Psychology of Sport & Exercise 43 (2019) 4–15

Contents lists available at ScienceDirect

Psychology of Sport & Exercise

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

Full Length Article

Internal consistency and validity of measures of automatic exercise associations

Zachary Zenko[1,*], Panteleimon Ekkekakis[2]

[1] *Department of Kinesiology, California State University, Bakersfield, USA*
[2] *Department of Kinesiology, Iowa State University, USA*

ARTICLE INFO

Keywords:
Dual-process models
Affect
Implicit
Automatic associations

ABSTRACT

Researchers increasingly recognize that exercise and physical activity may be determined by interacting explicit and implicit processes. While this trend has led to a substantial increase in the use of measures of implicit processes within exercise psychology, such measures are typically selected without a supporting rationale. To facilitate the refinement of theoretical models and the testing of interventions, investigating the internal consistency and validity of measures of implicit processes, such as automatic exercise associations, is crucial.

Objectives: To assess the internal consistency and validity of nine measures of automatic exercise associations.

Method: Participants ($N = 95$) completed an exercise session at the intensity of the ventilatory threshold, intended to generate heterogeneous affective responses. One week later, they also completed nine randomly ordered measures of automatic exercise associations. The slope of affect ratings during exercise, affect ratings at the end of exercise, recalled affect, explicit affective attitude, self-reported exercise, and situated decisions to exercise served as validation criteria.

Results: Three of the nine measures exhibited acceptable to good internal consistency. Only the Approach-Avoidance Task was significantly and meaningfully related to any of the validation criteria (i.e., self-reported exercise and situated decisions to exercise).

Conclusions: Most measures of automatic exercise associations exhibited unsatisfactory internal consistency and were unrelated to the validity criteria. For research on the role of implicit processes in exercise and physical activity to advance, further psychometric evaluation and refinement of measures is needed.

1. Introduction

A notable conceptual advance in the study of exercise and physical activity behavior over the past few years has been a wave of proposals for dual-process theoretical models (Bluemke, Brand, Schweizer, & Kahlert, 2010; Brand & Ekkekakis, 2018; Conroy & Berry, 2017; Ekkekakis, 2017; Ekkekakis & Dafermos, 2012; Ekkekakis & Zenko, 2016; Rebar et al., 2016; Williams & Evans, 2014). Dual-process models incorporate the type of reflective, deliberative, analytic, and explicit processes postulated by cognitivist theories but integrate them into a broader system that also includes nonreflective, nondeliberative, possibly nonrational and nonconscious, implicit processes. Such integrative models can expand the theoretical perspective beyond traditional information-processing models of the mind and acknowledge the unique challenges involved in understanding exercise and physical activity behavior. Specifically, a central feature of dual-process models is the assumption that explicit and implicit processes may result in conflicting motivational tendencies. Thus, as postulated in recent proposals, a possible scenario is that individuals who remain physically inactive may manifest positive explicit processes (e.g., report more perceived benefits than barriers, hold positive explicit attitudes, exhibit strong self-efficacy) but may be held back by a history of unpleasant affective experiences from exercise or physical activity that possibly includes memories of shame, guilt, boredom, pain, or exhaustion (Bluemke et al., 2010; Brand & Ekkekakis, 2018; Conroy & Berry, 2017; Ekkekakis, 2017; Ekkekakis & Dafermos, 2012; Ekkekakis & Zenko, 2016; Williams & Evans, 2014).

Measures of implicit processes have been shown to predict exercise and physical activity behavior beyond the variance accounted for by measures of explicit constructs (e.g., attitudes, intentions), even when behavior was measured prospectively by objective means (e.g., Cheval, Sarrazin, Isoard-Gautheur, Radel, & Friese, 2015; Chevance, Caudroit et al., 2018; Chevance, Stephan, Héraud, & Boiché, 2018; Conroy, Hyde, Doerksen, & Ribeiro, 2010). Moreover, Antoniewicz and Brand

* Corresponding author. Department of Kinesiology, California State University, Bakersfield, 9001 Stockdale Highway, Bakersfield, CA, 93311, USA.
E-mail address: zzenko@csub.edu (Z. Zenko).

https://doi.org/10.1016/j.psychsport.2018.12.005
Received 11 April 2018; Received in revised form 28 November 2018; Accepted 11 December 2018
Available online 21 December 2018

(2016a) found that implicit processes predicted exercise adherence over 14 weeks. These preliminary findings suggest that interventions designed to change relevant implicit processes may hold promise for increasing exercise and physical activity. Indeed, preliminary proof-of-concept studies have shown that short computer-based training interventions can modify implicit processes and, in turn, improve exercise behaviors, such as the choice of exercise intensity or duration (Antoniewicz & Brand, 2016b; Cheval, Sarrazin, Pelletier, & Friese, 2016).

Although these early studies illustrate the importance of implicit processes in regulating behavior and, therefore, the considerable potential of transitioning to dual-process models, problems are also evident. Perhaps the most challenging problem, and arguably the largest possible impediment to future progress, is the measurement of implicit processes. Researchers in exercise psychology have been selecting measures of implicit processes among a wide range of options originally developed for the study of social cognition (e.g., sequential priming tasks, response interference tasks). While all such measures are assumed to assess implicit processes, their essential features exhibit remarkable diversity, such that it would be unreasonable to assume that the measures are interchangeable or that they can all be interpreted as reflecting the same underlying implicit processes. According to Schinkoeth and Antoniewicz (2017):

> In order to decide which measurement tool to use, every researcher has to answer the questions (among others) what exactly should be measured, as well as how and when it should be assessed. Since each measure has characteristic features, the fit between the respective research question as well as the underlying theoretical assumptions and the implicit measure's specific procedure is decisive (p. 13).

An optimistic perspective perhaps could suggest that finding significant associations with physical activity or exercise behavior among a wide range of measures of implicit processes is evidence of the robustness and multifaceted nature of this relationship (Sheeran, Gollwitzer, & Bargh, 2013). On the other hand, a critical appraisal of the emerging literature on implicit processes that relate to exercise and physical activity indicates that it has become common practice to present the measure(s) in the Methods section unaccompanied by a specific rationale. This observation raises the possibility of bias, such as finding significant relationships by capitalizing on chance. Presumably, researchers (a) forego presenting a *conceptual* rationale for their measurement choices because there is no clear correspondence between underlying implicit processes and measures (i.e., the implicit processes that are being quantified through the numerous available measures remain enigmatic), and (b) are unable to offer concrete *psychometric* arguments for their choices because the psychometric properties of the measures (reliability, validity) are largely unknown.

The persistent inability to justify the selection of measures on the basis of conceptual or psychometric arguments has several consequences. First, the interpretation of results is made difficult when researchers are uncertain about what the tests are measuring and, in particular, the extent to which the constructs being measured indeed qualify as implicit processes. Conroy and Berry (2017) urged caution on this point, noting that "psychometric work within the physical activity domain is needed to establish what implicit measures actually measure" (p. 233).

Second, the great diversity of measures of implicit processes used in the study of exercise and physical activity, in conjunction with the inability to explain why each measure was chosen, makes the task of evidence synthesis precarious. In a recent critical review, Rebar et al. (2016) wrote that the use of a variety of disparate measures precludes "strong conclusions about effect sizes" and "conclusive summative work" (p. 403). Schinkoeth and Antoniewicz (2017) expressed the same concern, writing that, because of the "immense heterogeneity" of measures, "far-reaching conclusions are difficult to draw" (p. 17). Given the radically different approaches used in different measures, one cannot assume that results obtained with one measure would have also occurred with another. Therefore, attempts to reach conclusions that refer generically or collectively to the role of "implicit processes" from the early empirical literature on exercise and physical activity should be viewed with caution.

Third, given that preregistration is still rare in exercise psychology, the practice of using different measures in different studies without offering a justification may raise suspicions of selective outcome reporting and, therefore, the possible nonreplicability of published results. Along these lines, Schinkoeth and Antoniewicz (2017) warned that results showing significant associations between measures of implicit processes and exercise behavior "in the vast majority of studies" (p. 13) "may represent a publication bias due to the preferential publication of statistically significant results" (p. 16). This potential bias may hinder the forward momentum that is presently evident in this nascent, and promising, line of research.

2. The present study

The present investigation was conceived as a psychometric study that would serve as the foundation of future research on the role of implicit processes in exercise and physical activity. We conceptualize implicit processes as being unintentional, independent of executive resources, mostly outside of conscious awareness, and largely beyond volitional control. We use the adjective "automatic" to encapsulate these properties (Moors, Spruyt, & De Houwer, 2010). We further theorize that what the tests we evaluated reflect is "exercise associations." Specifically, we theorize that, through repeated life experiences, a pairing is formed between the stimulus concept of "exercise" with either positivity (i.e., something marked as pleasant, liked, desirable, attractive, to be pursued) or negativity (i.e., something marked as unpleasant, disliked, unwanted, repulsive, to be avoided). We presume that both the formation of such linkages over time, and the subsequent triggering of the associated positivity or negativity whenever the stimulus concept of "exercise" is activated, are automatic processes (i.e., mostly outside conscious awareness, unintentional, uncontrollable, effortless). Therefore, we refer to the measures evaluated in this study as measures of "automatic exercise associations."

The study was intended as a comparative evaluation. Therefore, we aimed to assess, in the same sample, the internal consistency and validity of nine measures of automatic exercise associations. Each measure is theorized to incorporate one or more features of automaticity, including unintentionality, uncontrollability, efficiency, and processing speed (Moors & De Houwer, 2006). Evidence of automaticity for each of the tasks is presented in the Supplementary Material (Section 1). Furthermore, among possible options, we selected measures on the basis of their feasibility (i.e., noninvasive, low-cost, computer-based, as opposed to requiring specialized equipment) and prior application in exercise psychology and other research fields.

The greatest conceptual challenge that an investigation of this sort faces is the selection of appropriate validation criteria. In accordance with conceptualizations of construct validation as a multifaceted process aimed at characterizing how well a measure fits into a network of theory-predicted relationships (Cronbach & Meehl, 1955), we opted to follow multiple validation approaches.

