Limited validity of breath-counting as a measure of mindfulness in ruminative adolescents

Isaac N. Treves[a], Anna O. Tierney[b,c], Simon B. Goldberg[d,e], Nancie Rouleau[b,f], Nicholas Carson[b,f], Zev Schuman-Olivier[b,f], Christian A. Webb[b,c]

[a]Department of Brain and Cognitive Sciences and McGovern Institute for Brain Research, Massachusetts Institute of Technology, 43 Vassar Street, Cambridge, MA 02139, USA
[b]Harvard Medical School, Department of Psychiatry, Boston, MA
[c]McLean Hospital, Belmont, MA
[d]Center for Healthy Minds, University of Wisconsin–Madison, Madison, WI
[e]Department of Counseling Psychology, University of Wisconsin–Madison, Madison, WI
[f]Department of Psychiatry, Cambridge Health Alliance, Center for Mindfulness and Compassion

Author Note

IN Treves https://orcid.org/0000-0001-7802-6145
Correspondence should be addressed to Isaac Treves, 43 Vassar Street 46-4077, Cambridge, MA 02139, 1-608-770-6805, treves@mit.edu

Isaac N. Treves: Isaac.Treves@nyspi.columbia.edu
Anna O. Tierney: atierney4@gmail.com
Simon B. Goldberg: sbgoldberg@wisc.edu
Nancie Rouleau: nrouleau@cha.harvard.edu
Nicholas Carson: ncarson@challiance.org
Zev Schuman-Olivier: zschuman@cha.harvard.edu
Christian A. Webb: cwebb@mclean.edu

Funding

This research was supported by NCCIH R01AT011002 (CAW). CAW was partially supported by NIMH R01MH116969, NCCIH R01AT011002, the Tommy Fuss Fund and a Young Investigator Grant from the Brain & Behavior Research Foundation. SBG was partially supported by NCCIH K23AT010879.

Abstract

Objective measurement of mindfulness could help us understand mechanisms of meditation interventions and how individuals vary in their disposition to be mindful. One proposed measure is the breath-counting task (BCT), which measures how accurately one can count cycles of their breath. Breath counting, which involves sustained attention, meta-awareness, and an internal locus of attention, has been shown in adults to be related to measures of mindfulness even when controlling for established attentional measures. In this study, we test the psychometrics of the BCT in a convenience sample of 78 adolescents with elevated rumination. In preregistered analyses, we related breath-counting measures, including novel objective respiration measures, to a suite of self-report measures as well as the sustained attention to response task. While breath-counting performance showed fair split-half reliability and similar distributions to studies in adults, it did not show the expected positive associations with self-reported mindfulness measures (neither trait nor EMA). Surprisingly, breath-counting accuracy showed negative correlations with a subscale measuring observing of emotions and body sensations, negative correlations with nonreactivity, and performance decrements were larger for individuals scoring more highly on mindfulness in general. The sustained attention to response task showed a small negative correlation with breath-counting resets (an index of mind-wandering). Finally, breath-counting performance was not related to other theoretically relevant clinical, personality and executive functioning criteria. Our results suggest that, at least in ruminative adolescents, breath-counting may measure a very narrow, contextual form of sustained attention, may not capture other qualities of mindfulness, and may lack predictive validity.

Introduction

The word mindfulness originated from Eastern contemplative traditions, specifically, as a translation of the term *sati* from Pali or *smrti* from Sanskrit, which is typically defined as remembering or being aware (Bodhi, 2013). Western psychologists have adapted and operationalized mindfulness in the service of quantifiable measurement. Some adaptations privilege awareness and attention (Brown & Ryan, 2003), others awareness coupled with non-judgement (Bishop et al., 2004; Lindsay & Creswell, 2017), and still others self-transcendence (Vago & Silbersweig, 2012). It is incontrovertible that measurement of mindfulness could help us understand the mechanisms of mindfulness meditation interventions (Goldberg et al., 2019), how people may vary in mindful disposition (Baer et al., 2022), and the relationship between mindfulness and mental health (e.g. Tomlinson et al., 2018). However, there is debate over how to measure mindfulness and which definitions are privileged. Traditional approaches use self-report questionnaires (Goodman et al., 2017; Rau & Williams, 2016), but there is emerging interest in objective measures, e.g., through behavioral tasks (Hadash & Bernstein, 2019) or neuroimaging (Ganesan et al., 2022; Rahrig et al., 2022; Treves et al., in press). Thus far, few proposed objective measures of mindfulness have been deployed in youth. In this study, we examine the validity of the breath-counting task (BCT; Levinson et al., 2014) in adolescents.

Breath-counting is a common element of meditative traditions like Zen and Tibetan Buddhism (Buddhaghosa, 2010). Meditators count the number of breath cycles (on the inhale or exhale), and if they lose count, they restart again. Over time, meditators gradually increment the number of cycles they attempt to count – from 5, to 7, until 21 or more. This practice is intended to train sustained attention and form the base of more advanced practices like insight. The first validation study of breath-counting as a psychometric task was inspired by this meditative practice (Levinson et al., 2014). The task consists of counting each breath using one button, and pressing a second button when a cycle of 9 breaths is completed. A third button may be used to 'reset' the count when the participant has lost count. Thus, the BCT allows for the extraction of correct cycles, incorrect cycles or miscounts, and resets. Importantly, Levinson and colleagues theorized that this task would not only capture attentional processes but also mindfulness more broadly, which they define as present moment awareness with facets of non-attachment and non-judgement. They state "although counting is not necessary for mindfulness, we propose mindfulness contributes to accurate breath counting" (pg. 2). Links between attentional processes and affective processes are common (Okon-Singer et al., 2015); indeed, it may be difficult to attend to the feelings of the breath if one is self-critical, anxious, or ruminating. In their validation studies of the BCT, Levinson and colleagues found evidence of these links. Breath-counting performance was related to greater meta-awareness, less mind wandering, better mood, and greater non-attachment.

This study led to extensive interest in breath-counting as an objective measure of mindfulness. Objective measures may be useful because there are well-established limitations of traditional self-report mindfulness. Two influential self-report measures of mindfulness are the Five-Facet Mindfulness Questionnaire (FFMQ)(Baer et al., 2006) and the Mindful Attention and Awareness Scale (MAAS)(Brown & Ryan, 2003). The MAAS involves negatively coded questions about the frequency of attentional lapses during daily life. More mindful individuals should be more present and more attentive to their thoughts, feelings and interactions with others. One main criticism of the MAAS is that it relies on introspective ability (Grossman, 2011). Those who are unaware may believe that they are present and attentive. Indeed, MAAS scores may actually be lower in some individuals who practice mindfulness and thus are frequently introspecting (Christopher & Gilbert, 2007; de Bruin et al., 2014; MacKillop & Anderson, 2007). On a similar scale (the Freiburg Mindfulness Inventory), binge drinkers rated themselves higher in mindfulness than healthy individuals (Leigh et al., 2005). An additional concern is that the negative coding of the MAAS results in a measure that reflects 'mindlessness' not 'mindfulness.'

The FFMQ captures additional dimensions of the mindfulness construct, with subscales of *Observing, Non-judging, Acting with Awareness, Non-reactivity,* and *Describing.* Although some subscales may be less reliant on introspective ability, the behavior of these various subscales can be unexpected. Meditators show different correlations between subscales than non-meditators; for this reason, *Observing* in particular is often left out in empirical studies (Baer et al., 2022; Pang & Ruch, 2019). Researchers have also cast doubt on whether the mindfulness subscales make up a cohesive construct by rigorously evaluating their constituent mechanisms and relationships to established psychological questionnaires. Bednar et al. (2020) created a model of mindfulness with factors of body awareness, attention regulation, emotion regulation, decentering, and nonattachment. They found that other questionnaires loaded more strongly on each factor than the FFMQ, and, in addition, the created factors predicted mental health more robustly than the FFMQ total score or subscales. Similarly, Tran et al. (2020) found that the FFMQ did not predict mental health when controlling for the Big Five personality measures. Altgassen et al. (2023) used advanced psychometric methods to create a new single-factor model of mindfulness, and found that mindfulness did not predict real-world outcomes, life satisfaction and health behaviors beyond the Big Five personality measures. Another recent study found that mindfulness may be subject to a *jangle* fallacy, wherein it is considered to be distinct from other constructs but is actually composed of preexisting constructs in domains like attention regulation and emotional awareness (Beloborodova & Brown, 2023). When examined in the context of training, it is also not clear that the FFMQ is uniquely sensitive to mindfulness training but may increase even without training in mindfulness (e.g., Goldberg et al., 2016).

The breath-counting task was introduced in the context of concern over the validity of mindfulness scales. Objective measures are commonly used in other domains of psychological inquiry – e.g. the Wechsler test of adult intelligence , executive function batteries (Zelazo et al., 2013), motor tasks involved in diagnosis of Parkinson's (Desrosiers et al., 1995), and tests of reading ability (Torgeson et al., 1999). These measures have demonstrated strong validity and reliability, including the ability to predict external behavior. Levinson et al. (2014) provided preliminary evidence that the breath-counting task could be one such measure.

Subsequent validation efforts have proceeded in three directions. First, researchers have employed the BCT in mindfulness interventions, finding that accuracy typically improves after mindfulness practice (Djernis et al., 2021; Isbel et al., 2019; Stieger et al., 2021; Tortella-Feliu et al., 2020; c.f. Ganesan et al., 2023; Yue et al., 2023). Second, researchers have correlated the BCT with trait mindfulness measures like the FFMQ and the MAAS. An individual participant mega-analysis of 430 adult participants found that *Acting with Awareness* showed a small but significant negative correlation ($\beta = -0.26$) with BCT resets (which may index awareness of attentional lapses) (Ching & Lim, 2023), although another study in 96 adults found no relationship between BCT and MAAS or FFMQ (Goldberg et al., 2020). A final arm of research has consisted of associating BCT with other objective measures. Psychomotor vigilance (the PVT) negatively correlates with miscount rates (Wong et al., 2018), but is unrelated to resets. This suggests that miscounts may measure attentional lapses, whereas resets measure meta-awareness of attentional lapses. Two studies have found working memory does not correlate with the BCT (Clapper et al., 2021; Levinson et al., 2014), which is typically interpreted as meaning that the counting itself is not cognitively taxing. High performers on the BCT may also show more time in attentional brain states at rest (Lim et al., 2018), and periods of inaccurate counting show distinct brain states (Treves et al., 2025). In summary, the established research on breath counting in adults suggests it may be sensitive to mindfulness training and related to other measures of interest, although relationships with self-report measures are modest. This does not invalidate the BCT, given concerns about self-report mindfulness, and given well-established evidence that self-report and behavioral tasks show limited convergence (Dang et al., 2020; Eisenberg et al., 2019). BCT validity is best explored by integrating evidence from behavioral measures and self-report scales that encompass diverse processes and effects of mindfulness.

