Computers in Human Behavior 27 (2011) 1937–1941

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Validity and reliability of the IAT: Measuring gender and ethnic stereotypes

Ali R. Rezaei *

College of Education, California State University, Long Beach, 1250 Bellflower Blvd., Long Beach, CA 90840, USA

ARTICLE INFO

Article history:
Available online 23 May 2011

Keywords:
Assessment
Implicit association
IAT
Reliability
Validity
Stereotypes

ABSTRACT

The Implicit Association Test (IAT) was developed in response to reports of low validity of explicit (self-report) measures of attitudes, stereotypes, and prejudices. Usually, people are unwilling to report what they think and feel about other races, groups, and nationalities. The IAT has been written about in many books, newspapers, journal articles, websites, and has been featured frequently on radio and television many times; its web site has now reached a peak of 5 million visits. However, despite this popularity its validity and particularly its reliability is under question. This article reports on the validity and reliability of the IAT. Four different experiments were conducted on 150 students at California State University, Long Beach to investigate the temporal reliability of IAT. Also students' opinion (trust) about the validity and reliability of the test was evaluated. The results showed that while there are numerous reports of moderate validity of the test, its reliability as measured in this study, particularly for the first time users, is relatively low. Familiarity with similar tests, however, improves its reliability.

1. Introduction

The unconscious mind has been identified by psychologists since the 18th century, and later it was widely introduced to the public through Sigmund Freud's publications (Breuer & Freud, 1895). Psychologists believe that people may not say what is on their minds, either because they are unwilling or because they are unable to do so (Greenwald & Banaji, 1995). This is a major problem in psychological assessment, particularly when people are asked about their biases or prejudices (self-report methods). They may not be willing or able to tell you about the extent of their biases, and they may lie, hide, or dissimulate. "Explicit measurements are subject to various limitations and therefore it is desirable to search for measurement tools that are less susceptible to distortions" (Grumm & von Collani, 2007, p. 2206).

Therefore, psychologists started developing the so-called implicit measures based primarily on performance data. In these scales the researchers hide the purpose of their measurement. Thus, when responding to these measures, participants are unaware of the way in which conclusions are drawn from their responses. According to Egloff, Schwerdtfeger, and Schmukle (2005, p. 82), the main reason for the construction of implicit tests was twofold: First, this method prevents the operation of response factors such as self-presentation, demand characteristics, and faking. Second, a distinction is drawn between an explicit and an implicit mode of processing.

In his recent book "hidden brain", Vedantam (2010) illustrates several situations in which people do something against their intentions. To explain this phenomenon he introduces the concept of "unconscious brain" or "the hidden brain" (the hidden forces that influence us in everyday life). He explains further that our conscious mind is like the pilot of the plane, and the hidden brain is the autopilot function of the plane or the co-pilot function of the plane. According to Vedantam, people regularly transfer functions back and forth between the pilot and the autopilot functions. The problem arises, he suggests, when we do this without our awareness, and the autopilot ends up flying the plane, when we should be flying the plane. That is how stereotypes, prejudices, and discrimination control people's behavior (they are not aware of it). He suggests that racial categorization begins at a very early age. He cites an experimental study from a day-care center in Canada that found children as young as three showed negative stereotypes against black faces (Aboud, 2003).

As reported by Greenwald and Banaji (1995), there is considerable evidence supporting the view that social behavior often operates in an implicit or unconscious fashion. Therefore, the IAT was developed to examine thoughts and feelings that exist either outside of conscious awareness or outside of conscious control (Project Implicit, 2010; Greenwald, McGhee, & Schwartz, 1998; Greenwald, Nosek, & Banaji, 2003). Since its inception, the IAT has been used in more than 300 published articles and cited in more than 1000 articles. According to Hofmann, Gawronski, Gschwendner, Le, and Schmitt (2005), one of the most important contributions in social cognition research within the last decade was the development of implicit measures of attitudes, stereotypes, self-concept, and self-esteem.