First, we examined criterion-related validity by investigating the associations of the nine measures with self-reported exercise behavior. Although exercise behavior is presumably determined by a multitude of factors, of which automatic exercise associations may be only one cluster, it is nonetheless the ultimate "downstream" outcome that measures of automatic exercise associations are aimed to predict. As such, exercise behavior (including self-reported exercise behavior) is *de facto* the principal criterion by which the validity of measures of automatic exercise associations is commonly judged (Antoniewicz & Brand, 2014; Calitri, Lowe, Eves, & Bennett, 2009).

Second, we examined convergent validity by investigating the associations of the nine measures with a measure of explicit attitudes. Using a measure of an explicit construct, such as attitude, to validate a measure of implicit processes has a precedent in the exercise psychology literature (e.g., Conroy et al., 2010; Rebar, Ram, & Conroy, 2015). However, it should be recognized that, according to dual-process theoretical models, explicit and implicit paths form through different processes (e.g., absorbing and evaluating information versus deriving pleasant or unpleasant affective experiences) and, consequently, measures presumed to reflect explicit and implicit processes may not necessarily converge.

Given the limitations of exercise behavior and explicit attitude as validation criteria, we considered additional criteria based on recent theoretical proposals (Bluemke et al., 2010; Brand & Ekkekakis, 2018; Conroy & Berry, 2017; Ekkekakis, 2017; Ekkekakis & Dafermos, 2012; Ekkekakis & Zenko, 2016; Williams & Evans, 2014). Specifically, we considered experienced and recalled affective responses to a bout of moderately challenging exercise as additional validation criteria. While automatic exercise associations may be influenced by several factors, including perceived societal values and extrapersonal associations (Rudman, 2004), the aforementioned theoretical proposals posit that the basis of automatic exercise associations is whether the stimulus concept of "exercise" has registered in associative memory as a positively or negatively valenced behavior, presumably as a result of multiple, consistent pairings of exercise experiences with pleasure or displeasure.

We used self-reports of pleasure versus displeasure (i.e., affective valence) responses to a laboratory-based exercise session as an indication of (possible) usual experiences with exercise. However, we recognize that the many factors that may have shaped such experiences (e.g., hurtful comments by physical education teachers or peers, personally meaningful praise by significant others) cannot be comprehensively recreated in a laboratory setting. Nevertheless, we aimed to create a meaningful proxy of usual exercise experiences by calibrating intensity to maximize interindividual variation in affective responses. Specifically, we set exercise intensity at the level of the individually determined ventilatory threshold (VT) based on evidence that, at this intensity, some individuals report feeling progressively better whereas others report feeling progressively worse during exercise (Ekkekakis, 2013; Ekkekakis, Parfitt, & Petruzzello, 2011).

Both the level of pleasure-versus-displeasure reported during exercise (henceforth "experienced affect") and subsequently recalled pleasure-versus-displeasure (henceforth "recalled affect") were considered as validation criteria. Experienced affect was operationalized in two ways, namely (a) the direction (progressive improvement versus decline) and the rate of change in pleasure-versus-displeasure during the exercise bout (henceforth "affective slope") and (b) the pleasure or displeasure reported just before the end of the exercise bout (henceforth "affective end"). These decisions were based on evidence that, while the total and the average degree of pleasure-displeasure experienced during an episode appear to be inconsequential in shaping subsequent behavior, both the pleasure-displeasure experienced at the end of an episode (e.g., Kahneman, Wakker, & Sarin, 1997) and the slope of pleasure-displeasure during an episode (see Zenko, Ekkekakis, & Ariely, 2016 and references therein) are especially influential.

Further, given that a myriad of variables may intervene between automatic exercise associations and exercise behavior, thus weakening any estimates of association, Brand and Schweizer (2015) have proposed the concept of "situated decisions" as a "functional link" between implicit (as well as explicit) processes and exercise behavior. Their Situated Decisions to Exercise Questionnaire (SDEQ) assesses decisions made in hypothetical but common scenarios in which the individual is called to choose between an exercise and a nonexercise option (e.g., "You're leaving work and you are just about to go to the gym. Now you hear that your colleagues plan to go for a drink. Do you exercise, or not?"). Arguably, such situated decisions may be a more meaningful validation criterion compared to exercise behavior over the previous week or over a "typical" week. Thus, situated decisions, as measured by the SDEQ (henceforth "situated decisions"), were also considered as a validation criterion.

The literature on implicit processes includes many speculations about the underlying implicit processes that various measures reflect (e.g., see the tentative classification proposed by Sheeran et al., 2013). For example, it has been suggested that the Approach-Avoidance Task may measure automatic approach and avoidance tendencies, the Single-Category Implicit Association Test may be influenced by personal and extrapersonal associations (Olson & Fazio, 2004), and the Visual Probe Procedure may be influenced by motivational relevance that drives attention toward or away from stimuli (Müller, Rothermund, & Wentura, 2016). In the absence of consensus or compelling evidence to support these suggestions, arguably the most prudent position is to accept that, at least at the present stage of knowledge development, it is still not possible to match specific underlying implicit processes to performance on specific tests. For the purposes of the present investigation, we assumed that (a) no measure of automatic associations relies entirely on automatic processes (Bargh, 1994), (b) performance on the measures of automatic exercise associations we evaluated relies on various (unknown) combinations of implicit processes, and, thus, (c) measures of implicit exercise associations may not relate to each other (e.g., Calitri et al., 2009) or to all of our validation criteria. Therefore, while there was a reasonable basis (conceptual reasoning as outlined in the previous paragraphs, prior published evidence, or both) to suggest that the selected measures of automatic exercise associations should relate to one or more of the validation criteria, there was no genuine *a priori* basis for hypothesizing which measure would relate to which criterion or criteria.

3. Methods

3.1. Participants

In the absence of a prior reliable estimate of effect size for the relationship between measures of automatic exercise associations and the validation criteria employed in the present investigation, we considered a "medium" correlation of $r \geq 0.30$ (Cohen, 1992, p. 157) as indicating a meaningful relationship (i.e., ≥9% shared variance). Thus, a power calculation for a two-tailed bivariate correlation, with $r = 0.30$, $\alpha = 0.05$, and $1\text{-}\beta = 0.80$, indicated that 84 participants were needed. To compensate for potential participant attrition and an elimination rate of 20%, the target sample size was 101.

Participants were deemed eligible if they were between the ages of 18 and 45 years (if male) or 55 years (if female), to prevent age from being considered a risk factor (American College of Sports Medicine, 2018). In addition, potential participants were deemed eligible if they had normal or corrected vision (due to the use of visual stimuli). Exclusionary criteria included pregnancy, a history of heart disease, unexpected pain or dizziness during exercise, any conditions that could be worsened by physical exercise, an affirmative answer to any of the items of the Physical Activity Readiness Questionnaire, and nonfluency in the English language.

Following Institutional Review Board Approval, participation was solicited via an e-mail message sent to all students, faculty, and staff of a large state university in the midwestern United States. Though 104 individuals met the inclusion/exclusion criteria, five failed to enroll in the study because they either had scheduling conflicts or failed to appear for scheduled appointments. After the first laboratory session (described below), two individuals dropped out for personal reasons unrelated to the study and two other individuals were eliminated because their VT could not be determined (due to early termination of the test in one case, and due to excessive noise in gas exchange data in the other case). Thus, 95 participants (i.e., more than the 84 needed for 80% statistical power) completed all three sessions and were included in the analyses. Participant characteristics are presented in Table 1.

Table 1
Participant characteristics by sex.

Variable	Men (n = 37)		Women (n = 58)	
	Mean ± SD		Mean ± SD	
Age (years)	23 ± 7		26 ± 10	
Height (cm)	178.5 ± 7.0		166.3 ± 7.1	
Body Mass (kg)	81.4 ± 13.0		68.1 ± 15.1	
Body Mass Index (kg·m^{-2})	25.51 ± 3.63		24.70 ± 5.47	
VO_2 peak (ml·kg^{-1}·min^{-1})	35.63 ± 8.39		29.26 ± 7.96	
Exercise Behavior (min·week^{-1})	279 ± 183		193 ± 149	
	Men (n)	**(%)**	**Women (n)**	**(%)**
Weight Status				
Underweight	0	0.0%	2	3.4%
Normal Weight	16	43.2%	33	56.9%
Overweight	18	48.6%	18	31%
Obese	3	8.1%	5	8.6%
Education				
High School Diploma or Equivalent	30	81.1%	34	58.6%
BS/BA/BFA	4	10.8%	10	17.2%
MS/MA/MFA	1	2.7%	9	15.5%
PhD/Terminal	2	5.4%	5	8.6%
Race/Ethnicity				
White/Caucasian	32	86.4%	45	77.3%
Black/African American	1	2.7%	4	6.9%
Asian/Pacific Islander	2	5.4%	3	5.2%
Latino/Latina/Hispanic	2	5.4%	3	5.2%
Multiracial	0	0%	2	3.4%
Undisclosed	0	0%	1	1.7%
Current Student	30	81.1%	37	63.8%

3.2. Measures

3.2.1. Measures of automatic exercise associations

Nine measures of automatic exercise associations were selected: (a) Single-Category Implicit Association Test (SC-IAT), (b) Personalized SC-IAT (PSC-IAT), (c) Approach-Avoidance Task (AAT), (d) Go/no-go Association Task (GNAT), (e) Extrinsic Affective Simon Task (EAST), (f) Evaluative Decision Task (EDT), (g) Affect Misattribution Procedure (AMP), (h) Visual Probe Procedure (VPP), and (i) Affective Stroop Task (AST). All measures were administered using Inquisit Lab (Version 4; Millisecond Software, Seattle, WA). During the Approach-Avoidance Task, participants moved a joystick (Extreme 3D Pro; Logitech, Newark, CA) with their dominant hand (Rinck & Becker, 2007). For all tasks, we used a 500 ms interval between trials and a 60 s rest period between tasks. However, participants could request a longer rest period if desired. Evidence of automaticity and additional procedural details, including stimuli descriptions, are provided in the Supplementary Material (Sections 1 and 2, Tables S1 and S2).

