Journal of Behavior Therapy and Experimental Psychiatry 82 (2024) 101918

Contents lists available at ScienceDirect

Journal of Behavior Therapy and Experimental Psychiatry

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

Reliability (or lack thereof) of smartphone ecological momentary assessment of visual dot probe attention bias toward threat indices

Nur Hani Zainal [a,b,*], Nicholas C. Jacobson [c]

[a] Harvard Medical School, Department of Health Care Policy, USA
[b] National University of Singapore, Department of Psychology, Singapore
[c] Center for Technology and Behavioral Health, Geisel School of Medicine, Dartmouth College, USA

ARTICLE INFO

Keywords:
Attention bias
Reliability
Visual dot probe
Generalized anxiety disorder
Social anxiety disorder
Ecological momentary assessment

ABSTRACT

Background and objectives: Cognitive bias theories posit that generalized anxiety disorder (GAD) and social anxiety disorder (SAD) are entwined with attention bias toward threats, commonly indexed by faster response time (RT) on threat-congruent (vs. threat-incongruent) trials on the visual dot probe. Moreover, although smartphone ecological momentary assessment (EMA) of the visual dot probe has been developed, their psychometric properties are understudied. This study thus aimed to assess the reliability of 8 smartphone-delivered visual dot probe attention bias and related indices in persons with and without GAD and SAD.

Methods: Community-dwelling adults ($n = 819$; GAD: 64%; SAD: 49%; Mixed GAD and SAD: 37%; Non-GAD/SAD Controls: 24%) completed a five-trial smartphone-delivered visual dot probe for a median of 60 trials (12 sessions x 5 trials/session) and an average of 100 trials (20 sessions x 5 trials/session).

Results: As hypothesized, Global Attention Bias Index, Disengagement Effect, and Facilitation Bias had low-reliability estimates. However, retest-reliability and internal reliability were good for Trial-Level Bias Scores (TLBS) (Bias Toward Treat: intra-class correlation coefficients (ICCs) = 0.626–0.644; split-half $r = 0.640$–0.670; Attention Bias Variability: ICCs = 0.507–0.567; split-half $r = 0.520$–0.580) and (In)congruent RTs. Poor retest-reliability and internal reliability estimates were consistently observed for all traditional attention bias and related indices but not TLBS.

Limitations: Our visual dot probe EMA should have administered ≥320 trials to match best-practice guidelines based on similar laboratory studies.

Conclusions: Future research should strive to examine attention bias paradigms beyond the dot-probe task that evidenced meaningful test-retest reliability properties in laboratory and real-world naturalistic settings.

Generalized anxiety disorder (GAD) is marked by persistent excessive and uncontrollable worry (American Psychiatric Association, 2013). The most frequent worry content in GAD involves interpersonal matters, a feature observed in social anxiety disorder (SAD; Zainal & Newman, 2018). GAD and SAD are common mental disorders that have been consistently related to diverse health problems (Tully et al., 2016; Vieira et al., 2010; Zainal & Newman, 2021). Relatedly, many individuals with GAD and SAD reportedly frequently experience low levels of relationship and job satisfaction (Bouwmans et al., 2014; Newman et al., 2013; Zainal & Magiati, 2019). Improving understanding of the correlates of GAD and SAD is thus essential.

SAD and GAD have been theorized to often co-occur with the mental habit of focusing on surrounding threats (Bar-Haim et al., 2007). *Cognitive bias models* (Mathews & MacLeod, 2005) posit that such attention bias or hypervigilance toward threats creates and maintains anxiety disorders, as successfully reducing attention bias corresponds to changes in symptomology (Mogoaşe et al., 2014; Teng et al., 2019). Accordingly, scientists have advocated the use of attention bias modification (ABM) to prevent the development of anxiety and related disorders (Bar-Haim, 2010; Hsu et al., 2021; Lazarov & Bar-Haim, 2021), to augment existing treatment, such as empirically-supported first-line cognitive behavior therapies (Lazarov et al., 2018; Shechner et al., 2014; Zainal et al., 2023), or to serve as second-line mental health interventions for treatment-resistant cases (Hallion & Ruscio, 2011; Pettit et al., 2020).

As these theories propose that attention bias toward threats has a

* Corresponding author. Harvard Medical School, Department of Health Care Policy, USA.
E-mail addresses: nhzainal@hcp.med.harvard.edu, hanizainal@nus.edu.sg (N.H. Zainal).

https://doi.org/10.1016/j.jbtep.2023.101918
Received 30 March 2022; Received in revised form 5 July 2023; Accepted 9 October 2023
Available online 20 October 2023

causal effect on the etiology and maintenance of GAD and SAD, reliable attention bias indices are needed to properly construe attention bias as a prevention and treatment target. The *global attention bias indices* based on response time (RT) differences are the most frequently used marker in attention bias paradigms. Global attention bias indices are premised on the logic that persons with (vs. without) GAD and SAD respond quicker to threat-congruent material. In the prominent visual dot probe task (MacLeod et al., 1986), participants respond to a probe while disregarding the prior foil (e.g., angry face (threat cue) or neutral face (non-threat cue)). However, based on evolutionary theories (Lazarov & Bar-Haim, 2021; Vaish et al., 2008), our human visual system cannot entirely discard the foils. Instead, it immediately primes visual attention and leads to faster RT when the probe replaces the foil with more fear-inducing and cognitive resource-absorbing material (such as a snake instead of a mushroom or a negative (e.g., angry, fearful) vs. positive (e.g., happy) or neutral facial emotional expression; Lipp & Derakshan, 2005). The global attention bias indices are usually computed as the average RT of trials wherein the probe substituted the threatening foil (e.g., snakes, angry face) subtracted by the average RT of trials wherein the probe substituted the neutral foil (e.g., mushrooms, neutral face).

Straightforwardly interpretable results with such global attention bias indices have been observed. Supporting *cognitive bias* and *evolutionary* theories and consistent with neuroscientific data (LeDoux & Pine, 2016), speedier responses toward threatening (vs. neutral) material were displayed by unselected college students (i.e., a sample that comprised those with low, moderate, and high levels of anxiety symptoms; Koster et al., 2004) and persons with GAD and SAD (Mogg & Bradley, 2005; Mogg et al., 2004). Alternatively, other researchers have been critiquing the reliability of the visual dot probe task (Rodebaugh et al., 2016). For example, in two initial studies (Schmukle, 2005; Staugaard, 2009), many reliability estimates of the attention bias indices fell within the negative range, and the highest reliability estimate from either study was 0.32. Subsequently, the visual dot probe showed low internal reliability (Cronbach αs = 0.00–0.48) and split-half reliability (r = -.09–0.32) in healthy controls who underwent a CO_2 challenge (a paradigm that induces anxiety symptoms by instructing inhalation of 7.5% CO_2; Cooper et al., 2011). Similarly, Waechter et al. (2014) noted that the global attention bias indices and other attention bias markers (e.g., slower RT to disengage from threatening vs. neutral material) had low split-half test-retest reliability (mean r = −0.10–0.21) and internal reliability (αs = −0.35–0.42) in participants with heightened SAD. However, individual trial-level attention bias toward threat average and variability scores (derived from subtracting the response time (RT) on threat-congruent trials from the nearest threat-incongruent trials in a trial-by-trial manner) have shown good retest-reliability among those with and without specific phobias (Spearman r = 0.58 to 0.67; Zvielli et al., 2015) and SAD (average measures intra-class correlation (ICC) = 0.64–0.83; Gade et al., 2022; Meissel et al., 2022; Packard et al., 2022; Rodebaugh et al., 2016). Collectively, evidence is accruing those traditional attention bias indices derived from visual dot probe laboratory tasks lack retest-reliability and internal reliability, but trial-level bias scores can enhance the psychometric properties of those scores.

Little is known, however, about the reliability of *smartphone-delivered ecological momentary assessments (EMA)* of visual dot probe tasks. Increasingly, such EMA apps are used to measure various psychological phenomena, including cognitive and related processes (Dennis & O'Toole, 2014), and benefits researchers by allowing repeated real-time assessments of variables across multiple ecologically pertinent contexts (Teng et al., 2019). Simultaneously, as most dot probe studies failed to provide reliability estimates (Parsons et al., 2019), we know of only one such study to date that has tested the reliability of a smartphone-delivered visual dot probe EMA in an anxiety-disordered sample. A 4-week ABM study wherein participants engaged in the visual dot probe thrice daily found that the Spearman-Brown adjusted internal reliability of the global attention bias indices ranged from 0.18 to 0.53 (Enock et al., 2014); however, the study did not test other forms of attention bias indices. This is important as poor reliabilities have also been found with other attention bias indices in laboratory settings, such as the facilitation bias and disengagement effect (Waechter et al., 2014). Thus, initial evidence suggests that while the global attention bias indices assessed by an EMA visual dot probe task have poor reliability, other forms of attention bias indices remain to be determined.

Therefore, this study aims to comprehensively examine the retest-reliability and internal consistency of six distinct attention bias indices derived from a smartphone visual dot probe app. This goal contributes to clinical science in several ways. Recruiting a large sample (n = 829) offers greater power than most prior studies that recruited small sample sizes (e.g., n = 20–29; Cooper et al., 2011; Staugaard, 2009) to detect any existing group differences based on diagnostic status. A larger sample also offers more unbiased and reliable parameter estimates of any group differences. Also, attention bias toward threat cuts across many distinct and overlapping forms of psychopathology (cf. negative valence domain in the Research Domain Criteria; Cuthbert & Insel, 2013). Better understanding the reliability of attention bias indices derived from visual dot probe EMA tasks is also essential, given calls to improve the methodology and replicability of studies on the efficacy of ABM for clinical depression, anxiety, and related disorders (Cristea et al., 2015; Fodor et al., 2020). Further, such psychometric inquiries can facilitate future clinically pertinent studies linked to individual differences (e.g., associations among shifts in symptom and shifts in purported therapy mechanisms, optimizing individualized therapy outcomes).

