “regular” condition IAT scores.

Results

Convergent Validity

Descriptive statistics for all variables investigated in this study are presented in Table 2. Hypothesis 1 stated that the IAT-sp would significantly correlate with the EII. The correlation coefficient produced by this relationship was significant ($r[208] = .198$, $p < .01$), thus providing support for Hypotheses 1.

To test Hypothesis 2 ($H2a$ through $H2c$: The IAT-sp is significantly correlated with conscientiousness, agreeableness, and neuroticism) correlation coefficients were computed between the variables (see Table 3). The IAT-sp was significantly correlated with conscientiousness, but not agreeableness or neuroticism. These data only support hypotheses 2a and provide partial evidence for convergent validity.

Discriminant Validity

To test Hypothesis 3 ($H3a$ & $H3b$: The IAT-sp will exhibit discriminant validity with extraversion and openness to experience), correlation coefficients were computed between the IAT-sp and the two personality variables (see Table 3). The IAT-sp was bit significant correlated
with extraversion or openness to experience. These results provide evidence supporting the discriminant validity of the IAT-sp.

**Criterion-Related Validity**

Hypothesis 4 states that integrity, as measured through the IAT-sp, should be positively correlated with admission of CWBs. The IAT-sp was significantly related to CWBs (see Table 3), supporting Hypothesis 4.

**Research Question 1**

This study’s research question asks whether or not the IAT-sp accounts for more variance in CWBs than the EII. The variance explained in CWBs by the IAT-sp was compared to that produced by the EII. In answer to this research question, a Fisher’s r to z transformation identified a significant difference in predictive ability of the IAT-sp and EII ($z = 3, p < .01$); specifically, identifying the EII as more predictive than the IAT-sp.

**Faking**

Hypothesis 5 stated that the difference in faking and non-faking conditions would be significantly lower for the IAT-sp than the EII. In order to test this, difference scores were computed for the faking and non-faking conditions of the IAT-sp and the faking and non-faking conditions of the EII. As these assessments were on different scales, the difference scores were compared using a Wilcoxon signed-rank test. Results of the Wilcoxon sign-ranks test indicate that the IAT-sp was significantly less fakable than the EII ($z = -7.09, p < .001$).

In addition, a dependent samples t-test indicated a significant and strong difference between the faking and non-faking EII conditions ($t[91] = 31.48, p < .001, d = -1.12$) but not the IAT-sp conditions ($t[212] = .273, p = .785, d = -0.02$); wherein EII scores are higher in the

---

1 These analyses were conducted only on data collected from participants who had successfully completed the IAT-sp, EII, and IODS with no omitted answers.
faking condition. All findings support the hypothesis that the IAT-sp is significantly less fakable than the EII sample. An interesting finding was that under the instructions to pretend they are job applicants, neither the EII nor the IAT-sp proved ineffective in predicting CWBs (EII, r[92] = - .133, p = .21; IAT-sp, r[213] = .072, p = .29). A Fisher r to z transformation revealed no significant difference between these correlations (z = -1.63, p = .10).

Discussion

IAT-sp Validation

This study’s primary purpose was to validate the IAT-sp and to compare the measure with the EII with respect faking. This study's first result indicated that the IAT-sp significantly related to the EII, providing convergent validity evidence. These findings are contrary to those of Fischer and Bates (2008) and Fischer, Osafo, and Turner (2010) who found no significant correlation when comparing their IATi to the EII. It is possible that altering the IAT target categories (self/other to me/not me) improved convergent validity, though as no data were collected on the original IATi, there are no data to support this supposition.

Past research has identified the three Big Five personality variables conscientiousness, agreeableness, and neuroticism to relate to the integrity construct; with conscientiousness exhibiting the strongest relationship (Berry et al., 2007). In line with expectation, the IAT-sp was found to correlate significantly with conscientiousness, though it did not correlate with agreeableness or neuroticism. Furthermore, as hypothesized, the IAT-sp did not correlate with either openness to experience or extraversion. On the whole, these findings indicate evidence for convergent and discriminant validity.

This study’s results also indicate that the IAT-sp is capable of predicting CWBs. These findings support those of Fischer, Osafo, and Turner (2010), who found a relationship between
an IAT for workplace integrity and cheating behavior. These results were notable, as research suggests the IAT will perform poorly in the prediction of scores on explicit measures (Fazio & Olsen, 2003). However, the IAT-sp showed a weaker correlation with these behaviors than did the EII. Explicit assessments have been shown to have higher correlations with other explicit measures than with implicit measures in general (Berry et al., 2007); though this effect is more pronounced when assessing a sensitive construct like deviance (Fischer & Bates, 2008). With these particular measures, the EII and IODS are both explicit measures that directly ask about previous CWBs. To some extent, participants are reporting on the same past behaviors in both measures. Using observed, rather than self-reported, CWBs as the criterion would likely show an attenuated correlation between the two explicit measures. Similarly, predicting future criterion behaviors using the EII’s report of past behaviors would also undoubtedly attenuate the correlation between the two.

Though the EII exhibited a stronger correlation with self-reported CWBs, the IAT-sp was much more resistant to faking. Faking is exacerbated when the assessment's construct is of a sensitive nature and the respondent perceives a benefit that can be derived by faking (Fischer & Bates, 2008). In the administration of this study, it was noted that participants’ attempts to fake the IAT were largely unsuccessfully. The FreeIAT program automatically flags respondents that respond faster than information processing allows, or unusually slow which would be consistent with deliberate faking (Greenwald et al., 2003). In attempting to fool the program by this method, approximately 11% of participants’ data were discarded by the program for being too fast. It is possible that, along with being more difficult to fake, attempting to fake the IAT carries more risk of being caught than with explicit measures. The transparency of the EII has been shown susceptible to faking (Schhnabel et al., 2008). In faking explicit measures, a respondent only
needs to strongly agree with socially desirable responses and strongly disagree with undesirable responses. In faking the IAT, in the event that a respondent understands the program's latency method, they must concentrate on tactically controlling their response speed.