3.2.2. Self-reported exercise

Exercise behavior was measured with the short form of the International Physical Activity Questionnaire (IPAQ-SF; Craig et al., 2003). Participants were asked to answer questions about their vigorous-intensity, moderate-intensity, and walking exercise behavior during a “usual week.” Participants were asked “only about those physical activities that [they] did for at least 10 min at a time” in their leisure time, which eliminated spontaneous physical activity and emphasized deliberate exercise behavior. This was necessary because the focus of this investigation was on automatic associations of exercise rather than physical activity in general. Craig et al. (2003) have reported fair agreement between the IPAQ-SF and accelerometers ($\rho = 0.30$), concluding that the IPAQ is “at least as good as other established self-reports" (p. 1381). Exercise behavior was quantified as the combined minutes of moderate- and vigorous-intensity exercise per week. Two participants who reported extreme amounts of exercise (more than 750 min·week^{-1}) were eliminated from analyses as outliers (outside Tukey's fences), as was one participant who did not provide exercise behavior data.

3.2.3. Explicit affective attitude

Ten items based on the theory of planned behavior (Ajzen, 1991) were used to assess explicit affective attitude toward exercise. The stem “For me, exercising at least 30 min per day for at least 5 days over the next week would be … " was followed by seven-point scales anchored by: Pleasant-Unpleasant; Enjoyable-Unenjoyable; Exciting-Boring; Soothing-Agitating; Distressing-Peaceful; Fatiguing-Energizing; Tempting-Uninviting; A lot of fun-Not fun at all; Good for my ego-Bad for my ego; Something that picks me up-Something that brings me down. Internal consistency was high (Cronbach's α = .90).

3.2.4. Experienced affect

The core affective dimension of valence, theorized to range from pleasure to displeasure (Russell, 2003), was assessed with the Feeling Scale (FS; Hardy & Rejeski, 1989), a single-item, 11-point bipolar rating scale ranging from +5 (I feel “very good”) to −5 (I feel “very bad”). Verbal anchors are provided at zero (“neutral”) and odd numbers throughout the scale. Hardy and Rejeski (1989) have reported concurrent validity data for the FS. The affective slope for each participant was calculated by fitting a linear regression over the FS ratings acquired from baseline to the last during-exercise time point (see below). The final FS rating reported during exercise was designated as the affective end.

3.2.5. Recalled affect

Recalled affect from the exercise session was assessed with the Empirical Valence Scale (EVS; Lishner, Cooter, & Zald, 2008). The EVS was used instead of the FS for the assessment of recalled affect in order to reduce common method variance. The EVS is a bipolar rating scale ranging from −100 (“Most unpleasant imaginable”) to +100 (“Most pleasant imaginable”). Thirteen empirically spaced verbal anchors (e.g., extremely, strongly, barely, neutral) are located throughout the scale. Participants were asked to respond to the question: “How did the exercise session make you feel?"

3.2.6. Situated decisions to exercise

Situated decisions were assessed using the Situated Decisions to Exercise Questionnaire (SDEQ; Brand & Schweizer, 2015). The SDEQ consists of eight items assessing whether participants would choose exercise or not in various hypothetical scenarios (e.g., “After a long day of work, you have just arrived at home and you feel tired. However, you planned to go workout tonight. Do you exercise, or not?”). A higher score on the SDEQ indicates lower likelihood of exercising (i.e., scores correlate negatively with exercise behavior). Brand and Schweizer (2015) have provided data on criterion and structural validity. Internal consistency in the present sample was acceptable (Cronbach's α = 0.71).

3.3. Procedures

3.3.1. Internet survey

One week prior to the first laboratory visit, participants received an invitation via e-mail to complete an online version of the IPAQ-SF (Qualtrics, Provo, UT).

3.3.2. Maximal exercise session

Upon entering the laboratory for their first visit, participants read and signed an informed consent form. Next, participants answered questions about their demographic characteristics and explicit attitude. Height was measured using a wall-mounted stadiometer and weight was measured with an electronic scale (BF-626, Tanita, Tokyo, Japan). Participants were then fitted with a heart rate monitor (Polar, Kempele, Finland) and nose-and-mouth facemask for the collection of expired gases (Hans Rudolph, Kansas, MO), and had the instructions for the FS read to them.

The incremental (ramp) exercise test was conducted using a computer-controlled, electronically braked recumbent cycle ergometer (Corival Recumbent, Lode, Groningen, The Netherlands). The workload was set at 0 W for 5 min (as a warm-up) and was subsequently progressively increased by 1 W every 4 s until volitional termination. Oxygen uptake and carbon dioxide production were measured continuously with a metabolic cart (TrueOne 2400, Parvomedics, Salt Lake City, UT) that was calibrated before each use. Upon termination, the facemask was removed, and participants completed a 5-min cool-down at 0 W.

Two to four experienced investigators, working independently, determined the VT of each participant with the aid of a software program (WinBreak 3.7, Epistemic Mindworks, Ames, IA) that combines three methods (V-slope, ventilatory equivalents, and excess carbon dioxide), as recommended by Gaskill et al. (2001). Disagreements were resolved by consensus. The VT was used to set the intensity of the individualized exercise session (described below).

The maximal exercise session was also used as an opportunity to familiarize the participants with the measures of experienced (FS) and recalled affect (EVS) that were to be used during the subsequent individualized exercise session. The FS was administered before, during, and after the exercise test. Participants were asked to indicate their responses by pointing on a poster-sized version of the FS (removed from the field of vision while not being used). Recalled affect of the exercise session was assessed with the EVS 5 min after the test. Participants were invited to complete an online SDEQ following their maximal exercise session.

3.3.3. Individualized exercise session

One week after the maximal exercise session, participants returned to the laboratory for their second session. An investigator reminded participants of the instructions for the FS. Participants were again fitted with the heart rate monitor and exercised on the same recumbent cycle ergometer. Participants exercised for a total of 15 min. Of these, the first 5 min were a warm-up, with the workload set at 20% of the workload corresponding to the VT of each participant. For the next 10 min, the workload was set to 100% of the workload corresponding to the VT. The FS was administered immediately before exercise, at min 1, 3, and 5 during the 5-min warm-up, and every 2 min during the 10-min exercise at the VT. Based on dual-mode theory (Ekkekakis, 2003, 2013), affective valence responses to exercise at the VT were expected to exhibit interindividual variability, presumably enhancing the utility of this variable as a validation criterion. Finally, recalled affect of the exercise session was assessed using the EVS 5 min after the end of exercise.

Based on the "end rule" (Kahneman et al., 1997), affective valence at the end of exercise was expected to weigh heavily on recalled affect of the exercise session. Because a cool-down would have likely led to homogenously pleasant affective responses (Ekkekakis et al., 2011), no cool-down was provided, in order to preserve interindividual variability in affective valence at the end of the session and, presumably, in ratings of recalled affect.

3.3.4. Automatic associations session

One week after the individualized exercise session, participants returned to the laboratory for their final session. The one-week interval was dictated by the need to strike a balance between two competing considerations, namely (a) the need to assess the validity criteria and the measures to-be-validated at different times in order to attenuate common method bias (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003; Podsakoff, MacKenzie, & Podsakoff, 2012) and thus prevent the artificial inflation of estimates of association (Scherer & Schott, 2012), and (b) the need to reduce random variance (e.g., different routines prior to testing), which could have negatively impacted the reliability of measurements. Thus, the final session was scheduled at the same day of the week and the same time of day as the individualized exercise session, to account for possible diurnal effects (adjustments were necessary for nine participants, who returned 3–9 days later rather than 7 because of scheduling conflicts). The nine measures of automatic exercise associations were completed in random order to prevent an order effect.

A memory-load task was completed concurrently with each measure of automatic exercise associations. This quality-control procedure was intended to engage explicit processes (working memory) and thus increase the likelihood that task responses would reflect implicit processes (Hermans, Crombez, & Eelen, 2000). Specifically, participants were instructed to remember six random digits during each task. After each task, they had to identify the six random digits from a sheet of paper containing 26 number strings. To provide additional motivation, $0.75 was given for each successfully remembered string of digits.

3.4. Data analysis

3.4.1. Data reduction and scoring

The data from one participant who could read Chinese and thus was not allowed to participate in the AMP (Payne, Cheng, Govorun, & Stewart, 2005), and another participant who could not understand the AMP and quit before completing the task, were missing from the AMP responses. Four participants, deemed to have deficiencies in color perception, were eliminated from analyses involving the EAST. Finally, data from nine participants were eliminated from analyses involving the EDT because of a script error that caused the trials to have a response deadline of 300 ms instead of the intended 600 ms.

The SC-IAT and PSC-IAT had extreme trials (e.g., very short response latencies) and participants with high error rates were treated following the *D*-score algorithm (Greenwald, Nosek, & Banaji, 2003), which is built into the Inquisit software. Processing efficiency was calculated for the SC-IAT using the data preparation and reduction procedures followed by Rebar et al. (2015). The necessary inputs (i.e., variance in reaction time, proportion correct) were calculated for the compatible and incompatible blocks of the SC-IAT and processing efficiency was calculated for each block following the EZ-diffusion algorithm (Wagenmakers, van der Maas, & Grasman, 2007). One incorrect trial was created for each block, if needed, because the EZ-diffusion algorithm requires at least one erroneous response (Rebar et al., 2015). The difference in processing efficiency between blocks was calculated to determine if participants had more (or less) processing efficiency for compatible blocks compared to incompatible blocks.