There are a number of open questions about breath counting, united by the overarching question of whether BCT really measures mindfulness – not just sustained attention, but also facets of meta-awareness and non-judgement. We address these questions in a sample of adolescents. To date, there is no research on the validity of the BCT in youth, including relationships with self-report mindfulness (FFMQ) and attentional tasks. Second, it is unclear whether breath-counting performance is associated with positive mental health outcomes, personality, or other criteria. Finally, there are a number of potentially relevant measures that can be derived from the BCT that have not been assessed. These include performance decrements, and individual-level correspondences with objective physiological recordings of respiration. To address these open questions, we conducted a preregistered study of breath counting in 78 adolescents. We note that this was a convenience sample from a larger study of non-normative adolescents who scored high on rumination measures.

Methods

This study was preregistered at https://osf.io/yfsev. We made the following deviations: (1) No longitudinal assessments were able to be conducted as the NIH did not permit unblinding, (2) We did not compute a composite of the attitudinal questions on EMA, as the internal consistency was low, (3) Beta regressions were not assessed, only spearman's rank correlations, (4) We conducted Bayes Factor analyses to identify the strength of evidence for the null hypothesis, (5) The sample size was 78 participants.

Participants

Participants were 78 adolescents (59 females, 19 males, classified by sex at birth) recruited from the greater Boston area for an ongoing app-based mindfulness intervention study (NCT04697966). Eligible participants were fluent in English, between the ages of 13-18 years old (m_{age} = 15.7, *SD* = 1.66), had a personal smartphone, and had elevated levels of rumination, as defined by a total score ≥ 13 on the Children's Response Style Questionnaire (CRSQ; Abela et al., 2002). Demographic and clinical characteristics are provided in **Table 1**. Main BCT measures showed no significant relationships with age, race, sex, or puberty ($p > .05$).

Participants were ineligible if they had a past or present DSM-5 diagnosis of schizophrenia spectrum or other psychotic disorder, bipolar disorder, ADHD, substance or alcohol use disorder within the past 12 months, or current chronic depression (episode > 2 years). Participants were also ineligible if they had serious or unstable medical illnesses or a diagnosis of neurodevelopmental disorders that would interfere with the ability to complete study tasks. Participants could be on a stable (≥ 2 months) dose of psychotropic medication but current use of stimulants was exclusionary. Given the fMRI component of the study, participants were ineligible if they had any of the following: systemic medical or neurological illnesses that could impact measures of cerebral blood flow, history of seizure disorder or previous head trauma with loss of consciousness > 2 minutes, cardiac or neural pacemakers, surgically implanted metal devices near the head / neck, metal shrapnel in their body, braces, cochlear implants, aneurism clips, claustrophobia, or a known pregnancy (validated through a positive urine pregnancy test). Given the mindfulness intervention portion of the study, participants were ineligible if they were currently in regular therapy, had past or present treatment with mindfulness-based psychotherapy, or had extensive mindfulness or meditation practice (≥ 300 minutes). Participants with active suicidal ideation were excluded and immediately referred to appropriate clinical treatment.

Measures

Breath-counting Task

We administered a 20-minute breath-counting task (BCT) (Lim & Doshi, 2022) concurrent with fMRI (not detailed here) and breath recordings (utilized in analyses). The BCT consists of counting each breath using one button, and pressing a second button when a cycle of 9 breaths is completed **(Figure 1)**. If the participant loses count, they may press a third button to 'reset' the count. The primary outcomes consist of correct cycles, incorrect cycles or miscounts, and resets, where each is calculated as a proportion of total cycles. The proportion of correct cycles is typically used as a measure of overall accuracy. These outcomes are intercorrelated but show distinct relationships to attention measures (Wong et al., 2018). Miscounts reflect attentional lapses, and resets reflect meta-awareness of attentional lapses. As miscounts may occur close to 9 breaths, or more variably, we extracted 'miscount distance,' i.e., how many counts away from 9 the miscount was. We calculated a similar measure for resets. We also assessed the performance decrement in accuracy, consisting of the slope of performance over the course of the task (generated using a logistic regression, *glm(Correct~Cycle,family="binomial")*. In a novel step, we also analyzed breath belt recordings from the participants on an individual level. We were able to thus extract additional outcomes of respiration. One outcome was physiologically verified accuracy: accuracy where a correct cycle must have 9 breaths. We also assessed breath variability, as well as physiological-behavioral correspondence, e.g. breath-button distances (at what point in the breath cycle participants were pressing). Finally, there was a single attention probe at the end of the BCT, where higher ratings represent more time off-task. The full list of metrics can be found in **Supplement Text 1.**

Sustained Attention to Response Task

To assess whether the BCT relates to attention in adolescents, we administered the sustained attention to response task (SART), which has previously been found to correlate with breath-counting in adults (Levinson et al., 2014). The SART (Robertson et al., 1997) consisted of pressing the spacebar as fast and accurately as possible when a digit (0-9) was presented on the screen (go trials), and withholding responses to the '3' (no-go trials). The SART thus involves sustained attention and inhibition. The present version of the task includes 260 trials in each of 4 blocks (Eusebio et al., 2022), with duration 20 minutes. The no-go trial accuracy and reaction time variability (CV) (Manly et al., 2000) were used. As exploratory measures, we calculated performance decrement in accuracy and reaction time, as well as probed self-report attentional lapses.

Self-reported Mindfulness Measures

We administered the FFMQ-SF (Bohlmeijer et al., 2011) before the breath-counting task. The FFMQ has five subscales: *Observing, Non-judging, Acting with Awareness, Non-reactivity,* and *Describing,* which are examined separately and summed for an overall mindfulness score. In our sample, internal consistency of the FFMQ total score was adequate (Cronbach's alpha = .82), and adequate for most of the subscales: *Observing* (α = .64), *Non-judging* (α = .77), *Acting with Awareness* (α = 0.79), *Non-reactivity* (α = .84), *Describing* (α = .73).

We additionally administered ecological momentary assessments (EMAs) of mindfulness three times daily for three days after the breath-counting task.[1] Five of the self-report questions probed participants' mindful awareness (Aizik-Reebs et al., 2021; Shoham et al., 2017). Participants were given statements such as, "Since the last survey, I was aware of what was going on in my mind," and they were asked to rate the extent to which that statement was true for them on a scale from 0 (Not at All) to 100 (As Much As Possible). The full list of questions is in **Supplement Table 1**. These five answers were averaged across instances to calculate an average for each participants' *Mindful Awareness* (Cronbach's alpha = .90, across items). Further, we measured *Non-judgement*, which is an important part of mindfulness scales and definitions (Baer et al., 2006; Bishop et al., 2004; Lindsay & Creswell, 2017), specifically "Since the last survey, I told myself that I shouldn't be feeling the way I'm feeling, or shouldn't be thinking the way I'm thinking." The item was reverse-coded. We also measured curiosity, non-reactivity, self-compassion, and disidentification, shown in **Supplement Table 2.** Response rates to EMA were

---

[1] These data are derived from a larger clinical trial (https://classic.clinicaltrials.gov/ct2/show/NCT04697966) of a mindfulness app which included 30-days of EMA over the 30-day intervention period. We selected the first 3 days as our baseline scores to include here to minimize the confounding influence of intervention-related change.

high (90.0%). Exploratory measures consisted of variability in *Non-judgement* and *Mindful Awareness* scores over the probes, as well as slope of change.

Criterion Measures

Criterion measures we define as measures theoretically related to mindfulness (Cronbach & Meehl, 1955)[2]. We collected measures of mental health symptoms including rumination (Children's Response Styles Questionnaire, CRSQ; Abela et al., 2002), depression (Center for Epidemiological Studies Depression Scale, CESD; Radloff, 1977), perceived stress (Perceived Stress Scale, PSS; Cohen et al., 1983). The Big Five was administered (Costa & McCrae, 2010). Anxiety was calculated as a composite across Multidimensional Anxiety Scale for Children (MASC) dimensions (March et al., 1997). Self-report executive function was assessed using the Behavioral Rating Inventory of Executive Function (BRIEF2) (Gioia et al., 2015).

Analyses

Outliers and Exclusions

For the BCT, the following rules were applied (Lim & Doshi, 2022): exclude participants who have an overall accuracy of less than 10%, and exclude subjects scoring above or below 2.5 SD of the mean on overall cycles or breath rate. For all other variables we removed scores above or below 2.5 SD of the mean.

Reliability and Correlations between BCT Variables

We conducted several analyses to assess the reliability and intercorrelations of the BCT metrics. We analyzed the primary breath-counting measures, examining the proportions of correct cycles, miscounts and resets, and their distributions (with additional Kolgomorov-Smirnov tests to formally evaluate normality). For reliability, we conducted a split-half, odd-even analysis, computing measures for odd cycles (1,3,5) and even cycles (2,4,6), calculating ICC(2,1) (Cicchetti, 1994), where 0–0.4 = poor, 0.4–0.6 = fair, 0.6–0.75 = good and 0.75–1 = excellent. To assess interrelationships between BCT metrics (including physiological recordings of respiration and attention probes), we computed Pearson's correlations. For visualization of the correlations between all BCT metrics, we constructed network graphs using *qgraph* in R with a "spring" layout, which clusters metrics in space based on the strength of their correlations. All correlations higher than 0.3 were shown as lines between nodes, and clustered nodes are more inter-correlated. Finally, we examined whether there were any relationships to demographic variables (age, gender, puberty, race and ethnicity).

Convergent validity

We conducted analyses to understand whether BCT converges with more established measures of mindfulness (i.e., convergent validity). We examined Pearson's correlations between breath-counting metrics with self-reported mindfulness measures (traditional and EMA). False Discovery Rate (FDR)-correction (Benjamini & Hochberg, 1995) was applied across the mindfulness measures, separated by conventional self-report questionnaires and EMA. In the case of non-normal breath-counting metrics (as defined by inspection and Kolgomorov-Smirnov tests), we tested sensitivity of the results using Spearman's correlations.