The results identified the IAT-sp as being significantly less fakable than the EII. While the EII explains more variance in deviant behavior than does the IAT-sp under low-stakes conditions, it is considerably more susceptible to faking. Neither measure correlated with CWBs under instructions of responding as though they were a job applicant. While it is unclear as to whether or not a respondent will purposefully fake responses to integrity tests (Berry et al., 2007), it is clear that the IAT-sp is better equipped to prevent or identify faking attempts.

**Implications.** This study's findings do not provide evidence to suggest that the IAT-sp can replace existing explicit assessments like the EII. While the IAT-sp was related to CWBs, there is no evidence to suggest that it is more effective than the EII. However, the IAT-sp could prove to be a more effective predictor of observed future CWBs, rather than self-reported past deviance. One firm conclusion from this study is that the IAT-sp is less fakable than the EII and as such the IAT-sp could prove to be a useful tool when faking is a concern. As it is less fakable than the EII and possesses sufficient validity, in such instances the IAT may be an appropriate assessment. Further research should be conducted to clarify the relationships identified in this study and determine the utility of the IAT-sp as an integrity assessment.

**Limitations and Future Research**

This study's most significant limitations relate to the sample and design used in this study. Participants were primarily unemployed undergraduate students between the ages of 18 and 21. Though most were currently unemployed, all had held some job within the past two years. Despite this work history, past research has suggested that students are poor employee surrogates
(Singer, 1989). While all participants had been employed, most jobs were likely temporary or part-time. Furthermore, the study itself was conducted in a lab setting without any incentives, whereas in an applied setting job applicants could be denied employment based on their responses. This likely proved detrimental to the study’s external validity.

More importantly, the study could greatly benefit from a predictive rather than concurrent validity design with these measures. As stated previously, participants self-reported previous instances of CWB on both the EII and the IODS. To this extent the predictive validity of these measures is almost overestimated with a concurrent design. Future research is needed with applicant samples and a predictive validity design to further understand the utility of the IAT-sp.


doi:10.1111/j.1468-2389.2006.00335.x


doi:10.1111/1467-9280.00328


doi:10.1016/j.jesp.2006.10.007


doi:10.1207/s15324834basp2704_3


to Morgeson, Campion, Dipboye, Hollenbeck, Murphy, and Schmitt (2007).


doi:10.1037/0021-9010.82.5.756
Table 1

*Implicit Association Test of integrity (IAT-sp)*

<table>
<thead>
<tr>
<th>Person (original IATi)</th>
<th>Person (IAT-sp)</th>
<th>Attribute (IAT-sp)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Self</strong></td>
<td><strong>Other</strong></td>
<td><strong>Self</strong></td>
</tr>
<tr>
<td>Me</td>
<td>Them</td>
<td>Me</td>
</tr>
<tr>
<td>My</td>
<td>Their</td>
<td>Not Me</td>
</tr>
<tr>
<td>Mine</td>
<td>Theirs</td>
<td>Honest</td>
</tr>
<tr>
<td>Self</td>
<td>Other</td>
<td>Dishonest</td>
</tr>
<tr>
<td>I</td>
<td>They</td>
<td>Fair</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sincere</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Trustworthy</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Moral</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lie</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Corrupt</td>
</tr>
</tbody>
</table>
### Table 2

*Descriptive statistics*

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>IAT-sp</td>
<td>223</td>
<td>.34</td>
<td>.22</td>
</tr>
<tr>
<td>EII</td>
<td>236</td>
<td>4.42</td>
<td>.56</td>
</tr>
<tr>
<td>CWB</td>
<td>251</td>
<td>1.95</td>
<td>.67</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>236</td>
<td>4.91</td>
<td>.88</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>250</td>
<td>5.06</td>
<td>.83</td>
</tr>
<tr>
<td>Neuroticism</td>
<td>249</td>
<td>3.26</td>
<td>1.14</td>
</tr>
<tr>
<td>Extraversion</td>
<td>251</td>
<td>4.95</td>
<td>1.07</td>
</tr>
<tr>
<td>Openness to experience</td>
<td>251</td>
<td>5.04</td>
<td>.99</td>
</tr>
</tbody>
</table>
Table 3

*Bivariate correlation matrix*

<table>
<thead>
<tr>
<th>Variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>IAT-sp</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EII (Total)</td>
<td>.198**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CWB</td>
<td>-.176**</td>
<td>-.496**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>.202**</td>
<td>.457**</td>
<td>-.342**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agreeableness</td>
<td>.051</td>
<td>.449**</td>
<td>-.481**</td>
<td>.442**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neuroticism</td>
<td>-.036</td>
<td>-.168*</td>
<td>.195**</td>
<td>-.369**</td>
<td>-.392**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extraversion</td>
<td>-.123</td>
<td>.023</td>
<td>.015</td>
<td>.275**</td>
<td>.213**</td>
<td>-.378**</td>
<td></td>
</tr>
<tr>
<td>Openness</td>
<td>.046</td>
<td>-.009</td>
<td>-.008</td>
<td>.129*</td>
<td>.185**</td>
<td>-.006</td>
<td>.138</td>
</tr>
</tbody>
</table>

*Note.* **p < .01;* Openness stands for the variable Openness to Experience. Sample sizes ranged from 219 to 251. Differences are attributed to missing data in the relevant scale; wherein incomplete data were discarded. Sample sizes for each scale are presented in Table 3.