Care was taken to standardize the definition of an outlier by using Tukey's fences. Outlier trials were defined as trials with response latencies that were outside of Tukey's fences for any participant, based on the interquartile range of that participant. Different fences were calculated for each participant to account for idiosyncratic response patterns. Outlier trials were not calculated for the AAT because unintentional movements of the joystick may or may not represent automatic associations, so all movements were retained regardless of latency. Timed out trials (if applicable) were defined as trials that did not have a response prior to the response deadline (AMP: 600 ms, EDT: 600 ms, VPP: 1000 ms). Only valid trials were used to calculate the scores for each measure. Valid trials were defined as trials that were not outliers and did not time out (if applicable). In addition, for the AST, a valid trial had to be accurate (Berry, 2006), so valid trials were defined as trials that were correct and not an outlier. Accuracy was calculated using the built-in voice analysis utility of the Inquisit 4 software. The percentage of trials that were outliers, timed out, and were valid for each task are displayed in the Supplementary Material (Table S5). Participants with an extremely low percentage of valid trials (defined by calculating Tukey's fences at the sample level) were eliminated from further analysis.

For the AAT, an index of automatic exercise associations was calculated by first determining the bias to approach rather than avoid (pull rather than push), or "approach bias," for each type of trial (exercise, neutral control). Positive values were taken to indicate a tendency to approach (pull) rather than avoid (push). Next, automatic associations with exercise were calculated by subtracting the mean approach bias for neutral trials from the approach bias for exercise trials. A positive value was taken to indicate that, when controlling for the bias to approach neutral images, participants had a faster automatic tendency to approach exercise images. For example, if a participant had a 5 ms approach bias for neutral images, and a 12 ms approach bias for exercise images, then the participant was determined to approach exercise images 7 ms faster than neutral images (12 ms − 5 ms = 7 ms).

For the AMP, an index of automatic exercise associations was defined as the proportion of targets rated as pleasant following exercise primes minus the proportion of targets rated as pleasant following neutral primes. Positive values were taken to indicate a positive automatic association with exercise, such that participants were more likely to rate exercise-primed targets as positive, compared to neutral-primed targets. For example, if participants rated 50% of the exercise-primed targets as positive and 30% of the neutral-primed targets as positive, then they were determined to have a positive automatic association with exercise (50% − 30% = 20%).

The EAST, EDT, and GNAT yielded both error rates and mean response latencies as outcome variables. For the EDT, the accuracy rate for responding to exercise-primed trials with negative targets was subtracted from the accuracy rate for responding to exercise-primed trials with positive targets. A positive score indicated that participants were more accurate in responding to trials in which exercise primes were paired with positive targets compared to trials in which exercise primes were paired with negative targets; this was called “ExercisePrime-Accuracy." The same procedure was followed to determine if participants were more accurate in responding to trials with neutral primes and positive targets compared to neutral primes with negative targets; this was called “NeutralPrime-Accuracy." Finally, “NeutralPrime-Accuracy" was subtracted from “ExercisePrime-Accuracy." Positive scores indicated that participants responded more accurately to positive targets than negative targets following exercise primes, while controlling for the accuracy in responding to positive targets versus negative targets following neutral primes. Positive scores indicated lower error rates and more favorable automatic exercise associations.

Similarly, for the EAST, positive accuracy scores indicated that participants were more accurate at responding to exercise target words when they were paired with the “positive” key compared to when exercise words were paired with the “negative” key, while controlling for the accuracy in responding to neutral target words when paired with the “positive” key versus neutral target words when paired with the “negative key.” Accuracy rate was also calculated for both the incompatible and compatible blocks of the GNAT. Positive accuracy scores for the GNAT indicated that participants made fewer errors when exercise was paired with a positive attribute (compatible block) compared to when exercise was paired with a negative attribute (incompatible block).

Response latencies for the EAST, EDT, and GNAT were calculated in similar fashion. For the EDT, positive response latency scores indicated that participants had facilitated mean responses (lower reaction times) for positive targets following exercise primes compared to negative targets following exercise primes, while controlling for the facilitation following positive targets and neutral primes versus negative targets and neutral primes. Likewise, for the EAST, a positive score indicated that, controlling for neutral words and neutral trials, participants were faster on average at responding to exercise words when they were paired with the “positive” key than when paired with the “negative” key. Finally, for the GNAT, a positive score for the response latency indicated that participants had faster mean reaction times for the compatible block (when exercise was paired with a positive attribute) than the incompatible block (when exercise was paired with a negative attribute).

For the AST, the mean reaction time when responding to neutral trials was subtracted from the mean reaction time when responding to exercise trials. A positive score indicated that reaction time for exercise trials was slower and participants had attentional bias toward exercise. The VPP is also indicative of attentional bias. The mean reaction time when responding to trials in which the exercise target was congruent with the probe was subtracted from the mean reaction time when responding to trials in which the exercise target was incongruent with the probe. A positive score indicated that participants were faster to locate the probe when it was congruent with an exercise image, or that they exhibited an attentional bias toward the exercise images.

3.4.2. Internal consistency

The internal consistency of each measure of automatic exercise associations was estimated as split-half reliability. First, the trials of each task were randomized for each participant to account for practice and fatigue effects. Second, two equivalent halves of trials were created (e.g., equivalent number of compatible and incompatible trials, or an equivalent number of exercise and neutral trials). Third, an index of automatic exercise associations was calculated for each half of each task. Finally, the indices of automatic exercise associations for both halves were correlated, and the Spearman-Brown prophecy formula was applied to estimate the reliability of scores on the task as a whole. This follows the process for determining internal consistency outlined previously (Hyde, Doerksen, Ribeiro, & Conroy, 2010). Calculating internal consistency of the SC-IAT is not relevant when using the EZ-diffusion algorithm as a scoring procedure (Chevance, Héraud, Guerrieri, Rebar, & Boiché, 2017).

3.4.3. Validity

To assess the validity of the nine measures of automatic exercise associations, bivariate correlations were calculated between the scores on measures of automatic exercise associations and the six validation criteria, namely (a) self-reported exercise, (b) explicit attitude, (c) affective slope, (d) affective end, (e) recalled affect, and (f) situated decisions. To address the multiplicity problem, a False Discovery Rate of 5% was applied to avoid the inflation of the Type I error rate while preserving statistical power (Benjamini & Hochberg, 1995; Benjamini & Yekutieli, 2005; Keselman, Cribbie, & Holland, 2002). This procedure was preferred over the Bonferroni method because it has been found to prevent the unnecessary loss of statistical power and inflation of the Type II error rate (Nakagawa, 2004; Rothman, 1990), thus providing a “much better compromise between Type I and Type II errors” (Nakagawa, 2004, p. 1045).

4. Results

4.1. Individualized exercise intensity

During the warm-up, exercise intensity remained below the range that the American College of Sports Medicine (2018) considers “moderate” intensity (i.e., 64–76% of maximal heart rate). During the 10-min individualized exercise at VT, intensity remained within the “moderate” range, increasing from 64.6% to 74.5% of maximal heart rate on average.

4.2. Experienced affect

Consistent with expectations, affective responses during exercise varied considerably between participants (see Figure 1). The FS ratings of most participants (61.1%) exhibited a negative slope during exercise, the slope was positive for 25.2%, and the slope was zero for 13.7%. Most participants (75.8%) had a positive affective end (i.e., FS > 0), 15.8% reported a negative affective end (i.e., FS < 0), and 8.4% reported a neutral (i.e., FS = 0) affective end.

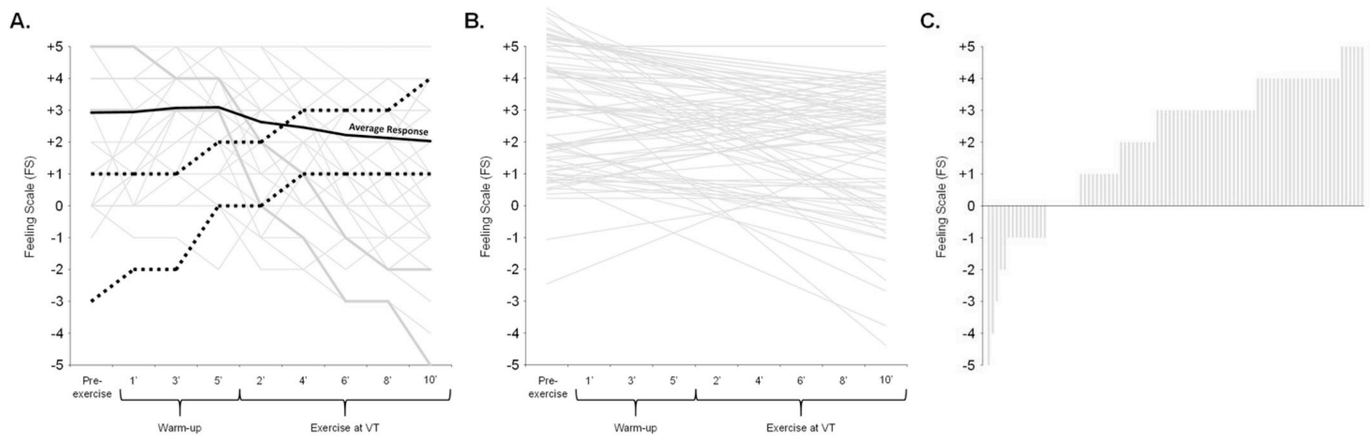

Figure 1. Illustration of the interindividual variability in affective responses during the exercise session at VT. Panel (A) shows the raw Feeling Scale (FS) ratings ($N = 95$), highlighting two individuals who exhibited progressive improvements (dotted black lines), two individuals who exhibited progressive declines (solid grey lines), and the average response (solid black line). Panel (B) shows the lines of best fit through each of the 95 individual sets of FS ratings, demonstrating the variability in "affective slopes." Panel (C) shows the interindividual variability in affective ends, namely the FS ratings at the final time point of the 10-min exercise session at VT.

4.3. Recalled affect

Recalled affect was positive for most participants. Mean recalled affect for the sample was 40.68 ± 24.78 units on the EVS (with a possible range from −100 to +100). Like experienced affect, recalled affect also exhibited considerable heterogeneity. Individual ratings ranged from −37 to +100, and the 25th, 50th, and 75th percentiles had values of 24.4, 38.38, and 60.71 units, respectively. One participant, with an extremely negative rating of recalled affect (−37 units) was deemed an outlier (outside Tukey's fences) and was eliminated from further analyses involving recalled affect. The affective slope and affective end accounted for 8.8% and 29.2% of the variance in recalled affect, respectively, indicating that these validation criteria were related but distinct.

