Psychology of Sport and Exercise 31 (2017) 70–78

Contents lists available at ScienceDirect

Psychology of Sport and Exercise

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

Measuring implicit attitudes toward physical activity and sedentary behaviors: Test-retest reliability of three scoring algorithms of the Implicit Association Test and Single Category-Implicit Association Test

Guillaume Chevance [a, b, *], Nelly Héraud [b], Agata Guerrieri [b], Amanda Rebar [c], Julie Boiché [a]

[a] Laboratory Epsylon "Dynamics of Human Abilities and Health Behaviors", University of Montpellier, Montpellier, France
[b] Les Cliniques du Souffle®, Groupe 5 Santé, Lodève, France
[c] Central Queensland University, School of Human, Health, and Social Sciences, Rockhampton, Queensland, Australia

ARTICLE INFO

Article history:
Received 18 October 2016
Received in revised form
10 April 2017
Accepted 10 April 2017
Available online 12 April 2017

Keywords:
Implicit measure
Indirect measure
Automatic evaluation
Dual process

ABSTRACT

Objectives: The Implicit Association Test (IAT) and the Single-Category IAT (SC-IAT) are two frequently used measures of implicit attitudes. Nonetheless, the test-retest reliability of these measures has not been investigated. The aim of this study was to examine the test-retest reliability of a physical activity *versus* sedentary behavior IAT, a physical activity SC-IAT, and a sedentary behavior SC-IAT.
Method: A total of 111 older adults living with chronic diseases were recruited. They either completed a physical activity *versus* sedentary behavior IAT ($N = 54$) or two independent SC-IATs of physical activity and sedentary behavior ($N = 57$). These tests were administered twice in a one-hour interval. Three scores were computed for each test (*D-Score*, *DW-Score*, *IP-Score*). Both absolute and relative test-retest reliability was computed.
Results: Regarding absolute reliability, the tests were comparable regardless of the scoring algorithm (Coefficients of Repeatability ranged from 1.27 for the two SC-IATs with the *D-Score*, to 1.36 for the IAT with the *D-Score* and *DW-Score*). Regarding relative test-retest reliability, the IAT systematically showed better reliability than the two SC-IATs. The *DW-Score* systematically exhibited better reliability compared to other scores (Intraclass Correlation Coefficient ranged from 0.20 for the sedentary behavior SC-IAT with the *D-Score* to 0.78 for the IAT with the *DW-Score*).
Conclusion: Adequate test-retest reliability for the IAT was supported independently from the scoring algorithms. Test-retest reliability for the two independent SC-IATs was not supported in this study. The IAT is more sensitive to change than the SC-IATs, which needs to be accounted for in future research on physical activity and sedentary behavior implicit attitudes.

1. Introduction

Contemporary theories in social and health psychology focus on psychological processes described as either controlled or automatic, depending on their degree of awareness, intentionality, efficiency and controllability (Bargh, 1994; Hofmann, Friese, & Wiers, 2008; Moors & De Houwer, 2006; Strack & Deutsch, 2004). In the physical activity context, past studies have been mainly investigated controlled behavioral precursors (Gourlan, Bernard, Bortolon, Romain, Lareyre & Carayol, 2015). Nonetheless, the total amount of physical activity variance explained by these determinants appears limited (Rhodes & de Bruijn, 2013), and investigating the automatic antecedents of physical activity has become increasingly more popular in the last few decades (Rebar et al., 2016; Sheeran et al., 2016). Currently, there is a call for more studies experimentally manipulating automatic processes; however the usefulness of these efforts will hinge on evidence of good psychometric properties of the measures of these illusive constructs (Rebar et al., 2016).

1.1. Implicit attitudes and physical activity

Amongst the different automatic processes, implicit attitudes

* Corresponding author. Laboratoire Epsylon, 4 Boulevard Henri IV, 34 000 Montpellier, France.
E-mail address: guillaumechevance@hotmail.fr (G. Chevance).

http://dx.doi.org/10.1016/j.psychsport.2017.04.007
1469-0292/

have received considerable attention in the literature (Sheeran, Gollwitzer, & Bargh, 2013). Implicit attitudes reflect automatic evaluations of a concept or object as pleasant or unpleasant, which are mostly outside of a person's awareness, and which result in behavioral tendencies toward or away from the concept/object (Chen & Bargh, 1999; Greenwald & Banaji, 1995). To date, research on implicit attitudes has been mostly correlational, showing that implicit attitudes are significantly associated to objectively-measured physical activity (Conroy, Hyde, Doerksen, & Ribeiro, 2010; Rebar, Ram, & Conroy, 2015) and self-reported physical activity in university students (Berry, Spence, & Clark, 2011; Bluemke, Brand, Schweizer, & Kahlert, 2010; Calitri, Lowe, Eves, & Bennett, 2009; Eves, Scott, Hoppé, & French, 2007), as well as self-reported physical activity in obese persons (Chevance, Caudroit, Romain, & Boiché, 2016) and adults with respiratory conditions (Chevance, Héraud, Varray, & Boiché, 2017). In addition, some studies have started to explore the malleability of physical activity implicit attitudes in quasi-experimental designs, such as single group pre- and post-test and post-test only comparison group studies (Antoniewicz & Brand, 2016; Berry, 2016; Hyde, Elavsky, Doerksen, & Conroy, 2012; Markland, Hall, Duncan, & Simatovic, 2015). Recently, a systematic review highlighted that more rigorous experiments manipulating implicit attitudes are required to further understand their role as potential intervention targets for increasing physical activity behavior (Rebar et al., 2016). Prior to achieving this goal, studies investigating the psychometric properties of the methods used to assess this construct are essential. Importantly, satisfactory test-retest reliability appears as a fundamental prerequisite to interpret changes in scores as the reflection of changes in a construct (Weir, 2005).

1.2. Measuring implicit attitudes: the Implicit Association Test

Implicit attitudes have been measured with many different measures (see Gawronski & De Houwer, 2012). The original and most employed measure in the literature is the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998). This test evaluates the relative strength of a person's mentally-held automatic associations of two opposing attributes (e.g., positive and negative) with two opposing conceptual targets (e.g., physical activity *versus* sedentary behavior). During the IAT, participants are required to sort stimuli (i.e., words or images) representing four categories with only two response keys, each assigned to two of the four categories (e.g., physical activity + positive *versus* sedentary behavior + negative; physical activity + negative *versus* sedentary behavior + positive). If two categories are highly associated mentally for that person, the sorting task is expected to be easier when they share the same response key than when they do not. Hence, ease of sorting can be estimated by the speed and/or accuracy of responding (Greenwald, Nosek, & Banaji, 2003).

