Addictive Behaviors 135 (2022) 107456

Contents lists available at ScienceDirect

Addictive Behaviors

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

Trial-level bias score versus mean bias score: Comparison of the reliability and external validity using dot-probe task among daily smokers

Min-Jeong Yang [a,b,*], Allison M. Borges [c], Noah N. Emery [d], Teresa M. Leyro [a]

[a] Department of Psychology, Rutgers University, Piscataway, NJ 08854, USA
[b] Department of Health Outcomes and Behavior, Moffitt Cancer Center, 4115 E. Fowler Ave, Tampa, FL 33617, USA
[c] Department of Behavioral Health, The Corporal Michael J. Crescenz VA Medical Center, 3900 Woodland Ave, Philadelphia, PA 19104, USA
[d] Department of Psychology, Colorado State University, 210 Behavioral Sciences Building, 1876, Campus Delivery, Fort Collins, CO 80523-1876, USA

ARTICLE INFO

Keywords:
Cigarette smoking
Attentional bias
Trial-level bias score
Reliability
External validity

ABSTRACT

Introduction: Attentional bias (AB) is an individual difference risk factor that represents the extent to which cigarette cues capture one's attention. AB is typically indexed by mean bias score (MBS), theoretically assuming that AB is static. However, poor reliability of MBS has threatened valid interpretation of the results on AB. Based on observed trial-by-trial temporal fluctuation and variability of attentional allocation, trial-level bias score (TLBS) has been introduced as an alternative index with evidence of better psychometric properties in various populations, as compared to MBS. However, such evidence is limited among daily smokers. The current study aimed to replicate and extend extant findings in a sample of daily smokers by hypothesizing that TLBS, as compared to MBS, would demonstrate superior reliability and external validity.
Methods: Forty-eight daily smokers completed self-reports, ad-libitum smoking, and a dot-probe task three times, which was comprised of 36 pairs of pictorial stimuli of cigarette and neutral cues, yielding 144 total trials.
Results: The TLBS demonstrated superior internal (range intra class correlation *[ICC]* = 0.79–0.95) and test-retest reliability (range *ICC* = 0.64–0.88) compared to MBS (range *ICC* = 0.31–0.40 and 0.06–0.16, respectively). However, few significant relations between either the MBS or TLBS and measures of biobehavioral and self-report indices of smoking reinforcement were observed.
Conclusions: The current findings demonstrate that TLBS, as compared to MBS, is a more reliable measure of AB among daily smokers, while evidence of its external validity is limited.

1. Introduction

Cigarette-related attentional bias (AB), i.e., biased allocation of attention to cigarette cues over neutral cues, is observed among current and former smokers (Masiero et al., 2019; Rehme et al., 2018). AB has traditionally been conceptualized as relatively stable individual difference risk factor that represents the extent to which cigarette cues capture one's attention as a result of classical conditioning and sensitization of the brain mesolimbic system eliciting a "wanting" response to cigarette-related cues (Berridge & Robinson, 2016). The literature in both psychopathology and addiction has extensively documented the evidence of AB in both psychopathology (e.g., bias to threat-related cues; Bardeen & Orcutt, 2011) and addiction (Field & Cox, 2008), moving the field to the development of AB modification as a novel targeted treatment approach (Boffo et al., 2019; Bunnell et al., 2013). Nevertheless, an observed lack of consistency in the relation between AB and smoking, and the impact of AB modification in changing smoking (Begh et al., 2015) is likely due to poor psychometric properties of conventional AB index (Ataya et al., 2012; Drobes et al., 2019).

Mean bias score (MBS) is a widely used index of AB. However, its poor reliability and its theoretical underpinnings of AB being static over the course of the AB task have been criticized (Zvielli et al., 2015). Despite efforts to enhance the reliability of MBS (e.g., outlier methods; Price et al., 2015), psychometric properties of the MBS have continued to raise questions (Emery & Simons, 2015; Jones et al., 2018; Molloy & Anderson, 2020; Rodebaugh et al., 2016; Schäfer et al., 2016), although

* Corresponding author at: Tobacco Research and Intervention Program, H. Lee Moffitt Cancer Center & Research Institute, 4115 E Fowler Avenue, Tampa, FL 33617, USA.
E-mail addresses: min-jeong.yang@moffitt.org (M.-J. Yang), allison.borges@va.gov (A.M. Borges), noah.emery@colostate.edu (N.N. Emery), teresa.leyro@rutgers.edu (T.M. Leyro).

https://doi.org/10.1016/j.addbeh.2022.107456
Received 8 February 2022; Received in revised form 29 July 2022; Accepted 1 August 2022
Available online 4 August 2022
0306-4603/© 2022 Elsevier Ltd. All rights reserved.

some recent findings demonstrated novel ways to improve the psychometric properties of the MBS using variants of AB task (Gladwin et al., 2021; Grafton et al., 2021; Heitmann et al., 2021). Using unreliable and unstable MBS is problematic, as it threatens valid interpretation of the results on AB (Rodebaugh et al., 2016), in particular, if a change in AB is proposed as an intervention target or mechanisms (Price et al., 2015; Rodebaugh et al., 2016). Given growing concerns, an alternative index, trial-level bias score (TLBS) has been developed (Zvielli et al., 2015).

The TLBS is a novel AB measure that reflects the temporal fluctuation and variability of attentional allocation trial-by-trial (Zvielli et al., 2015). The major differences between MBS and TLBS are threefold: (1) The theoretical assumption (stable vs fluctuating), (2) the capability to capture various phenomena of AB in TLBS (e.g., AB toward vs away), and (3) its superior psychometric properties (TLBS > MBS; Carlson & Fang, 2020; Jones et al., 2018; Molloy & Anderson, 2020; Vervoort et al., 2021; Zvielli et al., 2015). In particular, evidence supports the superior reliability of TLBS (versus MBS) in a nonclinical sample (Carlson & Fang, 2020), individuals with phobias (Zvielli et al., 2015), social anxiety disorder (Molloy & Anderson, 2020), and depression (Beevers et al., 2019), youth with and without obesity (Vervoort et al., 2021), and among alcohol users (Jones et al., 2018). Notably, TLBS is relatively unaffected by various data cleaning procedures, while MBS is sensitive to such procedures, which raises concern for inflated reliability (Molloy & Anderson, 2020). Recently, the validity of TLBS indices has been called into question given the potential impact of general reaction time variability on TLBS, suggesting findings should be interpreted with caution (Carlson & Fang, 2020; Kruijt et al., 2016; Vervoort et al., 2021). Nevertheless, reliability of the TLBS in smokers is limited to documentation of its split-half reliability in a sample of deprived smokers (Zvielli et al., 2015).