4.4. Administration of measures of automatic exercise associations

Administration times for the nine measures of automatic exercise associations are shown in the Supplementary Material (Table S3). No participant spent more than 11 min responding to any given task. Efficiency was quantified by the percentage of correctly identified random number strings after each task. Overall efficiency was high, suggesting automaticity. Participants identified 89.35% ± 12.01% of the random number strings correctly. Efficiency ratings and the percentages of trials for each task that were valid, outliers, or timed out are presented in the Supplementary Material (Tables S4 and S5).

4.5. Internal consistency of measures of automatic exercise associations

Internal consistency results are presented in Table 2. The AMP, PSC-IAT, and EDT had outcomes with acceptable to good internal consistency. Several measures, especially the EAST, VPP, and AST, demonstrated poor internal consistency, so caution should be taken in interpreting any results associated with these tasks.

4.6. Validity of measures of automatic exercise associations

Correlation coefficients between each measure of automatic exercise associations and the validation criteria are presented in Table 3 (note that some tasks yield two outcome variables). Most correlation coefficients (63.1%) were near zero ($r \leq 0.10$). The two outcome variables of the AAT were significantly (after applying a False Discovery Rate of 5%) and meaningfully (i.e., $r \geq 0.30$) correlated with both self-reported exercise and situated decisions to exercise. In addition, the

Table 2
Indices of internal consistency (split-half) of measures of automatic exercise associations.

Task	Outcome Variable		
	Response Latency	Error Rate	Other
Affect Misattribution Procedure (AMP)			.85[a]
Affective Stroop Task (AST)	.01		
Approach-Avoidance Task (AAT)	.56[b]		.45[c]
Evaluative Decision Task (EDT)	.25	.74	
Extrinsic Affective Simon Task (EAST)	-.13	.28	
Go/no-go Association Task (GNAT)	.12	.62	
Personalized Single-Category Implicit Association Test (PSC-IAT)	.85		
Single-Category Implicit Association Test (SC-IAT), *D*-score	.18		
Visual Probe Procedure (VPP)	-.05		

Note:
[a] Proportion of "pleasant" categorizations vs. "unpleasant" categorizations.
[b] Initial Response Latency: The time it took for the participant to make the first movement.
[c] Final response latency: The time it took for the participant to complete the movement.

PSC-IAT exhibited a correlation with the affective slope that closely approached the $r \geq 0.30$ criterion of meaningfulness ($r = 0.298$) but was not deemed reliable after applying a False Discovery Rate of 5%.

5. Discussion

The primary purpose of this investigation was to assess the internal consistency and validity of nine measures of automatic exercise associations. The broader objective was to assist researchers interested in studying dual-process models in the context of exercise by providing an empirical basis for selecting measures of automatic exercise associations among a wide array of available, but disparate, options. This undertaking was predicated on the idea that the continued absence of a conceptual or psychometric rationale for selecting measures of automatic associations in published reports can undermine the evolution of this otherwise promising line of research (Conroy & Berry, 2017; Rebar et al., 2016; Schinkoeth & Antoniewicz, 2017). The present study is the first known head-to-head comparison of measures of automatic exercise associations in the field of exercise psychology. Thus, the data reported here represent a unique resource for researchers interested in building

Table 3
Validity coefficients of measures of automatic exercise associations.

Task (Outcome Variable)	Affective Slope	Affective End	Recalled Affect	Explicit Attitude	Exercise Behavior	Situated Decisions
Approach-Avoidance Task (Initial Response Latency)	$r = -.170$	$r = -.033$	$r = -.002$	$r = -.096$	**$r = -.323$**	**$r = .331$**
	$n = 95$	$n = 95$	$n = 94$	$n = 95$	**$n = 92$**	**$n = 95$**
	$p = .100$	$p = .749$	$p = .989$	$p = .355$	**$p = .002$**	**$p = .001$**
Approach-Avoidance Task (Final Response Latency)	$r = -.143$	$r = -.066$	$r = .026$	$r = -.136$	**$r = -.330$**	**$r = .352$**
	$n = 95$	$n = 95$	$n = 94$	$n = 95$	**$n = 92$**	**$n = 95$**
	$p = .166$	$p = .525$	$p = .802$	$p = .190$	**$p = .001$**	**$p < .001$**
Affect Misattribution Procedure	$r = .074$	$r = .076$	$r = .043$	$r = .173$	$r = .156$	$r = -.096$
	$n = 85$	$n = 85$	$n = 84$	$n = 85$	$n = 82$	$n = 85$
	$p = .501$	$p = .489$	$p = .696$	$p = .114$	$p = .163$	$p = .384$
Affective Stroop Task (Response Latency)	$r = .000$	$r = .073$	$r = .038$	$r = .019$	$r = -.126$	$r = .055$
	$n = 91$	$n = 91$	$n = 90$	$n = 91$	$n = 88$	$n = 91$
	$p = .999$	$p = .491$	$p = .723$	$p = .858$	$p = .241$	$p = .605$
Evaluative Decision Task (Response Latency)	$r = -.131$	$r = .017$	$r = .196$	$r = -.071$	$r = -.092$	$r = -.115$
	$n = 82$	$n = 82$	$n = 81$	$n = 82$	$n = 80$	$n = 82$
	$p = .240$	$p = .878$	$p = .079$	$p = .528$	$p = .418$	$p = .305$
Evaluative Decision Task (Error Rate)	$r = -.059$	$r = -.191$	$r = -.158$	$r = -.099$	$r = -.140$	$r = -.016$
	$n = 82$	$n = 82$	$n = 81$	$n = 82$	$n = 80$	$n = 82$
	$p = .599$	$p = .086$	$p = .159$	$p = .376$	$p = .216$	$p = .887$
Extrinsic Affective Simon Task (Response Latency)	$r = -.112$	$r = .037$	$r = -.094$	$r = .104$	$r = .027$	$r = -.032$
	$n = 84$	$n = 84$	$n = 83$	$n = 84$	$n = 81$	$n = 84$
	$p = .311$	$p = .737$	$p = .400$	$p = .346$	$p = .813$	$p = .776$
Extrinsic Affective Simon Task (Error Rate)	$r = .126$	$r = .114$	$r = .069$	$r = .063$	$r = .054$	$r = -.134$
	$n = 84$	$n = 84$	$n = 83$	$n = 84$	$n = 81$	$n = 84$
	$p = .254$	$p = .302$	$p = .536$	$p = .567$	$p = .630$	$p = .225$
Go/no-go Association Task (Response Latency)	$r = -.062$	$r = -.071$	$r = -.226$	$r = .004$	$r = -.073$	$r = .046$
	$n = 82$	$n = 82$	$n = 81$	$n = 82$	$p = 80$	$n = 82$
	$p = .577$	$p = .524$	$p = .043$	$p = .970$	$n = .518$	$p = .684$
Go/no-go Association Task (Error Rate)	$r = .025$	$r = .053$	$r = .011$	$r = -.082$	$r = .115$	$r = .109$
	$n = 82$	$n = 82$	$n = 81$	$n = 82$	$n = 80$	$n = 82$
	$p = .820$	$p = .634$	$p = .924$	$p = .463$	$p = .308$	$p = .329$
PSC-IAT (*D*-score)	**$r = .298$**	$r = .051$	$r = -.111$	$r = -.020$	$r = -.083$	$r = -.030$
	$n = 95$	$n = 95$	$n = 94$	$n = 95$	$n = 92$	$n = 95$
	$p = .003$	$p = .621$	$p = .289$	$p = .850$	$p = .432$	$p = .774$
SC-IAT (*D*-score)	$r = -.029$	$r = .047$	$r = -.017$	$r = .034$	$r = -.005$	$r = -.069$
	$n = 95$	$n = 95$	$n = 94$	$n = 95$	$n = 92$	$n = 95$
	$p = .778$	$p = .650$	$p = .869$	$p = .743$	$p = .959$	$p = .509$
SC-IAT (Processing Efficiency)	$r = -.220$	$r = -.101$	$r = -.054$	$r = -.025$	$r = -.138$	$r = .117$
	$n = 90$	$n = 90$	$n = 89$	$n = 90$	$n = 87$	$n = 90$
	$p = .037$	$p = .343$	$p = .614$	$p = .812$	$p = .202$	$p = .273$
Visual Probe Procedure (Response Latency)	$r = .024$	$r = .008$	$r = -.059$	$r = .042$	$r = .197$	$r = .024$
	$n = 92$	$n = 92$	$n = 92$	$n = 92$	$n = 89$	$n = 92$
	$p = .821$	$p = .936$	$p = .579$	$p = .692$	$p = .064$	$p = .823$

Abbreviations: PSC-IAT: Personalized Single-Category Implicit Association Test. SC-IAT: Single-Category Implicit Association Test. *Notes:* Correlation coefficients approaching or exceeding |0.30| are boldfaced. The sample size is provided for each correlation.

an evidence-supported justification for selecting such measures. Therefore, an important potential contribution of this dataset is that it offers the means by which this line of research can overcome the questionable current practice of selecting measures without a supporting rationale.

It is essential to clarify that we did not conceive this psychometric evaluation as a competition to establish the "best" measure of automatic exercise associations and have avoided framing the issue in such terms. As stated in the introduction, although all the measures tested in the present study exhibit features of automaticity, we have no reason to believe that they all tap the same implicit processes and are, therefore, interchangeable. Thus, we concur with Schinkoeth and Antoniewicz (2017), who wrote that "although … the temptation is high to hand out advice that one might be the 'best' implicit measure," it is "neither possible nor advisable" to do so (p. 13). Nevertheless, the relative strength of the correlations of the measures with the validation criteria would be useful information for researchers interested in selecting a measure and performing power calculations to determine the appropriate sample size for their investigations.