An important feature of the IAT is that it provides an indication of a relative tendency between two targets (e.g., sedentary behavior *versus* physical activity) and does not enable the measurement of attitudes about independent targets (e.g., only sedentary behavior or only physical activity). For investigations of implicit attitudes toward physical activity, this test feature may be problematic because the concept has no clear opposite (Conroy et al., 2010). Sedentary behavior is frequently employed as a contrast category *versus* physical activity; however much of the field agrees that sedentary behavior (i.e., time spent sitting), and the motivation toward sedentary behavior, should be studied as distinct constructs from physical activity behavior and motivation (Biddle, Mutrie, & Gorely, 2015). Indeed, previous studies highlighted that both explicit and implicit motivational constructs of sedentary behavior could impede physical activity practice independently from physical activity motivation (Cheval et al., 2015; Rhodes & Blanchard, 2008). Therefore it is important to develop specific and independent measures of implicit attitudes for physical activity and sedentary behavior. To this aim, some researchers utilized a variant of the IAT, the Single Category Implicit Association Test (SC-IAT), to assess implicit attitudes toward physical activity or sedentary behavior only (Karpinski & Steinman, 2006). This test is based on the IAT structure, but only includes one conceptual target and two attributes (e.g., physical activity + positive *versus* negative; physical activity + negative *versus* positive), while the IAT includes two conceptual targets and two attributes.

1.3. Internal consistency and test-retest reliability of the IAT and SC-IAT

Satisfactory internal consistency (i.e., reliability of the outcome score of trials within the same test) has been established for both the IAT and SC-IAT in various research fields (Karpinski & Steinman, 2006; Nosek, Greenwald, & Banaji, 2007), including physical activity and sedentary behavior research (Chevance et al., 2016; Rebar et al., 2015). Indeed, there is evidence supporting the internal consistency of both the IAT and SC-IAT, with split-half correlations and Cronbach alphas usually ranging from 0.70 to 0.90 (Karpinski & Steinman, 2006; Nosek et al., 2007). These values are satisfactory according to current standards and are better than those obtained when using other indirect measures (Gawronski & De Houwer, 2012).

Regarding test-retest reliability however, less information is available in the literature. For the IAT, Egloff, Schwerdtfeger, and Schmukle (2005) reported correlations of 0.58, 0.62 and 0.47 between two administrations of an IAT measuring anxiety in one-week, one-month and one-year intervals, respectively (Egloff et al., 2005). Moreover, a review including 20 studies in which IATs were administered twice to the same individuals showed weak to moderate levels of test-retest reliability (r range from 0.25 to 0.69, with a mean of 0.50; Lane, Banaji, Nosek, & Greenwald, 2007). In this review, the time intervals ranged between 10 min and one year, and the length of time between assessments did not significantly impact test-retest reliability (Lane et al., 2007). Concerning the SC-IAT, a study reported significant but modest correlations for two versions of an anxiety SC-IAT ($r = 0.24$; $r = 0.33$) and non-significant correlations between the administrations of two calmness SC-IATs in a 5-month interval (Stieger, Göritz, & Burger, 2010). In the physical activity context, Hyde et al. (2012) reported a significant but low correlation ($r = 0.22$) between two administrations of a physical activity SC-IATs across one week.

In summary, internal consistency of the two tests has been established for the IAT toward physical activity *versus* sedentary behaviors (Chevance et al., 2016) and the SC-IAT toward physical activity (Conroy et al., 2010; Hyde et al., 2012; Rebar et al., 2015). However, with the exception of the physical activity SC-IAT (Hyde et al., 2012), the test-retest reliability of the physical activity *versus* sedentary behavior IAT and the sedentary behaviors SC-IAT remains unknown. Based on previous studies in other research fields, test-retest reliability appears variable for the IAT and weak for the SC-IAT (Egloff et al., 2005; Hyde et al., 2012; Lane et al., 2007; Stieger et al., 2010).

1.4. Measurement error of the IAT/SC-IAT and scoring algorithms

Past research has indicated that implicit attitudes toward physical activity have both stable and time-varying components (Hyde et al., 2012). The stable component of implicit attitudes has been proposed to reflect the impact of early experiences with the behavior (Greenwald & Banaji, 1995), such as physical activity in childhood (Thompson, Humbert, & Mirwald, 2003). On the other hand, the time-varying component of implicit attitudes has been theorized to reflect both true changes in the construct, in part due to recent experiences with physical activity (Bluemke et al., 2010), as well as measurement artifact (Hyde et al., 2012; Rebar et al., 2015). Thus, when a same individual completes an IAT/SC-IAT twice, change in score may reflect true change in implicit attitudes or merely measurement error. Recently, in an effort to reduce construct-irrelevant measurement artifact, new scoring algorithms for the IAT/SC-IAT have been proposed (Rebar et al., 2015; Richetin, Costantini, Perugini, & Schönbrodt, 2015). Indeed, some researchers have postulated that the criticisms of these tests actually concerned the scoring procedure more than the measure itself (Richetin et al., 2015) and that new scoring procedures may improve test-retest reliability for these tests in the physical activity context (Rebar et al., 2015).

Traditionally, the IAT/SC-IAT effect is computed by subtracting mean response time in the two contrasting test blocks, and dividing the result by the pooled standard deviation across both blocks, which results in the *D-score* (Greenwald et al., 1998, 2003). Recent studies have proposed two main alternatives to this traditional scoring procedure. First, Richetin et al. (2015) compared the psychometric properties of more than 400 variants of the algorithm and recommended (a) to replace the 10% fastest and slowest latencies (i.e., 10% winsorizing) instead of trimming at a pre-defined fixed value; and (b) to compute the difference for practice and test trials together, rather than separately. Other than that, the calculation was the same as the traditional *D-score*. This procedure, referred to in this study as the *DW-Score*, exhibited better internal consistency than any of the other tested algorithms (Richetin et al., 2015). Second, using the E-Z diffusion model (Wagenmakers, van der Maas, & Grasman, 2007), it was proposed to decompose the IAT/SC-IAT data, to disentangle construct-irrelevant variability due to measurement error from the relevant variability of the implicit attitudes construct (Klauer, Voss, Schmitz, & Teige-Mocigemba, 2007). In the context of physical activity, Rebar et al. (2015) observed that one score derived from this mathematical model, the *Information Processing* score (*IP-Score*), did not share variability across distinct SC-IATs, suggesting that this indicator would constitute a construct-relevant parameter. Moreover, this score was a significant predictor of objective physical activity behavior, above and beyond the *D-Score*. In an advancement of traditional scoring procedures, these scoring algorithms might reduce measurement artifact; however the test-retest reliability of these scores remains entirely unknown.

1.5. The present study

In past research, the IAT and SC-IAT have been frequently used to assess implicit attitudes in the physical activity context (Berry, 2016; Chevance et al., 2016; Conroy et al., 2010; Hyde et al., 2012; Rebar et al., 2015). Nonetheless, except for the SC-IAT toward physical activity, which has shown weak correlation between two administrations in one week interval (Hyde et al., 2012), the test-retest reliability of these tasks is unknown. Theoretically, the changes in the score between two administrations of these tests may reflect both true changes in the construct as well as measurement error (Egloff et al., 2005; Hyde et al., 2012). Recently, to deal with this issue, researchers have proposed new scoring algorithms for the IAT and SC-IAT (Rebar et al., 2015; Richetin et al., 2015). This study sought to investigate the test-retest reliability of the IAT and the SC-IAT in the context of physical activity and sedentary behavior, for the *D-Score*, *DW-Score* and *IP-Score*. It was hypothesized that (1) the physical activity *versus* sedentary behavior IAT would display better test-retest reliability compared to the two independent SC-IATs of physical activity and sedentary behavior; and that (2) the *DW-Score* and *IP-Score* would display better test-retest reliability than the *D-Score* across all tests.

