Measures of Implicit Attitudes

Ali H. Al-Hoorie
Published online on: 31 May 2022


Please scroll down for document
Overview

In the 19th century, Donders (1868/1969) argued that although psychologists could not directly observe mental processes, they could still infer them through performance speed in response to different stimuli. Donders described several experiments reporting different latencies depending on, for example, whether an object was placed to the right or left and whether the participant was using their right or left hand. A century later, researchers started investigating automatic stereotypes (Gaertner & McLaughlin, 1983) and automatic attitudes (Fazio et al., 1986) through sequential priming tasks. In a sequential priming task, the participant is first presented with a prime stimulus (e.g., Pleasant) and then with a target stimulus (e.g., Rose). The participant is to make a quick decision regarding the second stimulus (e.g., to classify it as Flower or Insect). A priming effect occurs when the similarity between the two stimuli makes the response speed faster than if the first prime stimulus was, for example, Disgusting.

Against this historical backdrop, implicit social cognition (Greenwald & Banaji, 1995) came to hold a central position in the map of the psychological sciences. Implicit social cognition is an area that investigates the extent to which traces of some past experiences may affect behavior unconsciously (or implicitly), in the sense that the individual may not recognize the link between those past experiences and their current behavior, and may instead believe that they are behaving purely rationally. Because of its potential contribution to the understanding of the human mind and human behavior, implicit social cognition has become so influential that “virtually every intellectual question in social psychology, and many outside of it, has been shaped by [its] theories and methods” (Payne & Gawronski, 2010, p. 1).

A fundamental assumption of implicit cognition is that, since some processes are unconscious, direct self-report may not be the best approach to tap into them. This view has led to the development of various “implicit measures” in an attempt to better understand the unconscious mind (Gawronski & De Houwer, 2014; Gawronski & Payne, 2010; Petty et al., 2009). These measures generally try to uncover implicit cognition indirectly. While a direct measure relies on the participant’s own evaluation of the attribute in question (e.g., through a questionnaire or an interview), “a signature feature” (Bar-Anan & Nosek, 2014, p. 668) of indirect measures is that social cognition is inferred from the response speed on a specially designed task.

Psychologists have developed different families of implicit measures. One important measure is the Implicit Association Test (IAT; Greenwald et al., 1998). This test has also been adapted and extended; examples of its adaptations include the Single-Target Implicit Association Test.
Measures of Implicit Attitudes

(Wigboldus et al., 2004), the Single-Category Implicit Association Test (Karpinski & Steinman, 2006), and Single-Attribute Implicit Association Test (Penke et al., 2006). These variants involve only one target category in order to avoid certain limitations of the standard IAT (see later in this chapter).

In this chapter, I will focus specifically on the IAT. According to one survey (Nosek et al., 2011a), the IAT accounts for about half of all research using indirect measures. Some of this research has also targeted language learners and teachers (Al-Hoorie, 2016a, 2017a, 2017b, 2019; Harrison & Lakin, 2018a, 2018b; McKenzie & Carrie, 2018). This chapter describes the IAT, its psychometric properties, and its contribution to research, before making suggestions for future language research.

**Technical Features**

**Description of the IAT**

The IAT is a computerized, reaction-time measure. It is flexible and can be adapted to infer implicit cognition in different domains, such as gender stereotypes (male–work vs. female–home), political attitudes (Democrat–pleasant vs. Republican–unpleasant), and self-esteem (me–good vs. not me–bad). As reviewed below, the IAT has also been extended to investigate language learners’ implicit attitudes toward L2 speakers as well as teachers’ implicit attitudes toward certain learner minorities.

To perform the IAT, the participant sees a series of stimuli in the middle of the screen. Depending on the specific instructions they read, their task is to decide whether each belongs to the right or to the left by clicking on one of two designated keyboard buttons. They are asked to perform the task as fast as possible as their reaction time will be measured.