We did not conduct a-priori power analyses. Instead, for main comparisons between BCT measures and overall mindfulness, we conducted Bayes Factor analyses to test evidence for the null hypothesis, using the 'BayesFactor' package in *R* (Morey et al., 2015); factors between 1-3

---

[2] They are not 'gold standard' tests of mindfulness in the traditional sense of criterion validity, and it is unclear what that would mean in the context of mindfulness.

reflected equivocal evidence for the null, and factors above 3 reflected substantial evidence (Wagenmakers et al., 2011, 2018).

We additionally correlated SART measures with breath-counting metrics.

Criterion Validity

We assessed the relationships between BCT with criteria of interest: mental health symptoms, personality traits, and self-reported executive function. In the case of significant relationships between breath-counting and SART and/or demographics, we conducted additional partial correlations with the criteria of interest, controlling for SART/demographics (incremental validity). For visualization of the correlations, we again used network graphs using *qgraph* in R.

Results

Reliability and Correlations between BCT Variables

We removed 1 participant for low accuracy (percentage of cycles correct), and 2 participants for too many or too few cycles (Lim & Doshi, 2022). The average participant completed 29.14 cycles over 20 minutes, mean accuracy was 56.66%, mean miscount rate was 26.52%, and the mean reset rate was 14.71%. Distributions are shown in **Figure 2.** Reset rates were right skewed and non-normal (Z = 4.33, p < .001), with 16% of participants never responding with resets. ICCs based on an odd-even split were fair for correct cycle rates and reset rates (ICC(2,1) = .61; ICC(2,1) = .57), but lower for miscounts (ICC(2,1) = .40). Accuracy was strongly negatively correlated with the percentage of miscounts and percentage of resets (**Table 2)**, but the miscounts and resets were not correlated. The majority of miscounts (96%) occurred 1-5 presses away from a correct cycle (M = 1.82, SD = 0.66), and resets occurred at variable press numbers (M = 6.02, SD = 1.97), but 81.3% occurred at less than 10 presses.

Exploratory BCT Variables

The network graph revealed three loose clusters based on the patterns of correlations among measures – accuracy on the task, breath/breath-button variability, and pressing/breathing rates (**Figure 3)**. There were some correlations between clusters. As expected, higher objectively-measured breathing rates corresponded to higher button pressing rates (r(75) = .73, p < .001). As expected, more off-task self-report rating correlated with worse objective accuracy (r(75) = -.40, p < .001), but was specifically related to reset rate (r(75) = .33, p = .003) and not miscount rate (p = .1). Higher breath variability was correlated with worse accuracy (r(75) = -.42, p < .001). Greater performance decrements were associated with worse accuracy (r(75) = .34, p = .002) (even stronger for physiologically verified measures of accuracy, *ps* < .001), more resets (*rho*(75) = -.32, p = .007), and more variable breathing (r(75) = -.34, p = .003).

Convergent Validity

We assessed correlations between the main BCT measures and traditional self-report mindfulness questionnaires (**Table 3, Supplement Figure 1**). Relationships between overall FFMQ scores and BCT measures were non-significant, and there was substantial Bayesian evidence for the null hypothesis (**Supplement Table 3**). There was a trend-level negative correlation between FFMQ-*Observing* and breath-counting accuracy (r(78) = -.27, p = .018, FDR-p = .109), and a positive correlation between FFMQ-*Nonreactivity* and breath-counting miscounts (r(78) = .30, p = .0075, FDR-p = .045). We then tested exploratory relationships between all the breath-counting metrics and mindfulness (**Supplement 2**). The relationship between *Observing* and breath-counting accuracy was larger magnitude when examining physiologically-verified accuracy (*r*(75)= -.33, *p* =.007, FDR-*p* = .025). *Observing* was positively associated with breath variability (*r*(75) = .28, *p* =.017, FDR-*p* = .099). Surprisingly, larger performance decrements over time on the task were positively associated with *Observing*, *Non-reactivity*, and total mindfulness (*rs* = .33, .29, .28, FDR-*p*s = .024, .028, .028). Off-task ratings on the task were positively correlated with *Observing* (*r*(75) = .26, *p* = .024, FDR-*p* = .115).

Second, we examined the relationships of the breath-counting measures with EMA mindfulness measures (**Supplement Table 4)**. No relationships were found between mindful awareness and miscounts, resets, or correct cycles (*ps* > .05). A significant positive relationship was found for EMA *Non-Judgement* and miscounts (*r*(75) = 0.25, *p* = 0.039, FDR-*p* = 0.078). More variance in mindful awareness reports was associated with slower button presses (*r*(75)=.31, *p* = .009, FDR-*p* = .05).

Relations to SART

Finally, we examined the relationship of the main breath-counting measures with SART measures. Overall SART no-go accuracy was negatively correlated with breath-counting resets (*rho* = -.26, *p* = .038) (**Table 4**), but the association became trend-level after correction (FDR-*p* = .076). Further, performance decrements on breath-counting were not associated with performance decrements on SART (*p* > .05). No relationship was found between breath-counting measures and SART self-report probes (**Supplement Table 5**).

Criterion Validity

The main breath-counting measures showed no significant associations with mental health symptoms or self-reported executive function (**Table 5, Figure 4**). There was a negative association between extraversion and accuracy (*r*(75) = -.26, *p* = .023), and between agreeableness and resets (*rho*(75) = -.28, *p* =.020); neither were significant when correcting for multiple comparisons (respectively, FDR-*p* = .230, FDR-*p* = .197). This held true when conducting partial correlations accounting for SART no-go accuracy.

Discussion

Mindfulness as a Western psychological construct typically encompasses both stable, present-focused attention and an attitude of acceptance and non-judgement (Bishop et al., 2004). Both of these attitudinal and attentional components are fundamental for mental health, constructive emotional experience, and a balanced sense of self (Lindsay & Creswell, 2017). The breath-counting task has been proposed as an objective measure of mindfulness, as it may involve not just attention to the breath but meta-awareness and non-reactivity as well (Levinson et al., 2014). As yet, no research has been conducted on its validity in youth. Here we conducted a preregistered cross-sectional study in 78 ruminative adolescents examining the relationships between breath-counting measures, self-report mindfulness (assessed via both conventional self-report and EMA), the SART task, and theoretically relevant measures of mental health, personality and executive function. Despite the constrained sample, and limitations of cross-sectional designs, the lack of conclusive support for previous findings across these different domains raise serious questions regarding the validity of the breath-counting task in adolescents.

Breath-counting measures

The BCT consists of counting each breath using one button, and pressing a second button when a cycle of 9 breaths is completed. If the participant loses count, they may press a third button to 'reset' the count. Main breath-counting measures consist of accuracy, miscounts, and resets. Previous work has indicated that miscounts and resets may capture attentional lapses and meta-awareness of attentional lapses, respectively (Wong et al., 2018). In our sample, miscounts and resets were both negatively correlated with accuracy, but not each other (as found in Wong et al., 2018; Ching & Lim, 2023). Miscounts tended to occur close to the targeted 9 breaths, whereas resets were more dispersed. Resets were right skewed, as found previously (Ching & Lim, 2023; Clapper et al., 2021; Wong et al., 2018). In addition, resets negatively correlated with self-reported attention (collected at the end of the task), as found previously (Clapper et al., 2021). Reliability (ICC) of the measures was fair and comparable to adults (Levinson et al., 2014; Wong et al., 2018; Ching & Lim, 2023), but lowest for miscounts. In general, there was evidence that the task was performing as expected. It should be noted that overall accuracy varies widely based on task demands and populations (in our sample of adolescent ruminators, 57%; in college students, 72%, Wong et al., 2018; in adult MBSR participants, 96%, Goldberg et al., 2020).

Validity

We expected breath-counting performance to correlate with other measures relevant to mindfulness. We examined the FFMQ short-form (Cortazar et al., 2020), and ecological momentary assessments (EMAs) of mindfulness (Aizik-Reebs et al., 2021; Baer et al., 2006; Shoham et al., 2017). A previous mega-analysis of correlations between the FFMQ and breath-counting found a small but significant relationship (β = -0.26) between *Acting with Awareness* and resets on the task (Ching & Lim, 2023). We saw no relationships between better breath-counting performance (i.e. higher accuracy or lower resets) and mindfulness. Indeed, we noted a set of relationships consistent with self-report mindfulness leading to worse performance on the task (**Table 6).** We observed a negative correlation between accuracy on the task and the *Observing* scale on the FFMQ, which only strengthened when using breath recordings to estimate accuracy. Notably, *Observing* did show the lowest internal consistency of the subscales (Cronbach's alpha = 0.64), consistent with prior work (De Bruin et al., 2012). Also unexpectedly, *Non-reactivity,* and EMA *Non-judgement* were positively related to miscounts. Further, performance decrements over time on the task were higher for individuals scoring higher on the FFMQ. It is unclear why these unexpected relationships were found. Prior work has shown that *Observing* scores correlate *positively* with psychological symptoms (for review, see Rudkin et al., 2018), and perhaps this is because attending to experiences in non-meditators may also involve judgement (Baer et al., 2008; Bergomi et al., 2013). For example, anxious individuals may show heightened body awareness but also heightened catastrophizing (Dunn et al., 2010). Speculatively, heightened awareness and sensitivity to breathing sensations (more reactivity) may have interfered with effective counting in the current task. A larger concern is that the adolescents may have had introspective biases that led them to misrepresent how mindful they are (Grossman, 2011), and the current results reflect the disconnect between their narratives about themselves and their behavior in the moment. However, even when looking at EMAs, which may be less susceptible to these retrospective biases (Webb et al., 2023), we did not find a relationship with breath-counting.