Based on cognitive bias theories and empirical data outlined, we hypothesized that the following four attention bias indices would display poor retest-reliability and internal reliability: (1) *Global Attention Bias Index*; (2) *Disengagement Effect* (i.e., difficulty disengaging from threatening material; Mogg et al., 2008) (3) *Facilitation Bias* (i.e., quicker RT with threat (vs. neutral); foils during trials where the foil and matched; Waechter & Stolz, 2015); (4) *Attention Bias Variability* (ABV; i.e., degree of instability of attention bias across time; Iacoviello et al., 2014). Conversely, based on prior evidence (e.g., Waechter et al., 2014), we expected moderate-to-good test-retest reliability and internal consistency for (5) *Congruent RT* (congruent trials raw RTs) and (6) *Incongruent RT* (incongruent trials raw RTs). Further, as RT tends to generally be more stable within- (vs. between-) sessions and persons for EMAs (Salthouse & Berish, 2005), we predicted the reliability scores to be higher within versus across sessions. Moreover, we hypothesized that the Trial-Level Bias Mean and Variability Attention Bias Toward Threat Scores would show moderate-to-good levels of test-retest reliability (Davis et al., 2016; Huppert et al., 2018). To further clarify the degrees of overlap and distinction of attention bias indices in GAD, SAD, and their comorbidities (Amir et al., 2009; Molloy & Anderson, 2020), we determined the reliabilities of these unique attention bias indices in the entire sample and within separate diagnostic subgroups.

1. Method

1.1. Participants and eligibility criteria

1.1.1. Participants

Community-dwelling adolescents and adults (n = 829, 83.3% female, M of age = 25.59 years (SD = 16.36), range = 13 to 65, 59.2% White/Caucasian, 6.3% Black/African American, 2.8% Asian/Asian American, 6.5% Hispanic/Latino/Latina, 1.4% Native American, 23.6% Multiracial/Other) were recruited for this study using the Google Play Store from the Android Market and undergraduate subject pool in a university in the Northeastern part of the U.S. Potential participants were screened for GAD and SAD based on the GAD-Questionnaire-Fourth Edition (GAD-Q-IV; Newman et al., 2002) and Social Phobia Diagnostic Questionnaire (SPDQ; Newman et al., 2003). Individuals from the community or university subject pool with GAD and SAD were

oversampled to capture the full range of anxiety at the moderate-to-severe end of the distribution. Individuals who had diagnosable GAD and SAD consistent with the Diagnostic and Statistical Manual–Fifth Edition (DSM-5) criteria assessed by the GAD-Q-IV (Newman et al., 2002) and SPDQ (Newman et al., 2003) baseline measures were invited to participate in the study at similar rates to non-anxiety-disordered controls. Accordingly, high proportions of participants met criteria for GAD and SAD (GAD: 64.34%; SAD: 48.60%; Mixed GAD and SAD: 36.51%; Non-GAD and SAD Controls: 23.57%). Table 1 presents sociodemographic and clinical variables across diagnostic groups at baseline. **Inclusion Criteria:** To be eligible, participants had to be at least 13 years of age and own an Android phone, consent to participate, and complete baseline assessments. **Exclusion Criteria**: None. Non-anxious controls had to display scores 1.5 standard deviations below the sample mean on the GAD-Q-IV and SPDQ. Ethics approval and participant informed consent were obtained before data collection.

1.2. Measures

1.2.1. Dot probe task

The dataset is from a visual dot probe attention bias task using the Mood Triggers app (Jacobson, 2020). Participants were given instructions on completing the task and then completed 5 trials of the visual dot probe. Participants were asked to view a "+" sign in the middle of a landscape screen in this task. After 500 ms (ms), participants were then shown a pair of 1 neutral face and 1 angry face (randomized) from the NIMSTIM faces (note that only faces that sufficiently elicited the desired emotions were shown; Tottenham et al., 2009). The pair of faces could either be open-mouthed or closed-mouthed (each open-mouthed photo was paired with an open-mouthed photo from the same person, and each closed-mouth photo was paired with a closed-mouth photo of the same person). After 500ms, they were shown a dot replacing either the neutral or angry face and asked to click the dot as quickly as possible. The image set that they were shown (of 52), the side where the angry face appeared, the side where the dot appeared, the side that they clicked on, and the RT in milliseconds (ms) were recorded. This trial was repeated a total of 5 times. Participants could complete the task as often as they liked during the study duration, which occurred across 14.62 days on average ($SD = 28.66$). On average, each participant engaged in the five-trial visual dot probe across 19.85 sessions ($SD = 20.35$, range = 1–110), but the median was 12 sessions. No differences in number of trials completed were observed between college student and community adult participants. Note that our study did not instruct participants to complete the visual dot probe EMA for a set number of trials because it aimed to determine the generalizability of attention bias phenomena outside laboratory settings and in real-world *in situ* contexts.

Nonetheless, our observed median of 60 trials (12 sessions x 5 trials/session) and observed mean of 100 trials (20 sessions x 5 trials/session) were consistent with established practices of administering EMA dot-probe tasks in psychiatric samples (i.e., ranging between 60 and 80 trials; Moskal et al., 2022; Smith et al., 2020). Moreover, data collection for this study began before published recommendations to increase the number of trials for dot-probe tasks to advance clinical science on this topic (Price et al., 2015). In addition, the number of participants who completed 2, 5, and 12 sessions was 294, 151, and 92, respectively. In the following sections, we examined the reliabilities of the attention bias indices *within* and *across* 2, 5, and 12 sessions.

1.2.2. Generalized anxiety disorder Questionnaire–IV (GAD-Q-IV; 14-item; Newman et al., 2002)

The GAD-Q-IV self-report was developed to measure DSM-IV GAD criteria, equivalent to the DSM-5 criteria (American Psychiatric Association, 2013). Based on a clinical interview, it shows good sensitivity and specificity across multiple samples (Moore et al., 2014; Newman et al., 2002). Convergent and discriminant validity was evidenced by large associations with trait anxiety ($r = 0.58$) and worry ($r = 0.63$) and small links to unique constructs (e.g., $r = 0.23$ with depression; Newman et al., 2002; Zainal et al., 2019). It also has good two-week test-retest reliability ($r = 0.81$) and strong internal consistency ($\alpha = 0.94$) (Moore et al., 2014; Newman et al., 2002). An advantage of the GAD-IV over other measures is that it assesses the full DSM-5 GAD symptom constellation, i.e., items capturing the presence, severity, worry topics, distress, interference, and six worry-linked symptoms (Zainal et al., 2021). Further, it can be scored dimensionally or categorically based on diagnosis (Moore et al., 2014). The GAD sample was chosen based on the DSM-5 diagnostic algorithm. The level of agreement with the Anxiety Disorder Interview Schedule (ADIS; Brown & Barlow, 2014; Di Nardo et al., 1994) is also good (Cohen's $\kappa = 0.67$; 88% correct classification), with high sensitivity (0.81) and specificity (0.89) with the ADIS using a cut-off score of $\geq$5.7 (Newman et al., 2002).

1.2.3. Social Phobia Diagnostic Questionnaire (SPDQ; Newman et al., 2003)

The SPDQ self-report captures excessive fearfulness and avoidance of various observational and evaluative social situations and the extent of distress and interference related to those fears. It has been shown to have strong internal consistency ($\alpha = 0.95$), test-retest reliability ($r = 0.92$), and split-half reliability ($r = 0.90$; Newman et al., 2003). Additionally, the SPDQ has strong convergent validity (e.g., $r = 0.66$ with another SAD measure; Watson & Friend, 1969) and good discriminant validity (e.g., $r = 0.28$ with the GAD-Q-IV; Newman et al., 2003). Further, SPDQ

Table 1
Sociodemographic comparisons across various diagnostic groups.

	GAD		SAD		GAD and SAD		Controls	
	($n = 533$)		($n = 402$)		($n = 302$)		($n = 195$)	
	M	(*SD*)	*M*	(*SD*)	*M*	(*SD*)	*M*	(*SD*)
Age (years)	25.90	(9.77)	30.62	(97.20)	25.84	(10.18)	25.82	(13.45)
	n	(%)	*n*	(%)	*n*	(%)	*n*	(%)
Sex								
Female	468	(87.80)	341	(84.83)	264	(87.42)	145	(74.36)
Male	65	(12.20)	61	(15.17)	38	(12.58)	50	(25.64)
Ethnicity								
African American	28	(5.25)	23	(5.72)	14	(4.64)	15	(7.69)
Asian American	7	(1.31)	7	(1.74)	4	(1.32)	13	(6.67)
Caucasian	326	(61.16)	251	(62.44)	193	(63.91)	107	(54.87)
Hispanic/Latino	35	(6.57)	21	(5.22)	16	(5.30)	14	(7.18)
Multiracial/multiethnic	42	(7.88)	24	(5.97)	19	(6.29)	5	(2.56)
Native American	10	(1.88)	9	(2.24)	9	(2.98)	2	(1.03)
Other	85	(15.95)	67	(16.67)	47	(15.56)	39	(20.00)

Note. GAD = generalized anxiety disorder; SAD = social anxiety disorder.

items paralleled DSM-5 SAD diagnosis. When compared to the DSM-5 ADIS SAD diagnosis (Brown & Barlow, 2014; Di Nardo et al., 1994), the SPDQ has high level of sensitivity (0.85) and specificity (0.82) with the ADIS (Newman et al., 2003).