Further, despite overall evidence on the superior reliability of the TLBS (vs MBS), evidence on its external validity remains limited. For example, findings between the TLBS and theoretically relevant constructs such as trait or state anxiety measures (Carlson & Fang, 2020), depressive symptoms (Beevers et al., 2019), alcohol consumption and craving (Jones et al., 2018), and weight and body mass index (Vervoort et al., 2021) have been equivocal despite some evidence in spider phobia (Zvielli et al., 2015) and posttraumatic stress disorder (Schäfer et al., 2016). Thus, contrary to its overall superior reliability, evidence of the external validity of the TLBS is still in question. Therefore, examining psychometric properties of both the TLBS and MBS in smokers is warranted and may help to clarify discrepancies in the existing literature.

The overarching aim of the current study is to examine the reliability and external validity of the TLBS versus MBS among daily smokers. Daily smokers completed a modified dot-probe task three times; each trial was separated by an affect manipulation. To examine the possible effect of the affect manipulation on AB, the effect of the affect manipulation condition on AB indices (i.e., TLBS and MBS) was examined. Next, internal reliability of each AB trial and test–retest reliability were tested. Finally, the associations between both TLBS and MBS and smoking variables were tested to further examine the relative external validity of each AB index. Here, consistent with the previous literature and the incentive sensitization theory (Berridge & Robinson, 2016), smoking variables included both self-report and biobehavioral indices of smoking reinforcement as well as cigarette dependence. We hypothesized that the TLBS would demonstrate superior internal and test–retest reliability compared to MBS. We also hypothesized that the TLBS would evidence stronger associations to smoking variables relative to the MBS.

2. Methods

2.1. Participants

Fifty daily smokers were recruited for a parent study that primarily aimed to examine the relation between cognitive-affective and physiological processes and smoking behaviors. The current study included all recruited 50 participants for the parent study. Inclusion criteria established for the parent study were; 1) $\geq$ 5 cigarettes per day (CPD) for the past one year; 2) 19–50 years old; 3) normal or corrected-to-normal color vision; 4) fluent in English; 5) ability to work with computer; 6) verification of smoking status via carbon monoxide (CO) analysis of breath sample > 8 ppm (Javors et al., 2005). Exclusion criteria established for the parent study included; 1) $\geq$ 35 body mass index; 2) evidence of current or past substance use disorder; 3) use of any smoking cessation aids or medication; 4) use of other tobacco or nicotine products for regular use; 5) current or past psychosis; 6) current suicidal or homicidal ideation; 7) inability to provide written informed consent; 8) visual or hearing impairments that interfere with the completion of computerized tasks. Participants were excluded on the basis of age (i.e., > 50) and body mass index (i.e., > 35) given the effect these parameters may have on a central physiological predictor of interest (i.e., heart rate variability) included in the parent investigation (e.g., age; Abhishekh et al., 2013; Sinnreich et al., 1998; body mass index; Koenig et al., 2014; Molfino et al., 2009). Of note, due to assessment errors, one participant with body mass index > 35 was accidentally enrolled in the parent study but retained for the current investigation given our analyses did not employ the physiological index affected. After data cleaning, data from 48 participants were included in the analyses of the current study. See supplementary material 1 for sample size justification.

2.2. Procedure

Following a brief phone screen, eligible participants were instructed to smoke as usual and asked to complete an online survey via Qualtrics, which included a separate electronic consent for participation in online survey completion and Cigarette Purchase Task (MacKillop et al., 2008) prior to their laboratory visit, and were subsequently invited to the lab for their appointment. Upon the arrival to the lab, participants completed written informed consent for the in-person portion of the study. Smoking status was confirmed through both Timeline Followback and biochemical verification via carbon monoxide (CO) analysis of breath sample ($\geq$8 ppm; Javors et al., 2005). Next, participants completed self-report assessments to gather additional demographic information, past one-month substance use, and smoking patterns and history. Following completion of self-report forms, participants were asked to smoke a cigarette of their usual brand in a designated smoking room to standardize nicotine withdrawal and craving. Participants smoked the cigarette using the Clinical Research Support System (CReSS) pocket device that recorded smoking topography. Next, participants completed a battery of self-reports that measured state affect, withdrawal symptoms, and craving, and the modified Cigarette Evaluation Questionnaire (Cappelleri et al., 2007). Participants then completed two brief behavioral tasks and were affixed to physiological sensors. At this point, participants were randomized to one of two affect manipulation conditions (conditions A and B), completed two brief computerized tasks, and then the AB tasks three times, separated by affect manipulations (supplementary Fig. 1). Finally, participants completed a smoking analog task. Self-report ratings of state affect and craving were collected between each task phase (i.e., AB + affect manipulation). The full protocol of the study lasted approximately 3.5 h and participants were compensated up to $76. The current study reports the results of the measures listed in the Measures section.

2.3. Apparatus and material

2.3.1. Modified dot-probe task

The task included 164 experimental trials. A set of 36 pairs of smoking and non-smoking neutral images (S-N trial) and 5 pairs of neutral images (NT-NT trial) were randomly presented four times each (41 trials × 4 blocks). The image presentation order was randomized across participants. Trials began with a presentation of a central fixation cross (500 ms), followed by a left–right bilateral presentation of paired images (500 ms), after which, a target probe appeared on either right or left side of the screen and remained until the participants identified which side the target probe was presented on using the keyboard (e.g., Beevers et al., 2019; Zvielli et al., 2015). A distractor probe (one dot) appeared opposite the target probe (two dots) in order to increase the AB task demand (Garland et al., 2012). Responses were recorded using a keyboard by pressing either 'f (left)' or 'j (right)' keys according to the location of target probe presentation. For the S and N image pairs, the number of presentations was fixed to nine in each cell of the 2 × 2 condition (location [right vs left] × congruency [congruent vs incongruent]) in each block. That is, the locations of the target probe and the S images were counterbalanced within each block. The intertrial interval randomly varied (i.e., 500 ms or 1500 ms) to prevent habituation. The interstimulus interval was set to 250 ms. There was no break between the blocks. The images used for S-N and NT-NT trials were selected from the International Smoking Image Series (Gilbert & Rabinovich, 1999) and the International Affective Picture System (see supplementary Table 1; Lang et al., 2008). E-Prime 2.0 software (Psychology Software Tools, Pittsburgh, PA) was used to program and conduct the AB task (see supplementary material 2 for details).