A crucial consideration in a comparative psychometric evaluation is to ensure that all measures are given an equal chance by ensuring stable and consistent testing conditions. In addition, in planning this investigation, we were mindful of the need to establish some degree of standardization in the procedural details of the testing protocols. Schinkoeth and Antoniewicz (2017) commented on this problem, stating that "even when studies used the same measure, the specific application varied" (p. 13). Therefore, in the present study, great care was taken in selecting stimuli and structuring the tasks. Specifically, unless an empirical justification for a different number of trials existed in the literature, each task consisted of 120 test trials and was preceded by 24 practice trials. Likewise, the number of stimuli used in each task was based on the number of trials and task structure for each measure (see Supplementary Material, Section 2).

5.1. Administration time and task efficiency

On average, each task took less than 7.5 min to complete. This administration time is comparable to most self-report questionnaires, underscoring the practicality of using measures of automatic exercise associations. The tasks were also efficient, as indicated by the high rates of random number strings identified correctly after each task (86.3%–92.6%). This finding lends empirical support to the claim that the tasks exhibit automaticity and can thus be presumed to tap underlying implicit processes (Hermans et al., 2000).

5.2. Internal consistency

The internal consistency of the AMP in this sample (i.e., 0.85) was comparable with previous research (i.e., > 0.80; Payne et al., 2005). The internal consistencies of the EDT, using error rate as the outcome variable, and the PSC-IAT were also acceptable. In contrast, other measurements yielded poor internal consistency, suggesting susceptibility to random measurement error. While it is not possible to ascertain the extent to which the unreliability of these measurements is due to the nature of the tasks or to the particular measurement conditions of the present study, it should be noted that the internal consistency coefficients found in this study were, in several cases, lower than previously reported (Calitri, et al., 2009; Gawronski, Deutsch, & Banse, 2011; Karpinski & Steinman, 2006; Rebar et al., 2015).

In evaluating our results, however, readers should note that we took several precautions aimed to help attenuate random measurement error and thus promote reliability. These included efforts to reduce ambient noise and visual distractions, a private laboratory environment, standardized instructions, and highly scripted experimenter-participant interactions. On the other hand, other methodological elements may have inadvertently contributed to random measurement error by increasing mental fatigue. These include the use of the memory-load task and the intensive nature of data collection (i.e., total of nine measures completed during the same session). Both of these methodological elements were introduced by design, the former to increase the likelihood that task responses would reflect implicit processes and the latter to enable the evaluation of the measures in the same sample while avoiding both participant attrition and error variance due to multiple visits (e.g., variation in sleep patterns, hydration and nutrition, health status, life stress, environmental conditions). However, it is possible that participant fatigue was the downside of these choices. This point is discussed further below.

In interpreting these internal-consistency estimates, readers should also keep in mind that reliability is not an inherent and invariant property of the measures; rather, it is measurement-specific. Therefore, the coefficients obtained from the present sample cannot be used as evidence that a certain measure "has" or "does not have" reliability, "is" or "is not" reliable. Rather, researchers are encouraged to calculate and report estimates of internal consistency (and test-test reliability, where possible) from their samples and under the unique measurement conditions of their studies. Finally, readers should keep in mind that unreliability limits validity. According to classic test theory, only the reliable portion of the variance of scores on a given test can be valid.

5.3. Validity

The AAT was found to be significantly and meaningfully correlated with self-reported exercise behavior and situated decisions to exercise. This was true for both the response time to make an initial movement and the response time to complete the pulling or pushing motion. However, it should be noted that the direction of the association raises an interpretational challenge (also see van Dantzig, Pecher, & Zwaan, 2008). Faster pulling movements were associated with less exercise behavior. When participants pulled the joystick, the image on the computer screen zoomed in, creating the illusion that the image was moving closer to the participant. Thus, it was originally postulated that pulling would indicate an approach tendency, as is commonly assumed in the literature (Cacioppo, Priester, & Berntson, 1993; Chen & Bargh, 1999; Cheval et al., 2016; Rinck & Becker, 2007; Wiers, Rinck, Dictus, & van den Wildenberg, 2009). However, it is plausible that pulling may be indicative of an avoidance tendency, despite the zooming-in of the on-screen image, because, when a participant pulls on the joystick, she or he pulls *away* from the computer screen. In contrast, pushing the joystick may indicate an effort to lean toward the image on the screen. This would be analogous to postural sway in the anterior (forward) direction when viewing pleasant images, a finding typically interpreted as indicative of approach (Eerland, Guadalupe, Franken, & Zwaan, 2012). To avoid this interpretational ambiguity, investigators may consider using a manikin variant of the AAT, which has demonstrated evidence of criterion validity and acceptable internal consistency in the context of physical activity and sedentary behavior (Cheval et al., 2015, 2016) and may be preferable to the joystick variant used in the present study (Krieglmeyer & Deutsch, 2010).

The possibility that the associations of the AAT variables with exercise behavior and situated decisions to exercise may still be spurious, despite being statistically significant and meaningful ($r \geq 0.32$, $p \leq .002$), cannot be discounted. However, *post hoc* Bayesian analysis, suggested by a reviewer, indicates that, compared to the null hypotheses (i.e., no associations), the data are 16.83–53.63 times more likely under the hypotheses that the AAT variables are correlated with self-reported exercise behavior and situated decisions to exercise. It should also be noted that the correlations of the AAT with exercise and situated decisions in this study were higher than many, if not most, previous correlations between measures of automatic exercise associations and validation criteria (e.g., Antoniewicz & Brand, 2014; Calitri et al., 2009; Conroy et al., 2010; Eves, Scott, Hoppé, & French, 2007; Hyde et al., 2010; Rebar et al., 2015). Conceivably, the AAT used in the present study (and manikin variants used in other studies; Cheval et al., 2015, 2016) may be specifically indicative of *behavioral* approach-avoidance tendencies, more so than other measures, and this may explain the relation of the AAT with self-reported exercise *behavior* and situated decisions to choose exercise *behavior*.

5.4. Critical reflections

While the measures of automatic exercise associations exhibited high efficiency (and can, therefore, be presumed to reflect automaticity), most showed low internal consistency and their correlations with the validation criteria were underwhelming. The AAT and the PSC-IAT were the only measures found to exhibit meaningful relations with any of the validation criteria. For the AAT, the direction of the association is ambiguous and will require additional research to establish whether pushing suggests approach and pulling suggests avoidance, or vice versa. The internal consistency of the PSC-IAT was satisfactory, but the negligible correlations with other validation criteria calls for a cautious interpretation of its correlation with the affective slope.

Overall, the absence of a consistent pattern of relationships between measures of automatic exercise associations and the validation criteria represents a critical challenge for proposals favoring a transition to dual-process models of exercise and physical activity behavior. It could be argued that the potential of the much-needed theoretical advancement could be undermined by a measurement technology that is still in need of considerable refinement.

At this point, the absence of a systematic pattern of associations across nine measures of automatic exercise associations with validation criteria in the present study raises some challenging questions. This finding clearly stands in contrast to the observation that significant associations between measures of automatic exercise associations and various operationalizations of exercise have been reported "in the vast majority of studies" in the published literature (Schinkoeth & Antoniewicz, 2017, p. 13). One possibility is that the findings reported here, despite a sample size that should entail some degree of stability of the correlational estimates, are statistically anomalous or can be attributed to certain methodological elements (as detailed below). Another possibility is that published studies showing consistent associations between measures of automatic exercise associations and exercise behavior reflect, at least in part, selective publication. The latter possibility is rendered perhaps more probable by the observation that different studies (even studies conducted by the same research group) have employed different measures without articulating a rationale for switching from one measure to another.

5.5. Possible methodological explanations for null correlations

It is noteworthy that no reliable correlations were found between the measures of automatic exercise associations and the novel theory-derived validation criteria of experienced and recalled affect, implemented for the first time in the present study. If true, this finding challenges theoretical proposals suggesting that affective responses to exercise shape automatic exercise associations (Brand & Ekkekakis, 2018; Conroy & Berry, 2017; Schinkoeth & Antoniewicz, 2017). Before drawing the conclusion that this theoretical proposition is wrong, however, it would be prudent to consider methodological explanations.

First, as the first endeavor to evaluate the relation between affective responses to exercise and automatic exercise associations, the present study relied on a simplifying assumption, namely that affective responses to a level of exercise intensity (i.e., VT) known to result in substantial interindividual variability in affective responses could serve as a meaningful proxy of a history of affective experiences with exercise in general. It is certainly possible that this simplifying assumption is false. In that case, a reasonable next step would be to sample affective responses from multiple instances of exercise performed under ecologically valid and diverse conditions using an experience sampling protocol (e.g., ecological momentary assessment; see Shiffman, Stone, & Hufford, 2008).

Second, as mentioned earlier, it is possible that the validity estimates of measures of automatic exercise associations in the present study were underestimated due to participant fatigue. In order to create an equitable comparison, each participant completed all nine tasks in random order during the same session. It is possible that this caused fatigue, increasing random measurement error, and consequently a decline in validity estimates as testing progressed. To explore this possibility, we conducted a *post hoc* analysis (see Supplementary Material, Table S6) on the subsample of participants who completed each task early (first, second, third, or fourth). The results did not indicate that fatigue was a factor since there was no evidence that validity coefficients were systematically stronger when the tasks were completed early. While some validity coefficients were slightly higher (six of eight correlations between the AAT and validation criteria approached or exceeded $|0.30|$), nearly half (48.8%) were negligible ($r \leq 0.10$).

Still, the possibility that validity estimates were attenuated cannot be fully discounted. In the future, researchers may want to contemplate the pros and cons of alternative approaches. These include comparative validation studies that have a more narrow scope, requiring participants to complete a smaller number of measures of automatic exercise associations, or studies in which participants are required to visit the laboratory on multiple occasions while attempting to control extraneous confounds (e.g., sleep, nutrition, health, stress). These are options that researchers may consider, with the understanding that such decisions require tradeoffs.