2. Methods

2.1. Participants

A total of 111 participants were recruited for this study; they were randomly allocated to (*i*) a physical activity *versus* sedentary behavior IAT ($N = 54$), or (*ii*) two SC-IATs toward physical activity and sedentary behavior, respectively ($N = 57$). There were no significant differences in participants across conditions regarding age ($M_{AGE} = 61.62 \pm 5.92$ years *versus* $M_{AGE} = 62.26 \pm 5.50$, $p = 0.62$) or physical fitness [VO_2peak % theoretical: $M_{VO2} = 44.52 \pm 16.93$ *versus* $M_{VO2} = 47.50 \pm 12.53$, $p = 0.43$; Wasserman, Hansen, Sue, Casaburi, & Whipp, 1999]. Participants were recruited during a clinic stay for chronic disease management (i.e., respiratory or metabolic diseases) but were not involved in an acute care procedure. They were not included in the study if they were unable to complete computer-based tests, had a medical contra-indication to exercise, or had a psychiatric disorder that affected their judgment. Participants gave their written consent before enrolling in the study. The study was performed according to the principles of the Declaration of Helsinki. The protocol was approved by the local institutional committee. The study was conducted in France between October 2015 and October 2016.

2.2. Procedure

Tests were performed on a 15′ monitor and computer equipped with the *Inquisit Millisecond 3.0*® software. There are no specific rules to determine an appropriate time interval for test-retest reliability study (Vaz, Falkmer, Passmore, Parsons, & Andreou, 2013). Nonetheless, to prevent potential true changes in implicit attitudes between the two administrations of the tests (Hyde et al., 2012), the time-interval was fixed at one hour to reduce the likelihood that physical activity was performed between administrations. In order to control for potential effects from the context, the complete protocol was conducted in the same room for all participants, with the same experimenter at both times. Last, to control for potential order effect, the two SC-IATs were randomly assigned in a 1:1 ratio.

2.3. Measures

2.3.1. Physical activity versus sedentary behavior IAT

The IAT procedure comprised seven blocks (see Table 1). In block 1, the two conceptual categories “physical activity” and “sedentary behavior” were displayed on the left and right sides of the window. Participants were asked to sort words into either the category “physical activity” or “sedentary behavior”. Each trial consisted of a stimulus appearing in the center of the computer screen which had to be classified into the correct category. The word remained on the screen until the participant made a categorization choice. Participants used the letter “Q” on the left side on the keyboard, and the number “5” on the right side on the numeric keypad to select their category choice for each stimuli. If a word was incorrectly categorized (e.g., the word “run” in the category “sedentary behavior”), an indication (‘X’) appeared on the screen, and the participant had to fix his/her error by pressing the correct response key before going on with the test. In block 2, participants were asked to sort words corresponding to the attributes “positive” or “negative”, displayed in the left and right side of the screen, and following the same procedure as in block 1. In blocks 3 (i.e., practice block) and 4 (i.e.,

Table 1
Structure of the IAT physical activity versus sedentary behavior and the two SC-IATs.

			Category Label	
			Physical activity *versus* sedentary behavior IAT	
Block	Trials	Task	Left key « Q »	Right key « 5 »
1	20	Behavior discrimination	Physical activity	Sedentary behavior
2	20	Attribute discrimination	Negative	Positive
3	20	Attribute + behavior	Physical activity + negative	Sedentary behavior + positive
4	40	Attribute + behavior	Physical activity + negative	Sedentary behavior + positive
5	40	Behavior discrimination (reversed)	Sedentary behavior	Physical activity
6	20	Attribute + behavior (reversed)	Sedentary behavior + positive	Physical activity + Negative
7	40	Attribute + behavior (reversed)	Sedentary behavior + positive	Physical activity + Negative
			Physical activity SC-IAT	
Block	Trials	Task	Left key « Q »	Right key « 5 »
1	24	Attribute discrimination	Negative	Positive
2	24	Attribute + behavior	Negative + Physical activity	Positive
3	72	Attribute + behavior	Negative + Physical activity	Positive
4	24	Attribute + behavior (reversed)	Negative	Positive + Physical activity
5	72	Attribute + behavior (reversed)	Negative	Positive + Physical activity
			Sedentary behavior SC-IAT	
Block	Trials	Task	Left key « Q »	Right key « 5 »
1	24	Attribute discrimination	Negative	Positive
2	24	Attribute + behavior	Negative + Sedentary behavior	Positive
3	72	Attribute + behavior	Negative + Sedentary behavior	Positive
4	24	Attribute + behavior (reversed)	Negative	Positive + Sedentary behavior
5	72	Attribute + behavior (reversed)	Negative	Positive + Sedentary behavior

test block), participants were asked to sort the stimuli corresponding to the four categories combined (e.g., "physical activity" + "positive" in the right side of the screen *versus* "sedentary behavior" + "negative" in the left side of the screen). Block 5 was similar to block 1 but the categories were reversed in position (i.e., if "physical activity" was previously displayed on the right side, the category was next placed on the left side and *vice versa*). In blocks 6 (i.e., practice block) and 7 (i.e., test block), participants were asked to sort stimuli in the four categories combined in a reversed version (e.g., "physical activity" + "negative" in the right side of the screen *versus* "sedentary behavior" + "positive" in the left side of the screen). Following the recommendation from Greenwald et al. (2003), practice blocks comprised 20 trials and test blocks comprised 40 trials. Before starting, participants were told that they would be making a series of category classifications. The instructions were to sort the stimuli as quickly as possible and to make as few mistakes as possible, insisting on the fact that these two parameters were equally important.

Stimuli for the category "positive" and "negative" were: pleasant/unpleasant; happy/sad; favorable/unfavorable; beneficial/harmful (in French: plaisant/déplaisant; joyeux/triste; favorable/défavorable; bénéfique/néfaste). Stimuli selected to represent the conceptual category "physical activity" were: run, walk, hiking, dancing, stairs, swimming, bike, lift, gardening, effort (in French: courir, marcher, randonnée, danser, escaliers, nager, vélo, soulever, jardiner, effort). Stimuli selected to represent the conceptual category "sedentary behavior" were: sitting, armchair, chair, television, reading, computer, couch, lying, desk, read (in French: assis, fauteuil, chaise, télévision, lire, ordinateur, canapé, allongé, bureau, lecture).

2.3.2. Physical activity and sedentary behavior SC-IATs

The SC-IAT procedure comprised 5 blocks (see Table 1). In block 1, participants were required to sort the words corresponding to the attributes "positive" and "negative". In blocks 2 and 3, participants were required to sort the words corresponding to the conceptual target and attributes combined in a practice and critical test block. The block 4 and 5 were similar to blocks 2 and 3 in reversed position. According to current recommendations (Karpinski & Steinman, 2006), practice blocks comprised 24 trials and test blocks 72 trials. Stimuli words as well as the error management strategy and time interval between trials were the same as those used in the IAT (Greenwald et al., 1998).

