Reliability and Validity of the Brief Attention and Mood Scale of 7 Items (BAMS-7): A Self-Administered, Online, and Real-World Measure

Kevin P. Madore[1], Allen M. Osman[1], Kelsey R. Kerlan[1], & Robert J. Schafer[1]

[1] Lumos Labs, Inc.

I. Abstract

Given changes in technology, regulatory guidance, and COVID-19, there has been an explosion in the number of online studies in the social and clinical sciences. This, in turn, has led to a need for brief and accessible instruments that are designed and characterized with self-administered, online research in mind. To fulfill this need, the Brief Attention and Mood Scale of 7 Items (BAMS-7) was developed and validated in five cohorts and four experiments to assess real-world attention and mood. In Experiment 1, an exploratory factor analysis was run on data from a large, healthy, adult sample (*N*=75,019, ages 18-89 years). Two subscales were defined and further characterized: one for Attention, the other for Mood. In Experiment 2, convergent validity (concordance) with existing questionnaires was established in a separate sample (*N*=150). Experiment 3 used a receiver-operating characteristic (ROC) analysis to demonstrate known-groups validity of the Attention and Mood subscales using a large sample (*N*=58,411) of participants reporting a lifetime diagnosis of ADHD, anxiety, or depression, as well as the healthy sample from Experiment 1. Experiment 3 also showed that the BAMS-7 Attention subscale provided superior classification performance for ADHD, and the Mood subscale provided superior classification for anxiety and depression. Finally, Experiment 4 applied the BAMS-7 definition to reanalyze data (*N*=3,489) from a previously published cognitive training study (Hardy et al., 2015), finding that the Attention and Mood subscales were sensitive to the intervention (compared to an active control) to different degrees. In sum, the elucidated psychometric properties and large normative dataset (*N*=75,019) for BAMS-7 may make it a useful instrument for assessing real-world attention and mood.

II. Introduction

Cognition and mood are impacted by numerous medical conditions (Armstrong & Okun, 2022; Bar, 2009; Eyre et al., 2015; Fast et al., 2023), lifestyle choices (Santos et al., 2014; Sarris et al., 2020; van Gool et al., 2007), healthy development and aging (Fernandes & Wang, 2018; Mather & Carstensen, 2005; Tomaszewski Farias et al., 2024; Yurgelun-Todd, 2007), and medications or other interventions (Keshavan et al., 2014; Koster et al., 2017; Reynolds et al., 2021; Skirrow et al., 2009). Conditions principally defined by impaired cognition – such as ADHD or mild cognitive impairment – are often associated with concomitant changes in mood status, either directly or indirectly (Chen et al., 2018; D'Agati et al., 2019; Ismail et al., 2017; Retz et al., 2012; Yates & Woods, 2013). Similarly, conditions principally

defined by one's mood or emotions – such as anxiety or depression – often have a corresponding impact on cognition (Gkintoni & Ortiz, 2023; Gulpers et al., 2016; Keller et al., 2019; Williams, 2016). Given the intimate relationship between cognition and mood, the ability to measure both in one scale may be both convenient and important.

Due to advances in technological capabilities, changes in emphases in regulatory guidance, and a shifting research landscape from the COVID-19 pandemic, there has been a recent surge of online studies in the social and clinical sciences (Arechar & Rand, 2020; Goodman & Wright, 2022; Lourenco & Tasimi, 2020; Saragih et al., 2022). As a result, instruments that are designed and characterized for online research are paramount. To collect reliable responses from large numbers of participants via their own internet-connected devices, instrument qualities like brevity and accessibility of language are required. For studies using relatively brief interventions, the time interval for evaluation is also important: for example, using Broadbent's Cognitive Failures Questionnaire (CFQ) (e.g., Bridger et al., 2013; Broadbent et al., 1982; Rast et al., 2009) to evaluate cognitive failures over the past six months may not be appropriate for measuring change over a shorter period of time. Furthermore, instruments that have been normed and validated based on traditional, in-person administration may have different characteristics with at-home, self-administration on one's own computer or smart device.

To provide a needed tool and a very large normative dataset to the research community, we present and describe a brief, seven-item scale of real-world attention and mood: the BAMS-7. The BAMS-7 moves beyond existing instruments in the literature by providing brevity, accessibility, and measures of multiple constructs (attention and mood) within one scale. Given that attention and mood are correlated in healthy (Carriere et al., 2008; Hobbiss et al., 2019; Irrmischer et al., 2018) and clinical populations (Bar, 2009; D'Agati et al., 2019; Retz et al., 2012; Skirrow et al., 2009), it may be advantageous to adopt one scale with separable measures of attention and mood. The scale may be useful both as an outcome measure (i.e., a dependent variable) or covariate (i.e., an independent variable). The BAMS-7 could be used to interpret the attention and mood of an individual by comparing with the large, normative population. Alternatively, the BAMS-7 could be used to analyze a population study-wide.

The Initial Nine-Item Survey

In 2015, Hardy et al. published the results of a large, online study evaluating an at-home, computerized cognitive training program (described also in Ng et al., 2021). As a secondary outcome measure, the authors created a nine-item survey of "cognitive failures and successes as well as emotional status" (p. 6; Hardy et al., 2015). This original survey is shown in **Table 1** and consisted of two parts. In a first section of four items, participants responded to questions about the frequency of real-world cognitive failures

***Table 1**. Original Hardy et al. (2015) nine-item survey.*

Item	**During the past month, how often have you…** *Valid response options: "Never", "1-2 times during the month", "1-2 times per week", "Several times per week", "Almost every day"*	**Reverse-coded**
1	. . lost track of details as you were reading and needed to go back and reread sections?	X
2	. . . misplaced items (e.g. reading glasses, keys) around the house?	X
3	. . . found yourself losing concentration during a conversation?	X
4	. . . remembered someone's name who had just been introduced to you?	
	Rate your experience over the last week… *Valid response options: "Strongly disagree", "Disagree", "Neither agree nor disagree", "Agree", "Strongly agree"*	
5	I felt creative.	
6	My ability to concentrate was good.	
7	I felt anxious.	X
8	I was in a bad mood.	X
9	I felt sad for no obvious reason.	X

or successes within the last month. In a second section of five items, participants responded to questions about the extent of agreement with statements relating to feelings of positive or negative mood and emotions, creativity, and concentration within the last week. Hardy et al. (2015) created a composite measure – the "aggregate rating" – by averaging across items.