We also examined relationships between breath-counting and another objective attentional measure, the sustained attention to response task (SART). The SART is a standard measure of inhibition and sustained attention commonly used in the mindfulness and meditation literature; it involves responding to targets as fast as possible while withholding responses to distractors. In the validation study of the breath counting tasks, SART errors and RT variability were found to negatively correlate with breath-counting accuracy in adults (*rs* = -0.19, -0.33; Levinson et al., 2014). A small indication of this relationship was found here in terms of resets (not overall BCT performance), which negatively correlated with SART accuracy (*rho* = -0.26), but this relation became trend-level following correction for multiple comparison. There were no relationships between probe self-report attention during the SART and breath-counting (making it unclear whether BCT measures meta-awareness). This result indicates that the breath-counting task may index a very specific form of attention that does not transfer to other attentional measures. There is conflicting empirical literature about the relationships between attention and breath meditation (MacCoon et al., 2014) – if practicing breath meditation improves sustained attention, the improvement is slight (for meta-analysis, see Whitfield et al., 2022). Our results contribute to this discussion by suggesting, at least in adolescents, there may not be overlap between paying attention to the breath and an externally oriented attentional task.

Finally, we examined relationships between breath counting and criteria of interest like self-report depression, anxiety, executive function, and personality. Few studies have examined the criterion validity of breath-counting. One study found that higher trait anxiety may be related to more resets during the task, but found no relationships between performance and the Big Five personality traits (Clapper et al., 2021). In our study, we found no statistically significant relationships between our criterion measures and breath-counting. Taken together, our results suggest that breath-counting in adolescents may index a specific skill with limited relevance to dispositional mindfulness, sustained attention, or external outcomes.

Physiological recordings

An innovative aspect of our study is that we collected and analyzed objective breath recordings from the task using a respiration belt. This provided additional insight into breath-counting behavior. First, we confirmed that button pressing rates highly corresponded to breathing rate indicating that adolescents were following the task instructions. We also verified correct cycles by extracting the number of breath peaks in the cycle, and verified cycles corresponded strongly to reported cycles. It should be mentioned that participants were not instructed to press the button at a particular part of the breath cycle (e.g. peak vs trough), and thus there was variability in button press timing. Novelly, we found that breath variability was associated with worse performance on the task and higher performance decrements over time. Analogous findings have been found with reaction times in sustained attention tasks, where participants experience periods of variable response times and worse performance (Brink et al., 2016; Esterman et al., 2013) or variable brain connectivity and worse performance (Fong et al., 2019). One suggestion is that periods of stable responding represent 'flow,' where participants are stably engaged but not reactively monitoring their performance (Fortenbaugh et al., 2017; Norsworthy et al., 2021). It could also be that variable stimuli (variable breaths) are harder to attend to. The current study adds to this literature by relating *physiological* variability to sustained attention performance. We found that FFMQ- *Observing* correlated positively with breath variability. Otherwise, physiological measures did not correlate with FFMQ or EMA mindfulness.

Sources of discrepancies

There are a number of possible reasons for our largely null findings. First, the breath-counting task utilized here did not contain probes during the task, contrasting with the original breath-counting study (Levinson et al., 2014). Probes may change the degree of monitoring and meta-awareness that occurs during the task. Indeed, the SART task contained probes and the BCT did not, possibly explaining why they showed a non-robust correlation. However, the tradeoff is that probe frequency negatively correlates with the degree of mind-wandering in attention tasks (Zanesco et al., 2024), and many recent implementations of breath-counting have excluded them (Clapper et al., 2021; Lim & Doshi, 2022). Across studies, the BCT shows a wide range of overall accuracy (57%-96%) dependent on participant characteristics, task characteristics like cycle length, etc. (Goldberg et al., 2020; Lim & Doshi, 2022), and perhaps the lack of standardization of the BCT hurts replicability.

Second, the adolescents in the current study were selected for elevated levels of rumination, making this a non-representative sample of adolescents. It may be that these adolescents showed diminished mindfulness scores compared to non-ruminative adolescents, leading to floor effects. It is not clear whether rumination moderates the relationship between self-report mindfulness and the BCT – studies show that mindfulness practice can be effective in ruminative adolescents. Developmental differences could also play a role, e.g. self-report measures may show limited convergent validity in adolescents (e.g., de Bruin et al., 2014; Mazefsky et al., 2011). In one study, adolescents with meditation and yoga experience showed lower self-report mindfulness (de Bruin et al., 2014). Of course, introspective awareness is conflated with outcomes on self-report, as more trait aware adolescents (who have practiced meditation) may be more aware of their attentional lapses (Grossman, 2011).

Relatedly, the relationships between breath-counting and self-report mindfulness that have been indicated in adults are small in magnitude and may require larger sample sizes for statistical significance (e.g., $n > 400$) (Ching & Lim, 2023). One reason for these small effects may be limited test-retest reliability of the breath counting measures given low reliability hinders the detection of effects (Nunnally, 1970). The current study may have been underpowered to detect modest effects, although Bayesian models show significant evidence in favor of null effects for self-report mindfulness. Future studies could employ large samples of normative adolescents, perhaps with online administration of breath-counting tasks (although this makes physiological validation difficult).

Finally, unlike the current study, previous studies have not been preregistered. Preregistration constrains researcher degrees of freedom and may reduce spurious findings (Nosek et al., 2018).

Future directions

Given the challenges with self-report measures of mindfulness, research on behavioral measures of mindfulness is vitally important. Researchers have increasingly conceptualized trait self-report mindfulness as relating to sets of interrelated skills and faculties (Altgassen et al., 2023; Bednar et al., 2020; Beloborodova & Brown, 2023; Tran et al., 2020). Behavioral mindfulness measures are no different. We should study specific behavioral skills, their malleability to intervention, and their relationships to health and everyday functioning.

The results here indicate that it is unlikely, at least in adolescents, that breath-counting generally correlates with mental health and mindful acceptance and other constructs. Of course, it is possible that with meditation experience, these associations may emerge. In adults, breath-counting skill may be acquired through breath meditation practice (Djernis et al., 2021; Isbel et al., 2019; Stieger et al., 2021; Tortella-Feliu et al., 2020; c.f. Ganesan et al., 2023; Yue et al., 2023). Future longitudinal work in children and adolescents is necessary. It could also simply be worthwhile to track breath-counting skill development if it predicts who maintains meditation practice.

Other measurement domains may be both malleable to mindfulness interventions and related to trait mindfulness and mental health. Body awareness is an important component of mental health (Khalsa et al., 2018), and it is targeted by many mindfulness practices (e.g. body scans). There are overall indications that mindfulness is associated with improved body awareness, whether measured objectively (Treves et al., 2019) or subjectively (Hanley et al., 2017). However, heartbeat perception specifically is not associated with mindfulness (Khalsa et al., 2020), while tasks involving somatosensory perception or 'emotional coherence' are more strongly related (Treves et al., 2019). Relationships with mental health symptom improvements still need to be established.

Another critical component of mindfulness is a non-judgmental curiosity towards experience (Bishop et al., 2004). This may be difficult to measure objectively. One approach is to elicit emotional experiences through video or audio and then use implicit association tests to measure attitudes (Hadash & Bernstein, 2019). Another approach may be to adapt self-referential encoding tasks which measure response biases to self-relevant adjectives (Derry & Kuiper, 1981). These measures should be employed in future mindfulness research.

Conclusion

The breath-counting task has received substantial attention as a possible objective measure of mindful awareness or attention. In our study with ruminative adolescents, we found that it has limited convergent validity with self-report mindfulness and a sustained attention task, and does not relate to other criteria like mental health symptoms. Breath-counting, at least in adolescents, may be a specific skill with narrow relevance to mindfulness and well-being.

Ethics Statement:

We obtained informed consent / assent from all participants. All procedures were approved by the Mass General Brigham IRB (#2020P004016).

Data Availability:

Data and code will be made available upon acceptance.

Acknowledgements:

We would like to thank David Vago, Gareth Parry, Javier Barria, and Hannah Goodman who were contributing members of the CHA study team.

Funding statement:

This research was supported by NCCIH R01AT011002 (CAW). CAW was partially supported by NIMH R01MH116969, NCCIH R01AT011002, the Tommy Fuss Fund and a Young Investigator Grant from the Brain & Behavior Research Foundation.

CREDIT:

INT contributed conceptualization, data curation, formal analysis, investigation, methodology, visualization, writing-original draft, writing-review & editing. AOT & NC contributed data curation, and project administration. NR contributed writing-review & editing. SBG & ZSO contributed conceptualization and writing-review & editing. CAW contributed funding acquisition, supervision, resources, conceptualization and writing-review & editing.

References

Abela, J. R. Z., Brozina, K., & Haigh, E. P. (2002). An Examination of the Response Styles Theory of Depression in Third- and Seventh-Grade Children: A Short-Term Longitudinal Study. *Journal of Abnormal Child Psychology*, *30*(5), 515–527. https://doi.org/10.1023/A:1019873015594

Aizik-Reebs, A., Shoham, A., & Bernstein, A. (2021). First, do no harm: An intensive experience sampling study of adverse effects to mindfulness training. *Behaviour Research and Therapy*, *145*, 103941.

Altgassen, E., Geiger, M., & Wilhelm, O. (2023). Do you mind a closer look? A jingle-jangle fallacy perspective on mindfulness. *European Journal of Personality*, 089020702311745. https://doi.org/10.1177/08902070231174575

Baer, R. A., Smith, G. T., Hopkins, J., Krietemeyer, J., & Toney, L. (2006). Using self-report assessment methods to explore facets of mindfulness. *Assessment*. https://doi.org/10.1177/1073191105283504

Baer, R. A., Smith, G. T., Lykins, E., Button, D., Krietemeyer, J., Sauer, S., Walsh, E., Duggan, D., & Williams, J. M. G. (2008). Construct validity of the five facet mindfulness questionnaire in meditating and nonmeditating samples. *Assessment*. https://doi.org/10.1177/1073191107313003

Baer, R., Gu, J., & Strauss, C. (2022). Five Facet Mindfulness Questionnaire (FFMQ). In O. N. Medvedev, C. U. Krägeloh, R. J. Siegert, & N. N. Singh (Eds.), *Handbook of Assessment* *in Mindfulness Research* (pp. 1–23). Springer International Publishing. https://doi.org/10.1007/978-3-030-77644-2_15-1

Bednar, K., Voracek, M., & Tran, U. S. (2020). Common Factors Underlying the Five Facets of Mindfulness and Proposed Mechanisms: A Psychometric Study Among Meditators and Non-meditators. *Mindfulness*, *11*(12), 2804–2817. https://doi.org/10.1007/s12671-020-01492-6

Beloborodova, P., & Brown, K. W. (2023). *The mismeasurement of mindfulness: Evidence from network analysis of a jangle fallacy in popular mindfulness scales*. https://psyarxiv.com/b9yjz/

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Benson, N., Hulac, D. M., & Kranzler, J. H. (2010). Independent examination of the Wechsler Adult Intelligence Scale—Fourth Edition (WAIS-IV): What does the WAIS-IV measure? *Psychological Assessment*, *22*(1), 121.