1.2.4. Data analyses

To handle outliers, we used a rescaling (Winsorizing) approach to minimize missing data frequency, a procedure recommended in studies using the visual dot probe (Price et al., 2015). Across all trials and persons, anomalous values 1.5 times above and below the 25th and 75th percentiles of the entire spread of RT values were substituted with the first and last valid values within that spread of values, i.e., data points were retained as new non-anomalous values. For instance, a spread of RT values with 500ms and 700ms as the 25th and 75th percentile values, respectively, has an interquartile range of 200ms. In this example, values < 200ms and >1,000ms would be rescaled to 200ms and 1,000ms, respectively, to reflect the smallest and largest valid values within that spread of values.

Next, based on the literature, we computed the various attention biases toward threats and related indices for each participant at each session. For all indices, congruent trials referred to trials where the location of the angry or neutral face was the same as the dot probe. Incongruent trials referred to trials where the location of the angry or neutral face was opposite to the dot probe. Threat (or neutral) trials were defined as trials where the angry (or neutral) face preceded and superimposed the dot probe location. Each session included 2 congruent-treat, 1 congruent-neutral, 1 incongruent-threat, and 1 incongruent-neutral trial(s) (randomized). This design ensured that participants engaged in either type of trial sufficiently to compute attention bias and related indices. Further, individuals can display attention bias away or toward threats within a single measurement session (Zvielli et al., 2015).

Global Attention Bias Index was computed using the following formula: Global attention bias indices = (Average RT from incongruent-threat trial(s) - Average RT from congruent-threat trial(s)) - (Average RT from incongruent-neutral trial(s) - Average RT from congruent-neutral trial(s)) (Price et al., 2015). *Disengagement Effect* was indexed by comparatively slower RT on incongruent threat relative to incongruent neutral trials (i.e., *Disengagement Effect* = Average RT from incongruent-threat trial(s)–Average RT from incongruent-neutral trials; Mogg et al., 2008). The *Facilitation Bias* was measured by faster correct RT for congruent threat trials than congruent neutral trials (i.e., *Facilitation Bias* = Average RT for congruent-neutral trial(s)–Average RT for congruent-threat trial(s)) (Waechter & Stolz, 2015). *Congruent RT* and *Incongruent RT* were based on RT during congruent and incongruent trials, respectively, regardless of response accuracy. *Attention Bias Variability* was the average standard deviation (*SD*) of Global attention bias indices across five trials within each session divided by the mean RT across all trials (i.e., Attention Bias Variability = *SD* of Global attention bias indices per session/(Average RT from incongruent-threat trial(s) + Average RT from incongruent-neutral trial(s) + Average RT from congruent-threat trial(s) + Average RT from congruent-neutral trial(s)) (Iacoviello et al., 2014). Last, participant-level Trial-Level Mean and Variability Bias scores were computed by subtracting the RT of threat-congruent trials from the nearest threat-incongruent trial inside a five-trial structure (Zvielli et al., 2015). *R* syntaxes for calculating attention bias indices were adapted from various online tutorials (e.g., https://osf.io/smzd5; https://tinyurl.com/ynk36662). Higher values on these indices denote higher degree of these constructs.

Test-test-retest reliability was indexed with the intraclass correlation coefficient (ICC). The ICC is a standard marker of reliability that simultaneously includes data from all time points and reflects proportion of variance of an assessment attributable to between-person variance (Weir, 2005). ICC values theoretically range from 0 to 1, where 0 denotes absence of reliability, <0.5 poor reliability, 0.5 to 0.75 moderate reliability, 0.75 to 0.9 good reliability, above 0.9 excellent reliability, and 1 perfect reliability (Koo & Li, 2016). Nonetheless, empirical ICC estimates can be negative as all estimates have a maximum limit value of 1 but no minimum limit value—negative ICCs parallel ICC values of 0, signifying absence of reliability. We computed ICC values using the *irr R* package (Gamer et al., 2007) using a two-way random-effects model (participant and time-point (trials and sessions) were regarded as random factors). Average-measures ICC values offer a global index of the proportion of variance attributable to between-person variance across all time points. To determine internal reliability, split-half reliability (across odd and even trials or sessions denoted by Pearson's *r*) was estimated. Split-half reliability *r* values of 0.10, 0.30, and 0.50 were considered substantially small, moderate, and large, respectively (Cohen, 1988). In addition, refer to the online supplemental materials for hierarchical linear modeling for interaction of trial and diagnostic status predicting for 8 attention bias and related indices.

2. Results

2.1. Reliability analysis

2.1.1. Across 5 trials per session

Table 2 displays the reliability analyses for attention biases toward

Table 2
Reliability analysis for five attention bias and related indices across five trials during each session.

	ICC	[95% CI]	Split-half reliability (*r*)
All participants (*n* = 829)			
Global ABI	0.04	[-0.003, 0.08]	0.04
Disengagement effect	0.21	[0.17, 0.24]	0.22
Facilitation effect	0.25	[0.22, 0.28]	0.25
Congruent RT	0.73	[0.70, 0.75]	0.75
Incongruent RT	0.96	[0.92, 0.98]	0.98
Trial-level bias score ABI	0.63	[0.61, 0.64]	0.65
Trial-level bias AB Variability	0.56	[0.48, 0.63]	0.58
Participants with GAD (*n* = 527)			
Global ABI	0.07	[0.02, 0.12]	0.07
Disengagement effect	0.28	[0.24, 0.32]	0.29
Facilitation effect	0.26	[0.22, 0.30]	0.27
Congruent RT	0.72	[0.70, 0.75]	0.76
Incongruent RT	0.95	[0.89, 0.97]	0.98
Trial-level bias score ABI	0.63	[0.61, 0.65]	0.65
Trial-level bias AB Variability	0.57	[0.46, 0.66]	0.58
Participants with SAD (*n* = 398)			
Global ABI	0.04	[-0.02, 0.10]	0.03
Disengagement effect	0.22	[0.17, 0.27]	0.23
Facilitation effect	0.23	[0.18, 0.28]	0.23
Congruent RT	0.72	[0.69, 0.75]	0.78
Incongruent RT	0.97	[0.92, 0.98]	0.98
Trial-level bias score ABI	0.64	[0.62, 0.66]	0.66
Trial-level bias AB Variability	0.54	[0.42, 0.65]	0.57
Participants with both GAD and SAD (*n* = 299)			
Global ABI	0.07	[-0.01, 0.13]	0.06
Disengagement effect	0.31	[0.26, 0.36]	0.32
Facilitation effect	0.22	[0.16, 0.27]	0.22
Congruent RT	0.72	[0.69, 0.76]	0.78
Incongruent RT	0.95	[0.89, 0.97]	0.98
Trial-level bias score ABI	0.65	[0.63, 0.68]	0.67
Trial-level bias AB Variability	0.51	[0.35, 0.64]	0.52
Non-GAD and SAD Controls (*n* = 193)			
Global ABI	0.01	[-0.08, 0.09]	0.01
Disengagement effect	0.17	[0.10, 0.24]	0.19
Facilitation effect	0.22	[0.16, 0.29]	0.22
Congruent RT	0.73	[0.69, 0.77]	0.76
Incongruent RT	0.97	[0.93, 0.98]	0.98
Trial-level bias score ABI	0.62	[0.59, 0.65]	0.64
Trial-level bias AB Variability	0.51	[0.34, 0.65]	0.57

Note. ABI = attention bias index; CI = confidence interval; GAD = generalized anxiety disorder; ICC = intra-class correlation; RT = response time; SAD = social anxiety disorder.

threat and related indices across 5 trials during each session for all participants and diagnostic subgroups (GAD, SAD, Mixed GAD and SAD, Non-GAD and SAD Controls). Global Attention Bias Index had poor ICC statistics (mean ICCs = 0.01–0.07) and split-half reliability (r = 0.01–0.02). Likewise, poor test-retest reliability and internal reliability outcomes were observed for Facilitation Bias (mean ICCs = 0.22–0.26; split-half r = 0.05–0.07) and Disengagement Effect (mean ICCs = 0.17–0.31) indices. However, analyses suggested moderate retest-reliability and internal reliability outcomes for Trial-Level Mean Bias Toward Threat Score (mean ICCs = 0.62–0.65; split-half r = 0.64–0.67) and Trial-Level Bias Attention Bias Variability (mean ICCs = 0.51–0.57; split-half r = 0.52–0.59). Conversely, analyses revealed moderate retest-reliability and internal reliability outcomes for Congruent RT (mean ICCs = 0.72–0.73; split-half r = 0.37–0.42). Relatedly, excellent retest-reliability and internal reliability were found for Incongruent RT (mean ICCs = 0.95–0.97; split-half r = 0.90–0.92).