Although the survey was not formally characterized, the first four items were similar to ones from the CFQ, and all items had a degree of face validity. Modifications of CFQ items reflected updates for modern-day relevance (e.g. removing "newspaper" as an example item that might be misplaced around the home) and for shortening the time interval of interest to make it possible to measure changes in a shorter study. Despite the reasonable set of items, it was clear that the survey was not designed to assess a single factor or construct. Although the authors reported that the average survey rating (and several individual items) improved as a result of a cognitive training intervention, more specificity may be warranted to interpret those changes.

The Current Research

Over the last several years, the same nine-item survey has been made available to hundreds of thousands of users of the Lumosity cognitive training program (Lumos Labs, Inc., San Francisco, CA) to inform future development of the program. Within this larger group, a subset of individuals has also provided demographic information (age, educational attainment, gender) and aspects of health history, including whether they have been diagnosed with any of a number of medical conditions.

Given this opportunity to psychometrically validate the survey, we conducted four experiments. Using data from 75,019 healthy individuals in this large, online cohort in Experiment 1, we explore the nine-item survey used by Hardy et al. (2015) and characterize a seven-item brief scale of real-world

attention and mood (the BAMS-7). Through factor analysis we identify two subscales, one for Attention and one for Mood, and identify distributional and psychometric properties. The two subscales are shown to have convergent validity with existing questionnaires of similar constructs in Experiment 2 in a separate sample of 150 individuals from Amazon Mechanical Turk (MTurk). Using cohorts of 12,976 individuals reporting an ADHD diagnosis, 20,577 individuals reporting an anxiety disorder diagnosis, and 24,858 individuals reporting a depression disorder diagnosis in Experiment 3, we show also that the BAMS-7 subscales have sound known-groups validity in three psychiatric conditions. We demonstrate through a double dissociation that each of the two subscales (Attention and Mood) are best at discriminating between healthy controls and individuals with different clinical diagnoses; the Attention subscale is more sensitive to ADHD and the Mood subscale is more sensitive to anxiety and depression. In Experiment 4, we use data from the study originally reported by Hardy et al. (2015) and show that the subscales are sensitive to a cognitive training intervention.

III. Methods

Participants

Data from five cohorts in four experiments were used in the following analyses. Survey responses from healthy participants ("Healthy cohort") were used to develop the BAMS-7 and its subscales in Experiment 1. Responses from MTurk participants were used to provide convergent validity with existing questionnaires in Experiment 2. Responses from participants who reported that they have been diagnosed with ADHD ("ADHD cohort"), anxiety disorder ("Anxiety cohort"), or depression disorder ("Depression cohort") were used to evaluate known-groups validity of the BAMS-7 subscales in Experiment 3. Responses from participants in the study run by Hardy et al. (2015; "Hardy cohort") were used to identify sensitivity to intervention effects in Experiment 4. See **Table 2** for demographic characteristics of each cohort in each experiment.

Ethics approval was provided by Western Institutional Review Board-Copernicus Group (WCGB IRB) for use of retrospective data from Lumosity users and for the collection and use of data from the MTurk sample.

Data from the Healthy, ADHD, Anxiety, and Depression cohorts were collected during normal use of a feature of the Lumosity training program. In the Lumosity Privacy Policy (www.lumosity.com/legal/privacy_policy), all participants agreed to the use and disclosure of non-personal data (e.g. de-identified or aggregate data) for any purpose. Participants were included if they were 18-89 years of age. The Healthy cohort included 75,019 participants who reported no diagnoses from a list of 34 options in an optional "information about you" survey. The ADHD cohort included 12,976 participants who reported a diagnosis of ADHD. The Anxiety cohort included 20,577 participants who reported a diagnosis of anxiety disorder. The Depression cohort included 24,858 participants who reported a diagnosis of depression disorder. Comorbid

***Table 2**. Demographics of each cohort in the current study.*

	Cohort					
	Healthy	**MTurk**	**ADHD**	**Anxiety**	**Depression**	**Hardy**
N	75,019	150	12,976	20,577	24,858	3,489
Age (range, mean, SD)	18-89yrs, 49.46, 16.32	20-63yrs, 40.16, 11.02	18-88yrs, 42.04, 15.93	18-89yrs, 47.38, 15.92	18-89yrs, 49.97, 15.87	18-80yrs, 37.73, 14.60
Gender (%female, %male, %unknown)	57.82%, 38.13%, 4.05%	49.33%, 50.00%, 0.67%	49.22%, 46.81%, 3.97%	68.48%, 27.66%, 3.86%	67.07%, 29.36%, 3.56%	54.46%, 45.54%, 0%
Education (%bachelor's or more, %less than bachelor's, %unknown)	59.15%, 34.36%, 6.50%	72.00%, 28.00%, 0%	53.61%, 40.90%, 5.49%	54.77%, 39.42%, 5.81%	54.40%, 40.08%, 5.52%	63.63%, 34.77%, 1.61%
Experiment #	1, 3	2	3	3	3	4

conditions were allowed in the ADHD, Anxiety, and Depression cohorts, such that a participant could be in multiple cohorts. Data from the additional MTurk cohort included 150 participants who were 18 or older from the general population and based in the United States without further restrictions.

The Hardy cohort included 3,489 participants who participated in the large, online, cognitive training study run by Hardy et al. (2015) and who provided complete responses on the nine-item survey used as a secondary outcome measure. Participants ranged in age from 18-80. Individuals completed the survey prior to randomization into a cognitive training intervention group or a crossword puzzle active control group, and completed the same survey following the 10-week intervention. A complete description of the cohort and study can be found in Hardy et al. (2015).

Survey Items

Individuals in all five cohorts in the four experiments took the original nine-item survey comprising items about cognitive failures, as well as mood, creativity, and concentration. As described in Hardy et al. (2015) and in the Introduction, the first four survey items related to a participant's cognitive performance over the past month, and the additional five items related to a participant's mood and emotional status over the past week. Response options for both sets of questions were on a five-point Likert scale, but the response options differed for the two sets. Response options for the first group of questions were "Never," "1-2 times during the month", "1-2 times per week", "Several times per week", "Almost every day". Response options for the second group of questions were "Strongly disagree", "Disagree", "Neither agree nor disagree", "Agree", "Strongly agree."