2.4. Data preparation and scoring procedures

2.4.1. D-score

The *D-score* was calculated following the recommendations provided by Greenwald et al. (2003): (1) latencies quicker than 400 ms and longer than 10 000 ms were eliminated; (2) the difference between the average latencies of the two critical blocks were divided by the pooled SD of the latencies; and (3) the score was computed with critical trials only. Scores were comprised between −2 and +2, with 0 representing a neutral score. For the IAT, positive scores indicated favorable implicit attitudes toward physical activity compared to sedentary behavior. For the two SC-IATs, positive scores represented favorable implicit attitudes toward the targeted behavior.

2.4.2. DW-score

The *DW-score* was calculated following the recommendations provided by Richetin et al. (2015): (1) for each participant, the 10% fastest and slowest latencies were replaced by the last untrimmed latencies for both error and correct responses; (2) the difference between the average latencies of the two critical blocks (i.e., practice and test blocks together) were divided by the pooled *SD* of all the latencies; and (3) the score was computed based on practice and critical trials together. Scores were also comprised between −2 and +2, with similar interpretations to the *D-score*.

2.4.3. IP-score

The *IP-Score* was computed for each critical block according to the EZ-diffusion model algorithm provided by Wagenmakers et al. (2007). The *IP-score* was calculated as the difference of the two test blocks (i.e., without practice blocks), positive values indicating a higher IP-Score during the compatible block, and hence more favorable implicit attitudes toward the targeted behavior. Data preparation was computed following the method used by Klauer et al. (2007) and Rebar et al. (2015). Response times shorter than 100 ms and individuals' outliers based on Tukey's outlier criterion were discarded (i.e., for each block, individuals' interquartile ranges were calculated; response times longer than the individual's third quartile plus 1.5 times their interquartile range in that block, or shorter than their first quartile minus 1.5 times the interquartile range were discarded).

All these scores were computed with the statistical software *R* (R Development Core Team, 2013) using the package *IAT.Score* provided by Richetin et al. (2015) for the *D-score* and the *DW-score*, and the function *SCIAT.Scores* developed by Rebar et al. (2015) for the *IP-Score*.

2.5. Data analysis

According to current recommendations (Vaz et al., 2013; Weir, 2005), indicators of both relative (i.e., consistency of the position of individuals in relation to others in a group) and absolute reliability (i.e., consistency of individuals' scores) were computed.

Relative reliability was estimated using the Pearson's correlation coefficient (r), and Intraclass Correlation Coefficient (ICC). Precisely, the two-way random effects model ($ICC_{2,1}$) was computed in this study with the R package *ICC*. ICCs represent the proportion of variance in a set of scores that is attributable to the true score variance; accordingly, the balance of the variance (1- ICC) is attributable to within-person change and measurement error (Weir, 2005). An ICC of 0.75 or greater is considered as reflecting acceptable reliability (Shrout & Fleiss, 1979).

Absolute reliability was estimated with the repeatability coefficient (CR) using the standard formula (Vaz et al., 2013). The CR quantifies absolute reliability in the unit of the measurement tool, and indicates the value below which the absolute difference between two measurements would lie within 95% CI. Mean biases between Time 1 and Time 2 were computed with a two-tailed paired sample t-tests.

Internal consistency (i.e., split-half reliability) was calculated using the function *SplitHalf* from Richetin et al. (2015) for the *D-Score* and *DW-Score*. As in previous research, the internal consistency of the *IP-Score* was not reported in this study given that the interpretation is not relevant for this scoring procedure (Klauer et al., 2007; Rebar et al., 2015). All other analyses were performed with R version 3.2.2 (R Development Core Team, 2013).

3. Results

3.1. Relative test-retest reliability

Pearson's correlations of the test-retests ranged from $r = 0.19$ for the sedentary behavior SC-IAT with the *D-Score* to $r = 0.80$ for the IAT with the *DW-Score*. All the correlations were statistically significant ($p < 0.05$), except for the sedentary behavior SC-IAT. Correlation coefficients were systematically higher for the IAT compared to the SC-IATs. All the ICCs were statistically significant ($p < 0.05$), except for the *D-Score* for the sedentary behavior SC-IAT ($p = 0.07$). ICCs were systematically higher for the IAT compared to the two SC-IATs, regardless of the scoring procedure. Moreover, the *DW-Score* systematically outperformed the other scoring procedures, regardless of the test. The highest test-retest reliability coefficient was found for the IAT with the *DW-Score* ($ICC_{2,1} = 0.78$, 95% CI [0.65; 0.87]), and the lowest coefficient for the sedentary behavior SC-IAT with the *D-Score* ($ICC_{2,1} = 0.20$, 95% CI [-0.075; 0.43]).

3.2. Absolute test-retest reliability

For the *D-Score* and *DW-Score*, CRs were comparable for all tests and scoring procedure ranging from 1.27 for the two SC-IATs with the *D-Score*, to 1.36 for the IAT with the *D-Score* and *DW-Score*. CRs of the *IP-Score* were also comparable for the IAT and the two SC-IATs. Statistically significant biases ($p < 0.05$) were observed between Time 1 and Time 2 for the IAT and physical activity SC-IAT, regardless of the scoring procedure. This indicates that the magnitude of scores significantly changed between the two administrations. For these two tests, mean biases were systematically negative, indicating a regression to the mean at Time 2. For the sedentary behavior SC-IAT, no statistically significant bias was observed across any of the scoring procedures.

Table 2 presents the indices of both relative (Pearson's correlations and ICC) and absolute (CRs, and mean biases) test-retest reliability for each test (i.e., IAT, physical activity SC-IAT, sedentary behavior SC-IAT) and scoring procedure (i.e., *D-Score*, *DW-Score*, *IP-Score*).[1]

3.3. Internal consistency

Internal consistency was high both for the IAT ($r_{s\ \mathrm{T1\ D\text{-}Score}} = 0.95$, $r_{s\ \mathrm{T2\ D\text{-}Score}} = 0.91$, $r_{s\ \mathrm{T1\ DW\text{-}Score}} = 0.97$, $r_{s\ \mathrm{T2\ DW\text{-}Score}} = 0.96$), and the two SC-IATs (physical activity: $r_{s\ \mathrm{T1\ D\text{-}Score}} = 0.64$, $r_{s\ \mathrm{T2\ D\text{-}Score}} = 0.77$, $r_{s\ \mathrm{T1\ DW\text{-}Score}} = 0.84$, $r_{s\ \mathrm{T2\ DW\text{-}Score}} = 0.90$; Sedentary behavior: $r_{s\ \mathrm{T1\ D\text{-}Score}} = 0.77$, $r_{s\ \mathrm{T2\ D\text{-}Score}} = 0.74$, $r_{s\ \mathrm{T1\ DW\text{-}Score}} = 0.90$, $r_{s\ \mathrm{T2\ DW\text{-}Score}} = 0.86$).