2.1.2. Across 2, 5, and 12 sessions

Table 3 shows the reliability analyses for attention bias toward threats and related indices across 2, 5, and 12 sessions for all participants and diagnostic subgroups. Across 2 sessions (10 trials), analyses generated moderate-to-excellent test-retest reliability for these 4 attention bias indices: (1) Trial-Level Mean Bias Toward Threat Score (mean ICCs = 0.610–0.655; split-half r = 0.630–0.670); (2) Trial-Level Bias Attention Bias Variability (mean ICCs = 0.507–0.567; split-half r = 0.520–0.580); (3) Congruent RT (mean ICCs = 0.692–0.732; split-half r = 0.680–0.710); (4) Incongruent RT (mean ICCs = 0.959–0.979; split-half r = 0.940–0.950). However, analyses yielded insufficient test-retest reliability and internal reliability in the remaining 4 visual dot probe indices: (1) Global Attention Bias Index (mean ICCs = −0.22–0.07; split-half r = 0.02–0.07); (2) Facilitation Bias (mean ICCs = −0.30–0.08; split-half r = 0.01–0.08); (3) Disengagement Effect (mean ICCs = −0.10–0.14; split-half r = 0.01–0.08); (4) Attention Bias Variability (mean ICCs = −0.04–0.19; split-half r = 0.0003–0.02).

Across 5 sessions (25 trials), analyses indicated moderate-to-excellent test-retest reliability for these 4 attention bias indices: (1) Trial-Level Mean Bias Toward Threat Score (mean ICCs = 0.604–0.641; split-half r = 0.630–0.660); (2) Trial-Level Bias Attention Bias Variability (mean ICCs = 0.507–0.567; split-half r = 0.520–0.580); (3) Congruent RT (mean ICCs = 0.694–0.744; split-half r = 0.680–0.730); (4) Incongruent RT (mean ICCs = 0.953–0.977; split-half r = 0.940–0.950). Despite that, the test-retest reliability and internal reliability for the other 4 attention bias and related indices were inadequate across 5 sessions: (1) Global Attention Bias Index (mean ICCs = −0.19–0.32; split-half r = 0.02–0.06); (2) Disengagement Effect (mean ICCs = −0.13–0.33; split-half r = 0.03–0.10); (3) Facilitation Bias (mean ICCs = −0.22–0.01; split-half r = 0.04–0.11); (4) Attention Bias Variability (mean ICCs = −0.36–0.19; split-half r = 0.02–0.04).

Across 12 sessions (60 trials), analyses suggested moderate-to-excellent test-retest reliability for these 4 attention bias indices: (1) Trial-Level Mean Bias Toward Threat Score (mean ICCs = 0.626–0.644; split-half r = 0.640–0.670); (2) Trial-Level Bias Attention Bias Variability (mean ICCs = 0.507–0.567; split-half r = 0.520–0.580); (3) Congruent RT (mean ICCs = 0.710–0.733; split-half r = 0.710–0.740); (4) Incongruent RT (mean ICCs = 0.948–0.969; split-half r = 0.930–0.940). However, across 12 sessions, test-retest reliability and internal reliability were poor for other 4 attention bias and related indices: (1) Global Attention Bias Index (mean ICCs = −0.81–0.03; split-half r = 0.05–0.08); (2) Disengagement Effect (mean ICCs = −0.42–0.19; split-half r = 0.04–0.07); (3) Facilitation Bias (mean ICCs = −0.38–0.03; split-half r = 0.03–0.07); (4) Attention Bias Variability (mean ICCs = −0.04–0.28; split-half r = 0.02–0.06).

3. Discussion

Supporting our hypotheses, results showed that within a five-trial

Table 3
Reliability analysis for six attention bias and related indices across two, five, and twelve sessions.

	ICC	[95% CI]	Split-half reliability (r)
Number of timestamps = 2 (Trials = 10)			
All participants (n = 294)			
Global ABI	−0.095	[-0.38, 0.13]	0.023
Disengagement effect	−0.037	[-0.31, 0.18]	0.019
Facilitation effect	−0.210	[-0.52, 0.04]	0.046
Congruent RT	0.710	[0.68, 0.74]	0.700
Incongruent RT	0.966	[0.92, 0.98]	0.940
AB Variability	0.002	[-0.26, 0.21]	0.014
Trial-level bias score ABI	0.630	[0.596, 0.661]	0.650
Trial-level bias score AB Variability	0.558	[0.478, 0.630]	0.580
Participants with GAD (n = 167)			
Global ABI	−0.219	[-0.66, 0.10]	0.069
Disengagement effect	−0.096	[-0.49, 0.19]	0.048
Facilitation effect	−0.303	[-0.77, 0.04]	0.084
Congruent RT	0.732	[0.70, 0.76]	0.710
Incongruent RT	0.959	[0.90, 0.98]	0.940
AB Variability	0.047	[-0.29, 0.30]	0.009
Trial-level bias score ABI	0.624	[0.58, 0.66]	0.640
Trial-level bias score AB Variability	0.567	[0.46, 0.66]	0.580
Participants with SAD (n = 74)			
Global ABI	0.030	[-0.37, 0.31]	0.050
Disengagement effect	0.125	[-0.23, 0.38]	0.069
Facilitation effect	−0.143	[-0.61, 0.19]	0.007
Congruent RT	0.692	[0.65, 0.73]	0.680
Incongruent RT	0.964	[0.92, 0.98]	0.940
AB Variability	0.070	[-0.31, 0.34]	0.011
Trial-level bias score ABI	0.655	[0.61, 0.70]	0.670
Trial-level bias score AB Variability	0.543	[0.42, 0.65]	0.570
Participants with both GAD and SAD (n = 88)			
Global ABI	0.034	[-0.48, 0.37]	0.036
Disengagement effect	0.138	[-0.32, 0.44]	0.077
Facilitation effect	−0.153	[-0.75, 0.24]	0.022
Congruent RT	0.731	[0.69, 0.77]	0.710
Incongruent RT	0.967	[0.92, 0.98]	0.940
AB Variability	0.190	[-0.24, 0.47]	0.0003
Trial-level bias score ABI	0.641	[0.59, 0.69]	0.660
Trial-level bias score AB Variability	0.507	[0.35, 0.64]	0.520
Non-GAD and SAD Controls (n = 81)			
Global ABI	0.070	[-0.45, 0.40]	0.018
Disengagement effect	0.008	[-0.55, 0.36]	0.005
Facilitation effect	−0.081	[-0.69, 0.31]	0.030
Congruent RT	0.723	[0.69, 0.77]	0.730
Incongruent RT	0.979	[0.95, 0.99]	0.950
AB Variability	−0.035	[-0.62, 0.34]	0.020
Trial-level bias score ABI	0.610	[0.54, 0.67]	0.630
Trial-level bias score AB Variability	0.511	[0.34, 0.65]	0.570
Number of timestamps = 5 (Trials = 25)			
All participants (n = 151)			
Global ABI	0.079	[-0.18, 0.29]	0.038
Disengagement effect	0.105	[-0.14, 0.31]	0.033
Facilitation effect	−0.080	[-0.38, 0.17]	0.040
Congruent RT	0.717	[0.69, 0.74]	0.710
Incongruent RT	0.961	[0.91, 0.98]	0.940
AB Variability	0.055	[-0.20, 0.27]	0.015
Trial-level bias score ABI	0.623	[0.60, 0.65]	0.640
Trial-level bias score AB Variability	0.558	[0.48, 0.63]	0.580
Participants with GAD (n = 87)			
Global ABI	−0.192	[-0.65, 0.17]	0.024
Disengagement effect	−0.128	[-0.56, 0.21]	0.037
Facilitation effect	−0.224	[-0.78, 0.17]	0.041
Congruent RT	0.744	[0.72, 0.77]	0.730
Incongruent RT	0.953	[0.89, 0.98]	0.940

(continued on next page)

Table 3 (*continued*)