2.3.2. Affect manipulation

Two affect manipulation conditions (i.e., A and B) were implemented to manipulate negative and neutral affect. Participants randomized to each condition completed three phases of affect manipulation that included speech preparation, reading a book, and watching a video clip (see supplementary material 3 for details).

2.4. Measures

2.4.1. Self-report measures

The self-report measures collected at baseline were included in the current study. A 30-item Smoking History Questionnaire (Brown et al., 2002) was used to assess the age of onset of smoking, and years of being a daily smoker. Timeline Followback (TLFB; Sobell & Sobell, 1992) and Fagerström Test for Cigarette Dependence (Heatherton et al., 1991) were used to assess CPD and nicotine dependence. To measure the reinforcing effect of cigarettes, 26-item Cigarette Purchase Task (CPT; MacKillop et al., 2008) and the 12-item Modified Cigarette Evaluation Questionnaire (mCEQ; Cappelleri et al., 2007) were used. CPT includes five measures (intensity, maximum price per cigarette [P_{max}], breakpoint, the largest cost to obtain cigarette [O_{max}], and elasticity). The mCEQ includes five subscales: 3-item satisfaction (Cronbach's $\alpha = 0.87$ in the current sample), 5-item psychological reward ($\alpha = 0.80$ in the current sample), 1-item enjoyment of respiratory tract sensation, 1-item craving reductions, and 2-item aversion. The aversion subscale was not used in the current study due to its low variability. For manipulation check, state positive and negative affect were assessed at baseline and after each affect manipulation with the 10-item Positive and Negative Affect Schedule (PANAS; Thompson, 2007; range $\alpha = 0.35$–0.55 and 0.89–0.92 for negative and positive affect in the current sample, respectively).

2.4.2. Smoking topography

The CReSS Pocket Device (Clinical Research Support System for Laboratories; Hauni Group) was used to measure smoking topography as a proxy for in vivo smoking reinforcement. The CReSS measured puff level data including: 1) volume of puff (mL), 2) duration of puff (seconds), 3) interval between puffs (IPI; seconds), and 4) velocity (mL/seconds). Velocity (mL/seconds) was calculated from the volume and duration values. In the current investigation, puff topography parameters were conceptualized as behavioral proxies of smoking reinforcement. For example, extant studies have shown that larger puff volumes, sustained duration of puff, and shorter IPI indicate greater reinforcement (Borges et al., 2019; Farris et al., 2017). Averages of these indices across smoking a single cigarette were used for the analyses.

2.4.3. Attentional Bias (AB) scores

After data cleaning (see supplementary material 2 for details), MBS was computed by subtracting mean RT across congruent trials (i.e., target behind cigarette image) from mean RT across incongruent trials (i.e., target behind neutral image). Greater MBS indicates greater AB toward cigarette cues. TLBS was computed by subtracting temporally adjacent pairs of congruent and incongruent trial RTs (i.e., no further than five trials away). This approach produces a time series of TLBSs per participant (see supplementary Fig. 2). There are five TLBS indices categorized into three domains. First, AB toward cigarette cue is indexed by mean toward (average of TLBS > 0 ms) and peak toward (maximum TLBS > 0 ms). A greater value of TLBS toward indicates greater AB toward cigarette cues. Second, AB away from cigarette cue is indexed by mean away (average of TLBS < 0 ms) and peak away (maximum TLBS < 0 ms). A greater absolute value of TLBS away indicates a greater tendency to move attention away from cigarette cues. Finally, the stability of AB was indexed by variability (the average of the absolute value of the sequential differences in TLBSs) that indicates temporal stability of AB toward or away over time. Greater TLBS variability indicates less stability in the expression of AB. TLBS indices were computed by utilizing the R package *itrak* with the "nearest" option (Beevers et al., 2019), consistent with Zvielli et al.'s method.

2.5. Analysis

Descriptive statistics, two-way mixed ANOVA, reliability, and correlation analyses were conducted in Rstudio (version 1.4.1106). Partial correlation analyses were conducted in JASP (version 0.16.2). Outliers and normality were examined. Supplementary material 4 presents data reduction procedures used for smoking topography and CPT data. Absolute values of TLBS were used in the current analysis. The TLBS indices as well as smoking topography and CPT task data were log transformed due to skew per the recommendation in the literature (Tabachnick & Fidell, 2018). To test the possible effect of affect manipulation on AB indices, two-way mixed ANOVA was conducted. Correlation analyses were conducted between theoretically relevant demographic/smoking variables (i.e., age, sex, income, CPD) and study variables. Split-half reliability was tested by both intra class correlation (ICC) and Pearson's r to examine internal consistency of MBS and TLBS aligned with previous methods (Jones et al., 2018; Molloy & Anderson, 2020). To compute the reliability indexed by average-measure ICC, *psych* R package was used. For MBS, split-half reliability was examined between odd and even-numbered trials in each AB task. For TLBS, due to the importance of temporal sequence in computing TLBS, split-half reliability was tested between first- and second-half of each AB task. Test-retest reliability of each MBS and TLBS was tested by both single-measure ICC and Pearson's r between AB1 and AB2, and then between AB2 and AB3. The 0.7 cut-off criterion (i.e., acceptable) was used to determine the strength of the internal reliability (Kline, 1999). To detect any potential impact of affect manipulation on each AB index, internal consistency and test–retest were also examined in each condition. To test the external validity of the AB indices, correlation analyses were conducted between the AB indices of AB1 and nicotine dependence, mCEQ, smoking topography, and CPT measures. Partial correlation analyses were run for the outcome variables that were significantly associated with demographic variables or CPD.

3. Results

3.1. Sample characteristics

Among fifty participants, two participants were removed from the main analyses. One was removed due to low CO level (<2ppm) and the other due to more than 50 % of their AB data being lost during the cleaning procedure. The current results are from the remaining n = 48 participants. Table 1 shows the demographic characteristics and

Table 1
Sample Characteristics and Descriptive Statistics (N = 48).