4. Discussion

The aim of this study was to examine the test-retest reliability of a physical activity *versus* sedentary behavior IAT, a physical activity SC-IAT, and a sedentary behavior SC-IAT. For these three tests, three scoring procedures were computed, resulting in the examination of test-retest reliability of nine distinct scores. Results showed that relative test-retest reliability was systematically higher for the IAT compared to the two SC-IATs. Moreover, the *DW-Score* systematically provided better relative test-retest reliability compared to the other scoring procedures. Absolute reliability was comparable between tests and scoring procedures, with the exception that the *IP-Score* referred to a different unit of measurement and thus was difficult to compare with other scores. Statistically significant biases were observed for the IAT and physical activity SC-IAT, with a systematic regression to the mean between the two administrations, regardless of the scoring procedure. No significant biases were found for the sedentary behavior SC-IAT.

4.1. Implicit Association Tests versus Single-Category Implicit Association Test

Regarding relative test-retest reliability, the IAT systematically showed better reliability (i.e., Pearson's r and ICC) than the two SC-IATs, regardless of the scoring algorithm. For the IAT, relative test-retest reliability in this study was generally better than previous studies (Lane et al., 2007). This could be explained by the fact that the present study was specifically designed to assess test-retest reliability, which was not the case in previous studies (Lane et al., 2007). Accordingly, the time between the two test

[1] A table providing Pearson's correlations between all versions of tests and scores at both times of data collection is available in supplementary data.

Table 2
Coefficients of test-retest reliability of three scoring algorithms of the Implicit Association Test and Single Category-Implicit Association Test.

Tests	Scores	Time 1	Time 2	*r*	*P*	Relative reliability					Absolute reliability					
		Mean (*SD*)	Mean (*SD*)			ICC (2,1)	*F*	*P*	CI 95% LB	CI 95% UB	WSV	SEM	CR	Bias	*t*	P
IAT ($N = 54$)	*D*	0.66 (0.61)	0.52 (0.61)	0.75	<0.001	0.73	7.0	<0.001	0.57	0.84	0.25	0.49	±1.36	−0.14	2.42	0.019
	DW	0.77 (0.67)	0.65 (0.70)	0.80	<0.001	0.78	8.7	<0.001	0.65	0.87	0.24	0.49	±1.36	−0.12	2.09	0.041
	IP	0.09 (0.10)	0.06 (0.11)	0.65	<0.001	0.64	4.7	<0.001	0.44	0.77	0.05	0.22	±0.61	−0.3	2.16	0.035
PA SC-IAT ($N = 57$)°	*D*	0.36 (0.28)	0.20 (0.29)	0.33	0.013	0.29	2.0	0.006	0.04	0.50	0.21	0.46	±1.27	−0.16	3.61	<0.001
	DW	0.44 (0.32)	0.26 (0.37)	0.38	0.003	0.34	2.2	0.002	0.09	0.55	0.23	0.48	±1.33	−0.18	3.43	0.001
	IP	0.08 (0.08)	0.05 (0.08)	0.36	0.007	0.33	2.1	0.003	0.09	0.54	0.05	0.23	±0.64	−0.3	2.74	0.008
SB SC-IAT ($N = 57$)°	*D*	0.11 (0.31)	0.09 (0.28)	0.19	0.147	0.20	1.5	0.073	-0.07	0.43	0.21	0.46	±1.27	−0.02	0.31	0.758
	DW	0.14 (0.37)	0.18 (0.35)	0.40	0.002	0.40	2.3	0.001	0.16	0.60	0.22	0.47	±1.30	0.04	−0.65	0.518
	IP	0.02 (0.08)	0.02 (0.08)	0.28	0.036	0.28	1.8	0.017	0.02	0.51	0.05	0.23	±0.64	0	−0.16	0.870

Note: PA = Physical Activity; SB = Sedentary behavior; D = *D-Score*; DW = *DW-Score*; IP = *IP-Score*; Mean = mean values for each IAT/SC-IAT scores depending on the algorithms. ICC (2, 1) = Intraclass Correlation Coefficient; CI 95% LB = 95% Confidence Interval Lower Boundary of the ICC; CI 95% UB = 95% Confidence Interval Upper Boundary of the ICC; WSV = Within Subject Variability; SEM = Standard Error Measurement ($\sqrt{}$ WSV); CR = 2,77 x SEM; Bias = mean difference between the two administration; ° = Same sample of participants.

administrations was short (one hour), reducing the risk of measurement errors due to environmental changes, for example (Gschwendner, Hofmann, & Schmitt, 2008). Moreover, given that implicit attitudes are theorized as a malleable construct (Hyde et al., 2012; Schmukle & Egloff, 2004), the time interval of this study may have prevented true change in implicit attitudes during the two administrations, and thus better test-retest reliability. The lower correlations observed for the two SC-IATs were also in line with results from past literature (Hyde et al., 2012; Stieger et al., 2010). Moreover, ICCs observed for the IAT were close to or higher than the acceptable threshold of reliability proposed in the literature (ICC > 0.75, Shrout & Fleiss, 1979), which was not the case for the SC-IATs.

Regarding the discrepancy in relative test-retest reliability between tests, post-hoc analyses indicated systematically lower response times for correct responses and percentage of errors for the SC-IATs compared to the IAT, which is likely a by-product of the characteristics of these tests. The SC-IAT only has three lexical categories to sort, whereas the IAT has four; therefore the SC-IAT is simpler to complete than the IAT. Thus, it could be that the SC-IAT is more subject to learning effects and hence to construct-irrelevant variability than the IAT. In future studies, it is important to test the impact of learning effects on psychometric properties of these two tests, and to test whether learning effects differ across different time intervals between assessments.

Regarding absolute test-retest reliability, CRs were comparable between tests, suggesting that the IAT/SC-IAT may not be appropriate to measure change on an individual level (see Fiedler, Messner, & Bluemke, 2006; Gray, MacCulloch, Smith, Morris, & Snowden, 2003). In previous research it has been proposed that the IAT may have clinical utility (Gray et al., 2003); accordingly, a physical activity coach could be interested in discussing change in implicit attitudes toward physical activity with a client or patient in a program. However, results from the present study suggest that a clinician or coach using an IAT or SC-IAT, in a context similar to the one of this study, would have to observe a change of around 1.30 units to be 95% confident that an individual has experienced true change in implicit attitudes. Considering that the *D-Score* and *DW-Score* derived from an IAT and SC-IAT are confined between [-2; +2], a change of 1.30 represents an unlikely large magnitude of change. Thus, this result questions the clinical utility of those tests to describe individual change in implicit attitudes (see Teachman, Cody, & Clerkin, 2010). Regarding mean biases (i.e., differences in mean scores between the two administrations), results exhibited significant regression to the mean for the IAT and physical activity SC-IAT between the first and the second administration, resulting in more neutral implicit attitudes at Time 2 compared to Time 1. These results are in line with previous research indicating that the magnitude of the IAT effects can be weakened by a prior experience with the tests (Greenwald et al., 2003). This phenomenon suggests that, in an experimental study aiming to enhance implicit attitudes toward physical activity or attitudes of physical activity compared to sedentary behavior, there is a risk of underestimation biases (i.e., underestimation of the score at Time 2). However, this result was not observed regarding the sedentary behavior SC-IAT and was not observed in a previous study investigating physical activity SC-IAT change (Hyde et al., 2012). Additional research is necessary to better understand this issue and to determine to what extent this phenomenon could differ between physical activity and sedentary behavior implicit attitudes.