	ICC	[95% CI]	Split-half reliability (*r*)
AB Variability	0.190	[-0.10, 0.42]	0.040
Trial-level bias score ABI	0.626	[0.59, 0.66]	0.640
Trial-level bias score AB Variability	0.567	[0.46, 0.66]	0.580
Participants with SAD (n = 74)			
Global ABI	0.009	[-0.40, 0.33]	0.026
Disengagement effect	0.028	[-0.37, 0.34]	0.070
Facilitation effect	−0.023	[-0.45, 0.31]	0.047
Congruent RT	0.694	[0.66, 0.73]	0.680
Incongruent RT	0.955	[0.90, 0.98]	0.930
AB Variability	−0.011	[-0.40, 0.30]	0.026
Trial-level bias score ABI	0.641	[0.60, 0.68]	0.660
Trial-level bias score AB Variability	0.543	[0.42, 0.65]	0.570
Participants with both GAD and SAD (n = 54)			
Global ABI	−0.160	[-0.75, 0.27]	0.028
Disengagement effect	−0.122	[-0.69, 0.30]	0.079
Facilitation effect	−0.166	[-0.76, 0.27]	0.053
Congruent RT	0.737	[0.70, 0.77]	0.720
Incongruent RT	0.961	[0.91, 0.98]	0.940
AB Variability	0.147	[-0.24, 0.45]	0.033
Trial-level bias score ABI	0.638	[0.59, 0.68]	0.660
Trial-level bias score AB Variability	0.507	[0.35, 0.64]	0.520
Non-GAD and SAD Controls (n = 44)			
Global ABI	0.319	[-0.06, 0.60]	0.058
Disengagement effect	0.326	[-0.05, 0.60]	0.098
Facilitation effect	−0.010	[-0.59, 0.40]	0.105
Congruent RT	0.722	[0.68, 0.76]	0.740
Incongruent RT	0.977	[0.95, 0.99]	0.950
AB Variability	−0.360	[-1.14, 0.19]	0.043
Trial-level bias score ABI	0.604	[0.55, 0.66]	0.630
Trial-level bias score AB Variability	0.511	[0.34, 0.65]	0.570
Number of timestamps = 12 (Trials = 60)			
All participants (n = 92)			
Global ABI	−0.236	[-0.65, 0.11]	0.045
Disengagement effect	−0.189	[-0.59, 0.14]	0.028
Facilitation effect	−0.132	[-0.51, 0.18]	0.033
Congruent RT	0.714	[0.69, 0.73]	0.710
Incongruent RT	0.954	[0.89, 0.98]	0.94
AB Variability	0.097	[-0.19, 0.34]	0.019
Trial-level bias score ABI	0.627	[0.605, 0.649]	0.650
Trial-level bias score AB Variability	0.558	[0.478, 0.630]	0.580
Participants with GAD (n = 52)			
Global ABI	−0.594	[-1.32, −0.02]	0.068
Disengagement effect	−0.636	[-1.38, −0.05]	0.048
Facilitation effect	−0.242	[-0.80, 0.20]	0.060
Congruent RT	0.725	[0.70, 0.75]	0.720
Incongruent RT	0.948	[0.88, 0.97]	0.930
AB Variability	0.126	[-0.25, 0.43]	0.048
Trial-level bias score ABI	0.626	[0.60, 0.65]	0.640
Trial-level bias score AB Variability	0.567	[0.46, 0.66]	0.580
Participants with SAD (n = 43)			
Global ABI	−0.550	[-1.35, 0.06]	0.058
Disengagement effect	−0.561	[-1.35, 0.05]	0.064
Facilitation effect	−0.320	[-1.00, 0.20]	0.048
Congruent RT	0.731	[0.69, 0.77]	0.710
Incongruent RT	0.950	[0.89, 0.97]	0.930
AB Variability	−0.041	[-0.54, 0.36]	0.031
Trial-level bias score ABI	0.637	[0.61, 0.67]	0.660
Trial-level bias score AB Variability	0.543	[0.42, 0.65]	0.570
Participants with both GAD and SAD (n = 31)			
Global ABI	−0.814	[-1.87, −0.01]	0.060
Disengagement effect	−0.917	[-1.99, −0.08]	0.094
Facilitation effect	−0.381	[-1.20, 0.23]	0.070
Congruent RT	0.710	[0.68, 0.74]	0.710
Incongruent RT	0.956	[0.90, 0.98]	0.930
AB Variability	0.076	[-0.44, 0.48]	0.042
Trial-level bias score ABI	0.644	[0.61, 0.68]	0.670
Trial-level bias score AB Variability	0.507	[0.35, 0.64]	0.520
Non-GAD and SAD Controls (n = 28)			
Global ABI	0.029	[-0.63, 0.49]	0.078
Disengagement effect	0.142	[-0.43, 0.55]	0.061
Facilitation effect	−0.034	[-0.73, 0.46]	0.067
Congruent RT	0.733	[0.70, 0.76]	0.740
Incongruent RT	0.969	[0.92, 0.98]	0.940
AB Variability	0.275	[-0.19, 0.61]	0.057
Trial-level bias score ABI	0.634	[0.59, 0.67]	0.650
Trial-level bias score AB Variability	0.511	[0.34, 0.65]	0.570

Note. ABI = attention bias index; CI = confidence interval; GAD = generalized anxiety disorder; ICC = intra-class correlation; RT = response time; SAD = social anxiety disorder.

session, all attention bias toward threats and related indices except for Trial-Level Mean and Variability Bias, Congruent RT, and Incongruent RT scores assessed by a smartphone visual dot probe EMA had poor psychometric properties. This pattern of test-retest reliability and internal consistency outcomes were similar across 2, 5, and 12 assessment sessions (10–60 trials). All reliability estimates were far from the minimum of 0.90 required to classify persons based on a uniform and stable attribute (Sattler, 2008) and to guide clinical decision-making (Rodebaugh et al., 2016). This pattern occurred for participants regardless of their GAD/SAD diagnostic status. Consequently, our smartphone visual dot probe EMA study adds to laboratory-based experimental evidence that the reliabilities of traditional attention bias and related indices (not based on the individual trial-level bias scores) derived from the visual dot probe tend to be very low (e.g., *r*s = 0.04 to 0.29 in children and adults; Brown et al., 2014; McNally et al., 2013). Traditional attention bias indices based on difference scores thus are likely not reliable markers of individual differences in attention bias toward threat. Therefore, findings strongly suggest that traditional attention bias toward threat markers derived from the visual dot probe EMA are poor indices of attention bias toward threat across anxiety and other mental disorders (Jones et al., 2018; MacLeod et al., 2019). Simultaneously, our EMA study replicates and extends previous reports of higher test-retest reliability of trial-level attention bias mean and variability scores than traditional attention bias indices among individuals with anxiety disorders (e.g., Carlson & Fang, 2020; Molloy & Anderson, 2020), PTSD, and other mental disorders (Schafer et al., 2016), highlighting the *dynamic* nature of attention bias constructs.

In addition, findings that Congruent RT and Incongruent RT showed within-session (ICCs = 0.72–0.96) and between-session stability (ICCs = 0.69–0.98) replicates prior studies that used key-press and eye-tracking visual dot probe laboratory assessments (Waechter et al., 2014) and emotional Stroop tasks (Eide et al., 2002). These two previous studies reported strong reliabilities for attention bias indices derived from raw RTs but weak reliabilities for attention bias indices based on RT difference scores (e.g., Stroop interference effect). Regrettably, raw RT scores are theoretically unimportant compared to attention bias indices based on RT difference scores because raw RT scores do not capture preferential orientation to threats or difficulties disengaging from threats (Carlson & Fang, 2020; Meissel et al., 2022). Moreover, stable individual differences in raw RT on dot-probe tasks might indicate consistent individual differences in attributes apart from attention bias toward threat, such as attentional regulation, executive functioning, and

related cognitive control processes (Paap & Sawi, 2016; Swick & Ashley, 2017).

Therefore, it may also be the case that attention bias models propose upper limits on the temporal stability of traditional attention bias indices. For instance, traditional attention bias indices based on RT difference scores derived by subtracting from 2 correlated assessments can inevitably limit reliability (Sipos et al., 2014). On that note, traditional attention bias indices based on variability scores have recently been challenged as these markers are not able to separate attention bias toward threat variability (signal) from measurement error (noise; Kruijt et al., 2016) and general attention control stability (Carlson et al., 2022). Moreover, traditional attention bias toward threat may not constitute an unchanging, trait-level feature but may vary due to mood states, focus, motivation, and other situational factors. For instance, high- (vs. low-) trait anxious persons showed greater attention bias toward threat marked by traditional indices after induced state anxiety (vs. calmness; Waechter & Stolz, 2015).

Relatedly, the low reliability of traditional Attention Bias Variability indices observed herein contrasts that of prior studies, which found good reliabilities in persons with posttraumatic stress disorder (PTSD) and persons with elevated SAD symptoms (Iacoviello et al., 2014; Price et al., 2015). Reasons for discrepancies might include differences in sample characteristics, dot probe task features (e.g., top-down vs. left-right display), and number of trials within and across sessions. For instance, Price et al. (2015) administered up to 320 trials per session and computed the Attention Bias Variability with a different formula (*SD* of Global Attention Bias Index across 16 bins/mean RT) than herein (*SD* of Global Attention Bias Index across 5 trials/mean RT). Moreover, the number of trials our visual dot probe EMA administered fell short of hundreds of trials needed per session to attain reliabilities of 0.85 and above (Miller & Ulrich, 2013). Future visual dot probe studies can test if increasing the number of trials improves reliability and determine ways to maximize protocol compliance using an EMA design. Nonetheless, our dot probe attention bias EMA generated trial-level attention bias variability scores with moderate test-retest reliabilities, outcomes similarly documented in prior laboratory studies that recruited patients with anxiety disorders and PTSD (Carlson & Fang, 2020; Caudek et al., 2017; Gade et al., 2022). Overall, using trial-level attention bias variability scores is an essential step toward accurately measuring this phenomenon in naturalistic settings.

Simultaneously, persons with GAD were, on average, more likely to display higher trial-level mean attention bias toward threat scores than their non-anxious counterparts, with a small effect size. On average, those with (vs. without) SAD were likelier to show greater traditional attention bias and facilitation effect scores. These outcomes are consistent with abundant evidence that chronic worriers tend to show high attention bias toward threats depicted in facial expressions (vs. words) across different age groups (Goodwin et al., 2017; Waters et al., 2015; Zainal & Newman, 2018). Results also parallel meta-analytic data that highly self-conscious and socially anxious people tend to show increased vigilance to negative than neutral faces displayed in 500ms, with small effect sizes (Bantin et al., 2016; Bar-Haim et al., 2007). Examining if trait empathy, executive functioning, initial attention deployment, and transient mood states make people with GAD and SAD susceptible to attention bias toward threats are intriguing avenues for future visual dot probe EMA studies.

Study limitations must be acknowledged. We only used one form of an attention bias measure, i.e., the button-press smartphone-delivered visual dot probe EMA task. Other attention bias paradigms (e.g., multi-stimulus free-viewing test, visual search task, eye-tracking visual dot probe task) and the novel drift-diffusion modeling computational techniques might yield higher reliability in samples with various anxiety disorders (Lazarov et al., 2016; Price et al., 2019). For example, it is plausible that people with GAD and SAD persist in visually fixating on threatening objects or people in the room (Lazarov et al., 2019). Additionally, harnessing social stimuli (e.g., faces conveying negative emotions) than non-social stimuli (e.g., predatory animals such as snakes) is likely to capture the attention bias toward threat phenomena more strongly (Leppanen et al., 2018). RT-based attention bias indices that depend on button presses as markers of attention may introduce confounds linked to the need for fine hand motor responses and preclude straightforward interpretation of findings (Olatunji et al., 2013). Thus, testing alternative smartphone attention bias measures could advance the field. Also, our visual dot probe EMA did not administer over 200 trials per session as done in previous studies (e.g., Price et al., 2015). Future visual dot probe EMA research should thus ensure adequate trials are administered. Clinical psychological science can also profit from using multimodal biomarker assessments of attention bias toward threats, such as electromyography (Fani et al., 2012) and electrophysiology (Wieser et al., 2018). Nonetheless, study strengths include the large sample size, use of a community sample, and examination of attention bias toward threats across multiple diagnostic groups. Further, we used DSM-5 criteria-consistent instruments to diagnose GAD and SAD. Last, we selected an extensive age range to capture a wide variety of cognitive control abilities and RT variability that previous studies suggested were differentially associated with attention bias toward threats (e.g., Namaky et al., 2017) and to draw inferences from our results that were generalizable across various age groups.