Participants were able to skip an item by selecting "N/A" in Experiments 1, 3, and 4; the MTurk sample in Experiment 2 did not have this option. Only participants who responded to all items were included; scores are only considered valid or complete if there are responses to all items (i.e., no N/A values).

Scoring involves numerically coding each response option on a scale from 0 to 4, where 0 represents the most negative response and 4 represents the most positive response. Items 1 (losing track of details reading), 2 (misplacing keys), 3 (losing concentration), 7 (anxious), 8 (bad mood), and 9 (sad) are reverse scored to preserve consistency of the scale with 0 representing the most negative response and 4 representing the most positive response. Thus, with this scale, a higher item score denotes better attention or more positive mood, depending on the focus of the question.

Concordance With Existing Questionnaires

To establish convergent validity with existing questionnaires in Experiment 2, data from 150 participants via MTurk were collected and analyzed for the BAMS-7. The additional questionnaires included: the 18-item Adult ADHD Self-Report Scale with a 6-month time interval (ASRS; Kessler et al., 2005), 12-item Attention-Related Cognitive Errors Scale with an open (unspecified) time interval (ARCES; Cheyne et al., 2006), 9-item Patient Health Questionnaire with a 2-week time interval (PHQ-9; Kroenke et al., 2001), 20-item Positive and Negative Affect Schedule with a "few-weeks" time interval (PANAS; Watson et al., 1988), and 7-item Generalized Anxiety Disorder questionnaire with a 2-week time interval (GAD-7; Spitzer et al., 2006). Standard scoring was adopted for each questionnaire:

ASRS (Kessler et al., 2005). The three outcome variables are the sum of 9 items for Part A, the sum of 9 separate items for Part B, and the sum of all 18 items for Parts A+B. Each item is rated on a five-point Likert scale (0=Never, 1=Rarely, 2=Sometimes, 3=Often, 4=Very Often). A higher sum (ranging from 0-36) for Part A denotes more inattention, a higher sum (ranging from 0-36) for Part B denotes more hyperactivity/impulsivity, and a higher sum (ranging from 0-72) for Parts A+B denotes more inattention and hyperactivity/impulsivity.

ARCES (Cheyne et al., 2006). The outcome variable is the item mean score across the 12 items. Each item is rated on a five-point Likert scale (1=Never, 2=Rarely, 3=Sometimes, 4=Often, 5=Very Often). A higher average (ranging from 1-5) denotes more inattention.

PHQ-9 (Kroenke et al., 2001). The outcome variable is the sum of the 9 items. Each item is rated on a four-point Likert scale (0=Not at all, 1=Several days, 2=More than half the days, 3=Nearly every day). A higher sum (ranging from 0-27) reflects greater severity of depression, where a score of 1-4 is minimal, 5-9 is mild, 10-14 is moderate, 15-19 is moderately severe, and 20-27 is severe.

PANAS (Watson et al., 1988). The two outcome variables are the sum of the 10 positive items and the sum of the 10 negative items, respectively. Each item is rated on a five-point Likert scale (1=Very slightly or not at all, 2=A little, 3=Moderately, 4=Quite a bit,

5=Extremely). A higher sum (ranging from 10-50) denotes more positive affect or more negative affect, respectively.

GAD-7 (Spitzer et al., 2006). The one outcome variable is the sum of the 7 items. Each item is rated on a four-point Likert scale (0=Not at all, 1=Several days, 2=More than half the days, 3=Nearly every day). A higher sum (ranging from 0-21) reflects greater severity of anxiety, where a score of 0-4 is minimal, 5-9 is mild, 10-14 is moderate, and 15-21 is severe.

The Supplemental Materials contain additional details regarding the MTurk approach and questionnaires.

Statistical Approaches

Four core approaches (Beavers et al., 2013; Brysbaert, 2024; Costello & Osborne, 2005; Knetka et al., 2019) were implemented to establish the BAMS-7. First, exploratory factor analysis was run for the Healthy cohort in Experiment 1 to identify the items for the BAMS-7 and its subscales. For this approach, latent factors were extracted from the nine items of the survey by maximizing item loadings on one factor and minimizing loadings on the other factors via varimax rotation. To address collinearity and sampling distribution adequacy, correlations among the items were examined, along with the Bartlett sphericity test statistic for factorability. A Scree plot was then qualitatively assessed to understand the inflexion point on its curve and thereby justify the number of factors to retain. More formally, factors with eigenvalues greater than or equal to 1 were retained (based on Kaiser, 1960). The items that loaded onto each factor with a communality above 0.4 were identified, and Cronbach's alpha was utilized to determine the internal consistency of each factor. Second, after appropriate removal of two of the nine items (Guvendir & Ozkan, 2022), convergent validity (concordance) was established in Experiment 2 by examining correlations between the two subscales and existing questionnaires of attention and mood from the MTurk sample. Third, the exploratory factor analysis informed the predictive approach for the ADHD, Anxiety, and Depression cohorts to evaluate known-groups validity of the subscales in Experiment 3. Fourth, an analysis of sensitivity to intervention effects was conducted with the Hardy cohort in Experiment 4. Model fit and covariance between factors were examined.

Statistical Analysis

All statistical analyses were conducted in Python (version 3.9.7) using Pandas (version 1.3.5) and NumPy (version 1.20.3) and the following freely available libraries. Exploratory factor analysis used the factor_analyzer library (version 0.3.1). Cronbach's alpha was computed with Pingouin (version 0.5.2), as was the ANCOVA analysis to reanalyze the original Hardy et al. (2015) intervention results given the newly defined BAMS-7. Distribution skewness and kurtosis were computed with SciPy (version 1.7.3), as were correlations among questionnaires from the MTurk sample. Receiver operating characteristic (ROC) analysis used Scikit-learn (version 1.0.2).

Unless otherwise stated, 95% confidence intervals and statistical comparisons were computed using standard bootstrap

procedures (Wright et al., 2011) with 10,000 iterations.

IV. Results

Experiment 1

Analysis of the Original Nine-Item Survey

Survey results from the Healthy cohort (*N*=75,019) had varying degrees of inter-item (pairwise) correlation, ranging from -0.05 to 0.53, as shown in **Figure 1**. All correlations were significantly different from 0 (with bootstrapped 95% confidence intervals) at the $p<.0001$ level. Item-total correlations ranged from 0.03 ("Remembered Names") to 0.52 ("Good Concentration"), with all p's<.0001 significantly different from 0. Cronbach's alpha for the full survey was 0.705 (0.702-0.708). Bartlett's test of sphericity was significant (T=136408.60, $p<.0001$), indicating strong factorability and sampling adequacy in the dataset.