Variable	M (SD)	Range
Sex (n females, %)	15 (31.3 %)	
Age	33.96 (7.73)	19–50
Hispanic/Latinx (yes; n, %)	7 (14.6 %)	–
Race (n, %)		
White	27 (56.3 %)	–
Black	13 (27.1 %)	–
Others	5 (10.4 %)	–
More than on race	3 (6.3 %)	–
Income (n, %)[1]		
< $5,000	19 (39.6 %)	–
< $35,000	19 (39.6 %)	–
≥ $35,000	6 (12.5 %)	–
CPD	14.60 (5.46)	7.25 – 30.00
Nicotine dependence	6.00 (1.83)	0.00 – 10.00
AB1 (n = 47)[2]		
MBS	4.15 (19.81)	−56.17–74.01
TLBS – mean toward	92.51 (46.71)	39.04 – 245.46
TLBS – mean away	90.47 (49.47)	34.11 – 258.20
TLBS – peak toward	327.83 (165.47)	139.00 – 758.00
TLBS – peak away	318.00 (170.26)	104.00 – 787.00
TLBS – variability	78.40 (39.80)	32.72 – 212.76
AB2 (n = 46)[2]		
MBS	10.26 (24.21)	−79.19 – 95.74
TLBS – mean toward	96.88 (53.67)	42.27 – 297.37
TLBS – mean away	94.49 (53.99)	41.06 – 260.65
TLBS – peak toward	348.61 (172.74)	112.00 – 828.00
TLBS – peak away	334.74 (185.49)	122.00 – 962.00
TLBS – variability	80.34 (47.12)	35.73 – 249.48
AB3 (n = 45)[2]		
MBS	2.56 (21.93)	−56.13 – 53.72
TLBS – mean toward	97.36 (51.92)	35.85 – 280.37
TLBS – mean away	93.19 (52.63)	31.88 – 238.37
TLBS – peak toward	352.38 (188.34)	117.00 – 864.00
TLBS – peak away	341.47 (187.82)	136.00 – 829.00
TLBS – variability	81.51 (43.92)	31.86 – 209.12

Note. [1]n = 4 missing. [2] data missing for n = 1 in AB1, n = 2 in AB2, and n = 3 in AB3 for procedural and/or technical reasons. CPD = Cigarette per day. AB = Attentional bias task. MBS = Mean bias score. TLBS = Trial-level bias score. The TLBS indices are presented in absolute values.

descriptive statistics of the study variables.

3.2. Testing the effect of affect manipulation condition on AB indices

The two-way mixed ANOVA results revealed no significant interaction between condition and time on subsequent AB (supplementary Table 2) despite evidence of a significant manipulation effect (supplementary Table 3). Therefore, the data from the two conditions were combined and used in the subsequent analyses.

3.3. Internal consistency across timepoints

As expected, the MBS exhibited low and mostly non-significant internal consistency across timepoints for both ICC and Pearson's *r* split-half reliability estimates (Table 2). All of the TLBS indices were significant. For ICC, the internal consistency of the TLBS indices ranged from 0.79 to 0.95, indicating high internal consistency. The overall level of internal consistency as indexed by Pearson's *r* stayed at the acceptable to high level, while TLBS peak toward at AB1 and AB3, peak away at AB2, and mean toward at AB3 demonstrated relatively low internal consistency with a wide confidence interval. Similar findings were observed in the analyses within each condition excepting for the MBS demonstrating significant internal consistency, albeit in nonacceptable ranges, at AB1 and AB3 in the condition B (Supplementary Table 4).

Table 2
Internal Consistency (N = 48).

Split-half reliability	ICC[+]	95 % CI	Pearson's *r*	95 % CI
AB 1				
MBS	.37	−.02, .61	.23	−.06, .49
TLBS – mean toward	.91***	.86, .95	.84***	.74, .91
TLBS – mean away	.92***	.87, .95	.85***	.75, .92
TLBS – peak toward	.79***	.66, .87	.65***	.45, .79
TLBS – peak away	.89***	.82, .93	.81***	.67, .89
TLBS – variability	.95***	.91, .97	.91***	.84, .95
AB 2				
MBS	.40*	.05, .62	.27	−.02, .52
TLBS – mean toward	.85***	.75, .91	.76***	.60, .86
TLBS – mean away	.86***	.77, .91	.75***	.59, .86
TLBS – peak toward	.84***	.74, .91	.75***	.59, .85
TLBS – peak away	.79***	.66, .87	.68***	.48, .81
TLBS – variability	.94***	.89, .96	.90***	.82, .94
AB 3				
MBS	.31	−.11, .58	.19	−.10, .46
TLBS – mean toward	.79***	.65, .88	.68***	.49, .81
TLBS – mean away	.88***	.81, .93	.79***	.65, .88
TLBS – peak toward	.80***	.66, .88	.68***	.48, .81
TLBS – peak away	.83***	.72, .90	.71***	.52, .83
TLBS – variability	.91***	.85, .94	.84***	.72, .91

Note. * $p < .05$, ** $p < .01$, *** $p < .001$. [+] Average-measure with random-effects model with absolute agreement. MBS = Mean bias score. TLBS = Trial-level bias score. ICC = Intraclass correlation. CI = Confidence interval. The TLBS variables were log-transformed.

3.4. Test-retest reliability

The results of the test–retest reliability analyses in each pair of AB1-AB2 and then AB2-AB3 were in the expected direction such that MBS, either indexed by ICC or Pearson's *r*, demonstrated nonsignificant and very low test–retest reliability (Table 3). Not surprisingly, the TLBS indices demonstrated significant, acceptable to high level of test–retest reliability while somewhat consistent with the results of internal consistency, the TLBS indices of peak toward demonstrated relatively low, yet significant, test–retest reliability for the comparison pair of AB1 and AB2. The results of the analyses within each condition did not substantially change with the exception of the MBS demonstrating significant test–retest reliability, albeit in nonacceptable ranges, for the AB2-AB3 comparison pair in the condition A (Supplementary Table 5).

3.5. External validity: Correlation results

Correlation analyses revealed significant associations between sex and MBS (*Spearman's rho* = −.33, p = .024, 95% confidence interval

Table 3
Test-Retest Reliability (N = 48).