Some clinical implications deserve mention. To offset constraints related to accessing face-to-face evidence-based cognitive behavior therapy, researchers have capitalized on using smartphones to disseminate ABM interventions for anxiety-disordered persons (Enock et al., 2014). However, the efficacy of ABM appears to be limited to controlled laboratory and academic clinical settings. Meta-analytic data have shown that home-based Internet-delivered ABMs produced dissatisfying null effects for reducing self-reported anxiety symptoms (Linetzky et al., 2015). Despite doubts about the extent to which ABMs are scalable, treatments guiding attention toward benign or positive cues across multiple contexts seem encouraging (Waters et al., 2015) and merit more empirical scrutiny.

To summarize, our findings suggest that future studies should not use our approach to allow participants to use the dot probe EMA as much as they would like but to explicitly instruct the completion of a fixed number of trials that include $\geq$320 trials. This strategy has been shown to enhance test-retest reliability for measuring the standard global attention bias indices in laboratory contexts (Aday & Carlson, 2019). Moreover, future research should strive to examine attention bias paradigms beyond the dot-probe task that evidenced meaningful test-retest reliability attributes in laboratory settings (e.g., proportion of dwell time on threat stimuli assessed by free-viewing eye-tracking-based attention bias measures; Clauss et al., 2022; van Ens et al., 2019) and self-report assessments (Azriel et al., 2022), which await testing in real-world naturalistic contexts deploying EMA-based intensive longitudinal designs. Last, trial-level mean and variability bias scores measured by the visual dot probe EMA generate moderate test-retest reliability compared to traditional attention bias scores that produce poor reliabilities among those with and without GAD and SAD.

CRediT authorship contribution statement

Nur Hani Zainal: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing. **Nicholas C. Jacobson:** Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no conflict of interest.

Data availability

The authors are unable or have chosen not to specify which data has been used.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jbtep.2023.101918.

References

Aday, J. S., & Carlson, J. M. (2019). Extended testing with the dot-probe task increases test-retest reliability and validity. *Cognitive Processing, 20*(1), 65–72. https://doi.org/10.1007/s10339-018-0886-1

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). American Psychiatric Association. https://doi.org/10.1176/appi.books.9780890425596

Amir, N., Beard, C., Burns, M., & Bomyea, J. (2009). Attention modification program in individuals with generalized anxiety disorder. *Journal of Abnormal Psychology, 118*(1), 28–33. https://doi.org/10.1037/a0012589

Azriel, O., Britton, J. C., Gober, C. D., Pine, D. S., & Bar-Haim, Y. (2022). Development and validation of the attention bias Questionnaire (ABQ). *International Journal of Methods in Psychiatric Research, 31*(2), Article e1905. https://doi.org/10.1002/mpr.1905

Bantin, T., Stevens, S., Gerlach, A. L., & Hermann, C. (2016). What does the facial dot-probe task tell us about attentional processes in social anxiety? A systematic review. *Journal of Behavior Therapy and Experimental Psychiatry, 50*, 40–51. https://doi.org/10.1016/j.jbtep.2015.04.009

Bar-Haim, Y. (2010). Attention bias modification (ABM): A novel treatment for anxiety disorders. *Journal of Child Psychology and Psychiatry, 51*(8), 859–870. https://doi.org/10.1111/j.1469-7610.2010.02251.x

Bar-Haim, Y., Lamy, D., Pergamin, L., Bakermans-Kranenburg, M. J., & van Ijzendoorn, M. H. (2007). Threat-related attentional bias in anxious and nonanxious individuals: A meta-analytic study. *Psychological Bulletin, 133*(1), 1–24. https://doi.org/10.1037/0033-2909.133.1.1

Bouwmans, C. A. M., Vemer, P., van Straten, A., Tan, S. S., & Hakkaart-van Roijen, L. (2014). Health-related quality of life and productivity losses in patients with depression and anxiety disorders. *Journal of Occupational and Environmental Medicine, 56*(4), 420–424. https://doi.org/10.1097/jom.0000000000000112

Brown, T. A., & Barlow, D. H. (2014). *Anxiety and related disorders interview schedule for DSM-5 (ADIS-5L): Client interview schedule*. Oxford University Press.

Brown, H. M., Eley, T. C., Broeren, S., MacLeod, C., Rinck, M., Hadwin, J. A., & Lester, K. J. (2014). Psychometric properties of reaction time based experimental paradigms measuring anxiety-related information-processing biases in children. *Journal of Anxiety Disorders, 28*(1), 97–107. https://doi.org/10.1016/j.janxdis.2013.11.004

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, 44*(5), 657–669. https://doi.org/10.1007/s11031-020-09834-6

Carlson, J. M., Fang, L., & Kassel, D. (2022). The questionable validity of attention bias variability: Evidence from two conceptually unrelated cognitive tasks. *Journal of Affective Disorders Reports, 10*. https://doi.org/10.1016/j.jadr.2022.100411

Caudek, C., Ceccarini, F., & Sica, C. (2017). Facial expression movement enhances the measurement of temporal dynamics of attentional bias in the dot-probe task. *Behaviour Research and Therapy, 95*, 58–70. https://doi.org/10.1016/j.brat.2017.05.003

Clauss, K., Gorday, J. Y., & Bardeen, J. R. (2022). Eye tracking evidence of threat-related attentional bias in anxiety- and fear-related disorders: A systematic review and meta-analysis. *Clinical Psychology Review, 93*, Article 102142. https://doi.org/10.1016/j.cpr.2022.102142

Cohen, J. (1988). *Statistical power for the social sciences*. Hillsdale, NJ: Laurence Erlbaum and Associates.

Cooper, R. M., Bailey, J. E., Diaper, A., Stirland, R., Renton, L. E., Benton, C. P., … Munafò, M. R. (2011). Effects of 7.5% CO2 inhalation on allocation of spatial attention to facial cues of emotional expression. *Cognition & Emotion, 25*(4), 626–638. https://doi.org/10.1080/02699931.2010.508887

Cristea, I. A., Kok, R. N., & Cuijpers, P. (2015). Efficacy of cognitive bias modification interventions in anxiety and depression: Meta-analysis. *British Journal of Psychiatry, 206*(1), 7–16. https://doi.org/10.1192/bjp.bp.114.146761

Cuthbert, B. N., & Insel, T. R. (2013). Toward the future of psychiatric diagnosis: The seven pillars of RDoC. *BMC Medicine, 11*(1), 1–8. https://doi.org/10.1186/1741-7015-11-126. Article 126.

Davis, M. L., Rosenfield, D., Bernstein, A., Zvielli, A., Reinecke, A., Beevers, C. G., … Smits, J. A. J. (2016). Attention bias dynamics and symptom severity during and following CBT for social anxiety disorder. *Journal of Consulting and Clinical Psychology, 84*(9), 795–802. https://doi.org/10.1037/ccp0000125

Dennis, T. A., & O'Toole, L. J. (2014). Mental health on the go: Effects of a gamified attention-bias modification mobile application in trait-anxious adults. *Clinical Psychological Science, 2*(5), 576–590. https://doi.org/10.1177/2167702614522228

Di Nardo, P. A., Brown, T. A., & Barlow, D. H. (1994). *Anxiety disorders interview schedule for DSM-IV: Lifetime version (ADIS-IV-L)*. Psychological Corp.

Eide, P., Kemp, A., Silberstein, R. B., Nathan, P. J., & Stough, C. (2002). Test-retest reliability of the emotional stroop task: Examining the paradox of measurement change. *Journal of Psychology, 136*(5), 514–520. https://doi.org/10.1080/00223980209605547

Enock, P. M., Hofmann, S. G., & McNally, R. J. (2014). Attention bias modification training via smartphone to reduce social anxiety: A randomized, controlled multi-session experiment. *Cognitive Therapy and Research, 38*(2), 200–216. https://doi.org/10.1007/s10608-014-9606-z

van Ens, W., Schmidt, U., Campbell, I. C., Roefs, A., & Werthmann, J. (2019). Test-retest reliability of attention bias for food: Robust eye-tracking and reaction time indices. *Appetite, 136*, 86–92. https://doi.org/10.1016/j.appet.2019.01.020

Fani, N., Tone, E. B., Phifer, J., Norrholm, S. D., Bradley, B., Ressler, K. J., … Jovanovic, T. (2012). Attention bias toward threat is associated with exaggerated fear expression and impaired extinction in PTSD. *Psychological Medicine, 42*(3), 533–543. https://doi.org/10.1017/S0033291711001565

Fodor, L. A., Georgescu, R., Cuijpers, P., Szamoskozi, Ş., David, D., Furukawa, T. A., & Cristea, I. A. (2020). Efficacy of cognitive bias modification interventions in anxiety and depressive disorders: A systematic review and network meta-analysis. *The Lancet Psychiatry, 7*(6), 506–514. https://doi.org/10.1016/S2215-0366(20)30130-9

Gade, P. R., Molloy, A., & Anderson, P. L. (2022). The use of trial-level bias scores to examine attention bias and attention bias variability among people with and without social anxiety disorder. *Journal of Clinical Psychology, 78*(5), 847–856. https://doi.org/10.1002/jclp.23259

Gamer, M., Lemon, J., Fellows, I., & Singh, P. (2007). Irr: Various coefficients of interrater reliability and agreement (2017. In H. Goodwin, J. Yiend, & C. R. Hirsch (Eds.), *Generalized anxiety disorder, worry and attention to threat: A systematic review. Clinical Psychology review* (pp. 107–122). https://doi.org/10.1016/j.cpr.2017.03.006, 54.