Scree and eigenvalue analyses were used to determine the number of factors to retain in an exploratory factor analysis as displayed in **Figure 2**. The analysis with an eigenvalue cutoff of 1.0 indicates that a 3-factor solution might be appropriate. The results of the 3-factor solution with

***Figure 1**. Inter-item correlation coefficients of original nine-item survey in Experiment 1. Warmer colors show stronger correlations between respective items. All correlations are significantly different from 0 with p's<.0001.*

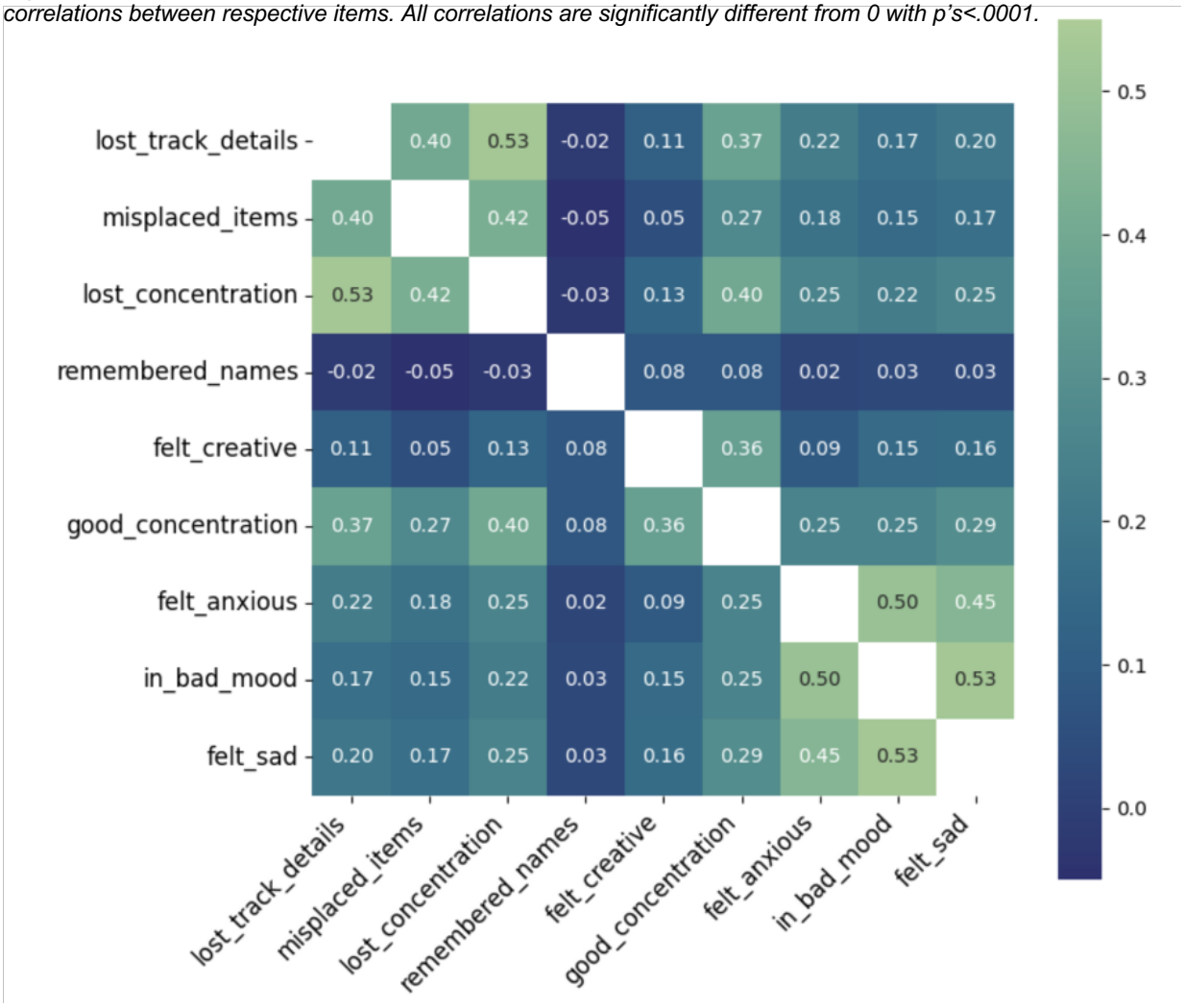
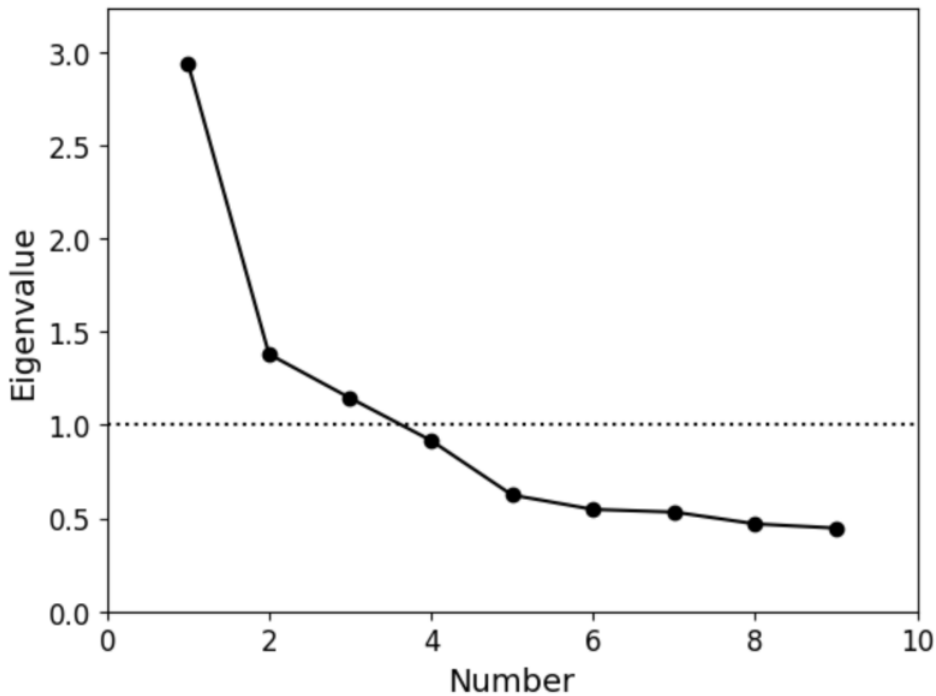