Finally, the discrepancy between accurate relative test-retest reliability (i.e., consistency of the position of individuals compared to others in a group) and inadequate absolute test-retest reliability (i.e., consistency of individuals' scores) for the IAT could have implications for future research. Indeed, as previously mentioned in the literature (Fiedler et al., 2006), this discrepancy discourages the creation of groups based on IAT/SC-IAT scores (e.g., participants who are 1 *SD* above the mean *versus* others), and encourages using IAT scores as a continuous variable when analyzing these tests (see Rucker, McShane, & Preacher, 2015). Accordingly, the creation of cut-off scores for the IAT is also not warranted based on results from the present study.

4.2. Scoring comparisons (D-score, DW-score, IP-score)

Regarding the coefficients of absolute test-retest reliability, scoring procedures were comparable across tests. However, regarding relative reliability, the *DW-Score* from Richetin et al. (2015) systematically exhibited better test-retest reliability compared to other scores. In their study, Richetin et al. (2015) showed that this scoring procedure outperformed the *D-Score* 80% of the time regarding validity (i.e., convergence with direct and indirect measures, and predictive validity), and 100% of the time regarding internal consistency. That the *DW-Score* also outperformed the *D-Score* regarding test-retest reliability in this study support those previous findings. These results may be explained, in part, by how outliers are handled. The individualized extreme latency treatment (i.e., 10% winsorizing) of the *DW-score* may be more suitable for the participants of this study (i.e., older adults with chronic diseases) than the traditional cut-off of the *D-Score* validated in young adults.

The *IP-Score* outperformed the *D-Score* regarding the two SC-IATs, but not for the IAT. Moreover, this scoring procedure did not provide better reliability than the *DW-Score*. Considering that the *IP-Score* is based on a diffusion model meant to represent different conceptual elements of decision making (Wagenmakers et al., 2007), we expected to observe strong test-retest reliability coefficients for this score compared to the others. It is possible that the data provided by the IAT and SC-IAT do not totally fit with the assumptions of the EZ-diffusion model, which may have limited its performance in this study. In fact, applying this model to a dataset is conditioned by the distribution of reaction time, the relative speed of errors responses, and the fact that the two alternative responses in the task are a priori equally attractive (Wagenmakers et al., 2007). According to the authors of the model, depending on the nature and the seriousness of the violation to these prerequisites, the results from the EZ-diffusion model should be interpreted with caution. Even if it has been successfully applied to the IAT/SC-IAT in past literature, these conditions were not always checked, and when it was the case, they were sometimes not met (Klauer et al., 2007; Rebar et al., 2015). Other mathematical models have been proposed (see Sherman, Klauer, & Allen, 2010), and some may be more suitable to handle IAT/SC-IAT data. Also, it is important to note that the present results specifically tested test-retest reliability, and do not take into account predictive validity (Rebar et al., 2015).

4.3. Limitations and future directions

The generalization of the study findings needs to be tested in other populations to rule out the influence of construct-irrelevant individual differences (e.g., general processing speed, reaction time). This study was conducted in a sample of older adults living with chronic conditions. Previously, it has been shown that individual differences could influence performances for the IAT (Greenwald & Nosek, 2001; Klauer, Schmitz, Teige-Mocigemba, & Voss, 2010), thus it is possible that population-specific differences impacted the findings. However, it should be noted that the IAT/SC-IAT means of this study were comparable with those previously observed in student populations for the most part (Egloff et al., 2005; Hyde et al., 2012; Lane et al., 2007; Stieger et al., 2010). Secondly, the two tests were conducted within a time span of one-hour. To the best of our knowledge, there are no specific rules to determinate an appropriate time interval for test-retest reliability (Vaz et al., 2013). Given that implicit attitudes toward physical activity were shown to have both stable and time-varying components (Hyde et al., 2012), we assume that in a one-day or one-week interval, it is theoretically possible to observe true change in the construct. Consequently we opted for a one-hour interval in this study to limit real change in implicit attitudes. The medium to long-term stability of implicit attitudes for physical activity and sedentary behavior is an essential unanswered question that needs further investigation (Gawronski, Morrison, Phills, & Galdi, 2017); however it is outside the scope of the present study. Thirdly, post-hoc power analyses (i.e., conducted for Pearson's correlations) revealed a lack of statistical power for the analyses concerning the two SC-IATs (i.e., 1- β range from 0.49 to 0.53 using the *Z* transformation of correlation coefficient, Cohen, 1988). For the aforementioned reasons, it would be interesting to replicate this study in other samples of participants, with varying time intervals and other indirect measures. Indeed, there is wide heterogeneity in the measurement of automatic processes toward physical activity, which is problematic for efforts to synthesize literature and draw theoretical conclusions (Rebar et al., 2016). Thus, this kind of methodological investigation is important in parallel with more theoretical studies and could be extended to other indirect measures.

Different paradigms, such as the Manikin Task (Cheval, Sarrazin, Isoard-Gautheur, Radel, & Friese, 2015), the Extrinsic Affective Simon Task (Calitri et al., 2009), and affective priming paradigm (Eves et al., 2007) have been used to measure automatic processes in the context of physical activity. Nonetheless, to our knowledge, the test-retest reliability of these measures remains unknown. It will be important for future research to test the test-retest reliability of implicit measures which provide independent measures of implicit attitudes for physical activity and sedentary behavior, given that these are independent behaviors with unique motivational profiles (Biddle et al., 2015). Indeed a benefit of SC-IATS is that the outcome scores represent attitudes of a single construct as opposed to the relative attitudes measured by IATs. However, the findings of the present study revealed that the two SC-IATs are not psychometrically-sound for measuring change in implicit attitudes of physical activity and sedentary behavior independently. The creation and validity testing of unique measures of implicit attitudes of physical activity and sedentary behavior is an imperative avenue of future research.

Finally, the use of the IAT/SC-IAT to study implicit attitudes requires more critical construct validity testing (see Fiedler et al., 2006). Indeed, IAT/SC-IAT scores reflect responses that are the result of a variety of processes which are not all entirely implicit, such as person's desire and capacity to overcome tasks along the test (Conrey, Sherman, Gawronski, Hugenberg, & Groom, 2005). Moreover, a series of studies recently demonstrated that people could accurately predict their own implicit attitudes toward different social groups, regardless of their explicit attitudes, which refutes the hypothesis that these tests are reflections of entirely automatic processes (Hahn, Judd, Hirsh, & Blair, 2014). Automaticity is characterized in terms of unawareness, non-intentionality, efficiency and non-controllability (Bargh, 1994). Although the IAT/SC-IAT appear more suitable to investigate automaticity compared to self-report questionnaires, it should be noted that these tasks do not provide process-pure measures of implicit associations (Fiedler et al., 2006).