Goodwin, H., Yiend, J., & Hirsch, C. R. (2017). Generalized anxiety disorder, worry and attention to threat: A systematic review. *Clinical Psychology Review, 54*, 107–122. https://doi.org/10.1016/j.cpr.2017.03.006.

Hallion, L. S., & Ruscio, A. M. (2011). A meta-analysis of the effect of cognitive bias modification on anxiety and depression. *Psychological Bulletin, 137*(6), 940–958. https://doi.org/10.1037/a0024355

Hsu, K. J., Shumake, J., Caffey, K., Risom, S., Labrada, J., Smits, J. A. J., … Beevers, C. G. (2021). Efficacy of attention bias modification training for depressed adults: A randomized clinical trial. *Psychological Medicine*, 1–9. https://doi.org/10.1017/S0033291721000702

Huppert, J. D., Kivity, Y., Cohen, L., Strauss, A. Y., Elizur, Y., & Weiss, M. (2018). A pilot randomized clinical trial of cognitive behavioral therapy versus attentional bias modification for social anxiety disorder: An examination of outcomes and theory-based mechanisms. *Journal of Anxiety Disorders, 59*, 1–9. https://doi.org/10.1016/j.janxdis.2018.08.002

Iacoviello, B. M., Wu, G., Abend, R., Murrough, J. W., Feder, A., Fruchter, E., … Charney, D. S. (2014). Attention bias variability and symptoms of posttraumatic stress disorder. *Journal of Traumatic Stress, 27*(2), 232–239. https://doi.org/10.1002/jts.21899

Jacobson, N. C. (2020). *Mood Triggers app*. http://www.nicholasjacobson.com/project/mood-triggers/.

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 : Journal of the Society of Psychologists in Addictive Behaviors, 32*(8), 922–932. https://doi.org/10.1037/adb0000414

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Koster, E. H. W., Crombez, G., Van Damme, S., Verschuere, B., & De Houwer, J. (2004). Does imminent threat capture and hold attention? *Emotion, 4*(3), 312–317. https://doi.org/10.1037/1528-3542.4.3.312

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), Article e0166600. https://doi.org/10.1371/journal.pone.0166600

Lazarov, A., Abend, R., & Bar-Haim, Y. (2016). Social anxiety is related to increased dwell time on socially threatening faces. *Journal of Affective Disorders, 193*, 282–288. https://doi.org/10.1016/j.jad.2016.01.007

Lazarov, A., & Bar-Haim, Y. (2021). Emerging domain-based treatments for pediatric anxiety disorders. *Biological Psychiatry, 89*(7), 716–725. https://doi.org/10.1016/j.biopsych.2020.08.030

Lazarov, A., Marom, S., Yahalom, N., Pine, D. S., Hermesh, H., & Bar-Haim, Y. (2018). Attention bias modification augments cognitive–behavioral group therapy for social anxiety disorder: A randomized controlled trial. *Psychological Medicine, 48*(13), 2177–2185. https://doi.org/10.1017/S003329171700366X

Lazarov, A., Suarez-Jimenez, B., Tamman, A., Falzon, L., Zhu, X., Edmondson, D. E., & Neria, Y. (2019). Attention to threat in posttraumatic stress disorder as indexed by eye-tracking indices: A systematic review. *Psychological Medicine, 49*(5), 705–726. https://doi.org/10.1017/S0033291718002313

LeDoux, J. E., & Pine, D. S. (2016). Using neuroscience to help understand fear and anxiety: A two-system framework. *American Journal of Psychiatry, 173*(11), 1083–1093. https://doi.org/10.1176/appi.ajp.2016.16030353

Leppanen, J. M., Cataldo, J. K., Bosquet Enlow, M., & Nelson, C. A. (2018). Early development of attention to threat-related facial expressions. *PLoS One, 13*(5), Article e0197424. https://doi.org/10.1371/journal.pone.0197424

Linetzky, M., Pergamin-Hight, L., Pine, D. S., & Bar-Haim, Y. (2015). Quantitative evaluation of the clinical efficacy of attention bias modification treatment for