Bergomi, C., Tschacher, W., & Kupper, Z. (2013). The Assessment of Mindfulness with Self-Report Measures: Existing Scales and Open Issues. *Mindfulness*, *4*(3), 191–202. https://doi.org/10.1007/s12671-012-0110-9

Bishop, S. R., Lau, M., Shapiro, S., Carlson, L., Anderson, N. D., Carmody, J., Segal, Z. V., Abbey, S., Speca, M., Velting, D., & Devins, G. (2004). Mindfulness: A proposed operational definition. *Clinical Psychology: Science and Practice*, *11*(3). https://doi.org/10.1093/clipsy/bph077

Bodhi, B. (2013). What does mindfulness really mean? A canonical perspective. In *Mindfulness* (pp. 19–39). Routledge. https://api.taylorfrancis.com/content/chapters/edit/download?identifierName=doi&identifierValue=10.4324/9781315874586-2&type=chapterpdf

Bohlmeijer, E., Klooster, P. M., Fledderus, M., Veehof, M., & Baer, R. (2011). Psychometric properties of the five facet mindfulness questionnaire in depressed adults and development of a short form. *Assessment*, *18*(3), 308–320. https://doi.org/10.1177/1073191111408231

Brink, R. L. van den, Murphy, P. R., & Nieuwenhuis, S. (2016). Pupil Diameter Tracks Lapses of Attention. *PLOS ONE*, *11*(10), e0165274. https://doi.org/10.1371/journal.pone.0165274

Brown, K. W., & Ryan, R. M. (2003). The Benefits of Being Present: Mindfulness and Its Role in Psychological Well-Being. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/0022-3514.84.4.822

Buddhaghosa, B. (2010). *Visuddhimagga: The path of purification*. https://buddhism.lib.ntu.edu.tw/en/search/search_detail.jsp?seq=374493

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review*, *97*(3), 404–431. https://doi.org/10.1037/0033-295X.97.3.404

Ching, A. S. M., & Lim, J. (2023). A Mega-Analysis of the Relationship Between Breath Counting Test Performance and Subscales of the Five Facet Mindfulness Questionnaire. *Mindfulness*, *14*(9), 2097–2110. https://doi.org/10.1007/s12671-023-02201-9

Christopher, M., & Gilbert, B. (2007). Psychometric properties of the Kentucky inventory of mindfulness skills (KIMS) and the mindful attention awareness scale (MAAS) among Thai Theravada Buddhist monks. *Faculty Scholarship (SPP)*, *2*.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*(4), 284.

Clapper, J., Ware, S., Martinez, F. J., Benitez, K., & Koshino, H. (2021). Breath counting as a measure of sustained attention in mindfulness meditation and its effect on mood. *Psychology of Consciousness: Theory, Research, and Practice*. https://doi.org/10.1037/cns0000311

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior*, 385–396.

Cortazar, N., Calvete, E., Fernández-González, L., & Orue, I. (2020). Development of a Short Form of the Five Facet Mindfulness Questionnaire–Adolescents for Children and Adolescents. *Journal of Personality Assessment*, *102*(5), 641–652. https://doi.org/10.1080/00223891.2019.1616206

Costa, P. T., & McCrae, R. R. (2010). NEO inventories for the NEO personality inventory-3, NEO five factor model 3, and NEO personality inventory-revised. *Psychological Assessment Resources: Lutz, FL*.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281.

Dang, J., King, K. M., & Inzlicht, M. (2020). Why Are Self-Report and Behavioral Measures Weakly Correlated? *Trends in Cognitive Sciences*, *24*(4), 267–269. https://doi.org/10.1016/j.tics.2020.01.007

De Bruin, E. I., Topper, M., Muskens, J. G. A. M., Bögels, S. M., & Kamphuis, J. H. (2012). Psychometric Properties of the Five Facets Mindfulness Questionnaire (FFMQ) in a Meditating and a Non-meditating Sample. *Assessment*, *19*(2), 187–197. https://doi.org/10.1177/1073191112446654

de Bruin, E. I., Zijlstra, B. J. H., & Bögels, S. M. (2014). The Meaning of Mindfulness in Children and Adolescents: Further Validation of the Child and Adolescent Mindfulness Measure (CAMM) in Two Independent Samples from The Netherlands. *Mindfulness*. https://doi.org/10.1007/s12671-013-0196-8

Derry, P. A., & Kuiper, N. A. (1981). Schematic processing and self-reference in clinical depression. *Journal of Abnormal Psychology*, *90*(4), 286–297. https://doi.org/10.1037/0021-843X.90.4.286

Desrosiers, J., Hébert, R., Bravo, G., & Dutil, E. (1995). The Purdue Pegboard Test: Normative data for people aged 60 and over. *Disability and Rehabilitation*, *17*(5), 217–224. https://doi.org/10.3109/09638289509166638

Djernis, D., O'Toole, M. S., Fjorback, L. O., Svenningsen, H., Mehlsen, M. Y., Stigsdotter, U. K., & Dahlgaard, J. (2021). A Short Mindfulness Retreat for Students to Reduce Stress and Promote Self-Compassion: Pilot Randomised Controlled Trial Exploring Both an Indoor and a Natural Outdoor Retreat Setting. *Healthcare*, *9*(7), 910. https://doi.org/10.3390/healthcare9070910

Dunn, B. D., Galton, H. C., Morgan, R., Evans, D., Oliver, C., Meyer, M., Cusack, R., Lawrence, A. D., & Dalgleish, T. (2010). Listening to your heart: How interoception shapes emotion experience and intuitive decision making. *Psychological Science*. https://doi.org/10.1177/0956797610389191

Eisenberg, I. W., Bissett, P. G., Zeynep Enkavi, A., Li, J., MacKinnon, D. P., Marsch, L. A., & Poldrack, R. A. (2019). Uncovering the structure of self-regulation through data-driven ontology discovery. *Nature Communications*, *10*(1), 2319.

Esterman, M., Noonan, S. K., Rosenberg, M., & Degutis, J. (2013). In the zone or zoning out? Tracking behavioral and neural fluctuations during sustained attention. *Cerebral Cortex*, *23*(11), 2712–2723. https://doi.org/10.1093/cercor/bhs261

Eusebio, J., Forbes, B., Sahyoun, C., Vago, D. R., Lazar, S. W., & Farb, N. (2022). Contemplating movement: A randomized control trial of yoga training for mental health. *Mental Health and Physical Activity*, *23*, 100483. https://doi.org/10.1016/j.mhpa.2022.100483

F. Wong, K., A. A. Massar, S., Chee, M. W. L., & Lim, J. (2018). Towards an Objective Measure of Mindfulness: Replicating and Extending the Features of the Breath-Counting Task. *Mindfulness*, *9*(5), 1402–1410. https://doi.org/10.1007/s12671-017-0880-1

Fong, A. H. C., Yoo, K., Rosenberg, M. D., Zhang, S., Li, C.-S. R., Scheinost, D., Constable, R. T., & Chun, M. M. (2019). Dynamic functional connectivity during task performance and rest predicts individual differences in attention across studies. *NeuroImage*, *188*, 14–25. https://doi.org/10.1016/j.neuroimage.2018.11.057

Fortenbaugh, F. C., DeGutis, J., & Esterman, M. (2017). Recent theoretical, neural, and clinical advances in sustained attention research. *Annals of the New York Academy of Sciences*, *1396*(1), 70–91. https://doi.org/10.1111/nyas.13318

Ganesan, S., A. Moffat, B., Van Dam, N. T., Lorenzetti, V., & Zalesky, A. (2023). Meditation attenuates default-mode activity: A pilot study using ultra-high field 7 Tesla MRI. *Brain Research Bulletin*, *203*, 110766. https://doi.org/10.1016/j.brainresbull.2023.110766

Ganesan, S., Beyer, E., Moffat, B., Van Dam, N. T., Lorenzetti, V., & Zalesky, A. (2022). Focused attention meditation in healthy adults: A systematic review and meta-analysis of cross-sectional functional MRI studies. *Neuroscience and Biobehavioral Reviews*, *141*. https://doi.org/10.1016/j.neubiorev.2022.104846

Gioia GA, Isquith PK, Guy SC, & Kenworthy L. (2015). *BRIEF2: Behavior Rating Inventory of Executive Function (2nd ed.).* [Computer software]. Psychological Assessment Resources, Inc.

Goldberg, S. B., Knoeppel, C., Davidson, R. J., & Flook, L. (2020). Does practice quality mediate the relationship between practice time and outcome in mindfulness-based stress reduction? *Journal of Counseling Psychology*, *67*(1), 115.