Before the test starts, the participant is given the opportunity to practice and become familiar with the test procedure. First, a number of stimuli (e.g., Roses or Cockroaches) may appear on the screen to classify as either Flower or Insect. Next, another series of stimuli to classify as either Good or Bad (e.g., Peace, Happy, Disease, Vomit) appears. The classification has to be correct. If Roses is classified as belonging to Insect, a red X appears on the screen. All stimuli may also be shown to the participants in advance so that their task is to classify each stimulus correctly, not to guess which category it should belong to.

Then the test starts. The two tasks above are combined into one. On the left of the screen, the participant may find Flower and Good, while Insect and Bad are on the right (see Figure 30.1A). The participant’s task is still the same, to classify each stimulus to the right or left. Until this stage, most participants find this test easy to perform. If the stimulus is either a pleasant adjective or a type of

![Figure 30.1](image-url) The Two Parts of the Implicit Association Test.

467
flower (i.e., all nice words), press the left button; if the stimulus is either an unpleasant adjective or a type of insect (i.e., repulsive words), press the right button.

The participant then moves to the second part of the test. Flower is now paired with Bad while Insect with Good (see Figure 30.1B). Suddenly, the task becomes difficult. Performance speed generally decreases and errors become more frequent. This is because Flower and Bad can no longer merge into an intuitive superordinate category (e.g., nice words). Nor do Insect and Good clearly share, for many people, a common higher-order attribute. In other words, the “association” between Flower and Good is stronger than that between Flower and Bad (hence the Implicit Association Test). Similarly, Insect associates more readily with Bad (i.e., faster performance) compared to Good (slower performance).

The test typically has 60 trials in each of these two parts, taking about ten minutes to complete. The reaction times of each participant are then used to compute an overall score. This score is based on the difference between the average speed in the first part (i.e., compatible block) and the second part (incompatible block). A score close to zero indicates a neutral position while a higher positive or negative score indicates a stronger association between the two categories for which faster reaction time was obtained.

**Psychometric Properties of the Implicit Association Test**

As the IAT is one of the most frequently used measures of implicit cognition, there has become an empirical basis to conduct synthetic research on its reliability and validity. Starting with reliability, research by Hofmann et al. (2005) showed that the IAT had an average reliability (internal consistency and split-half) of 0.79, while test–retest reliability was 0.51. In another, large-scale study by Bar-Anan and Nosek (2014), the IAT was shown to have an average internal consistency reliability of 0.88. These are some of the highest figures among implicit measures (see also Krause et al., 2011). In contrast to reliability, there is less consensus about the validity of the IAT. Validity refers to whether the instrument actually measures what it intends to measure, in this case implicit cognition.

In terms of construct validity, some researchers argued that IAT scores may represent shared cultural stereotypes rather than bias at the individual level (e.g., Arkes & Tetlock, 2004; Gehring et al., 2003; Karpinski & Hilton, 2001; though see Banaji et al., 2004). Another consideration in the IAT is that its scores are relative. In the Flower–Insect test above, a high score (using the IAT D score formula) may suggest that one has a positive attitude toward flowers and a negative attitude toward insects. However, logically, this score could mean that one may already have a positive attitude toward insects, but a more positive attitude toward flowers. A neutral score may also mean that the respondent likes the two categories equally, dislikes them equally, or is simply indifferent. This ambiguity led some researchers to develop variations of the standard IAT that involve one target category only (see Bar-Anan & Nosek, 2014; Nosek et al., 2011a). Nevertheless, findings by Bar-Anan and Nosek (2014) suggest that single-category attitude scores computed from the standard IAT still exhibit good relationships with other measures of implicit attitudes, a finding that points toward the validity of the standard IAT.

Some insight into the convergent and discriminant validity between the IAT and self-report measures comes from neuroscience. Findings by Cunningham et al. (2004) show that explicit attitudes are associated with activation in the frontal cortex, which is concerned with control and regulation. In contrast, implicit attitudes are correlated with activation in the amygdala, which is the brain area responsible for emotions (see also Cunningham et al., 2003; Phelps et al., 2000). Such results have been used to argue that implicit attitudes are not mere “cold cognition” and that “the IAT was capturing something warm and affect-laden” (Dasgupta et al., 2003, p. 241). Along the same lines, Izuma et al. (2018) showed that there was a robust association between neural signals in the reward-related regions with implicit, but not explicit, self-esteem, “thus providing unique evidence for the validity of the self-esteem IAT” (p. 349).
Measures of Implicit Attitudes