Comparison	ICC[+]	95 % CI	Pearson's *r*	95 % CI
AB 1 & AB 2				
MBS	.06	-.17, .29	.07	-.23, .35
TLBS – mean toward	.76***	.64, .84	.76***	.60, .86
TLBS – mean away	.84***	.76, .90	.83***	.72, .91
TLBS – peak toward	.64***	.48, .76	.64***	.43, .79
TLBS – peak away	.70***	.56, .80	.69***	.49, .82
TLBS – variability	.88***	.81, .92	.87***	.78, .93
AB 2 & AB 3				
MBS	.16	-.07, .37	.17	-.14, .44
TLBS – mean toward	.82***	.73, .89	.82***	.70, .90
TLBS – mean away	.81***	.71, .88	.81***	.68, .89
TLBS – peak toward	.76***	.64, .84	.76***	.60, .86
TLBS – peak away	.72***	.58, .82	.72***	.54, .84
TLBS – variability	.87***	.80, .92	.88***	.79, .93

Note. * $p < .05$, ** $p < .01$. [+] Single-measure with random-effects model with absolute agreement. AB = Attentional bias task. MBS = Mean bias score. TLBS = Trial-level bias score. ICC = Intraclass correlation. CI = Confidence Interval. The TLBS variables were log-transformed.

[CI] = −.56, −.05), CPT elasticity (*rho* = .40, p = .006, 95%CI = .13, .62), and mCEQ craving reductions (*rho* = −.29, p = .045, 95%CI = −.54, −.01), as well as between CPD and nicotine dependence (r = .59, p < .001, 95%CI = .37, .75), CPT Intensity (r = .69, p < .001, 95%CI = .50, .82), and mCEQ satisfaction (r = .43, p = .002, 95%CI = .17, .64). Thus, partial correlation analyses between AB indices and these outcome variables were conducted by entering sex and/or CPD as covariates.

Overall, the only significant association observed was between TLBS indices (mean away, peak toward, peak away, and variability) and the mCEQ satisfaction subscale (Table 4). These results indicate that experiencing satisfaction following smoking a cigarette was associated with AB away from cigarette cues at both mean and peak level, AB towards cigarette cues at peak level, and greater instability of AB. The TLBS mean toward was marginally associated with mCEQ satisfaction (p = .059). MBS did not demonstrate any association with mCEQ satisfaction subscale. Besides mCEQ satisfaction subscale, there was another marginal association between MBS and CPT O_{max} (p = .087). However, no other significant or marginal associations were found between AB indices and outcome variables.

4. Discussion

The current study examined the internal consistency, test–retest reliability, and external validity of MBS and TLBS measures derived from the visual dot-probe task among daily cigarette smokers. TLBS demonstrated superior reliability over MBS. However, AB indices were not significantly associated with the majority of the measures of biobehavioral and self-report indices of smoking reinforcement.

Consistent with our hypothesis, the TLBS indices demonstrated superior split-half reliability as indexed by both ICC and Pearson's r, as compared to MBS indices. These observations are consistent with extant findings among individuals with depressive symptoms (Beevers et al., 2019), social anxiety disorder (Molloy & Anderson, 2020), nicotine-deprived daily smokers (Zvielli et al., 2015), and youth with and without obesity (Vervoort et al., 2021). Notably, the magnitude of the split-half reliability of TLBS indices in the current study was higher than those in the previous studies, which is attributable to the higher number of S-N images (36 images) used in the current study (Jones et al., 2018).

The results on the test–retest reliability, as indexed by both ICC and Pearson's r, demonstrated a stark contrast between MBS and TLBS indices. None of the MBS demonstrated any significant stability over time, while all of the TLBS indices did, aligned with the extant findings (Carlson & Fang, 2020; MacLeod & Mathews, 1988; Price et al., 2015). It should be noted that our test–retest reliability was somewhat higher than previous studies that assessed AB on different days (Jones et al., 2018; Molloy & Anderson, 2020) and derived AB indices from a smaller number of trials per block within a same AB task (Carlson & Fang, 2020). In the current study, all of the three AB tasks were temporally close (i.e., ~8 min apart), which might explain high correlations. Another interesting finding was that the TLBS peak toward/away, compared to other TLBS indices, demonstrated relatively lower, albeit significant, test–retest reliability, consistent with previous observations among individuals who use alcohol (Jones et al., 2018) and those with social anxiety disorder (Molloy & Anderson, 2020). While TLBS peak measures are conceptualized as the maximum phasic expression of AB, their utility and implication require further investigation.

Finally, there was limited evidence on the external validity of both MBS and TLBS. The only significant finding was on the association between TLBS indices and perceived satisfaction indexed by mCEQ after controlling for CPD, although the pattern was mixed such that TLBS indices of both toward and away from cigarette cues were positively associated with mCEQ satisfaction subscale. While no studies, to the best

Table 4
Correlation Analyses (N = 48).