* Address: 6207 E 6th Street, Long Beach, CA 90803, USA. Tel.: +1 562 985 4532; fax: +1 562 985 4534.
E-mail address: arezaei@csulb.edu

0747-5632/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.chb.2011.04.018

Investigating the validity and reliability of the IAT is very important. The IAT has been applied in various disciplines including social and cognitive psychology, clinical psychology, developmental psychology, neuroscience, market research, and health psychology (Nosek, Greenwald, & Banaji, 2007, p. 267). The test examines disparate topics such as attitudes, stereotypes, self-esteem, phobias, and consumer behavior (Aberson & Beeney, 2007, p. 27). It is being used in courts to test the biased witnesses (David, 2008), to evaluate applicants based on university admission requirements (Bererji, 2005), and in religious studies to evaluate subjects religious faith (Ventis, Ball, & Viggiano, 2010). Nevertheless, many educators believe that this unusual excitement both among psychologists and the public stems from the kinds of associations (popular topics) that researchers have used the test to measure rather than validity or reliability of the test (Azar, 2008).

1.1. Scoring

The IAT measures associations between two concepts (e.g., *Asian American* and *European American*) and two attributes (e.g., *good* and *bad*). The participants are required to sort stimuli representing four concepts using just two responses, each assigned to two of the four concepts. The basic assumption here is that, if two concepts are highly associated, the sorting task will be easier and faster (Schmukle & Egloff, 2005, p. 101).

What is critical in IAT assessment is the reaction time to associate a concept (*old–young*, *Muslim–Christian*) to good and bad attributes. The resulting score is on a scale of −2.0 to 2.0, with all scores above 0.65 or below negative 0.65 indicating a "strong" link. It should be noted that these actual numbers are not displayed on screen to show the results of the test to the user. The IAT results categorizes (ranks) the users into seven groups with strong, moderate, slight, and little or no association with one of the two concepts (e.g., females association with science). Therefore, the user can see only the amount of his/her association (prejudice against) towards the two concepts (male with science, female with arts).

1.2. Validity

Nosek and Hansen (2008) studied the IAT on 158 samples (N = 107,709) and concluded that the test was reliably and variably related to explicit attitudes, and explicit attitudes accounted for the relationship between the IAT and cultural knowledge. In another comprehensive chapter Nosek et al. (2007) reported the accumulated evidence for the construct and predictive validity of the IAT in assessing individual differences and suggested that it is thriving as a research tool and will likely continue to do so. They provide numerous research examples of the relationship between the IAT and other implicit and explicit measures. They also reported that while some of the earlier research efforts found weak to absent correlations between IAT and other self-report scales, more recent research has shown that, in some cases, the IAT and self-report can be strongly related. At the most extreme, a large Internet sample revealed a high correlation between the IAT and self-report of 0.86 (Nosek, et al., 2007, p. 278). They provide evidence supporting the construct validity (divergent and convergent) of the IAT in a wide variety of domains. Surprisingly, they found relatively weak relations between IAT and other implicit measures. Therefore, the authors of the test believe that IAT can effectively predict human behavior. They also believe that is, in part, because of its ease of administration, adaptability to a variety of topics, large effect sizes, and good reliability compared to other implicit measures.

While the authors of the test are quite confident about the validity of the test, many researchers are raising serious concerns about its validity. Greenwald, Nosek, and Sriram (2006) admit that the IAT has attracted more scholarly criticism than have other measures designed for similar purposes. It is not clear whether these criticisms are due to the popularity of the test or due to its technical problems.

For example, Blanton and Jaccard (2006) have criticized the IAT for its arbitrary scoring and labeling. They believe that many psychological tests have arbitrary metrics, but the problem is that the IAT authors draw inferences about the true, absolute standing of a group or individual on the latent psychological dimension being measured, and this is a serious concern. Other researchers are also concerned about the premature publicity of the IAT and the fact that many studies have found that the IAT provides only modest predictions of behavior (Azar, 2008).

Furthermore, Han, Czellar, Olson, and Fazio (2010, p. 286) argue that the IAT results as conventionally implemented are open to multiple interpretations, and therefore, can provide contextually malleable measurement outcomes. They suggest that the IAT's sensitivity to extra-personal associations leaves it susceptible to momentary contextual influences. Consequently, this sensitivity has the potential to obscure the detection of change in a situation where there is every reason to believe that attitude change has occurred. Likewise, Karpinski, Steinman, and Hilton (2005) believe that the evidence for predictive validity of the IAT is weak. Their experiment with the IAT showed that explicit attitude measures are better predictors of deliberative behaviors than the IAT scores. They also found that as the importance of the attitude (to be measured) increases, the strength of the relationship between the IAT and explicit attitude measures also increases. Similarly, von Stülpnagel (2010) reported that faster and more intelligent participants have larger IAT effects in some IATs due to their reduced task-switch costs and, therefore, larger IAT effects are expected for more intelligent people. Finally, Arkes and Tetlock (2004) argue that the IAT data may reflect shared cultural stereotypes rather than personal beliefs. Therefore, even Jesse Jackson, the famous American civil rights activist, may fail the IAT race test.