Following a different approach, Schimmack (2020) used confirmatory factor analysis in order to revisit convergent and discriminant validity. He reanalyzed data from several studies that used both explicit and implicit measures. He found that 64% of the variance in the IAT scores represents valid variance in political orientation, 20% in racial preference, and only 14% in implicit self-esteem. This pattern of results suggests that understanding the “self” is a much harder and more complex endeavor than either political or racial attitudes. Further analyses showed that IAT scores demonstrated incremental validity in relation to both the political and race domains (Vianello & Bar-Anan, 2020). Incremental validity here means that implicit attitudes were able to explain variance in meaningful outcomes in these two domains over and above self-reported explicit attitudes.

When it comes to predictive validity, some evidence stems from known-groups validity, where the IAT was able to reliably predict the participant’s gender, political preference, nationality, and even affiliation to a group researchers artificially created in the laboratory (for a review, see Lane et al., 2007). Other psychologists argue that this predictive validity is rather modest, however. According to one meta-analysis, IAT scores predicted relevant criterion measures from different domains at an average correlation of 0.27 (Greenwald et al., 2009). In a subsequent meta-analysis, the average correlation dropped further to 0.14 (Oswald et al., 2013). Some critics have used these results to argue that “IAT studies often rely on small-sample studies that predict criteria far removed from meaningful forms of discrimination (e.g., performance on a Stroop test or amygdala activation) in artificial situations involving strangers” (Oswald et al., 2015, p. 565). The meta-analysis by Oswald et al. (2013) was in turn criticized by Greenwald et al. (2015) for including correlations that had no theoretical basis, thus suppressing the correlation. Certain methodological features, such as higher implicit–criterion correspondence (i.e., similarity between IAT content and predicted variables), also influenced the magnitude of the correlation, raising it to 0.37 (Kurdi et al., 2019).

Finally, Greenwald et al. (2015) further argued that a small magnitude per se is not sufficient grounds to dismiss findings. Some small effects can be of substantial societal significance over time (though see Oswald et al., 2015). This notion is related to what Messick (1995) referred to as consequential validity and to what Merton (1968) called the Matthew effect, according to which the poor get poorer while the rich get richer. In other words, even if the amount of bias (e.g., toward minorities) is small, it may lead to a substantial disadvantage in the long run. Considering such findings from different domains, some proponents of the IAT have argued that there is mounting evidence in favor of its validity, with some describing it as a “scientific certainty” (Rudman, 2008).

Contributions to ID Research

Research in Educational Psychology

A major stream of research in educational psychology centers on teachers’ unintentional bias and its influence on student behaviors and learning achievement. Teaching is a stressful job. The typical teacher is expected to deal with several groups of students, typically with different ability levels, while multitasking various responsibilities and trying to meet never-ending deadlines. Teachers also have to constantly make split-second decisions while dealing with students during a lesson, a situation that constrains thoughtful reflection; thus, implicit and automatic attitudes and biases can creep in even for well-meaning teachers (Al-Hoorie et al., 2021; Piit-ten Cate & Glock, 2019). The gravity of the teacher’s subtle biased behaviors toward different students was first illustrated in a seminal study by Rosenthal and Jacobson (1968). The researchers called the phenomenon they studied the Pygmalion effect after a sculptor who, according to Greek mythology, fell in love with one of his statues that then came to life. This metaphor symbolizes the notion that a teacher’s expectations of a student, even if false, can materialize through unintended mediating processes. For example, the teacher who (falsely) believes that a student is bright/dull might spontaneously engage in certain behaviors, interaction patterns, and feedback.
modes that would reflect this false belief. The student may in turn unconsciously internalize this belief, behave accordingly, and eventually manifest these expectations in socio-psychological variables, achievement, and other learning outcomes (see Wang et al., 2018 for a systematic review of evidence). Hattie (2009) estimated that teacher expectancy had an average effect size of $d = 0.43$.