***Figure 2**. Scree plot for exploratory factor analysis of original nine-item survey in Experiment 1.*



Note. X-axis shows factor number and Y-axis shows eigenvalue. The dotted line with 1.0 is the cutoff for inclusion in the factor solution.

varimax rotation are shown in **Table 3** with factor loadings of 0.4 or greater in bold type. As expected, several of the items related to cognitive successes and failures loaded together, as did several of the items related to mood. The "Remembered Names" item did not load significantly onto any of the three factors, and was dropped. The "Good Concentration" item was the only one to load strongly onto multiple factors: both the first factor, which included other items related to cognitive failures primarily associated with attention functioning, and the third factor, which included the "Felt Creative" item.

Cronbach's alpha was computed to assess the internal consistency of each factor in the 3-factor solution. Factors 1 and 2 both had acceptable Cronbach's alpha values of 0.728 and 0.745, respectively, with bootstrapped 95% confidence intervals of 0.725-0.731 and 0.742-0.748. Factor 3, however, had a lower Cronbach's alpha value of 0.529, with a bootstrapped 95% confidence interval of 0.522-0.535. The low internal consistency and lack of an obvious description of Factor 3 led us to eliminate this factor, and subsequently to eliminate the orphaned item ("Felt Creative") that no longer loaded onto a factor.

Characterization of the BAMS-7

The resulting seven-item, two-factor scale is the BAMS-7, shown in **Table 4**. On the basis of the factor analysis and the nature of the items, scores from items loading onto the first factor are averaged to

***Table 3**. Results of exploratory factor analysis of original nine-item survey in Experiment 1.*

Item	Factor 1	Factor 2	Factor 3
Lost track of details	**0.69**	0.13	0.04
Misplaced items	**0.56**	0.12	-0.04
Lost concentration	**0.72**	0.18	0.07
Remembered names	-0.06	0.02	0.18
Felt creative	0.11	0.09	**0.52**
Good concentration	**0.46**	0.19	**0.56**
Felt anxious	0.19	**0.63**	0.05
In bad mood	0.09	**0.77**	0.12
Felt sad	0.16	**0.65**	0.16

Note. Factor loadings of 0.4 or higher are in bold type.

***Table 4**. Characterization of the BAMS-7 and its subscales from Experiment 1.*

Item	**During the past month, how often have you…** *Valid response options: "Never", "1-2 times during the month", "1-2 times per week", "Several times per week", "Almost every day"*	Subscale
1	. . lost track of details as you were reading and needed to go back and reread sections?	Attention
2	. . . misplaced items (e.g. reading glasses, keys) around the house?	Attention
3	. . . found yourself losing concentration during a conversation?	Attention
	Rate your experience over the last week… *Valid response options: "Strongly disagree", "Disagree", "Neither agree nor disagree", "Agree", "Strongly agree"*	
4	My ability to concentrate was good.	Attention
5	I felt anxious.	Mood
6	I was in a bad mood.	Mood
7	I felt sad for no obvious reason.	Mood

compute an Attention subscale, and scores from items loading onto the second factor are averaged to compute a Mood subscale. Note that the two groups of question types in the BAMS-7 do not correspond directly to the two factors. Instead, the item on "Good Concentration," despite falling in the second group of questions, loads onto the first factor and therefore contributes to the Attention subscale.

Distributional and psychometric properties of the BAMS-7 subscales are shown in **Table 5** for the Healthy cohort.

***Table 5**. Distributional and psychometric properties of the BAMS-7 in Experiment 1.*

Characteristics	Factor 1: Attention subscale	Factor 2: Mood subscale
Mean	2.3856 (2.3795, 2.3919)	2.2359 (2.2288, 2.2428)
SD	0.8598	0.9836
Range	0-4	0-4
Skew	-0.4837 (-0.4955, -0.4721)	-0.1399 (-0.1508, -0.1291)
Kurtosis	-0.3624 (-0.3850, -0.3396)	-0.6942 (-0.7082, -0.6805)

Note. Bootstrapped 95% confidence intervals are shown in parentheses as applicable. Skew and kurtosis are significantly different from 0 with p's<.0001.

Both of the subscales have modest, but significant, negative skewness and kurtosis.

Both of the subscales are related to the demographic variables of gender and age in some way. With one-way ANOVAs, the Attention subscale significantly varied with gender ($Mean_{Male}$=2.3608, $Mean_{Female}$=2.4011, $Mean_{Unknown}$=2.3981; F=19.31, p<.0001) while the Mood subscale did not ($Mean_{Male}$=0.9727, $Mean_{Female}$=0.9912, $Mean_{Unknown}$=0.9785; F=1.33, p=.27). With correlation tests, both subscales are positively associated with age within the measured range (18-89 years) (Attention: r=0.1994, p<.001; Mood: r=0.2285, p<.001). While an age-related increase on the Attention scale may be surprising given the well-established decline in cognitive performance during aging, this finding is consistent with the characteristics of the CFQ (see de Winter et al. 2015; Rast et al., 2009; for similar results with additional questionnaires, see Cyr & Anderson, 2019; Tassoni et al., 2022). It is also consistent with the hypothesis that self-reported cognitive failures and successes may reflect something distinct from what is measured via objective cognitive tests (Eisenberg et al., 2019; Yapici-Eser et al., 2021).

Norms for the BAMS-7

An important goal of this paper is to provide a normative dataset for the BAMS-7. Norm distributions are shown across the whole population of 75,019 healthy participants for the Attention subscale (**Figure 3A**) and Mood subscale (**Figure 3D**), by gender (**Figure 3B** and **3E**), and by age in decade (**Figure 3C** and **3F**).