anxiety disorders. *Depression and Anxiety, 32*(6), 383–391. https://doi.org/10.1002/da.22344
Lipp, O. V., & Derakshan, N. (2005). Attentional bias to pictures of fear-relevant animals in a dot probe task. *Emotion, 5*(3), 365–369. https://doi.org/10.1037/1528-3542.5.3.365
MacLeod, C., Grafton, B., & Notebaert, L. (2019). Anxiety-linked attentional bias: Is it reliable? *Annual Review of Clinical Psychology, 15*(1), 529–554. https://doi.org/10.1146/annurev-clinpsy-050718-095505
MacLeod, C., Mathews, A., & Tata, P. (1986). Attentional bias in emotional disorders. *Journal of Abnormal Psychology, 95*(1), 15–20. https://doi.org/10.1037/0021-843X.95.1.15
Mathews, A., & MacLeod, C. (2005). Cognitive vulnerability to emotional disorders. *Annual Review of Clinical Psychology, 1*(1), 167–195. https://doi.org/10.1146/annurev.clinpsy.1.102803.143916
McNally, R. J., Enock, P. M., Tsai, C., & Tousian, M. (2013). Attention bias modification for reducing speech anxiety. *Behaviour Research and Therapy, 51*(12), 882–888. https://doi.org/10.1016/j.brat.2013.10.001
Meissel, E. E. E., Liu, H., Stevens, E. S., Evans, T. C., Britton, J. C., Letkiewicz, A. M., & Shankman, S. A. (2022). The reliability and validity of response-based measures of attention bias. *Cognitive Therapy and Research, 46*(1), 146–160. https://doi.org/10.1007/s10608-021-10212-w
Miller, J., & Ulrich, R. (2013). Mental chronometry and individual differences: Modeling reliabilities and correlations of reaction time means and effect sizes. *Psychonomic Bulletin & Review, 20*(5), 819–858. https://doi.org/10.3758/s13423-013-0404-5
Mogg, K., & Bradley, B. P. (2005). Attentional bias in generalized anxiety disorder versus depressive disorder. *Cognitive Therapy and Research, 29*(1), 29–45. https://doi.org/10.1007/s10608-005-1646-y
Mogg, K., Holmes, A., Garner, M., & Bradley, B. P. (2008). Effects of threat cues on attentional shifting, disengagement and response slowing in anxious individuals. *Behaviour Research and Therapy, 46*(5), 656–667. https://doi.org/10.1016/j.brat.2008.02.011
Mogg, K., Philippot, P., & Bradley, B. P. (2004). Selective attention to angry faces in clinical social phobia. *Journal of Abnormal Psychology, 113*(1), 160–165. https://doi.org/10.1037/0021-843X.113.1.160
Mogoaşe, C., David, D., & Koster, E. H. W. (2014). Clinical efficacy of attentional bias modification procedures: An updated meta-analysis. *Journal of Clinical Psychology, 70*(12), 1133–1157. https://doi.org/10.1002/jclp.22081
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), 883–888. https://doi.org/10.1037/pas0000912
Moore, M. T., Anderson, N. L., Barnes, J. M., Haigh, E. A. P., & Fresco, D. M. (2014). Using the GAD-Q-IV to identify generalized anxiety disorder in psychiatric treatment seeking and primary care medical samples. *Journal of Anxiety Disorders, 28*(1), 25–30. https://doi.org/10.1016/j.janxdis.2013.10.009
Moskal, K. R., Dvorak, R. D., Burr, E. K., El Helou, G., Cora, J. L., De Leon, A. N., & Leary, A. V. (2022). Attention bias and alcohol craving: Differential effects via biological sex and mood. *Alcoholism: Clinical and Experimental Research, 46*(8), 1580–1591. https://doi.org/10.1111/acer.14894
Namaky, N., Beltzer, M. L., Werntz, A. J., Lambert, A. E., Isaacowitz, D. M., & Teachman, B. A. (2017). Moderators of age effects on attention bias toward threat and its association with anxiety. *Journal of Affective Disorders, 216*, 46–57. https://doi.org/10.1016/j.jad.2016.10.048
Newman, M. G., Kachin, K. E., Zuellig, A. R., Constantino, M. J., & Cashman-McGrath, L. (2003). The Social Phobia Diagnostic Questionnaire: Preliminary validation of a new self-report diagnostic measure of social phobia. *Psychological Medicine, 33*(4), 623–635. https://doi.org/10.1017/S0033291703007669
Newman, M. G., Llera, S. J., Erickson, T. M., Przeworski, A., & Castonguay, L. G. (2013). Worry and generalized anxiety disorder: A review and theoretical synthesis of research on nature, etiology, and treatment. *Annual Review of Clinical Psychology, 9*(1), 275–297. https://doi.org/10.1146/annurev-clinpsy-050212-185544
Newman, M. G., Zuellig, A. R., Kachin, K. E., Constantino, M. J., Przeworski, A., Erickson, T., & Cashman-McGrath, L. (2002). Preliminary reliability and validity of the generalized anxiety disorder questionnaire-IV: A revised self-report diagnostic measure of generalized anxiety disorder. *Behavior Therapy, 33*(2), 215–233. https://doi.org/10.1016/S0005-7894(02)80026-0
Olatunji, B. O., Armstrong, T., McHugo, M., & Zald, D. H. (2013). Heightened attentional capture by threat in veterans with PTSD. *Journal of Abnormal Psychology, 122*(2), 397–405. https://doi.org/10.1037/a0030440
Paap, K. R., & Sawi, O. (2016). The role of test-retest reliability in measuring individual and group differences in executive functioning. *Journal of Neuroscience Methods, 274*, 81–93. https://doi.org/10.1016/j.jneumeth.2016.10.002
Packard, G., Molloy, A., & Anderson, P. L. (2022). Trait mindfulness is related to attention bias toward threat and attention bias variability in social anxiety disorder. *Current Psychology*. https://doi.org/10.1007/s12144-022-02817-5
Parsons, S., Kruijt, A.-W., & Fox, E. (2019). Psychological science needs a standard practice of reporting the reliability of cognitive-behavioral measurements. *Advances in Methods and Practices in Psychological Science, 2*(4), 378–395. https://doi.org/10.1177/2515245919879695
Pettit, J. W., Bechor, M., Rey, Y., Vasey, M. W., Abend, R., Pine, D. S., … Silverman, W. K. (2020). A randomized controlled trial of attention bias modification treatment in youth with treatment-resistant anxiety disorders. *Journal of the American Academy of Child & Adolescent Psychiatry, 59*(1), 157–165. https://doi.org/10.1016/j.jaac.2019.02.018
Price, R. B., Brown, V., & Siegle, G. J. (2019). Computational modeling applied to the dot-probe task yields improved reliability and mechanistic insights. *Biological Psychiatry, 85*(7), 606–612. https://doi.org/10.1016/j.biopsych.2018.09.022
Price, R. B., Kuckertz, J. M., Siegle, G. J., Ladouceur, C. D., Silk, J. S., Ryan, N. D., … Amir, N. (2015). Empirical recommendations for improving the stability of the dot-probe task in clinical research. *Psychological Assessment, 27*(2), 365–376. https://doi.org/10.1037/pas0000036
Rodebaugh, T. L., Scullin, R. B., Langer, J. K., Dixon, D. J., Huppert, J. D., Bernstein, 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
Salthouse, T. A., & Berish, D. E. (2005). Correlates of within-person (across-occasion) variability in reaction time. *Neuropsychology, 19*(1), 77–87. https://doi.org/10.1037/0894-4105.19.1.77
Sattler, J. M. (2008). *Assessment of children: Cognitive foundations*. J.M. Sattler.
Schafer, J., Bernstein, A., Zvielli, A., Hofler, M., Wittchen, H. U., & Schonfeld, S. (2016). Attentional bias temporal dynamics predict posttraumatic stress symptoms: A prospective-longitudinal study among soldiers. *Depression and Anxiety, 33*(7), 630–639. https://doi.org/10.1002/da.22526
Schmukle, S. C. (2005). Unreliability of the dot probe task. *European Journal of Personality, 19*(7), 595–605. https://doi.org/10.1002/per.554
Shechner, T., Rimon-Chakir, A., Britton, J. C., Lotan, D., Apter, A., Bliese, P. D., … Bar-Haim, Y. (2014). Attention bias modification treatment augmenting effects on cognitive behavioral therapy in children with anxiety: Randomized controlled trial. *Journal of the American Academy of Child & Adolescent Psychiatry, 53*(1), 61–71. https://doi.org/10.1016/j.jaac.2013.09.016
Sipos, M. L., Bar-Haim, Y., Abend, R., Adler, A. B., & Bliese, P. D. (2014). Postdeployment threat-related attention bias interacts with combat exposure to account for PTSD and anxiety symptoms in soldiers. *Depression and Anxiety, 31*(2), 124–129. https://doi.org/10.1002/da.22157
Smith, K. E., Mason, T. B., Juarascio, A., Weinbach, N., Dvorak, R., Crosby, R. D., & Wonderlich, S. A. (2020). The momentary interplay of affect, attention bias, and expectancies as predictors of binge eating in the natural environment. *International Journal of Eating Disorders, 53*(4), 586–594. https://doi.org/10.1002/eat.23235
Staugaard, S. R. (2009). Reliability of two versions of the dot-probe task using photographic faces. *Psychology Science, 51*, 339–350.
Swick, D., & Ashley, V. (2017). Enhanced attentional bias variability in post-traumatic stress disorder and its relationship to more general impairments in cognitive control. *Scientific Reports, 7*(1), Article 14559. https://doi.org/10.1038/s41598-017-15226-7
Teng, M. H., Hou, Y. M., Chang, S. H., & Cheng, H. J. (2019). Home-delivered attention bias modification training via smartphone to improve attention control in sub-clinical generalized anxiety disorder: A randomized, controlled multi-session experiment. *Journal of Affective Disorders, 246*, 444–451. https://doi.org/10.1016/j.jad.2018.12.118
Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., … Nelson, C. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research, 168*(3), 242–249. https://doi.org/10.1016/j.psychres.2008.05.006
Tully, P. J., Harrison, N. J., Cheung, P., & Cosh, S. (2016). Anxiety and cardiovascular disease risk: A review. *Current Cardiology Reports, 18*(12), 120. https://doi.org/10.1007/s11886-016-0800-3
Vaish, A., Grossmann, T., & Woodward, A. (2008). Not all emotions are created equal: The negativity bias in social-emotional development. *Psychological Bulletin, 134*(3), 383–403. https://doi.org/10.1037/0033-2909.134.3.383
Vieira, M. M. M., Ferreira, T. B., Pacheco, P. A. F., Barros, P. O., Almeida, C. R. M., Araújo-Lima, C. F., … Bento, C. A. M. (2010). Enhanced Th17 phenotype in individuals with generalized anxiety disorder. *Journal of Neuroimmunology, 229*(1), 212–218. https://doi.org/10.1016/j.jneuroim.2010.07.018
Waechter, S., Nelson, A. L., Wright, C., Hyatt, A., & Oakman, J. (2014). Measuring attentional bias to threat: Reliability of dot probe and eye movement indices. *Cognitive Therapy and Research, 38*(3), 313–333. https://doi.org/10.1007/s10608-013-9588-2
Waechter, S., & Stolz, J. A. (2015). Trait anxiety, state anxiety, and attentional bias to threat: Assessing the psychometric properties of response time measures. *Cognitive Therapy and Research, 39*(4), 441–458. https://doi.org/10.1007/s10608-015-9670-z
Waters, A. M., Zimmer-Gembeck, M. J., Craske, M. G., Pine, D. S., Bradley, B. P., & Mogg, K. (2015). Look for good and never give up: A novel attention training treatment for childhood anxiety disorders. *Behaviour Research and Therapy, 73*, 111–123. https://doi.org/10.1016/j.brat.2015.08.005
Watson, D., & Friend, R. (1969). Measurement of social-evaluative anxiety. *Journal of Consulting and Clinical Psychology, 33*(4), 448–457. https://doi.org/10.1037/h0027806
Weir, J. P. (2005). Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *The Journal of Strength & Conditioning Research, 19*(1), 231–240. https://doi.org/10.1519/15184.1
Wieser, M. J., Hambach, A., & Weymar, M. (2018). Neurophysiological correlates of attentional bias for emotional faces in socially anxious individuals – evidence from a visual search task and N2pc. *Biological Psychology, 132*, 192–201. https://doi.org/10.1016/j.biopsycho.2018.01.004
Zainal, N. H., Hellberg, S. N., Kabel, K. E., Hotchkin, C. M., & Baker, A. W. (2023). Cognitive behavioral therapy (CBT) plus attention bias modification (ABM) reduces anxiety sensitivity and depressive symptoms in panic disorder: A pilot randomized trial. *Scandinavian Journal of Psychology*. https://doi.org/10.1111/sjop.12902
Zainal, H., & Magiati, I. (2019). A comparison between caregiver-reported anxiety and other emotional and behavioral difficulties in children and adolescents with autism spectrum disorders attending specialist or mainstream schools. *Journal of Autism and*

Developmental Disorders, 49(7), 2653–2663. https://doi.org/10.1007/s10803-016-2792-7

Zainal, N. H., & Newman, M. G. (2018). Worry amplifies theory-of-mind reasoning for negatively valenced social stimuli in generalized anxiety disorder. *Journal of Affective Disorders, 227*, 824–833. https://doi.org/10.1016/j.jad.2017.11.084

Zainal, N. H., & Newman, M. G. (2021). Within-person increase in pathological worry predicts future depletion of unique executive functioning domains. *Psychological Medicine, 51*(10), 1676–1686. https://doi.org/10.1017/S0033291720000422

Zainal, N. H., Newman, M. G., & Hong, R. Y. (2019). Cross-cultural and gender invariance of transdiagnostic processes in the United States and Singapore. *Assessment, 28*(2), 485–502. https://doi.org/10.1177/1073191119869832

Zainal, N. H., Newman, M. G., & Hong, R. Y. (2021). Cross-cultural and gender invariance of transdiagnostic processes in the United States and Singapore. *Assessment, 28*(Suppl. 2), 485–502. https://doi.org/10.1177/1073191119869832

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