A similar picture appears from research using the IAT. Teachers may have implicit biases favoring majority over minority students, and these biases might be communicated to their students in subtle, unintended ways that could disadvantage these minorities. In one study by van den Bergh et al. (2010), the researchers used the IAT to examine the relationship between the implicit prejudice of teachers and the achievement of their students. The results showed that the higher the teacher’s implicit prejudice, the lower the achievement of their minority students—and sometimes the higher the achievement of majority students—in mathematics and text comprehension tests. The results also suggested that this effect may be mediated by teachers’ expectations. Teachers whose IAT scores indicated high implicit prejudice tended to view minority students as less intelligent and less promising. Not surprisingly, explicit measures of prejudice did not show any of these associations. Another study by Peterson et al. (2016) reached comparable conclusions. The researchers used the IAT to examine teacher implicit prejudiced attitudes and their relationship with minority student achievement in reading and mathematics. Overall, the results showed that, for reading, higher explicit expectations were associated with higher achievement. For mathematics, teachers’ implicit, but not explicit, expectations predicted achievement. According to a meta-analysis by Pitten Cate and Glock (2019), teachers’ implicit attitudes tend to favor non-marginalized groups, with an average effect size of $d = 0.56$.

Besides teachers, students might also have their own implicit biases. Being a student can similarly put pressure on them in relation to task engagement, course and major selection, and all the way to educational and career pathways. Nosek et al. (2002) reviewed evidence showing that research relying on explicit self-reports has not been able to explain certain educational patterns, such as the overrepresentation of males in the sciences and females in the arts. Nosek et al. (2002) further argued that responses to self-report measures might represent conscious belief in egalitarianism, while actual behavior might be guided more by implicit processes. Decisions people make “appear to reflect a free and individually determined choice when in fact they reflect group membership, the strength of identity with the group, and beliefs about the capability of the group” (Nosek et al., 2002, p. 44), thus making free will seem an illusion (Al-Hoorie, 2015). Using the IAT, Nosek and Smyth (2011b) showed that implicit stereotypes had stronger predictive validity than explicit stereotypes in explaining this gender gap. That is, women who exhibited stronger math = male stereotypes also had greater negativity toward math, weaker self-ascribed ability, less participation, and worse math achievement. In another study, Lane et al. (2012) used the IAT to examine the relationship between gender and academic participation in the sciences. They found that implicit stereotypes mediated the link between gender and pursuit of science over the humanities. In fact, once implicit stereotypes were controlled for, gender no longer predicted students’ academic plans, indicating that the gender gap was completely accounted for by implicit stereotypes in their sample.

Some research has examined the possibility of modifying these implicit biases. In a series of studies, Lai et al. (2014) tested the effectiveness of 17 theoretically grounded interventions in reducing implicit racial prejudice. They found nine of them could significantly reduce implicit bias. Interventions that were effective tended to involve multiple mechanisms. For example, the single most effective intervention in this study was using a Vivid Counterstereotypic Scenario. In this intervention, the participant reads an evocative second-person narrative placing the participants at the center of the story. The participant is told that a White man is assaulting the participant while a Black man rescues him/her. The participant is then explicitly told that the moral of this story is to affirm that White = Bad and Black = Good in order to reduce pro-White implicit bias. The more vivid the description, the stronger the effect. Lai and colleagues used the following narrative:
With sadistic pleasure, he beats you again and again. First to the body, then to the head. You fight to keep your eyes open and your hands up. The last things you remember are the faint smells of alcohol and chewing tobacco and his wicked grin.

(Lai et al., 2014, p. 1771)

The presence of multiple mechanisms (personal involvement of the participant, highly threatening life-and-death situation, counterstereotypical exemplars, and specific strategies to overcome bias) has apparently made this intervention highly effective. The meta-analytic effect size across Lai et al.’s (2014) four studies for this particular intervention was sizable: \( d = 0.49 \).