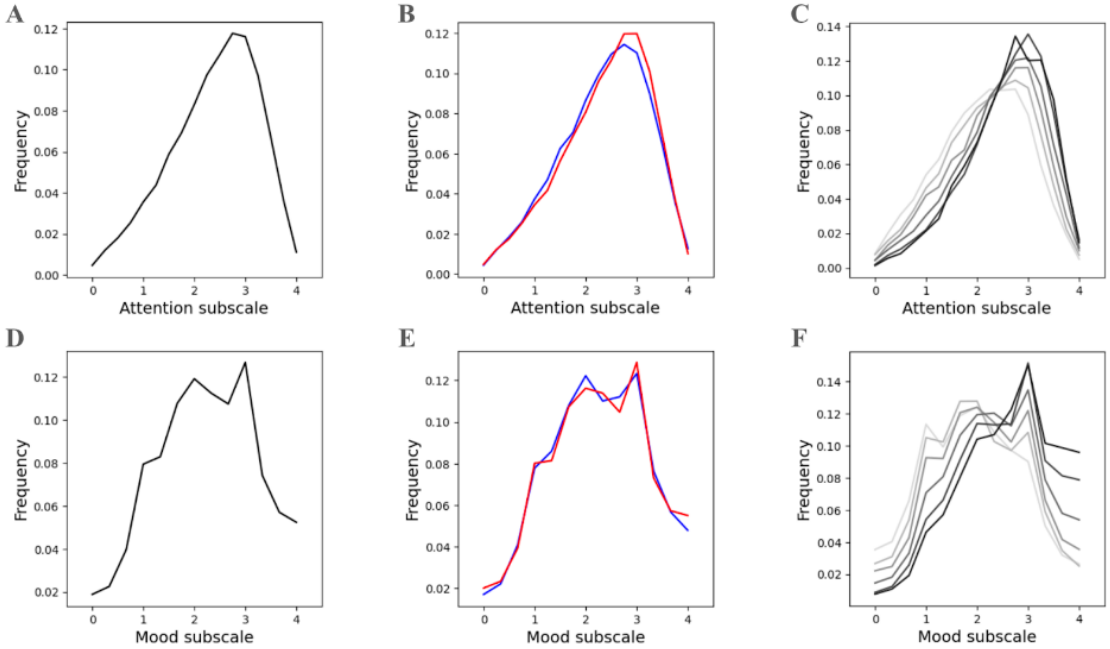
Norm tables across the whole population, by gender, and by age in decade are also provided in look-up format for the Attention subscale (**Table 6A**) and Mood subscale (**Table 6B**).

Experiment 2

Concordance with Existing Questionnaires

To establish convergent validity, a series of correlations were computed relating the BAMS-7 Attention and Mood subscales to five known instruments of attention and mood over various

***Figure 3**. Normed distribution of BAMS-7 Attention and Mood subscales across whole population, by gender, and by age in decade.*



Note. Panels (A) and (D) refer to the whole normed distribution of ages and genders for Attention and Mood subscales, respectively. Panels (B) and (E) refer to distribution by gender (red=female, blue=male) across ages. Panels (C) and (F) refer to distribution by age in decade (darker color is older) across genders.

***Table 6A**. Norm table for the Attention subscale across the whole population, by gender, and by age in decade.*

Population	N	0.0	0.33	0.67	1.0	1.33	1.67	2.0	2.33	2.67	3.0	3.33	3.67	4.0
Healthy cohort (all genders, ages 18-89)	75019	2	4	8	16	24	35	47	58	69	82	89	95	100
Ages (18-29), gender f	5059	5	10	17	31	42	53	65	75	84	92	95	98	100
Ages (18-29), gender m	7171	3	6	12	22	31	43	56	67	77	87	93	97	100
Ages (30-39), gender f	4456	3	7	13	25	35	48	60	70	79	89	95	98	100
Ages (30-39), gender m	4712	2	5	10	19	29	41	54	65	75	87	93	97	100
Ages (40-49), gender f	6650	3	5	10	20	29	41	53	65	74	86	92	97	100
Ages (40-49), gender m	4419	2	4	8	16	25	37	50	61	72	85	92	96	100
Ages (50-59), gender f	11662	2	3	7	14	21	32	44	56	67	81	89	95	100
Ages (50-59), gender m	5215	1	3	6	13	22	32	44	56	67	80	89	95	100
Ages (60-69), gender f	10931	1	2	5	10	17	26	37	49	60	75	84	92	100
Ages (60-69), gender m	4702	1	2	4	10	17	26	37	47	59	75	84	93	100
Ages (70-79), gender f	4034	1	2	4	8	14	23	33	44	56	71	81	90	100
Ages (70-79), gender m	2020	1	2	4	8	14	21	32	42	55	69	80	91	100
Ages (80-89), gender f	584	1	2	4	7	14	24	34	47	60	73	81	89	100
Ages (80-89), gender m	368	0	1	2	7	14	23	35	46	57	71	82	92	100

***Table 6B**. Norm table for the Mood subscale across the whole population, by gender, and by age in decade.*

Population	N	0.0	0.25	0.5	0.75	1.0	1.25	1.5	1.75	2.0	2.25	2.5	2.75	3.0	3.25	3.5	3.75	4.0
Healthy cohort (all genders, ages 18-89)	75019	0	2	3	6	10	14	20	27	35	45	55	67	79	89	95	99	100
Ages (18-29), gender f	5059	1	3	6	11	16	22	29	38	47	58	68	79	88	94	98	100	100
Ages (18-29), gender m	7171	1	2	6	9	15	21	30	39	49	59	69	79	88	94	97	99	100
Ages (30-39), gender f	4456	1	3	5	9	14	19	26	35	45	55	65	75	86	93	97	100	100
Ages (30-39), gender m	4712	1	2	4	7	11	17	24	32	41	51	62	73	84	91	97	99	100
Ages (40-49), gender f	6650	1	2	4	8	12	17	23	30	39	49	59	70	82	91	97	99	100
Ages (40-49), gender m	4419	0	1	3	5	9	13	20	26	35	45	56	68	79	89	95	99	100
Ages (50-59), gender f	11662	0	2	3	5	9	13	18	25	33	42	53	65	78	88	95	99	100
Ages (50-59), gender m	5215	0	1	3	5	8	11	16	22	30	40	51	63	75	86	94	99	100
Ages (60-69), gender f	10931	0	1	2	4	6	9	14	19	26	35	46	59	72	84	94	98	100
Ages (60-69), gender m	4702	0	1	2	4	6	9	13	18	26	36	47	59	72	84	93	98	100
Ages (70-79), gender f	4034	0	1	1	3	5	7	12	18	25	34	45	59	71	84	94	99	100
Ages (70-79), gender m	2020	0	1	2	4	6	9	14	20	28	37	49	61	72	83	93	98	100
Ages (80-89), gender f	584	0	1	2	4	7	9	12	19	28	37	47	60	74	84	93	99	100
Ages (80-89), gender m	368	1	1	2	4	4	8	14	20	27	39	52	67	78	87	95	98	100