Third, it is possible that the timing of the measurement of validation criteria and automatic exercise associations may have attenuated validity estimates. In the present study, self-reported exercise behavior, explicit affective attitudes, affective responses to exercise, and situated decisions to exercise were measured between one and three weeks prior to the measurement of automatic associations. In contrast, in other studies, it is common for measurements of validation criteria and automatic exercise associations to be obtained during the same session. We believe that this practice is ill-advised in the context of a validation study because it fails to create the "temporal separation" (Podsakoff et al., 2003, 2012) that is essential in order to attenuate common method bias. For example, back-to-back measurement of automatic associations and explicit constructs can inflate their congruence (Scherer & Schott, 2012). Given the aim of the present study, the attenuation of common method bias was deemed crucially important. While, in theory, neither the validation criteria (i.e., exercise behavior during a "usual" week, affective responses to a standardized exercise bout, situated decisions to exercise based on hypothetical scenarios) nor automatic exercise associations should be expected to change to a meaningful extent within a three-week time frame, it is possible that the relationship between the validation criteria and automatic exercise associations is time-dependent. If this is the case, the shape of this relationship remains unknown. In the future, investigators may wish to further explore this question.

Fourth, the memory-load task was used to engage controlled processes (working memory), thereby increasing reliance on efficient, implicit processes during task performance (Feldon, 2007; Friese, Hofmann, & Wänke, 2009; Hermans et al., 2000). Previous investigations have similarly incorporated cognitive-load tasks for the same reason (e.g., Gibbons, Seib-Pfeifer, Koppehele-Gossel, & Schnuerch, 2018; Klauer & Teige-Mocigemba, 2007). However, while the use of the memory-load task may have strengthened the argument that performance on the automatic exercise association tasks relied on implicit processes, it is possible that this parallel task also interfered with performance on measures of automatic associations. It should be noted that cognitive-load manipulations have been used previously mainly with the AMP and EDT (Gibbons et al., 2018; Hermans et al., 2000; Klauer & Teige-Mocigemba, 2007). Other procedures may depend to some extent, in addition to implicit processes, on executive or controlled processes (e.g., attentional resources). Therefore, it is plausible that the reduced cognitive capacity caused by the memory-load manipulation may have interfered with the execution and outcomes of these tasks, thereby limiting their validity. To our knowledge, no other investigation focusing on automatic exercise associations has used a concurrent cognitive-load task.

Fifth, as alluded to earlier, the possible mental fatigue resulting from the combination of the memory-load task and the completion of all nine measures of automatic exercise associations during the same session may have contributed to the low coefficients of internal consistency exhibited by six of the measures. These methodological elements differed from previous investigations in the context of exercise and may explain why the internal consistency of several measures was lower than previously reported. In turn, measurement error likely attenuated some of the validity estimates.

Finally, error of measurement in some of the validation criteria may have also led to attenuated correlations with indices of automatic exercise associations. In particular, although validated, the self-report measure of exercise behavior (IPAQ-SF) is still not ideal for research purposes (Lee, Macfarlane, Lam, & Stewart, 2011). Although encumbered by different challenges, it is possible that an objective measure of exercise behavior (i.e., accelerometer) would have yielded more significant and meaningul corrleations with measures of automatic exercise associations (e.g., Cheval et al., 2015; Chevance, Caudroit et al., 2018). It is also conceivable that the specific nature of the instructions used for the IPAQ-SF (i.e., reference to bouts with a minimum duration of 10 min) and the measure of explicit affective attitude (i.e., "exercising at least 30 min per day for at least 5 days") may have made these measures not directly aligned with the measures of automatic exercise associations, which referred to the stimulus concept of "exercise" in general, with no such restrictions.

6. Conclusions and suggestions for future investigations

Researchers willing to pursue future psychometric evaluations of measures of implicit exercise associations should contemplate the role of exercise behavior as a validation criterion. Exercise is not only an extremely complex and multidetermined behavior (such that automatic exercise associations may influence it only through interactions with other personal and extrapersonal factors) but it is also notoriously difficult to measure with any accuracy at the individual level. Researchers who insist on this variable due to its obvious public-health interest should consider employing a multimethod approach that combines self-reports and objective assessments (e.g., accelerometers, attendance records). Here, we also adopted the suggestion by Brand and Schweizer (2015) that "situated decisions" can be regarded as a more proximal "functional link" between implicit processes and the distal outcome of exercise behavior. While this suggestion has theoretical appeal, in the present sample, scores on the Situated Decisions to Exercise Questionnaire (SDEQ) exhibited meaningful and reliable correlations only with the AAT and these were of a magnitude comparable to those of the IPAQ-SF. Therefore, whether measures of implicit associations should be expected to relate to situated decisions more closely or robustly than to measures of exercise behavior remains to be demonstrated empirically.

It is possible that implicit processes manifest themselves primarily as behavioral urges and may thus be more strongly associated with general physical activity (i.e., unplanned and unstructured movement not specifically aimed at fitness promotion) or with sedentary behavior (Cheval et al., 2015) rather than exercise *per se*. In support of this position, Hagger and Chatzisarantis (2014) argued that "formal types of exercise that involve equipment, transport, venue, costs, and other considerations, require considerable planning" and are therefore likely to be "a behavior that is largely under the control of the reflective system" (p. 65). While this is a plausible argument, it should be kept in mind that several published studies have reported that measures of implicit associations correlated significantly with both exercise and physical activity (Schinkoeth & Antoniewicz, 2017).

It is also important to emphasize that, unlike most other validation scenarios, the correspondence between automatic exercise associations and validation criteria may not be direct and straightforward. Dual-process theories, for example, postulate that exercise behavior is always jointly determined by *interacting* implicit and explicit processes (Brand & Ekkekakis, 2018; Conroy & Berry, 2017; Williams & Evans, 2014) and these interactions may be moderated by a multitude of other situational and dispositional variables (Brand & Ekkekakis, 2018; Cheval, Sarrazin, Isoard-Gautheur, Radel, & Friese, 2016; Chevance, Stephan, et al., 2018; Friese, Hofmann, & Schmitt, 2009; Padin, Emery, Vasey, & Kiecolt-Glaser, 2017). Therefore, taking into account such factors as the availability of self-regulatory resources, executive capacity, impulsivity, need for cognition, and mood may provide useful insights about the conditions in which automatic associations relate to behavior and other validation criteria.

In conclusion, in the first study designed to examine the internal consistency and validity of nine measures of implicit exercise associations, we found that the internal consistency coefficients of the AMP, EDT, and PSC-IAT were satisfactory, and only the AAT was significantly and meaningfully related to validation criteria (i.e., self-reported exercise and situated decisions to exercise). These findings underscore the inherent challenge of tapping implicit processes via behavioral tasks, as well as the challenge of investigating emerging, and promising, dual-process theoretical models of exercise and physical activity behavior. They also highlight the importance of consistent and transparent reporting (e.g., stimuli, trial details, task details, scoring procedures) and the urgent necessity of additional foundational psychometric work.

Acknowledgements

The authors wish to express their gratitude to all those who participated in this study. They also recognize the thorough and thoughtful comments and suggestions offered eponymously by Dr Boris Cheval, Dr Guillaume Chevance, and Dr Amanda Rebar during the peer-review process. All remaining errors are solely the responsibility of the authors.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.psychsport.2018.12.005.

References

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*, 179–211.
American College of Sports Medicine (2018). *ACSM's guidelines for exercise testing and prescription* (10th ed.). Philadelphia, PA: Wolters Kluwer.
Antoniewicz, F., & Brand, R. (2014). Automatic evaluations and exercise setting preference in frequent exercisers. *Journal of Sport & Exercise Psychology, 36*, 631–636.
Antoniewicz, F., & Brand, R. (2016a). Dropping out or keeping up? Early-dropouts, late-dropouts, and maintainers differ in their automatic evaluations of exercise already before a 14-week exercise course. *Frontiers in Psychology, 7* 00838.
Antoniewicz, F., & Brand, R. (2016b). Learning to like exercising: Evaluative conditioning changes automatic evaluations of exercising and influences subsequent exercising behavior. *Journal of Sport & Exercise Psychology, 38*, 138–149.
Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition. In (2nd ed). R. S. Wyer, & T. K. Srull (Vol. Eds.), *Handbook of social cognition: vol. 1*, (pp. 1–40). Hillsdale, NJ: Erlbaum.
Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B, 57*, 289–300.
Benjamini, Y., & Yekutieli, D. (2005). False Discovery Rate–Adjusted multiple confidence intervals for selected parameters. *Journal of the American Statistical Association, 100*, 71–81.
Berry, T. R. (2006). Who's even interested in the exercise message? Attentional bias for exercise and sedentary-lifestyle related words. *Journal of Sport & Exercise Psychology, 28*, 4–17.
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.
Brand, R., & Ekkekakis, P. (2018). Affective-reflective theory of physical inactivity and exercise: Foundations and preliminary evidence. *German Journal of Exercise and Sport Research, 48*(1), 48–58.
Brand, R., & Schweizer, G. (2015). Going to the gym or to the movies? Situated decisions as a functional link connecting automatic and reflective evaluations of exercise with exercising behavior. *Journal of Sport & Exercise Psychology, 37*(1), 63–73.
Cacioppo, J. T., Priester, J. R., & Berntson, G. G. (1993). Rudimentary determinants of attitudes: II. Arm flexion and extension have differential effects on attitudes. *Journal of Personality and Social Psychology, 65*(1), 5–17.
Calitri, R., Lowe, R., Eves, F. F., & Bennett, P. (2009). Associations between visual attention, implicit and explicit attitude and behaviour for physical activity. *Psychology and Health, 24*, 1105–1123.
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.
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.
Cheval, B., Sarrazin, P., Isoard-Gautheur, S., Radel, R., & Friese, M. (2016a). How impulsivity shapes the interplay of impulsive and reflective processes involved in objective physical activity. *Personality and Individual Differences, 96*, 132–137.
Cheval, B., Sarrazin, P., Pelletier, L., & Friese, M. (2016b). Effect of retraining approach-avoidance tendencies on an exercise task: A randomized controlled trial. *Journal of Physical Activity and Health, 13*, 1396–1403.
Chevance, G., Caudroit, J., Henry, T., Guerin, P., Boiché, J., & Héraud, N. (2018a). Do implicit attitudes toward physical activity and sedentary behavior prospectively predict objective physical activity among persons with obesity? *Journal of Behavioral Medicine, 41*(1), 31–42.
Chevance, G., Stephan, Y., Héraud, N., & Boiché, J. (2018b). Interaction between self-regulation, intentions and implicit attitudes in the prediction of physical activity among persons with obesity. *Health Psychology, 37*(3), 257–261.
Chevance, G., Héraud, N., Guerrieri, A., Rebar, A., & Boiché, J. (2017). 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. *Psychology of Sport and Exercise, 31*, 70–78.
Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155–159.
Conroy, D. E., & Berry, T. R. (2017). Automatic affective evaluations of physical activity. *Exercise and Sport Sciences Reviews, 45*(4), 230–237.
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, 39*, 112–118.
Craig, C. L., Marshall, A. L., Sjöström, M., Bauman, A. E., Booth, M. L., Ainsworth, B. E., ... Oja, P. (2003). International physical activity questionnaire: 12-country reliability and validity. *Medicine & Science in Sports & Exercise, 35*, 1381–1395.
Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281–302.
van Dantzig, S., Pecher, D., & Zwaan, R. A. (2008). Approach and avoidance as action effects. *The Quarterly Journal of Experimental Psychology, 61*(9), 1298–1306.
Eerland, A., Guadalupe, T. M., Franken, I. H., & Zwaan, R. A. (2012). Posture as index for approach-avoidance behavior. *PLoS One, 7*(2), e31291.
Ekkekakis, P. (2003). Pleasure and displeasure from the body: Perspectives from exercise. *Cognition & Emotion, 17*, 213–239.
Ekkekakis, P. (2013). Pleasure from the exercising body: Two centuries of changing outlooks in psychological thought. In P. Ekkekakis (Ed.). *Routledge handbook of physical activity and mental health* (pp. 35–56). New York: Routledge.
Ekkekakis, P. (2017). People have feelings! Exercise psychology in paradigmatic