Nevertheless, successfully manipulating implicit attitudes during an intervention speaks to its variability and not necessarily to its long-term durability. In a subsequent study, Lai et al. (2016) tested the durability of the effects of nine interventions that demonstrated a significant reduction in implicit attitudes in Lai et al. (2014). Surprisingly, virtually none of these interventions remained effective even a few days after the intervention. These results might indicate that implicit attitudes exhibit elastic changes, akin to how a rubber band returns to its original shape once the intervention is complete (Banaji & Greenwald, 2013). Or to quote Lai et al. (2016), such change is “fleeting” (p. 1014).

Other research depicts a more optimistic picture. In contrast to the above lab-based interventions, some naturalistic investigations seem to have obtained more durable effects. For example, taking part in a six-month cultural music education program was effective in reducing implicit bias, a reduction that persisted two years later (Neto et al., 2015). Other research suggests that age is also an important factor in the changeability of implicit attitudes, as there might be an optimal developmental stage for implicit attitudes to change (Gonzalez et al., 2017). Still other research shows that, even if it does not manifest at the implicit level, change at the explicit level does occur as a result of such interventions and can be observed two years later, which might be more important in promoting long-term behavioral change (Forscher et al., 2017).

**Research in Language Learning and Teaching**

When it comes to language learning and teaching, some research has been conducted using the IAT. One line of research investigated the implicit attitudes of teachers in the US, where the student population is becoming more diverse while the teacher population is still predominantly White (Putman et al., 2016). Some of these students have first languages other than English (LOTEs; Dörnyei & Al-Hoorie, 2017), and so they are described as “English-learners”. Harrison and Lakin (2018b) used the IAT to investigate the implicit attitudes of teacher trainees toward mainstream students and English learners. The researchers found that teacher trainees had somewhat positive implicit attitudes toward working with English learners. However, in another study, Harrison and Lakin (2018a) found that the implicit attitudes of in-service teachers toward English learners were negative. One explanation for these contrasting results is that teacher trainees initially lack experience with English learners. Negative implicit attitudes start to form after teacher trainees start interacting with these learners in the field and experiencing firsthand the challenges associated with learners who need extra language support. Another explanation is that the newer generations of teachers have had exposure to a more diverse population than teachers who were students in the 1970s and 1980s. Along these lines, longitudinal research has shown that implicit attitudes are malleable and can change over time, perhaps more so than explicit attitudes (Gawronski et al., 2017), and that this change can sometimes be toward neutrality and sometimes away from neutrality (Charlesworth & Banaji, 2019).

A second line of research looked at the implicit attitudes of language learners. Attitudes of language learners toward the speakers of the target language have been argued to be an important factor determining the extent of success in L2 learning (Al-Hoorie & MacIntyre, 2020; Gardner,
1979, 1985, 2010). However, investigations of learners’ attitudes have mostly relied on self-report measures, notably questionnaires. Using the IAT, Al-Hoorie (2016b) compared the implicit and explicit attitudes toward British people in a sample of Arab learners of English in the UK. He compared learners with “congruent” attitudes (i.e., positive attitudes both at the explicit and implicit levels) with learners with “incongruent” attitudes (i.e., positive explicit attitudes but negative implicit attitudes). The results showed that congruent learners also scored lower in fear of assimilation, ethnocentrism, religious attitudes, and attitudes toward Arabs (i.e., their own in-group). These results suggest that implicit attitudes are related meaningfully to intergroup attitudes, even if explicit self-reports indicate otherwise. Using a variation of the IAT, called the Single-Target IAT, Al-Hoorie (2016c) examined the implicit attitudes of Saudi language learners, most of whom had not visited a foreign country. The results of this study replicated those from Al-Hoorie’s (2016b) study in the UK. They additionally showed that learners with more favorable implicit attitudes toward L2 speakers also achieved higher grades in their English classes. Furthermore, one feature of this study was that it involved two implicit measures: one toward L2 speakers and the other toward their L2 course. Explicit attitudes toward L2 speakers were related to implicit attitudes toward L2 speakers, but not to implicit attitudes toward the L2 course. This pattern seems to provide further evidence of the construct validity of implicit attitudes.