Goldberg, S. B., Tucker, R. P., Greene, P. A., Simpson, T. L., Hoyt, W. T., Kearney, D. J., & Davidson, R. J. (2019). What Can We Learn from Randomized Clinical Trials About the Construct Validity of Self-Report Measures of Mindfulness? A Meta-Analysis. *Mindfulness*, *10*(5), 775–785. https://doi.org/10.1007/s12671-018-1032-y

Goldberg, S. B., Wielgosz, J., Dahl, C., Schuyler, B., MacCoon, D. S., Rosenkranz, M., Lutz, A., Sebranek, C. A., & Davidson, R. J. (2016). Does the Five Facet Mindfulness Questionnaire measure what we think it does? Construct validity evidence from an active controlled randomized clinical trial. *Psychological Assessment*, *28*(8), 1009–1014. https://doi.org/10.1037/pas0000233

Goodman, M. S., Madni, L. A., & Semple, R. J. (2017). Measuring Mindfulness in Youth: Review of Current Assessments, Challenges, and Future Directions. *Mindfulness*, *8*(6), 1409–1420. https://doi.org/10.1007/s12671-017-0719-9

Grossman, P. (2011). Defining Mindfulness by How Poorly I Think I Pay Attention During Everyday Awareness and Other Intractable Problems for Psychology's (Re)Invention of Mindfulness: Comment on Brown et al. (2011). *Psychological Assessment*. https://doi.org/10.1037/a0022713

Hadash, Y., & Bernstein, A. (2019). Behavioral assessment of mindfulness: Defining features, organizing framework, and review of emerging methods. *Current Opinion in Psychology*, *28*, 229–237. https://doi.org/10.1016/j.copsyc.2019.01.008

Hanley, A. W., Mehling, W. E., & Garland, E. L. (2017). Holding the body in mind: Interoceptive awareness, dispositional mindfulness and psychological well-being. *Journal of Psychosomatic Research*. https://doi.org/10.1016/j.jpsychores.2017.05.014

Isbel, B., Lagopoulos, J., Hermens, D. F., & Summers, M. J. (2019). Mindfulness Induces Changes in Anterior Alpha Asymmetry in Healthy Older Adults. *Mindfulness*, *10*(7), 1381–1394. https://doi.org/10.1007/s12671-019-01106-w

Khalsa, S. S., Adolphs, R., Cameron, O. G., Critchley, H. D., Davenport, P. W., Feinstein, J. S., Feusner, J. D., Garfinkel, S. N., Lane, R. D., Mehling, W. E., Meuret, A. E., Nemeroff, C. B., Oppenheimer, S., Petzschner, F. H., Pollatos, O., Rhudy, J. L., Schramm, L. P., Simmons, W. K., Stein, M. B., … Zucker, N. (2018). Interoception and Mental Health: A Roadmap. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*. https://doi.org/10.1016/j.bpsc.2017.12.004

Khalsa, S. S., Rudrauf, D., Hassanpour, M. S., Davidson, R. J., & Tranel, D. (2020). The practice of meditation is not associated with improved interoceptive awareness of the heartbeat. *Psychophysiology*, *57*(2), 1–16. https://doi.org/10.1111/psyp.13479

Leigh, J., Bowen, S., & Marlatt, G. A. (2005). Spirituality, mindfulness and substance abuse. *Addictive Behaviors*, *30*(7), 1335–1341.

Levinson, D. B., Stoll, E. L., Kindy, S. D., Merry, H. L., & Davidson, R. J. (2014). A mind you can count on: Validating breath counting as a behavioral measure of mindfulness. *Frontiers in Psychology*, *5*(OCT). https://doi.org/10.3389/fpsyg.2014.01202

Lim, J., & Doshi, K. (2022). Breath Counting Task (BCT). In O. N. Medvedev, C. U. Krägeloh, R. J. Siegert, & N. N. Singh (Eds.), *Handbook of Assessment in Mindfulness Research* (pp. 1–13). Springer International Publishing. https://doi.org/10.1007/978-3-030-77644-2_48-1

Lim, J., Teng, J., Patanaik, A., Tandi, J., & Massar, S. A. A. (2018). Dynamic functional connectivity markers of objective trait mindfulness. *NeuroImage*, *176*(March), 193–202. https://doi.org/10.1016/j.neuroimage.2018.04.056

Lindsay, E. K., & Creswell, J. D. (2017). Mechanisms of mindfulness training: Monitor and Acceptance Theory (MAT). *Clinical Psychology Review*, *51*, 48–59. https://doi.org/10.1016/j.cpr.2016.10.011

MacCoon, D. G., MacLean, K. A., Davidson, R. J., Saron, C. D., & Lutz, A. (2014). No sustained attention differences in a longitudinal randomized trial comparing mindfulness based stress reduction versus active control. *PLoS ONE*, *9*(6). https://doi.org/10.1371/journal.pone.0097551

MacKillop, J., & Anderson, E. J. (2007). Further Psychometric Validation of the Mindful Attention Awareness Scale (MAAS). *Journal of Psychopathology and Behavioral Assessment*, *29*(4), 289–293. https://doi.org/10.1007/s10862-007-9045-1

Manly, T., Davison, B., Heutink, J., Galloway, M., & Robertson, I. H. (2000). Not enough time or not enough attention? Speed, error and self-maintained control in the Sustained Attention to Response Test (SART). *Clinical Neuropsychological Assessment: An International Journal for Research & Clinical Practice*, *3*, 167–177.

March, J. S., Parker, J. D., Sullivan, K., Stallings, P., & Conners, C. K. (1997). The Multidimensional Anxiety Scale for Children (MASC): Factor structure, reliability, and validity. *Journal of the American Academy of Child & Adolescent Psychiatry*, *36*(4), 554–565.

Mazefsky, C. A., Kao, J., & Oswald, D. (2011). Preliminary evidence suggesting caution in the use of psychiatric self-report measures with adolescents with high-functioning autism spectrum disorders. *Research in Autism Spectrum Disorders*, *5*(1), 164–174.

Morey, R. D., Rouder, J. N., Jamil, T., & Morey, M. R. D. (2015). *Package 'bayesfactor.'* http://cran.dcc.fc.up.pt/web/packages/BayesFactor/BayesFactor.pdf

Norsworthy, C., Jackson, B., & Dimmock, J. (2021). Advancing Our Understanding of Psychological Flow: A Scoping Review of Conceptualizations, Measurements, and Applications. *Psychological Bulletin*, *147*, 806–827. https://doi.org/10.1037/bul0000337

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, *115*(11), 2600–2606. https://doi.org/10.1073/pnas.1708274114

Nunnally Jr, J. C. (1970). *Introduction to psychological measurement.* https://psycnet.apa.org/PsycINFO/1970-19724-000

Okon-Singer, H., Hendler, T., Pessoa, L., & Shackman, A. J. (2015). The neurobiology of emotion-cognition interactions: Fundamental questions and strategies for future research.

Frontiers in Human Neuroscience, *9*(FEB), 1–14. https://doi.org/10.3389/fnhum.2015.00058

Pang, D., & Ruch, W. (2019). Scrutinizing the Components of Mindfulness: Insights from Current, Past, and Non-meditators. *Mindfulness*, *10*(3), 492–505. https://doi.org/10.1007/s12671-018-0990-4

Radloff, L. S. (1977). The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. *Applied Psychological Measurement*, *1*(3), 385–401. https://doi.org/10.1177/014662167700100306

Rahrig, H., Vago, D. R., Passarelli, M., Auten, A., Lynn, N. A., & Brown, K. W. (2022). Disrupting The Resting State: Meta-Analytic Evidence That Mindfulness Training Alters Default Mode Network Connectivity. *Scientific Reports*, *0123456789*, 1–21. https://doi.org/10.1038/s41598-022-15195-6

Rau, H. K., & Williams, P. G. (2016). Dispositional mindfulness: A critical review of construct validation research. *Personality and Individual Differences*, *93*, 32–43. https://doi.org/10.1016/j.paid.2015.09.035

Robertson, I. H., Manly, T., Andrade, J., Baddeley, B. T., & Yiend, J. (1997). "Oops!": Performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia*, *35*(6), 747–758. https://doi.org/10.1016/S0028-3932(97)00015-8

Rudkin, E., Medvedev, O. N., & Siegert, R. J. (2018). The Five-Facet Mindfulness Questionnaire: Why the Observing Subscale Does Not Predict Psychological Symptoms. *Mindfulness*, *9*(1), 230–242. https://doi.org/10.1007/s12671-017-0766-2

Shoham, A., Goldstein, P., Oren, R., Spivak, D., & Bernstein, A. (2017). Decentering in the process of cultivating mindfulness: An experience-sampling study in time and context. *Journal of Consulting and Clinical Psychology*, *85*(2), 123.

Stieger, J. R., Engel, S., Jiang, H., Cline, C. C., Kreitzer, M. J., & He, B. (2021). Mindfulness Improves Brain–Computer Interface Performance by Increasing Control Over Neural Activity in the Alpha Band. *Cerebral Cortex*, *31*(1), 426–438. https://doi.org/10.1093/cercor/bhaa234

Tomlinson, E. R., Yousaf, O., Vittersø, A. D., & Jones, L. (2018). Dispositional Mindfulness and Psychological Health: A Systematic Review. *Mindfulness*, *9*(1), 23–43. https://doi.org/10.1007/s12671-017-0762-6

Torgeson, J. K., Wagner, R. K., & Rashotte, C. A. (1999). Test review: Test of word reading efficiency (TOWRE). *Inc.: Austin, TX, USA*. https://sites.ualberta.ca/~lphillip/documents/Test%20of%20Word%20Reading%20Efficiency%20(TOWRE).doc

Tortella-Feliu, M., Luís-Reig, J., Gea, J., Cebolla, A., & Soler, J. (2020). An Exploratory Study on the Relations Between Mindfulness and Mindfulness-Based Intervention Outcomes. *Mindfulness*, *11*(11), 2561–2572. https://doi.org/10.1007/s12671-020-01471-x

Tran, U. S., Wasserbauer, J., & Voracek, M. (2020). Testing the incremental validity of dispositional mindfulness over and above the Big Five in accounting for mental health: A facet-level structural-equation modeling and predictor communality and dominance approach. *Personality and Individual Differences*, *156*, 109769.

Treves, I. N., Pichappan, Kannamai, Hammoud, Jude, Ehmann, Sebastian, Sacchet, Matthew, Bauer, Clemens, & Gabrieli, John. (in press). The Mindful Brain: A Systematic Review of the Neural Correlates of Trait Mindfulness. *Journal of Cognitive Neuroscience*.