5. Conclusion

Based on previous studies (Egloff et al., 2005; Hyde et al., 2012; Lane et al., 2007; Stieger et al., 2010), it was hypothesized that the IAT of physical activity versus sedentary behavior would display better test-retest reliability compared to the two SC-IATs of physical activity and sedentary behavior. The study results supported this hypothesis, with systematically higher relative test-retest reliability coefficients for the IAT compared to the two SC-IATs. It was also hypothesized that the *DW-Score* and *IP-Score* would display better test-retest reliability than the *D-Score* across all tests. In partial support of the hypothesis, results indicated that the *DW-Score* exhibited slightly better relative test-retest reliability than the *D-Score* and the *IP-Score*. Accordingly, the IAT of physical activity *versus* sedentary behavior appears to be more suitable for assessing implicit attitude change in the physical activity/sedentary behavior compared to the SC-IATs. Moreover, the *DW-Score* could constitute a more reliable alternative to the traditional scoring procedure of the IAT/SC-IAT.

Acknowledgments

The authors wish to thank Sofiane Ramdani and Alain Varray for their advices on the earlier draft on the manuscript. We also thank Giulio Costantini for his help concerning the *DW-Score*. GC was partially supported by a grant in aid from the French Agency for

Research and Technology (ANRT).

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.psychsport.2017.04.007.

References

Antoniewicz, F., & Brand, R. (2016). Learning to like Exercising: Evaluative conditioning changes automatic evaluations of exercising and influences subsequent exercising behavior. *Journal of Sport & Exercise Psychology*. http://doi.org/10.1123/jsep.2015-0125.

Bargh, J. A. (1994). The four horsemen of Automaticity: Awareness, intention, efficiency, and control in social cognition. In R. Wyer, & T. Srull (Eds.), *Handbook of social cognition*. Lawrence Erlbaum.

Berry, T. R. (2016). Changes in implicit and explicit exercise-related attitudes after reading targeted exercise-related information. *Psychology of Sport and Exercise, 22*, 273–278.

Berry, T. R., Spence, J. C., & Clark, M. E. (2011). Exercise is in! Implicit exercise and sedentary-lifestyle bias held by in-groups. *Journal of Applied Social Psychology, 41*(12), 2985–2998. http://doi.org/10.1111/j.1559-1816.2011.00857.x.

Biddle, S. J. H., Mutrie, N., & Gorely, T. (2015). *Psychology of sitting: New kid on the block. Psychology of physical Activity: Determinants, well-being and interventions* (3rd ed.). Routledge.

Bluemke, M., Brand, R., Schweizer, G., & Kahlert, D. (2010). Exercise might be good for me, but I don't feel good about it: Do automatic associations predict exercise behavior? *Journal of Sport & Exercise Psychology, 32*(2), 137–153.

Calitri, R., Lowe, R., Eves, F. F., & Bennett, P. (2009). Associations between visual attention, implicit and explicit attitude and behaviour for physical activity. *Psychology & Health, 24*(9), 1105–1123. http://doi.org/10.1080/08870440802245306.

Chen, M., & Bargh, J. A. (1999). Consequences of automatic Evaluation: Immediate behavioral predispositions to approach or avoid the stimulus. *Personality and Social Psychology Bulletin, 25*(2), 215–224. http://doi.org/10.1177/0146167299025002007.

Cheval, B., Sarrazin, P., Isoard-Gautheur, S., Radel, R., & Friese, M. (2015). Reflective and impulsive processes explain (in)effectiveness of messages promoting physical activity: A randomized controlled trial. *Health Psychology, 34*(1), 10–19. http://doi.org/10.1037/hea0000102.

Chevance, G., Caudroit, J., Romain, A. J., & Boiché, J. (2016). The adoption of physical activity and eating behaviors among persons with obesity and in the general population: The role of implicit attitudes within the theory of planned behavior. *Psychology, Health & Medicine*, 1–6. http://doi.org/10.1080/13548506.2016.1159705.

Chevance, G., Héraud, N., Varray, A., & Boiché, J. (2017). Change in explicit and implicit motivation toward physical activity and sedentary behavior in pulmonary rehabilitation and associations with post-rehabilitation. *Rehabilitation Psychology*.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale,NJ: Lawrence Erlbaum.

Conrey, F. R., Sherman, J. W., Gawronski, B., Hugenberg, K., & Groom, C. J. (2005). Separating multiple processes in implicit social cognition: The quad model of implicit task performance. *Journal of Personality and Social Psychology, 89*, 469–487. http://dx.doi.org/10.1037/0022-3514.89.4.469.

Conroy, D. E., Hyde, A. L., Doerksen, S. E., & Ribeiro, N. F. (2010). Implicit attitudes and explicit motivation prospectively predict physical activity. *Annals of Behavioral Medicine: A Publication of the Society of Behavioral Medicine, 39*(2), 112–118. http://doi.org/10.1007/s12160-010-9161-0.

Egloff, B., Schwerdtfeger, A., & Schmukle, S. C. (2005). Temporal stability of the implicit association test-anxiety. *Journal of Personality Assessment, 84*(1), 82–88. http://doi.org/10.1207/s15327752jpa8401_14.

Eves, F. F., Scott, E. J., Hoppé, R., & French, D. P. (2007). Using the affective priming paradigm to explore the attitudes underlying walking behaviour. *British Journal of Health Psychology, 12*(Pt 4), 571–585. http://doi.org/10.1348/135910706X153775.

Fiedler, K., Messner, C., & Bluemke, M. (2006). Unresolved problems with the 'I', the 'a', and the 'T': A logical and psychometric critique of the implicit association test (IAT). *European Review of Social Psychology, 17*, 74–147.

Gawronski, B., & De Houwer, J. (2012). Implicit measures in social and personality psychology. In H. T. Reis, & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (2nd ed.). New York, NY: Cambridge University Press.

Gawronski, B., Morrison, M., Phills, C., & Galdi, S. (2017). Temporal stability of implicit and explicit measures: A longitudinal analysis. *Personality and Social Psychology Bulletin, 43*(3), 300–312. http://dx.doi.org/10.1177/0146167216684131.

Gourlan, M., Bernard, P., Bortolon, C., Romain, A. J., Lareyre, O., Carayol, M., et al. (2015). Efficacy of theory-based interventions to promote physical activity. A meta-analysis of randomised controlled trials. *Health Psychology Review*, 1–17. http://doi.org/10.1080/17437199.2014.981777.

Gray, N. S., MacCulloch, M. J., Smith, J., Morris, M., & Snowden, R. J. (2003). Violence viewed by psychopathic murderers: Adapting a revealing test may expose those psychopaths who are most likely to kill. *Nature, 423*, 497–498.

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. (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. (2001). Health of the implicit association test at age 3. *Experimental Psychology, 48*(2), 85–93. http://doi.org/10.1026//0949-3946.48.2.85.

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

Gschwendner, T., Hofmann, W., & Schmitt, M. (2008). Differential stability: The effects of acute and chronic construct accessibility on the temporal stability of the implicit association test. *Journal of Individual Differences, 29*(2), 70–79. http://doi.org/10.1027/1614-0001.29.2.70.