Variable (p-value; 95 %CI)	MBS	TLBS Mean Toward	TLBS Mean Away	TLBS Peak Toward	TLBS Peak Away	TLBS Variability
Nicotine Dependence	.22¶+ (.141; −.15, .48)	.05+ (.726; −.33, .45)	−.01+ (.928; −.41, .38)	.01+ (.956; −.32, .41)	−.12+ (.445; −.45, .27)	.03+ (.823; −.35, .42)
Smoking Topography Variables						
Volume	.07¶ (.628; −.24, .32)	−.11 (.486; −.38, .19)	−.11 (.454; −.39, .18)	−.06 (.689; −.35, .23)	−.01 (.964; −.30, .28)	−.11 (.469; −.39, .19)
Duration	−.04¶ (.809; −.30, .20)	−.04 (.802; −.33, .26)	.02 (.906; −.27, .31)	−.01 (.937; −.30, .28)	.13 (.402; −.17, .40)	.01 (.928; −.28, .30)
IPI	.20¶ (.187; −.06, .44)	.05 (.738; −.24, .34)	−.00 (.991; −.29, .29)	.24 (.116; −.06, .49)	.01 (.948; −.28, .30)	.03 (.840; −.26, .32)
Velocity	.09¶ (.559; −.21, .33)	−.07 (.639; −.35, .22)	−.13 (.387; −.41, .17)	−.07 (.649; −.35, .23)	−.15 (.319; −.42, .15)	−.13 (.376; −.41, .16)
CPT Variables						
Elasticity	−.24¶ (.127; −.48, .13)	−.06¶ (.683; −.31, .30)	.07¶ (.680; −.25, .28)	−.01¶ (.942; −.28, .43)	.09¶ (.585; −.17, .29)	.01¶ (.949; −.22, .29)
Intensity	.04¶+ (.818; −.23, .26)	.11+ (.490; −.19, .41)	.23+ (.146; −.06, .49)	.04+ (.794; −.24, .33)	.21+ (.174; −.11, .48)	.18+ (.252; −.12, .47)
Omax	.26¶ (.087; .01, 0.50)	.02 (.896; −.28, .32)	−.04 (.804; −.33, .26)	−.02 (.878; −.32, .28)	−.12 (.457; −.40, .19)	−.04 (.778; −.34, .26)
Pmax	.20¶ (.189; −.08, .46)	−.00 (.999; −.30, .30)	−.08 (.625; −.36, .23)	−.04 (.813; −.33, .26)	−.18 (.249; −.45, .13)	−.07 (.657; −.36, .23)
Breakpoint	.22¶ (.164; −.04, .44)	−.01 (.968; −.30, .29)	−.09 (.561; −.38, .21)	−.05 (.727; −.35, .25)	−.19 (.223; −.46, .12)	−.09 (−.583; −.37, .22)
mCEQ Variables						
Satisfaction	.00¶+ (.990; −.25, .31)	.28+ (.059; .02, .54)	.36+* (.015, .07, .61)	.31+* (.040; .04, .56)	.31+* (.039; .02, .57)	.35+* (.019; .07, .60)
Reward	.10¶ (.523; −.14, .37)	−.11 (.473; −.39, 19)	−.06 (.699; −.34, .24)	−.05 (.723; −.34, .24)	−.11 (.458; −.39, .18)	−.10 (.531; −.38, .20)
Enjoyment	.13¶ (.408; −.11, .32)	.14 (.348; −.16, .42)	.19 (.208; −.11, .45)	.13 (.391; −.17, .41)	.17 (.257; −.13, .44)	.15 (.323; −.15, .42)
Craving Reductions	−.18¶ (.248; −.44, .20)	−.20¶ (.195; −.47, .09)	−.04¶ (.788; −.32, .23)	−.14¶ (.376; −.47, .18)	−.08¶ (.615; −.36, .20)	−.09¶ (.573; −.36, .20)

Note. * p <.05. ¶ Sex was covariate. + Cigarettes per day (CPD) was covariate. ¶+ Both sex and CPD were covariates. CI = Confidence interval, MBS = Mean bias score, TLBS = Trial-level bias score, CPT = Cigarette Purchase Task. mCEQ = Modified Cigarette Evaluation Questionnaire. The MBS and TLBS indices were derived from AB1. The variables of TLBS, smoking topography, and CPT were log-transformed.

of our knowledge, have examined the association between AB indices assessed by the dot-probe task and the mCEQ, it is notable that TLBS indices, but not MBS, were associated with the degree of satisfaction one experiences following smoking. It is also noteworthy that no relation between the AB indices and other smoking indices was observed. Although the small sample size warrants caution in the interpretation of results, our overall mixed findings are aligned with accumulating evidence on the inconsistent and limited associations between TLBS and theoretically relevant variables in various populations (Beevers et al., 2019; Carlson & Fang, 2020; Jones et al., 2018; Vervoort et al., 2021), as well as those between MBS and various smoking processes and behaviors (Drobes et al., 2019; Mogg et al., 2005; Rehme et al., 2018; Spiegelhalder et al., 2011; Waters et al., 2003). Pertaining to TLBS, one study has identified a significant relation between TLBS indices and CPD among deprived daily smokers (Zvielli et al., 2015). Our study participants were not required to be abstinent prior to smoking a cigarette using CReSS device, which might explain our null findings. Laboratory manipulation of withdrawal symptoms may strengthen these associations.

There are several limitations in the current study. First, the effect of general RT variability on the reliability of TLBS was not accounted for due to the limited number of NT-NT trials. Our AB task was carefully designed by reviewing existing AB designs (e.g., number of S-N and NT-NT trials; Jones et al., 2018; Zvielli et al., 2015). Given recent evidence on the potential impact of general RT variability on the TLBS reliability (Carlson & Fang, 2020; Vervoort et al., 2021), this possibility should be tested in future studies with a sufficient number of NT-NT trials. Second, there were 144 S-N trials used in the current analyses. It is possible that the reliability of the TLBS peak toward/away is improved with a higher number of trials. Third, the current study design may not be optimal for testing test–retest reliability given affect manipulation and especially given the significant time and condition interaction effect on state affect in the current sample. However, our analyses indicated that manipulation condition did not have a substantial impact on the AB indices. Replication of the current findings is warranted. Third, although all participants in our sample reported daily cigarette use and cigarettes as their primary form of nicotine consumption, it is notable that n = 4 reported vaping. Although post-hoc analysis indicated no AB score differences in our sample as a function of vaping, future work would benefit from separate examination of AB task scores in other nicotine or co-nicotine using samples. Finally, participants were not nicotine deprived, which might have contributed to our null findings on the association between AB indices and most of the smoking indices.

Taken together, our findings add to evidence that TLBS, as compared to MBS, is a more reliable measure of AB. With the limited evidence on the external validity of TLBS in mind, the current results add to the literature supporting that AB may not be a stable construct, but a construct that fluctuates even within a single session of AB task.

Role of funding source

This work was supported in part by the Society of Addiction Psychology (PI, Yang), Foundation for Education and Research in Biofeedback and Related Sciences (PI, Yang), National Cancer Institute (T32CA090314 to Yang), National Institute on Drug Abuse (F31-DA043934: PI, Borges), and National Institute on Alcohol Abuse and Alcoholism (R21AA029033, L30AA027041: PI, Emery). None of the funding sources noted played a role in the study design, collection, analysis or interpretation of the data, writing the manuscript, or the decision to submit the paper for publication.

CRediT authorship contribution statement

Min-Jeong Yang: Funding acquisition, Investigation, Project administration, Conceptualization, Data curation, Formal analysis, Writing – original draft. **Allison M. Borges:** Funding acquisition, Investigation, Project administration, Data curation, Writing – review & editing. **Noah N. Emery:** Writing – review & editing. **Teresa M. Leyro:** Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

The authors thank Xinyi Xu for her work as project coordinator.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.addbeh.2022.107456.