Treves, I. N., Tello, L. Y., Davidson, R. J., & Goldberg, S. B. (2019). The relationship between mindfulness and objective measures of body awareness: A meta-analysis. *Scientific Reports*, *9*(1), 1–12. https://doi.org/10.1038/s41598-019-53978-6

Vago, D. R., & Silbersweig, D. A. (2012). Self-awareness, self-regulation, and self-transcendence (S-ART): A framework for understanding the neurobiological mechanisms of mindfulness. *Frontiers in Human Neuroscience*, *6*, 296. https://doi.org/10.3389/fnhum.2012.00296

Wagenmakers, E.-J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., Selker, R., Gronau, Q. F., Šmíra, M., Epskamp, S., Matzke, D., Rouder, J. N., & Morey, R. D. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review*, *25*(1), 35–57. https://doi.org/10.3758/s13423-017-1343-3

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., & Van Der Maas, H. L. (2011). *Why psychologists must change the way they analyze their data: The case of psi: comment on Bem (2011).* https://psycnet.apa.org/fulltext/2011-01895-001.html

Webb, C. A., Hilt, L. M., Swords, C. M., Bolt, D. M., Fisher, H., & Goldberg, S. (2023). *Are ecological momentary assessment measures of intervention change worth the trouble? Evaluation in four digital mental health trials*. https://osf.io/preprints/psyarxiv/3xvck/

Wechsler, D. (1955). Wechsler adult intelligence scale–. *Archives of Clinical Neuropsychology*. https://psycnet.apa.org/doiLanding?doi=10.1037/t15169-000

Whitfield, T., Barnhofer, T., Acabchuk, R., Cohen, A., Lee, M., Schlosser, M., Arenaza-Urquijo, E. M., Böttcher, A., Britton, W., Coll-Padros, N., Collette, F., Chételat, G., Dautricourt, S., Demnitz-King, H., Dumais, T., Klimecki, O., Meiberth, D., Moulinet, I., Müller, T., … Marchant, N. L. (2022). The Effect of Mindfulness-based Programs on Cognitive Function in Adults: A Systematic Review and Meta-analysis. *Neuropsychology Review*, *32*(3), 677–702. https://doi.org/10.1007/s11065-021-09519-y

Yue, W. L., Ng, K. K., Koh, A. J., Perini, F., Doshi, K., Zhou, J. H., & Lim, J. (2023). Mindfulness-based therapy improves brain functional network reconfiguration efficiency. *Translational Psychiatry*, *13*(1), 345. https://doi.org/10.1038/s41398-023-02642-9

Zanesco, A. P., Denkova, E., & Jha, A. P. (2024). Mind-wandering increases in frequency over time during task performance: An individual-participant meta-analytic review. *Psychological Bulletin*. https://doi.org/10.1037/bul0000424

Zelazo, P. D., Anderson, J. E., Richler, J., Wallner-Allen, K., Beaumont, J. L., & Weintraub, S. (2013). II. NIH TOOLBOX COGNITION BATTERY (CB): MEASURING EXECUTIVE FUNCTION AND ATTENTION. *Monographs of the Society for Research in Child Development*, *78*(4), 16–33. https://doi.org/10.1111/mono.12032

Sample Characteristics	**N**	**%**
Biological Sex		
Female	59	75.6
Male	19	24.4
Race		
American Indian or Alaska Native	1	1.3
Asian	11	14.1
Black or African American	4	5.1
Native Hawaiian or Other Pacific Islander	1	1.3
White	48	61.5

Multiracial	10	12.8
Unknown	3	3.8
Ethnicity		
Hispanic or Latino	10	12.8
Not Hispanic or Latino	68	87.2
Current Diagnoses (DSM-V)		
Major Depressive Episode	7	9.0
Generalized Anxiety Disorder	34	43.6
Social Anxiety Disorder	19	24.4
Panic Disorder	7	9.0
Specific Phobia	5	6.4
Other Diagnoses	11	14.1
No Diagnosis	36	46.2
Medication		
Psychotropic Medication	6	7.7

	M (Range)	**SD**
Age (years)	15.7 (13-18)	1.7
Family Income (dollars)	182,573.8 (3,000 - 500,000)	113,768.0

Table 1: Demographic characteristics. M: mean. SD: standard deviation. Current Diagnoses: participants may be multiply diagnosed.

	BC Correct	BC Miscounts
BC Miscounts	-0.63****	
BC Resets	-0.57****	-0.08

Table 2: Relationships among the main breath-counting measures. **** $p < 0.0001$

	BC Correct	BC Miscounts	BC Resets
Observing	-0.27*	0.13	0.19
Describing	-0.04	0.1	-0.07
Non-Judging	0.1	0.06	-0.09
Non-Reactivity	-0.08	0.30**	-0.06
Acting-Awareness	0.04	-0.01	-0.04
Total FFMQ	-0.05	0.2	-0.05

Table 3: Relationships between breath-counting metrics and traditional self-report mindfulness. *, $p < 0.05$ **, $p < 0.01$

	BC Correct	BC Miscounts	BC Resets
SART Accuracy	0.07	-0.02	-0.22 (-0.26*)
SART RT_CV	0.09	0.04	-0.15 (-0.16)

Table 4: Relationships between breath-counting metrics and SART. RT_CV: reaction time coefficient of variation. Value in parentheses is Spearman's correlation.

	BC Correct	BC Miscounts	BC Resets
Rumination	-0.09	0.04	-0.08
Depression	-0.21	0.00	0.14
Stress	-0.13	-0.01	0.06
Openness	-0.15	0.09	-0.03
Conscientiousness	0.15	-0.11	-0.18
Extraversion	-0.26*	0.14	0.21
Agreeableness	0.19	0.00	-0.28 (-0.28)*
Neuroticism	-0.03	-0.09	-0.03

Anxiety	-0.01	-0.05	0.02
BRIEF	-0.11	0.01	0.15

Table 5: Relationships between self-report measures and breath-counting measures. BRIEF: self-reported executive function. Value in parentheses is Spearman's correlation. *, $p < 0.05$ **, $p < 0.01$

Domain	Outcome	Correlation	FDR *p*-value
Self-report mindfulness	BCT Miscounts positively correlate with *FFMQ Nonreactivity*	0.30	0.045
Self-report mindfulness	BCT Accuracy* negatively correlates with *FFMQ Observing*	-0.33	0.04
Self-report mindfulness	BCT Performance decrements positively correlate with *FFMQ Observing*, *Non-reactivity*, and *Total* scores.	0.33, 0.29, 0.28	0.024, 0.028, 0.028

Table 6: Correlations passing FDR correction for BCT outcomes. *listed statistic for physiologically-verified accuracy (participants breath belt recorded 9 breaths for 9 button presses).

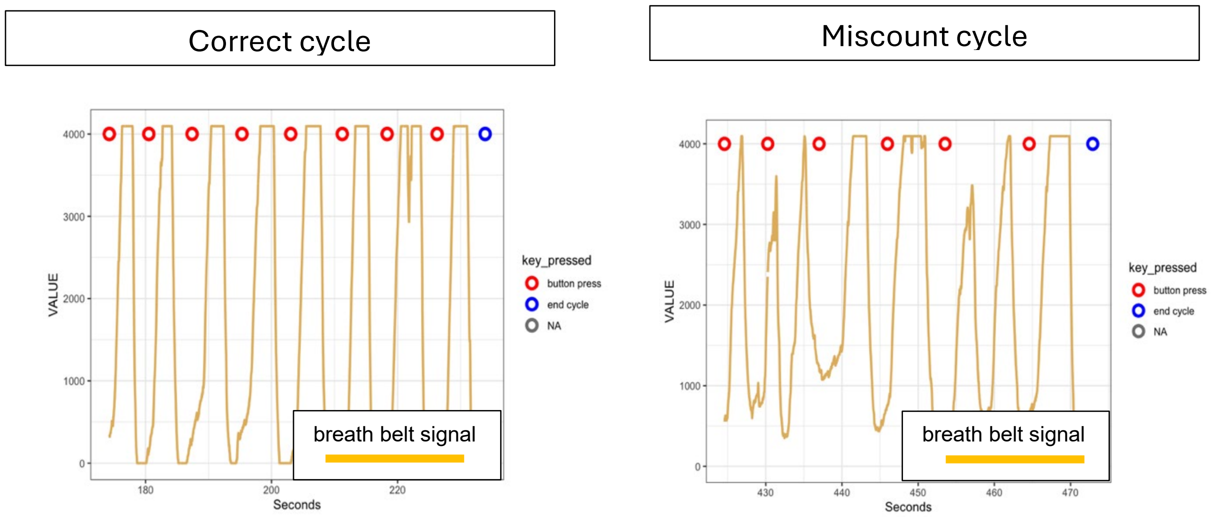

Figure 1: On left, correct cycle consisting of 8 breaths and then an ‘end cycle’ button press. On right, a miscount cycle consisting of only 6 cycles and then an ‘end cycle’ button press. Orange line is breath belt signal.