**Future Directions**

**Developing L2-Specific Implicit Measures**

This chapter has focused on one measure, the IAT, because it is the most commonly used measure in the implicit cognition literature. It also exhibits superior psychometric properties relative to most other implicit measures (Bar-Anan & Nosek, 2014). However, other implicit measures might shed additional light on different aspects of individual differences (for comparative analyses, see Bar-Anan & Nosek, 2014; Bosson et al., 2000). Language learning research could adapt some of these measures or devise completely new ones along their lines.

Implicit measures do not have to be limited to reaction time. For example, an early approach attempting to uncover implicit language attitudes without relying on reaction time is the Matched-Guise Technique (Lambert et al., 1960). In the standard approach, the participant is asked to listen to two individuals reading the same text in two different languages (or accents). The participant is then asked to rate the personality of the speaker based on their voice, just like people usually do when they listen to someone on the phone or on the radio. However, the catch is that the two “speakers” reading the same text are actually one person fluent in the two languages. If the participant evaluates the two differently, this might reflect a bias or a stereotype regarding the linguistic community in question. In one study using this technique, Giles (1971) found that speakers of the received pronunciation were rated as more prestigious (e.g., intelligent, ambitious) while Welsh speakers were rated as more socially attractive (e.g., humorous, good-natured). A helpful guide to designing language experiments using the match-guise technique is provided by Kircher (2016).

Another popular paradigm of implicit measures that do not rely on reaction measures is the Affect Misattribution Procedure (Payne et al., 2005). In this procedure, participants unfamiliar with Chinese are first briefly presented with a stimulus (e.g., Flower or Insect) and then a Chinese character. Their task is to indicate whether they consider this Chinese character to be visually more pleasant or less pleasant than the “average” Chinese character. The results show that individuals tend to evaluate it more positively after a pleasant prime, and more negatively after an unpleasant prime. Interestingly, this effect persists even when participants are informed in advance about the logic of this procedure and when they are explicitly instructed not to let the prime influence their decision. More recently, Cummins and De Houwer (2019) provided initial validation for a relational variant of this test, which they called the Truth Misattribution Procedure.
Measures of Implicit Attitudes

Such measures could be used to ask interesting questions related to individual differences in language learning and teaching. For example, research could investigate attitudes toward World Englishes, whether some varieties are more favorably appraised than others, and whether this is manifested at the implicit or explicit level (or both). Research might also consider cultural differences to determine if learners and teachers from different backgrounds have different implicit attitudes (e.g., due to political factors), if these differences are related to language learning and teaching, and if it is possible to mitigate their effects. For instance, negative implicit attitudes might reflect on learner behaviors and preferences, such as aversion to target language music and other cultural products.

Conducting Implicit Attitudes Interventions

Another important future direction for implicit measures is intervention research. Considering that investigation of implicit attitudes is a relatively new area, most of the research conducted to date has been correlational and observational. Intervention research would inform us whether it is possible to, first, change implicit attitudes (e.g., reduce implicit bias against L2 speakers or minority students) and, second, whether this change plays a role in the motivation, language processing, and eventual language learning success (Gardner & Tremblay, 1994). As reviewed above, intervention research that has targeted implicit attitudes makes a distinction between malleability and long-term change. Malleability refers to variability and sensitivity to contextual features; long-term change research is concerned with whether such variability is temporary and whether implicit attitudes revert to their original state once the intervention is over (Lai et al., 2013).

It is possible to see the relevance of this research for language learning. One fertile area for research is the effect of study-abroad programs (DuFon & Churchill, 2006). Many language learners participate in exchange programs or travel to other countries for relatively long periods. Such research could examine changes in implicit attitudes resulting from the study-abroad experience, the durability of this change, whether it varies by factors such as age, and whether it is related to motivation and success in language learning. Like learners, many language teachers also travel to teach abroad for extended periods. Little is known about how this experience influences these teachers’ attitudes either at the explicit or implicit level. Another avenue for research is capitalizing on the short-term malleability of implicit attitudes. For example, research could examine whether the use of the short-lasting implicit interventions during warm-up activities at the beginning of a class has an impact on the dynamics of the lesson, attention and motivation, and student uptake from the lesson.

References

Ali H. Al-Hoorie


474
Measures of Implicit Attitudes


Ali H. Al-Hoorie