1.3. Reliability

A few researchers who have examined the reliability of the IAT either have focused on the internal consistency or have found moderate reliability coefficients. Furthermore, those studies were performed on earlier IAT tests and, unfortunately, there is no recent evidence of temporal reliability of the new IAT scales. According to Aberson and Beeney (2007), although the IAT has been applied broadly, little is known about the psychometric properties of the measure.

Steffens and Buchner (2003, p. 33) noted the internal consistency of the IAT is very high. However, they reported "within-situation consistency is accompanied by considerable unexplained between-situation variability". For example, substance use may hinder the reliability of the IAT (Aberson & Beeney, 2007).

A few studies on the temporal reliability of the IAT came to the conclusion that the test results are subject to personal, situational, and environmental changes. Schmukle and Egloff (2005) concluded that use of the IAT for assessing personality is somewhat more affected by situations than explicit measures, because they are more occasion-specific. For example, changing the test-taking strategy might be a factor in changing the reliability. Likewise, learning effects could be another reason for concern (e.g., participants may be more familiar with the procedure on the second occasion).

Nonetheless, the authors of IAT believe that, in contrast to other implicit measures, the IAT routinely shows good reliability, which is a prerequisite for measuring individual differences. Nosek et al. (2007) report that the IAT has displayed satisfactory internal consistency, which is relatively rare for latency-based measures. They report a split-half internal consistency for the self-esteem IAT of r = 0.69 compared to r values of −0.05 to 0.28 for other

latency-based implicit self-esteem measures. They conclude that the internal consistency estimates (split-half correlations or alphas) for the IAT measures tend to range from 0.7 to 0.9.

1.4. Research questions

As shown in the literature review, current research on the IAT has focused mostly on validity, rather than the reliability of the test. The goal of this project was to re-examine the reliability of the test and the factors that may improve its reliability. The present study, also intends to evaluate the test from the users' (college students) perspective. Four studies are reported here that were designed to answer the following four research questions.

a. To what extent do students trust the results of IAT? Do they believe the test is reliable and valid?
b. Does it make a difference to train students about their hidden stereotypes and prejudices? Is there a difference in the trust in IAT between a group of students who receive such training and those who do not receive such training?
c. What is the reliability of the IAT based on the results reported (displayed) to the users? Is there a difference between the coefficients determined using a one week test–retest reliability and an immediate test–retest reliability index?
d. Does it make a difference if the users are familiar (trained) with the test? Does such training make the test more reliable?

2. Methodology

A total of 150 participants, all of whom were college students, took part in this study. The participants' ages ranged from 20 to 55, and about 73% of them were female. Subjects received course credit points for their participation. Most of the subjects were relatively new classroom teachers. All subjects had initial familiarity with the two concepts of reliability and validity.

In an initial pilot study it was observed that some students did not take the test seriously and some did not seem to know that the speed is an issue. To increase the accuracy of the measurement, all students were shown a slide show about the theoretical framework for the IAT and a brief explanation on how it works. This introduction was given to prepare them for the test and also to increase their active participation in the experiment. The content of the introduction was mainly adopted from the IAT web site.

At the end of each experiment, the subjects were asked to rate the validity and reliability of the test (based on their experience with the test, the given results, and what they know about validity and reliability). They used a Likert type scale (1 = no validity or reliability, 2 = low, 3 = moderate, 4 = strong) to do so. This was considered as their trust (perceived) in reliability and validity of the IAT.

2.1. Experiment 1

Thirty-two students participated in this experiment. Students watched a 20 min video clip about the unconscious mind and stereotypes and prejudices prior to taking the test. The video used was from a very informative TV documentary "Race and Sex: What We Think But Don't Say" (Mastropolo & Varney, 2006) which presents some social psychology experiments, and interviews with experts and psychologists about the hidden brain, unconscious mind, and implicit measures.

Students in this group took the Gender/Science IAT twice. This test is available online. The time interval between the pretest and posttest was only 15 min. This test measures the degree to which the users associate females and males with science and liberal arts.