References

Abhishekh, H. A., Nisarga, P., Kisan, R., Meghana, A., Chandran, S., Raju, T., & Sathyaprabha, T. N. (2013). Influence of age and gender on autonomic regulation of heart. *Journal of Clinical Monitoring and Computing, 27*(3), 259–264.

Ataya, A. F., Adams, S., Mullings, E., Cooper, R. M., Attwood, A. S., & Munafò, M. R. (2012). Internal reliability of measures of substance-related cognitive bias. *Drug and Alcohol Dependence, 121*(1–2), 148–151. https://doi.org/10.1016/j.drugalcdep.2011.08.023

Bardeen, J. R., & Orcutt, H. K. (2011). Attentional control as a moderator of the relationship between posttraumatic stress symptoms and attentional threat bias. *Journal of Anxiety Disorders, 25*(8), 1008–1018. https://doi.org/10.1016/j.janxdis.2011.06.009

Beevers, C. G., Mullarkey, M. C., Dainer-Best, J., Stewart, R. A., Labrada, J., Allen, J. J., McGeary, J. E., & Shumake, J. (2019). Association between negative cognitive bias and depression: A symptom-level approach. *Journal of Abnormal Psychology, 128*(3), 212–227. https://doi.org/10.1037/abn0000405

Begh, R., Munafò, M. R., Shiffman, S., Ferguson, S. G., Nichols, L., Mohammed, M. A., Holder, R. L., Sutton, S., & Aveyard, P. (2015). Lack of attentional retraining effects in cigarette smokers attempting cessation: A proof of concept double-blind randomised controlled trial. *Drug and Alcohol Dependence, 149*, 158–165. https://doi.org/10.1016/j.drugalcdep.2015.01.041

Berridge, K. C., & Robinson, T. E. (2016). Liking, wanting, and the incentive-sensitization theory of addiction. *American Psychologist, 71*(8), 670–679. https://doi.org/10.1037/amp0000059

Boffo, M., Zerhouni, O., Gronau, Q. F., van Beek, R. J., Nikolaou, K., Marsman, M., & Wiers, R. W. (2019). Cognitive bias modification for behavior change in alcohol and smoking addiction: Bayesian meta-analysis of individual participant data. *Neuropsychology Review, 29*(1), 52–78. https://doi.org/10.1007/s11065-018-9386-4

Borges, A. M., Leyro, T. M., Rosen, R. L., Zvolensky, M. J., & Farris, S. G. (2019). Negative urgency and ad-libitum smoking topography. *Drug and Alcohol Dependence, 201*, 220–226. https://doi.org/10.1016/j.drugalcdep.2019.04.021

Brown, R. A., Lejuez, C. W., Kahler, C. W., & Strong, D. R. (2002). Distress tolerance and duration of past smoking cessation attempts [Article]. *Journal of Abnormal Psychology, 111*(1), 180–185. https://doi.org/10.1037/0021-843X.111.1.180

Bunnell, B. E., Beidel, D. C., & Mesa, F. (2013). A randomized trial of attention training for generalized social phobia: Does attention training change social behavior? *Behavior Therapy, 44*(4), 662–673. https://doi.org/10.1016/j.beth.2013.04.010

Cappelleri, J. C., Bushmakin, A. G., Baker, C. L., Merikle, E., Olufade, A. O., & Gilbert, D. G. (2007). Confirmatory factor analyses and reliability of the modified cigarette evaluation questionnaire. *Addictive behaviors, 32*(5), 912–923. https://doi.org/10.1016/j.addbeh.2006.06.028

Carlson, J. M., & Fang, L. (2020). The stability and reliability of attentional bias measures in the dot-probe task: Evidence from both traditional mean bias scores and trial-level bias scores. *Motivation and Emotion, 1–13*. https://doi.org/10.1007/s11031-020-09834-6

Drobes, D. J., Oliver, J. A., Correa, J. B., & Evans, D. E. (2019). Attentional Bias and Smoking. In *Neuroscience of Nicotine* (pp. 145–150). Elsevier.

Emery, N. N., & Simons, J. S. (2015). Mood & alcohol-related attentional biases: New considerations for gender differences and reliability of the visual-probe task. *Addictive Behaviors, 50*, 1–5. https://doi.org/10.1016/j.addbeh.2015.06.007

Farris, S. G., Aston, E. R., Abrantes, A. M., & Zvolensky, M. J. (2017). Tobacco demand, delay discounting, and smoking topography among smokers with and without psychopathology. *Drug and Alcohol Dependence, 179*, 247–253. https://doi.org/10.1016/j.drugalcdep.2017.06.042

Field, M., & Cox, W. M. (2008). Attentional bias in addictive behaviors: A review of its development, causes, and consequences. *Drug and Alcohol Dependence, 97*(1), 1–20. https://doi.org/10.1016/j.drugalcdep.2008.03.030

Garland, E. L., Franken, I. H. A., & Howard, M. O. (2012). Cue-elicited heart rate variability and attentional bias predict alcohol relapse following treatment. *Psychopharmacology (Berl), 222*(1), 17–26. https://doi.org/10.1007/s00213-011-2618-4

Gilbert, D., & Rabinovich, N. (1999). *International Smoking Image Series, Version 1.2*. Carbondale, Integrative Neuroscience Laboratory, Department of Psychology, Southern Illinois University.

Gladwin, T. E., Halls, M., & Vink, M. (2021). Experimental control of conflict in a predictive visual probe task: Highly reliable bias scores related to anxiety. *Acta Psychologica, 218*, Article 103357.

Grafton, B., Teng, S., & MacLeod, C. (2021). Two probes and better than one: Development of a psychometrically reliable variant of the attentional probe task. *Behaviour Research and Therapy, 138*, Article 103805.

Heatherton, T. F., Kozlowski, L. T., Frecker, R. C., & Fagerstrom, K.-O. (1991). The Fagerstrom Test for Nicotine Dependence: A revision of the Fagerstrom Tolerance Questionnaire [Article]. *British Journal of Addiction, 86*(9), 1119–1127. https://doi.org/10.1111/j.1360-0443.1991.tb01879.x

Heitmann, J., Jonker, N. C., & De Jong, P. J. (2021). A promising candidate to reliably index attentional bias toward alcohol cues–an adapted odd-one-out visual search task. *Frontiers in Psychology, 12*, 102.