transition. *Current Opinion in Psychology, 16*, 84–88.
Ekkekakis, P., & Dafermos, M. (2012). Exercise is a many-splendored thing but for some it does not feel so splendid: Staging a resurgence of hedonistic ideas in the quest to understand exercise behavior. In E. O. Acevedo (Ed.). *The Oxford handbook of exercise psychology* (pp. 295–333). New York: Oxford University Press.
Ekkekakis, P., Parfitt, G., & Petruzzello, S. J. (2011). The pleasure and displeasure people feel when they exercise at different intensities: Decennial update and progress towards a tripartite rationale for exercise intensity prescription. *Sports Medicine, 41*(8), 641–671.
Ekkekakis, P., & Zenko, Z. (2016). Escape from cognitivism: Exercise as hedonic experience. In M. Raab, P. Wylleman, R. Seiler, A. M. Elbe, & A. Hatzigeorgiadis (Eds.). *Sport and exercise psychology research from theory to practice* (pp. 389–414). London: Academic Press.
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*, 571–585.
Feldon, D. F. (2007). Cognitive load and classroom teaching: The double-edged sword of automaticity. *Educational Psychologist, 42*(3), 123–137.
Friese, M., Hofmann, W., & Schmitt, M. (2009). When and why do implicit measures predict behaviour? Empirical evidence for the moderating role of opportunity, motivation, and process reliance. *European Review of Social Psychology, 19*(1), 285–338.
Friese, M., Hofmann, W., & Wänke, M. (2009). Control me or I will control you: Impulses, trait self-control, and the guidance of behavior. *Journal of Research in Personality, 43*, 795–805.
Gaskill, S. E., Ruby, B. C., Walker, A. J., Sanchez, O. A., Serfass, R. C., & Leon, A. S. (2001). Validity and reliability of combining three methods to determine the ventilatory threshold. *Medicine & Science in Sports & Exercise, 33*(11), 1841–1848.
Gawronski, B., Deutsch, R., & Banse, R. (2011). Response interference tasks as indirect measures of automatic associations. In K. C. Klauer, A. Voss, & C. Stahl (Eds.). *Cognitive methods in social psychology* (pp. 78–123). New York: Guilford Press.
Gibbons, H., Seib-Pfeifer, L.-E., Koppehele-Gossel, J., & Schnuerch, R. (2018). Affective priming and cognitive load: Event-related potentials suggest an interplay of implicit affect misattribution and strategic inhibition. *Psychophysiology, 55*, e13009.
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*, 197–216.
Hagger, M. S., & Chatzisarantis, N. L. D. (2014). An integrated behavior change model for physical activity. *Exercise and Sport Sciences Reviews, 42*(2), 62–69.
Hardy, C. J., & Rejeski, W. J. (1989). Not what, but how one feels: The measurement of affect during exercise. *Journal of Sport & Exercise Psychology, 11*, 304–317.
Hermans, D., Crombez, G., & Eelen, P. (2000). Automatic attitude activation and efficiency: The fourth horseman of automaticity. *Psychologica Belgica, 40*, 3–22.
Hyde, A. L., Doerksen, S. E., Ribeiro, N. F., & Conroy, D. E. (2010). The independence of implicit and explicit attitudes toward physical activity: Introspective access and attitudinal concordance. *Psychology of Sport and Exercise, 11*, 387–393.
Kahneman, D., Wakker, P. P., & Sarin, R. (1997). Back to bentham? Explorations of experienced utility. *Quarterly Journal of Economics, 112*, 375–406.
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*, 16–32.
Keselman, H. J., Cribbie, R., & Holland, B. (2002). Controlling the rate of Type I error over a large set of statistical tests. *British Journal of Mathematical and Statistical Psychology, 55*, 27–39.
Klauer, K. C., & Teige-Mocigemba, S. (2007). Controllability and resource dependence in automatic evaluation. *Journal of Experimental Social Psychology, 43*(4), 648–655.
Krieglmeyer, R., & Deutsch, R. (2010). Comparing measures of approach-avoidance behaviour: The manikin task vs. two versions of the joystick task. *Cognition & Emotion, 24*(5), 810–828.
Lee, P. H., Macfarlane, D. J., Lam, T. H., & Stewart, S. M. (2011). Validity of the international physical activity questionnaire short form (IPAQ-SF): A systematic review. *International Journal of Behavioral Nutrition and Physical Activity, 8*, 115.
Lishner, D. A., Cooter, A. B., & Zald, D. H. (2008). Addressing measurement limitations in affective rating scales: Development of an empirical valence scale. *Cognition & Emotion, 22*, 180–192.
Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin, 132*, 297–326.
Moors, A., Spruyt, A., & De Houwer, J. (2010). In search of a measure that qualifies as implicit: Recommendations based on a decompositional view of automaticity. In B. Gawronski, & K. B. Payne (Eds.). *Handbook of implicit social cognition: Measurement, theory, and application* (pp. 19–37). New York: Guilford Press.
Müller, S., Rothermund, K., & Wentura, D. (2016). Relevance drives attention: Attentional bias for gain- and loss-related stimuli is driven by delayed engagement. *Quartlery Journal of Experimental Psychology, 69*(4), 752–763.
Nakagawa, S. (2004). A farewell to Bonferroni: The problems of low statistical power and publication bias. *Behavioral Ecology, 15*, 1044–1045.
Olson, M. A., & Fazio, R. H. (2004). Reducing the influence of extrapersonal associations on the implicit association test: Personalizing the IAT. *Journal of Personality and Social Psychology, 86*(5), 653–667.
Padin, A. C., Emery, C. F., Vasey, M., & Kiecolt-Glaser, J. K. (2017). Self-regulation and implicit attitudes toward physical activity influence exercise behavior. *Journal of Sport & Exercise Psychology, 39*, 237–248.
Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology, 89*, 277–293.
Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879–903.
Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology, 63*, 539–569.
Rebar, A. L., Dimmock, J. A., Jackson, B., Rhodes, R. E., Kates, A., Starling, J., & Vandelanotte, C. (2016). A systematic review of the effects of non-conscious regulatory processes in physical activity. *Health Psychology Review, 10*(4), 395–407.
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*, 96–105.
Rinck, M., & Becker, E. S. (2007). Approach and avoidance in fear of spiders. *Journal of Behavior Therapy and Experimental Psychiatry, 38*, 105–120.
Rothman, K. J. (1990). No adjustments are needed for multiple comparisons. *Epidemiology, 1*, 43–46.
Rudman, L. A. (2004). Sources of implicit attitudes. *Current Directions in Psychological Science, 13*(2), 79–82.
Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review, 110*, 145–172.
Scherer, D., & Schott, J. P. (2012). The affect misattribution task impacts future implicit and explicit judgments. *Social Cognition, 30*(5), 537–563.
Schinkoeth, M., & Antoniewicz, F. (2017). Automatic evaluations and exercising: Systematic review and implications for future research. *Frontiers in Psychology, 8*, 2103.
Sheeran, P., Gollwitzer, P. M., & Bargh, J. A. (2013). Nonconscious processes and health. *Health Psychology, 32*(5), 460–473.
Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology, 4*, 1–32.
Wagenmakers, E.-J., van der Maas, H. L. J., & Grasman, R. P. P. P. (2007). An EZ-diffusion model for resposne time and accuracy. *Psychonomic Bulletin & Review, 14*(1), 3–22.
Wiers, R. W., Rinck, M., Dictus, M., & van den Wildenberg, E. (2009). Relatively strong automatic appetitive action-tendencies in male carriers of the OPRM1 G-allele. *Genes, Brain and Behavior, 8*, 101–106.
Williams, D. M., & Evans, D. R. (2014). Current emotion research in health behavior science. *Emotion Review, 6*(3), 277–287.
Zenko, Z., Ekkekakis, P., & Ariely, D. (2016). Can you have your vigorous exercise and enjoy it too? Ramping intensity down increases postexercise, remembered, and forecasted pleasure. *Journal of Sport & Exercise Psychology, 38*, 149–159.