The introductory screen of the test asks the users to complete a brief questionnaire before starting the test. The subjects in this study were allowed to skip this part if they wanted. As mentioned earlier, the users of the IAT are asked to sort words and pictures into categories as quickly as possible. The actual test usually takes about 5 min and at the end they see their results as well as a chart showing the results for the overall population of test-takers. The users do not receive any numeric results; all they receive is their category on a 7-unit scale as follows:

1 = Strong association of **male with science** and female with liberal arts
2 = Moderate association of **male with science** and female with liberal arts
3 = Slight association of **male with science** and female with liberal arts
4 = Little or **no association** between female and male with science and liberal arts
5 = Slight association of **female with science** and male with liberal arts
6 = Moderate association of **female with science** and male with liberal arts
7 = Strong association of **female with science** and male with liberal arts

The following category items are used in this test for the users to associate with the four categories.

Category	Items
Male	Man, Boy, Father, Male, Grandpa, Husband, Son, Uncle
Female	Girl, Female, Aunt, Daughter, Wife, Woman, Mother, Grandma
Science	Biology, Physics, Chemistry, Math, Geology, Astronomy, Engineering
Liberal arts	Philosophy, Humanities, Arts, Literature, English, Music, History

2.2. Experiment 2

Thirty-four students participated in this experiment. Unlike experiment 1, subjects in this experiment did not watch the video. Furthermore, subjects in this group were asked to practice with the IAT before they started the main test (not the same test). Although the IAT has some practice items implemented in the test, the practice is not long enough for the user to learn how to interact with the software. The pilot study revealed the fact that many first time users were either too slow or too fast, and consequently they received an error message and their test was not completed.

Students took the Asian/American – European/American IAT twice. The time interval between the pretest and posttest was one week. The results screen is similar to the Gender/Science IAT and informs the users about the extent to which they associate Asian/Americans to foreign and European/Americans to America.

2.3. Experiment 3

Thirty-eight subjects participated in this experiment. Similar to Group 1, this group used the Gender-Science IAT. This group also took the test twice in a one-week time interval. Unlike Group 1, however, this group had a chance to practice with the IAT before starting the main test.

2.4. Experiment 4

Forty-six subjects participated in this experiment. This group also had a chance to watch the introductory video, to practice with the IAT before starting the main test, and they did the Asian/America test twice in a 15-min interval. A summary of four experiments (four groups) is presented in Table 1.

3. Results

The first research question in this study investigated whether students consider the test to be both reliable and valid.

The results as shown in Table 2 indicate that about 73% of subjects believed the test was valid to some extent. About 29% of subjects believed that IAT was highly valid, and about 27% thought the test was not valid at all. Table 3 shows how much each group trusted in the reliability of the IAT. This table shows that about 82% of subjects believed the test was reliable to some extent. About 55% of subjects believed that IAT was highly reliable, and about 18% thought the test was not valid at all.

The second research question investigated whether watching the introductory video helped build students' trust in the validity and reliability of the IAT. The results as reflected in Table 2 showed that different groups showed different levels of trust in the validity of the ITA. Subjects in Group 2, who did not get a chance to see the video, mostly rated the test as having no or low validity. Comparing Group 2 and Group 4 (both evaluated the same IAT test) showed that watching the video significantly increased their trust in IAT ($t = 2.447$, $p = 0.017$).

Table 3 shows how much each group trusted in the reliability of the IAT. This table shows that subjects in Group 2, who did not have a chance to see the video, mostly rated the test as having no or low reliability. Comparing Group 2 and Group 4 (both evaluated the same IAT test) shows that watching the video significantly increased their trust in the IAT ($t = 2.039$, $p = 0.045$).

The third goal of this study was to re-examine the reliability of the IAT based on the results reported (displayed) to the users. Consequently, it was intended to check if one-week test–retest reliability was different from immediate test–retest reliability index. The reliability of the test for the control group (the group that did not practice before the test) was 0.32. However, the combined reliability (computed on all 150 subjects in all 4 groups together) was 0.52.

Comparing Groups 2 and 4 (both used the Asian/American IAT) showed that the time interval between pretest and posttest did not make a difference. Also comparing Groups 2 and 3 (both used one week interval between pretest and posttest) showed that there was no difference in the reliability of Asian/American IAT and the Gender/Science IAT.

The last research question was to investigate whether it made a difference in the reliability of the IAT when the users practice with other IATs before the actual test. Does such training make the test more reliable? The results showed that there was a large difference in reliability of the test between the first experiment and the other three experiments. The test–retest reliability index for the four experiments were 0.32 ($N = 32$), .57 ($N = 34$), 0.57 ($N = 38$), and 0.56 ($N = 46$) respectively. These numbers indicate that familiarity with the test did improve the reliability of the test. The only difference between Group 1 and Group 3 was that the first group did not practice with any IATs before the main test and so they were not familiar with it. These two groups used the same IAT test (Gender/Science) and the interval between test and retest was one week for both groups.