Javors, M. A., Hatch, J. P., & Lamb, R. J. (2005). Cut-off levels for breath carbon monoxide as a marker for cigarette smoking. *Addiction, 100*(2), 159–167. https://doi.org/10.1111/j.1360-0443.2004.00957.x

Jones, A., Christiansen, P., & Field, M. (2018). Failed attempts to improve the reliability of the alcohol visual probe task following empirical recommendations. *Psychology of Addictive Behaviors, 32*(8), 922. https://doi.org/10.1037/adb0000414

Kline, R. B. (1999). *The Handbook of Psychological Testing*. London: Routledge.

Koenig, J., Jarczok, M., Warth, M., Ellis, R., Bach, C., Hillecke, T., & Thayer, J. F. (2014). Body mass index is related to autonomic nervous system activity as measured by heart rate variability—a replication using short term measurements. *The Journal of Nutrition, Health & Aging, 18*(3), 300–302.

Kruijt, A.-W., Field, A. P., & Fox, E. (2016). Capturing dynamics of biased attention: are new attention variability measures the way forward? *PLoS One, 11*(11), e0166600. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0166600#sec033.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). *International affective picture system (IAPS): Technical manual and affective ratings*. University of Florida.

MacKillop, J., Murphy, J. G., Ray, L. A., Eisenberg, D. T., Lisman, S. A., Lum, J. K., & Wilson, D. S. (2008). Further validation of a cigarette purchase task for assessing the relative reinforcing efficacy of nicotine in college smokers. *Experimental and Clinical Psychopharmacology, 16*(1), 57. https://doi.org/10.1037/1064-1297.16.1.57

MacLeod, C., & Mathews, A. (1988). Anxiety and the allocation of attention to threat. *The Quarterly Journal of Experimental Psychology, 40*(4), 653–670. https://doi.org/10.1080/14640748808402292

Masiero, M., Lucchiari, C., Maisonneuve, P., Pravettoni, G., Veronesi, G., & Mazzocco, K. (2019). The Attentional bias in current and former smokers. *Frontiers in Behavioral Neuroscience, 13*, 154. https://doi.org/10.3389/fnbeh.2019.00154

Mogg, K., Field, M., & Bradley, B. P. (2005). Attentional and approach biases for smoking cues in smokers: An investigation of competing theoretical views of addiction. *Psychopharmacology (Berl), 180*(2), 333–341. https://doi.org/10.1007/s00213-005-2158-x

Molfino, A., Fiorentini, A., Tubani, L., Martuscelli, M., Fanelli, F. R., & Laviano, A. (2009). Body mass index is related to autonomic nervous system activity as measured by heart rate variability. *European Journal of Clinical Nutrition, 63*(10), 1263.

Molloy, A., & Anderson, P. L. (2020). Evaluating the reliability of attention bias and attention bias variability measures in the dot-probe task among people with social anxiety disorder. *Psychological Assessment, 32*(9), 883. https://doi.org/10.1037/pas0000912

Price, R. B., Kuckertz, J. M., Siegle, G. J., Ladouceur, C. D., Silk, J. S., Ryan, N. D., Dahl, R. E., & Amir, N. (2015). Empirical recommendations for improving the stability of the dot-probe task in clinical research. *Psychological Assessment, 27*(2), 365.

Rehme, A. K., Bey, K., Frommann, I., Mogg, K., Bradley, B. P., Bludau, J., Block, V., Sträter, B., Schütz, C. G., & Wagner, M. (2018). Selective attention to smoking cues in former smokers. *European Neuropsychopharmacology, 28*(2), 276–284. https://doi.org/10.1016/j.euroneuro.2017.12.003

Rodebaugh, T. L., Scullin, R. B., Langer, J. K., Dixon, D. J., Huppert, J. D., Bernstein, A., Zvielli, A., & Lenze, E. J. (2016). Unreliability as a threat to understanding psychopathology: The cautionary tale of attentional bias. *Journal of Abnormal Psychology, 125*(6), 840–851. https://doi.org/10.1037/abn0000184

Schäfer, J., Bernstein, A., Zvielli, A., Höfler, M., Wittchen, H. U., & Schönfeld, S. (2016). Attentional bias temporal dynamics predict posttraumatic stress symptoms: A prospective–longitudinal study among soldiers. *Depression and Anxiety, 33*(7), 630–639.

Sinnreich, R., Kark, J., Friedlander, Y., Sapoznikov, D., & Luria, M. (1998). Five minute recordings of heart rate variability for population studies: Repeatability and age–sex characteristics. *Heart, 80*(2), 156–162.

Sobell, L. C., & Sobell, M. B. (1992). Timeline follow-back. In *Measuring Alcohol Consumption* (pp. 41–72). Humana Press.

Spiegelhalder, K., Jahne, A., Kyle, S. D., Beil, M., Doll, C., Feige, B., & Riemann, D. (2011). Is Smoking-related Attentional Bias a Useful Marker for Treatment Effects? *Behavioral Medicine, 37*(1), 26–34. https://doi.org/10.1080/08964289.2010.543195

Tabachnick, B. G., & Fidell, L. S. (2018). *Using Multivariate Statistics* (6th ed.). Pearson Education, Limited.

Thompson, E. R. (2007). Development and validation of an internationally reliable short-form of the positive and negative affect schedule (PANAS). *Journal of Cross-cultural Psychology, 38*(2), 227–242. https://doi.org/10.1177/0022022106297301

Vervoort, L., Braun, M., De Schryver, M., Naets, T., Koster, E. H., & Braet, C. (2021). A pictorial dot probe task to assess food-related attentional bias in youth with and without obesity: Overview of indices and evaluation of their reliability. *Frontiers in Psychology, 12*, 561. https://doi.org/10.3389/fpsyg.2021.644512

Waters, A. J., Shiffman, S., Bradley, B. P., & Mogg, K. (2003). Attentional shifts to smoking cues in smokers. *Addiction, 98*(10), 1409–1417. https://doi.org/10.1046/j.1360-0443.2003.00465.x

Zvielli, A., Bernstein, A., & Koster, E. H. (2015). Temporal dynamics of attentional bias. *Clinical Psychological Science, 3*(5), 772–788. https://doi.org/10.1177/2167702614551572