Note. For each table, each row is a different slice of the population (or the whole population with all genders, at the top of each table). This shows the percentile (i.e., what percent of the population is less than or equal to the value shown) associated with each possible value of the respective subscale.

timescales from the independent online MTurk cohort. **Table 7A** shows *r*-values in a correlation matrix for the Attention subscale with the attention instruments, and **Table 7B** shows *r*-values for the Mood subscale with the mood instruments. All *p*-values for the correlations were < .001, meaning that the BAMS-7 Attention and Mood subscales showed significant relationships respectively with each existing questionnaire of attention and mood. The Attention subscale showed stronger relationships numerically with the attention instruments – ASRS and ARCES – while the Mood subscale showed stronger relationships with the mood instruments – GAD, PHQ, and PANAS. Note that many of the correlations are negative because a higher score on the BAMS-7 indicates higher attention or mood while a higher score on each of the known instruments (excluding PANAS positive affect) indicates higher inattention or lower mood. This pattern of results indicates that the BAMS-7 shows concordance with existing questionnaires. This can be seen in Supplementary Table 1 in the Supplementary Materials, which shows the entire set of correlations between both BAMS-7 subscales and all five pre-existing attention and mood instruments (for similar results, see Carriere et al., 2008; Franklin et al., 2017; Jonkman et al., 2017). The Supplemental Materials also contain an additional table demonstrating strong item-level correlations between each of the BAMS-7 questions and those from the existing questionnaires with similar descriptions, demonstrating additional concordance at the item level.

***Table 7A**. Convergent validity of BAMS-7 Attention subscale with existing attention questionnaires in Experiment 2.*

Questionnaire	BAMS-7 Attention Subscale
ASRS Part A	-0.750
ASRS Part B	-0.609
ASRS Total (A+B)	-0.720
ARCES	-0.788

***Table 7B**. Convergent validity of BAMS-7 Mood subscale with existing mood questionnaires in Experiment 2.*

Questionnaire	BAMS-7 Mood Subscale
PHQ	-0.770
PANAS Pos Aff	0.320
PANAS Neg Aff	-0.780
GAD	-0.765

Note. For each table, correlation coefficients are shown and p's<.001.

Experiment 3

Discriminatory Power of the Subscales in ADHD, Anxiety, and Depression

To evaluate the convergent and divergent validity of the BAMS-7, Attention and Mood subscale scores from the ADHD, Anxiety, and Depression cohorts were each compared to those from the Healthy cohort. A series of ROC analyses were performed to assess known-groups validity: (1) Attention subscale scores for ADHD vs Healthy, (2) Attention subscale scores for Anxiety vs Healthy, (3) Attention subscale scores for Depression vs Healthy, (4) Mood subscale scores for ADHD vs Healthy, (5) Mood subscale scores for Anxiety vs Healthy, and (6) Mood subscale scores for Depression vs Healthy. The resulting ROC curves are shown in **Figure 4A** for the Attention subscale and **4B** for the Mood subscale, and the corresponding areas under the curves (AUCs) are shown in **Table 8**.

Differences within each of the three psychiatric conditions vs healthy controls were assessed by subscale to examine discriminatory ability. Within ADHD vs Healthy, the Attention subscale had a significantly higher (p<.0001) AUC than the Mood subscale, which provides further evidence of the factor structure of the BAMS-7, given that ADHD is primarily a disorder of attention (Kessler et al., 2005). For both the Anxiety vs Healthy and Depression vs Healthy comparisons, it was instead the Mood subscale that was significantly better (p<.0001) at discriminating between populations compared to the Attention subscale. This profile provides additional validity of the meaning of the subscales

***Figure 4A**. Known-groups validity with ROC curves of BAMS-7 Attention subscale in ADHD, Anxiety, and Depression cohorts in Experiment 3.*

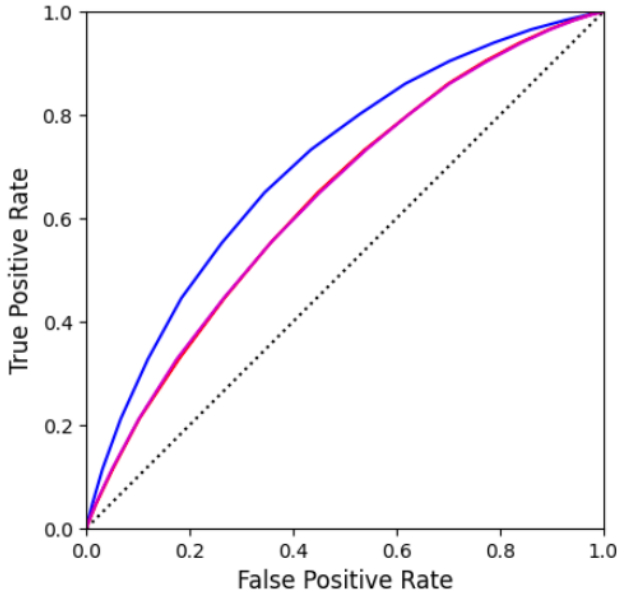
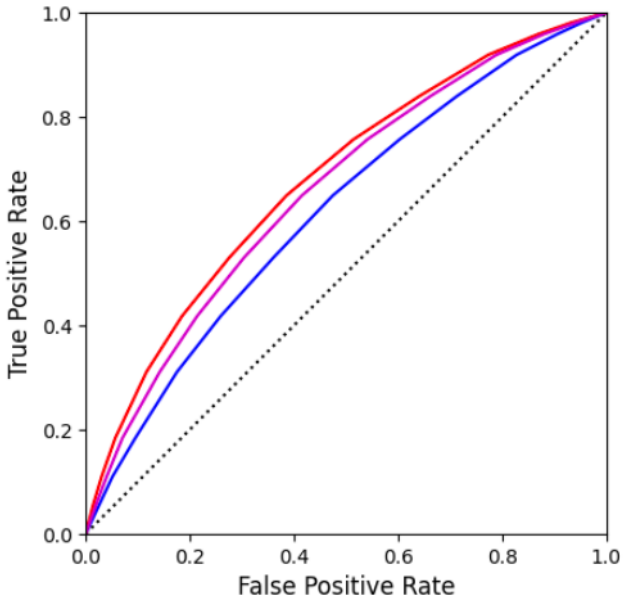


***Figure 4B**. Known-groups validity with ROC curves of BAMS-7 Mood subscale in ADHD, Anxiety, and Depression cohorts in Experiment 3.*



Note. In each figure, blue=ADHD cohort, red=Anxiety cohort, and magenta=Depression cohort.

because mood is a hallmark of anxiety and depression (Kroenke et al., 2001; Spitzer et al., 2006).