Table 2
Students' rating of the validity of the test.

		Rating of the validity			
		No	Low	Moderate	High
Group 1	Count	8	5	8	11
	%	25.0	15.6	25.0	34.4
Group 2	Count	13	8	9	4
	%	38.2	23.5	26.5	11.8
Group 3	Count	10	6	10	12
	%	26.3	15.8	26.3	31.6
Group 4	Count	10	8	12	16
	%	21.7	17.4	26.1	34.8
Total	Count	41	27	39	43
	%	27.3	18.0	26.0	28.7

Table 3
Students' rating of the reliability of the test.

		Rating of the reliability			
		No	Low	Moderate	High
Group 1	Count	6	5	5	16
	%	18.8	15.6	15.6	50.0
Group 2	Count	7	8	8	11
	%	20.6	23.5	23.5	32.4
Group 3	Count	6	3	4	25
	%	15.8	7.9	10.5	65.8
Group 4	Count	8	4	4	30
	%	17.4	8.7	8.7	65.2
Total	Count	27	20	21	82
	%	18.0	13.3	14.0	54.7

4. Discussion

The result of this study has a practical implication. The IAT website reports that the majority of people taking the race IAT (70%) have "automatic preference for Whites over Blacks" and 27% have a "strong automatic preference for Whites over Blacks." Similar statistics are presented about other social attitudes and stereotypes. According to Blanton and Jaccard (2008), these diagnoses probably lead many individuals to infer that that they possess hidden anti-black racist attitudes. The results of this study also indicates that subjects gained a high level of trust in validity and particularly the reliability of the IAT. Regarding the low reliability of the test as reported in this paper and the earlier discussion about the validity

Table 1
Experimental conditions of the 4 groups.

	N	Male	Female	Introductory video	Practice before	Test–retest	IAT test
Group 1	32	6	26	Yes	No	1 week	Gender/Science
Group 2	34	10	24	No	Yes	1 week	Asian/Americans
Group 3	38	10	28	Yes	Yes	1 week	Gender/Science
Group 4	46	14	32	Yes	Yes	15 min	Asian/Americans
Total	150	40	110				

of the test, one should be cautious about the implications, interpretations, and, particularly, in any kind of decision making.

This study focused on reliability, while validity of the IAT has also been the subject of hundreds of research studies. Lack or low correlation of the IAT results with parallel explicit (self-report) measures has made many researchers question the validity and reliability of the test. Reliability is a precondition for validity. As noted by Grumm and von Collani (2007, p. 2215), "the problem of low test–retest-reliability must be considered as a critical limitation for a diagnostic application". Without proving high levels of reliability any claims about the validity of the test should be considered with caution. Perhaps what makes this test vulnerable in any kind of assessment of its validity and reliability is the use of reaction time. Reaction time has been used for many other assessments, particularly, in recent computerized implicit measurements (e.g., cognitive styles) which has also been found to be unreliable (Rezaei & Katz, 2004). "For current reaction-time indices, a tenth of a second can have a consequential effect on a person's score, and such measurement sensitivity can lead to test unreliability" (Blanton & Jaccard, 2008, p. 289). Therefore, the difference between temporal reliability reported here ($r = 0.32$) and the internal consistency index reported earlier ($r = 0.89$) may also indicate that the reliability of the IAT is more affected by stability of users in their reaction time rather than the stability in their implicit attitudes.

Regarding the massive publicity of the IAT, higher level of reliability was expected in this study. The IAT is typically used to measure deeply seated personality traits or stereotypes rather than temporary states of mind. As observed in Steffens and Buchner (2003) and reported by Azar (2008), when compared to established tests of personality traits, the trans-situational component of the IAT looks quite small. Therefore, it is reasonable to consider the reliability coefficients in this study and other studies reported earlier to be unsatisfactory. As suggested by the authors of the IAT scholarly criticism of the IAT, increases the motivation to pursue questions about this test. The results of this study suggest that more research is needed to find out how the reliability and validity of the IAT could be improved. The theoretical framework of the test is quite powerful and the publicity of the test without any doubt has elevated social awareness in the areas of ethnical and gender studies. As reported by Azar (2008), the IAT has the potential to be a remarkably powerful tool.