Hahn, A., Judd, C., Hirsh, H., & Blair, I. (2014). Awareness of implicit attitudes. *Journal Of Experimental Psychology: General, 143*(3), 1369–1392. http://dx.doi.org/10.1037/a0035028.

Hofmann, W., Friese, M., & Wiers, R. W. (2008). Impulsive versus reflective influences on health behavior: A theoretical framework and empirical review. *Health Psychology Review, 2*(2), 111–137. http://doi.org/10.1080/17437190802617668.

Hyde, A. L., Elavsky, S., Doerksen, S. E., & Conroy, D. E. (2012). The stability of automatic evaluations of physical activity and their relations with physical activity. *Journal of Sport & Exercise Psychology, 34*(6), 715–736.

Karpinski, A., & Steinman, R. B. (2006). The single category implicit association test as a measure of implicit social cognition. *Journal of Personality and Social Psychology, 91*(1), 16–32. http://doi.org/10.1037/0022-3514.91.1.16.

Klauer, K. C., Schmitz, F., Teige-Mocigemba, S., & Voss, A. (2010). Understanding the role of executive control in the implicit association test: Why flexible people have small IAT effects. *Quarterly Journal of Experimental Psychology (2006), 63*(3), 595–619. http://doi.org/10.1080/17470210903076826.

Klauer, K. C., Voss, A., Schmitz, F., & Teige-Mocigemba, S. (2007). Process components of the implicit association test: A diffusion-model analysis. *Journal of Personality and Social Psychology, 93*(3), 353–368. http://doi.org/10.1037/0022-3514.93.3.353.

Lane, K. A., Banaji, M. R., Nosek, B. A., & Greenwald, A. G. (2007). Understanding and using the implicit association test: IV. What we know (so far). In B. Wittenbrink, & N. S. Schwarz (Eds.), *Implicit measures of attitudes: Procedures and controversies* (pp. 59–102). New York: Guilford Press.

Markland, D., Hall, C. R., Duncan, L. R., & Simatovic, J. (2015). The effects of an imagery intervention on implicit and explicit exercise attitudes. *Psychology of Sport and Exercise, 17*, 24–31. http://doi.org/10.1016/j.psychsport.2014.11.007.

Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin, 132*(2), 297–326. http://doi.org/10.1037/0033-2909.132.2.297.

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2007). Automatic processes in social thinking and behavior. In J. A. Bargh (Ed.), *The implicit association test at age 7: A methodological and conceptual review* (pp. 265–292). Psychology Press.

R Core Team. (2013). *A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.Rproject.org/.

Rebar, A. L., Dimmock, J. A., Jackson, B., Rhodes, R. E., Kates, A., Starling, J., et al. (2016). A systematic review of the effects of non-conscious regulatory processes in physical activity. *Health Psychology Review, 0*(0), 1–13. http://doi.org/10.1080/17437199.2016.1183505.

Rebar, A. L., Ram, N., & Conroy, D. E. (2015). Using the EZ-diffusion model to score a single-category implicit association test of physical activity. *Psychology of Sport and Exercise, 16*(3), 96–105. http://doi.org/10.1016/j.psychsport.2014.09.008.

Rhodes, R. E., & Blanchard, C. M. (2008). Do sedentary motives adversely affect physical activity? Adding cross-behavioural cognitions to the theory of planned behaviour. *Psychology & Health, 23*(7), 789–805. http://doi.org/10.1080/08870440701421578.

Rhodes, R. E., & de Bruijn, G.-J. (2013). How big is the physical activity intention-behaviour gap? A meta-analysis using the action control framework. *British Journal of Health Psychology, 18*(2), 296–309. http://doi.org/10.1111/bjhp.12032.

Richetin, J., Costantini, G., Perugini, M., & Schönbrodt, F. (2015). Should we stop looking for a better scoring algorithm for handling implicit association test Data? Test of the role of errors, extreme latencies treatment, scoring formula, and practice trials on reliability and validity. *PloS One, 10*(6), e0129601. http://doi.org/10.1371/journal.pone.0129601.

Rucker, D. D., McShane, B. B., & Preacher, K. J. (2015). A researcher's guide to regression, discretization, and median splits of continuous variables. *Journal of Consumer Psychology, 25*(4), 666–678.

Schmukle, S. C., & Egloff, B. (2004). Does the Implicit Association Test for assessing anxiety measure trait and state variance? *European Journal of Personality, 18*(6), 483–494. http://doi.org/10.1002/per.525.

Sheeran, P., Bosch, J. A., Crombez, G., Hall, P. A., Harris, J. L., Papies, E. K., et al. (2016). Implicit processes in health psychology: Diversity and promise. *Health Psychology, 35*(8), 761–766. http://dx.doi.org/10.1037/hea0000409.

Sheeran, P., Gollwitzer, P. M., & Bargh, J. A. (2013). Nonconscious processes and health. *Health Psychology: Official Journal of the Division of Health Psychology, American Psychological Association, 32*(5), 460–473. http://doi.org/10.1037/

a0029203.
Sherman, J. W., Klauer, K. C., & Allen, T. J. (2010). Mathematical modeling of implicit social cognition. In B. Gawronski, & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 156–175). New York: Guilford Press. The machine in the ghost.
Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*(2), 420–428. http://dx.doi.org/10.1037/0033-2909.86.2.420.
Stieger, S., Göritz, A. S., & Burger, C. (2010). Personalizing the IAT and the SC-IAT: Impact of idiographic stimulus selection in the measurement of implicit anxiety. *Personality and Individual Differences, 48*(8), 940–944. http://doi.org/10.1016/j.paid.2010.02.027.
Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review, 8*(3), 220–247. http://doi.org/10.1207/s15327957pspr0803_1.
Teachman, B. A., Cody, M. W., & Clerkin, E. M. (2010). Clinical applications of implicit social cognition theories and methods. In B. Gawronski, & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 117–139). New York: Guilford Press.
Thompson, A. M., Humbert, M. L., & Mirwald, R. L. (2003). A longitudinal study of the impact of childhood and adolescent physical activity experiences on adult physical activity perceptions and behaviors. *Qualitative Health Research, 13*, 358–377. http://dx.doi.org/10.1177/1049732302250332.
Vaz, S., Falkmer, T., Passmore, A. E., Parsons, R., & Andreou, P. (2013). The case for using the repeatability coefficient when calculating test–retest reliability. *PLoS One, 8*(9), e73990. http://doi.org/10.1371/journal.pone.0073990.
Wagenmakers, E.-J., van der Maas, H. L. J., & Grasman, R. P. P. P. (2007). An EZ-diffusion model for response time and accuracy. *Psychonomic Bulletin & Review, 14*(1), 3–22.
Wasserman, K., Hansen, J. E., Sue, D. Y., Casaburi, R., & Whipp, B. J. (1999). *Principles of exercise testing and Interpretation: Including pathophysiology and clinical applications* (3rd ed.). Philadelphia, Pa, USA: Lippincott: Williams & Wilkins.
Weir, J. P. (2005). Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *Journal of Strength and Conditioning Research/National Strength & Conditioning Association, 19*(1), 231–240. http://doi.org/10.1519/15184.1.