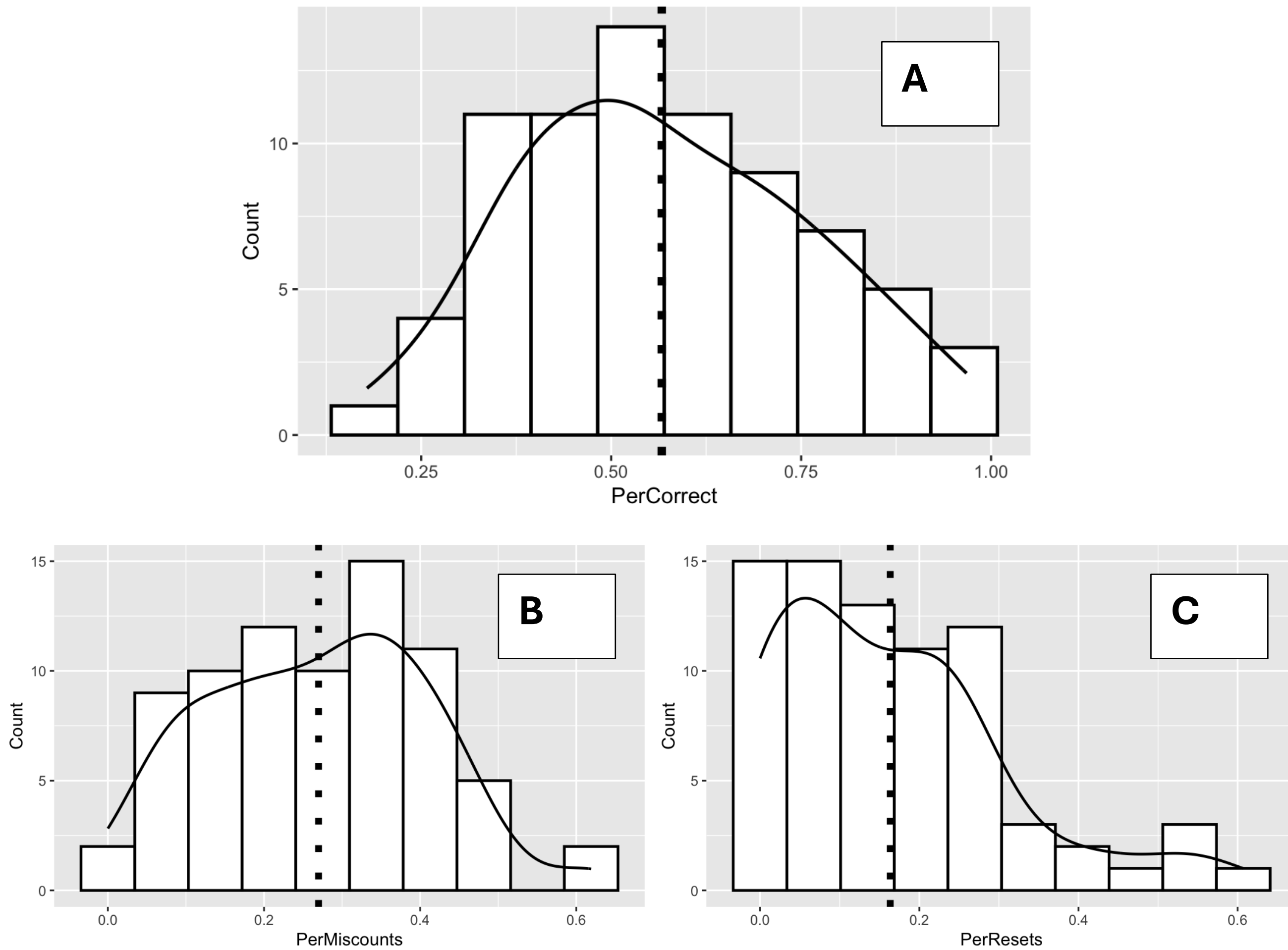

Figure 2: Rates of main breath-counting measures, where y-axis is number of participants with a given rate. A) Correct cycle rates per participant B) Miscount rates per participant C) Reset rates per participant. White bars are bins from histogram, black line is smoothed density, and dotted black line is mean.

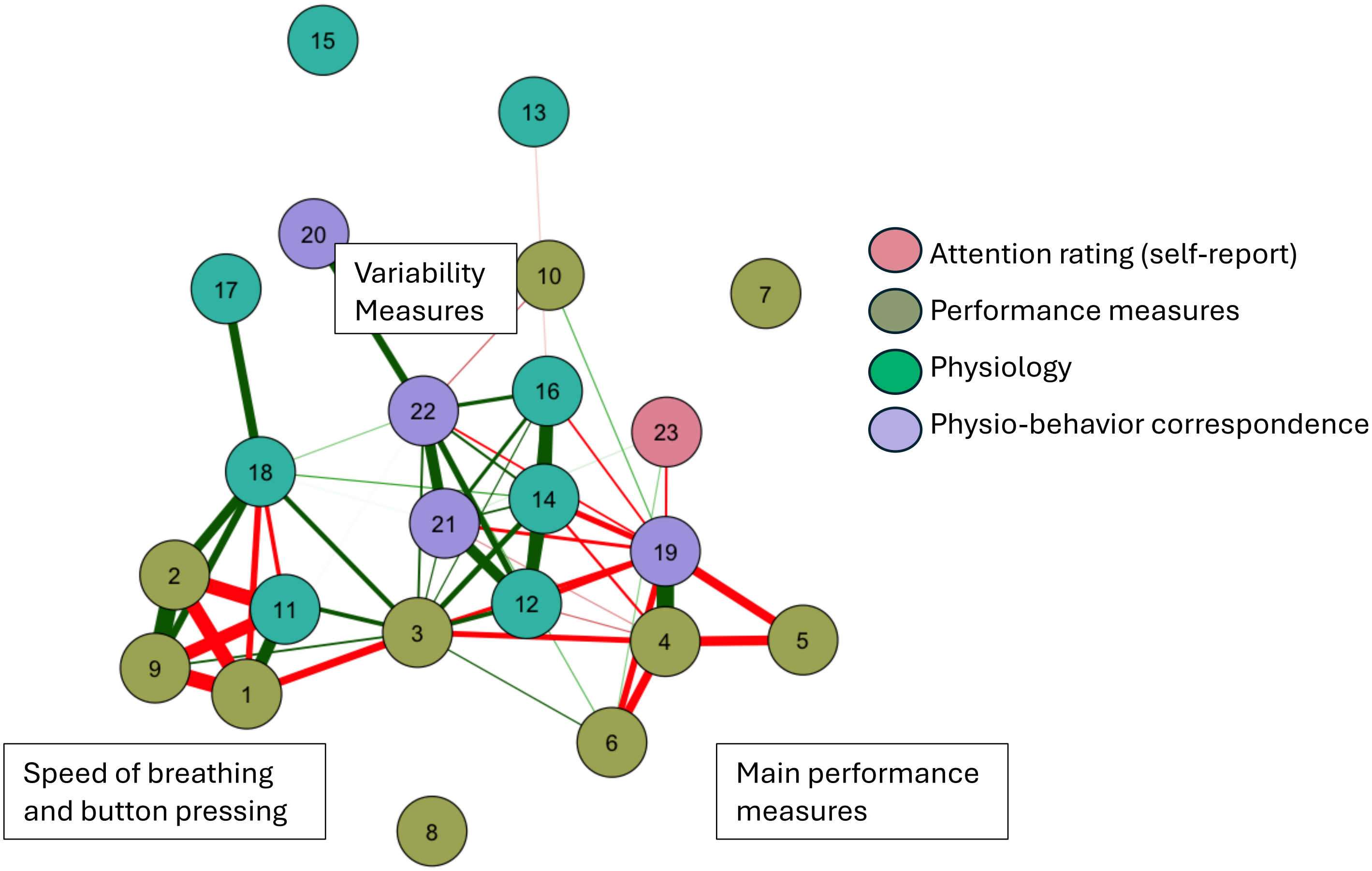

Figure 3: Network plot for all breath-counting metrics, generated using *qgraph* with spring layout with a correlation threshold of 0.3. Closer nodes reflect more related measures. 1: Number of Cycles, 2: Median Press Time, 3: Standard deviation of press time, 4: Correct Cycles, 5: Miscounts, 6: Resets, 7: Average Miscount Distance, 8: Average Reset Time, 9: Average Cycle Length, 10: Performance Decrement, 11: Mean Breath Rate, 12: Breath Variability, 13: Autoregressive relationship between peaks, 14: Coefficient of variation of peaks, 15: Autoregressive relationship between troughs, 16: Autoregressive relationship between troughs, 17: Breath in/out ratio, 18: Mean Pause Length, 19: Peak Percentage Correct, 20: Peak Mean Distance, 21: Peak Variance Distance, 22: Peak Change Distance, 23: attention rating. For calculations of metrics, see **Supplement Text 1.**

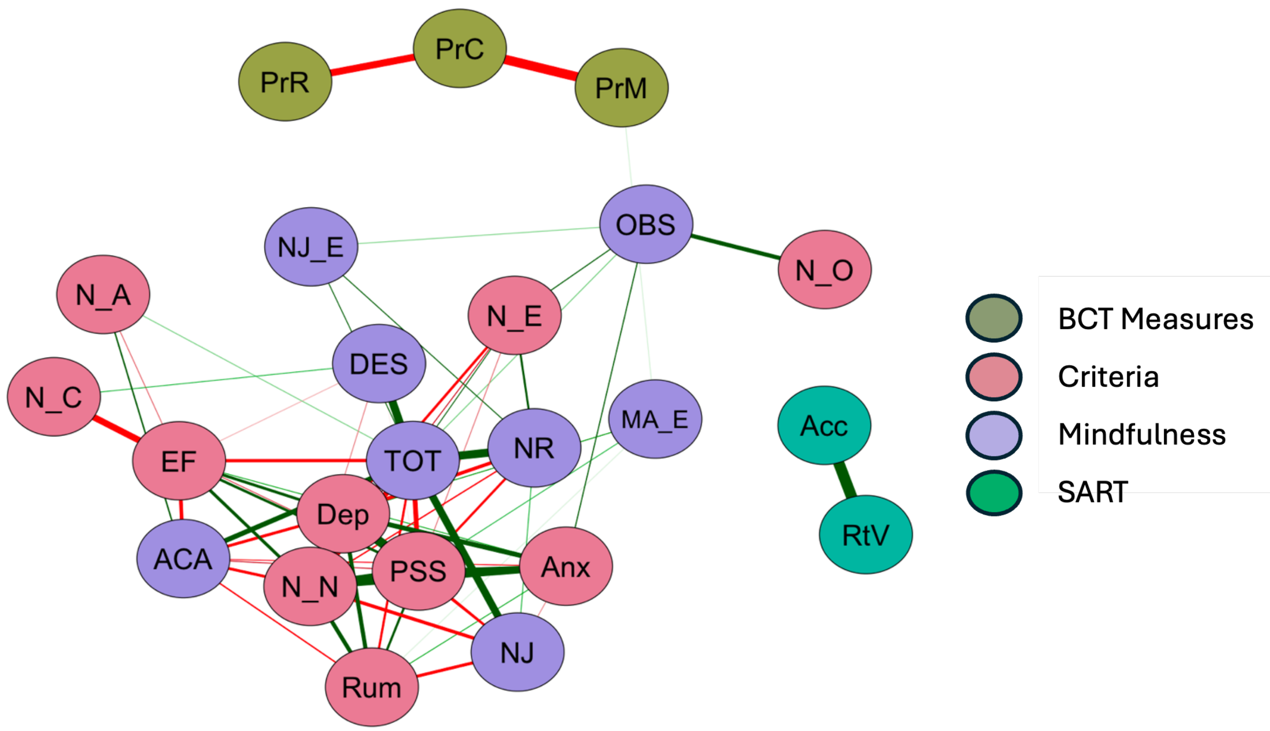

Figure 4: Network plot for main BCT measures, as well as mindfulness, SART, and criterion variables, generated using *qgraph* with spring layout with a correlation threshold of 0.3. Closer nodes reflect more related measures. PrC: Correct Cycles, PrM: Miscounts, PrR: Resets, N_O: open-mindedness, N_C: conscientiousness, N_E: extraversion, N_A: agreeableness, N_N: neuroticism, Dep: depression, Rum: rumination, PSS: perceived stress, Anx: anxiety, NJ_E: EMA non-judging (attitude), MA_E: EMA mindful awareness, ACA: acting with awareness, FFMQ, NJ: non-judging, FFMQ, NR: nonreactivity, FFMQ, DES: describing, FFMQ, OBS: observing, FFMQ, TOT: total score, FFMQ, Acc: SART accuracy, RT: SART reaction time variability.