References

Aberson, C. A., & Beeney, J. E. (2007). Does substance use impact implicit association test reliabilities? *Journal of Social Psychology, 147*, 27–40.

Aboud, F. E. (2003). The formation of in-group favoritism and out-group prejudice in young children: Are they distinct attitudes? *Developmental Psychology, 39*, 48–60.

Arkes, H. R., & Tetlock, P. E. (2004). Attributions of implicit prejudice, or "Would Jesse Jackson 'fail' the Implicit Association Test?". *Psychological Inquiry, 15*(4), 257–278.

Azar, B. (2008). IAT: Fad or fabulous? *Monitor in Psychology, 39*(7), 44.

Blanton, H., & Jaccard, J. (2008). Unconscious racism: A concept in pursuit of a measure. *Annual Review of Sociology, 34*, 277–297.

Blanton, H., & Jaccard, J. (2006). Arbitrary metrics in psychology. *American Psychologist, 61*, 27–41.

Bererji, S. (2005). Who do you think you are? *Issues in Higher Education, 22*(15), 32–34.

Breuer, J., & Freud, S. (1895). *Studies on hysteria*. New York: Collier.

David, G. (2008). Our hidden prejudices, on trial. *Chronicle of Higher Education, 54*(33), b12–b14.

Egloff, B., Schwerdtfeger, A., & Schmukle, S. C. (2005). Temporal stability of the implicit association test – anxiety. *Journal of Personality Assessment, 84*(1), 82–88.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review, 102*(1), 4–27.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology, 74*(6), 1464–1480.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: An improved scoring algorithm. *Journal of Personality and Social Psychology, 85*(2), 197–216.

Greenwald, A. G., Nosek, B. A., & Sriram, N. (2006). Consequential validity of the Implicit Association Test: Comment on the article by Blanton and Jaccard. *American Psychologist, 61*(1), 56–61.

Grumm, M., & von Collani, G. (2007). Measuring Big-Five personality dimensions with the implicit association test-Implicit personality traits or self-esteem? *Personality and Individual Differences, 43*, 2205–2217.

Han, H. A., Czellar, S., Olson, M. A., & Fazio, R. H. (2010). Malleability of attitudes or malleability of the IAT? *Journal of Experimental Social Psychology, 46*, 286–298.

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the implicit association test and explicit self-report measures. *Personality and Social Psychology Bulletin, 31*(10), 1369–1385.

Karpinski, A., Steinman, R. B., & Hilton, J. (2005). Attitude importance as a moderator of the relationship between implicit and explicit attitude measures. *Personality & Social Psychology Bulletin, 31*(7), 949–962.

Mastropolo, F., & Varney, A. (Producers) (2006). Race & Sex: what we think but don't say/ABC News. Originally broadcast as a segment on the television program 20/20 on September 15, 2006.

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2007). The implicit association test at age 7: A methodological and conceptual review. In J. A. Bargh (Ed.), *Social psychology and the unconscious: The automaticity of higher mental processes* (pp. 265–292). New York: Psychology Press.

Nosek, B. A., & Hansen, J. J. (2008). The associations in our heads belong to us: Searching for attitudes and knowledge in implicit evaluation. *Cognition and Emotion, 22*, 553–594.

Project Implicit (2010). Origins and measurement with IAT. Retrieved from the IAT web site. <https://implicit.harvard.edu/implicit/demo/background/posttestinfo.html>.

Rezaei, A. R., & Katz, L. (2004). An Evaluation of the reliability and validity of the cognitive styles analysis inventory and recommendations for improvement. *Journal of Personality and Individual Differences, 36*, 1317–1327.

Schmukle, S. C., & Egloff, B. (2005). A latent state-trait analysis of implicit and explicit personality measures. *European Journal of Psychological Assessment, 21*, 100–107.

Steffens, M. C., & Buchner, A. (2003). Implicit association test: Separating transsituationally stables and variable components of attitudes toward gay men. *Experimental Psychology, 50*(1), 33–48.

Vedantam, S. (2010). *The hidden brain: How our unconscious minds elect presidents, control markets, wage wars, and save our lives*. New York: Spiegel & Grau.

Ventis, W. L., Ball, C. T., & Viggiano, C. (2010). A Christian humanist implicit association test: Validity and test–retest reliability. *Psychology of Religion and Spiriuality, 2*(3), 181–189.

von Stülpnagel, R. (2010). Prejudiced or just smart? Intelligence as a confounding factor in the IAT effect. *Zeitschrift für Psychologie, 218*(1), 51–53.