The ability of each of the BAMS-7 subscales to discriminate between the three psychiatric populations was assessed. Indeed, for the Attention

***Table 8**. Known-groups validity with AUC properties of BAMS-7 Attention and Mood subscales in ADHD, Anxiety, and Depression cohorts in Experiment 3.*

AUC	Attention subscale	Mood subscale
ADHD vs Healthy	0.7040 (0.6993, 0.7088)	0.6196 (0.6144, 0.6247)
Anxiety vs Healthy	0.6392 (0.6351, 0.6435)	0.6795 (0.6755, 0.6835)
Depression vs Healthy	0.6391 (0.6217, 0.6408)	0.6587 (0.6549, 0.6626)

Note. Bootstrapped 95% confidence intervals are shown in parentheses.

subscale, the AUC was significantly greater (p<.0001) for the ADHD vs Healthy analysis relative to each of the Anxiety and Depression vs Healthy contrasts. There was no significant difference between the AUC of the Anxiety vs Healthy and Depression vs Healthy analysis for the Attention subscale. Conversely, the Mood subscale had the poorest discrimination in terms of AUC between ADHD vs Healthy. Instead the Mood subscale had the significantly highest (p<.0001) AUC for the Anxiety vs Healthy analysis followed by Depression vs Healthy followed by ADHD vs Healthy.

Experiment 4

Sensitivity to a Cognitive Intervention

To test whether the BAMS-7 might have utility as an outcome measure in studies, we re-analyzed the data from Hardy et al. (2015) using the new characterization of the BAMS-7 and its Attention and Mood subscales, excluding any participants with null responses. Following the original analysis of Hardy et al. (2015), statistical analysis used an ANCOVA with baseline score as a covariate to test for intervention effects. Age was also entered as a covariate to examine effects of intervention across the lifespan (for covariate results, see the Supplemental Materials). The intervention effect on a given measure was the change (post - pretest) in the Lumosity group vs change in the active control Crosswords group.

Table 9 shows the results of the sensitivity analysis of the Hardy cohort on the BAMS-7 Attention and Mood subscales. Consistent with the original analysis, there was a group effect on the change in both the Attention and Mood subscales with the cognitive training group improving more than the crosswords one (Attention: $F(1,3485)=53.73$, p<.0001; Mood: $F(1,3485)=17.57$, p<.001). However, as might be expected for a cognitive intervention, the effect size (Cohen's d of ANCOVA-adjusted change scores) was greater for the Attention subscale (0.247) than for the Mood subscale (0.148). These results demonstrate that the subscales of the BAMS-7 are sensitive to a cognitive

***Table 9**. Sensitivity analysis of intervention effects with BAMS-7 Attention and Mood subscales with original Hardy et al. (2015)*

Intervention Effect	Crossword Puzzles	Cognitive Training
Attention subscale Change mean (SD)	1.859 (2.901)	2.409 (3.020)
Mood subscale Change mean (SD)	1.027 (2.811)	1.317 (2.761)

intervention, and therefore may have utility as outcome measures in studies.

V. Discussion

We describe a brief, seven-item scale of real-world attention and mood established from five cohorts in four experiments: the BAMS-7. The scale is designed and characterized for at-home, self-administered use and shows promise for measuring multiple constructs with brevity and accessibility in mind.

The four experiments established the validity and reliability of the BAMS-7. The scale was developed using a very large data set of 75,019 healthy individuals who participated in the Lumosity cognitive training program in Experiment 1; Experiment 1 was also used to characterize the inter-item correlation coefficients of the original nine items of the initial survey, and to determine Cronbach's alpha and establish distributional and psychometric properties of the BAMS-7. Concordance with existing scales was indicated from an MTurk sample in Experiment 2 in the form of convergent validity. The resulting scale was further validated in Experiments 3 and 4. Experiment 3 established known-groups validity in cohorts of Lumosity users reporting diagnoses of conditions that might be expected to have specific impairments on one or the other subscale of the scale (ADHD on the Attention subscale and anxiety or depression on the Mood subscale). Experiment 4 re-examined data with the scale from a 2015 study published by Hardy on intervention effects with a formal sensitivity analysis.

Factor analysis indicates two latent factors in the seven-item scale. The first factor includes adaptations of three items from the CFQ that focus on real-world attention function, and one item that queries the extent to which the responder agrees with the statement "I had good concentration" over the past week. The second factor includes items related to mood and anxiety. The resulting subscales – Attention and Mood, respectively – have acceptable internal consistency and descriptive statistics that may make them particularly useful in research.

A strength of the BAMS-7 is the size and diversity of its normative data set. Age norms in the range 18-89 are provided, along with normed distribution and look-up tables across the whole population, by gender, and by age in decade for each subscale. These norms have potential to assist in comparisons from study to study and in standardized effect sizes, along with the identification of outliers (Brysbaert, 2024). Both the Attention and

Mood subscales are positively correlated with age, which may appear paradoxical given the extensive literature on age-related cognitive decline. However, this relationship with age is consistent with the CFQ (e.g, de Winter et al., 2015; Rast et al., 2009), suggesting a general divergence between objective and subjective measures of cognitive performance. It should be noted that the correlation with age is observed on a cross-sectional basis, so an alternative hypothesis is that there are generational differences in the perception of cognitive functioning. Future research should determine how subjective cognitive measures like the BAMS-7 change in longitudinal studies.

There are at least a few questions that stem from the current work on the BAMS-7. First, is there really a need for another scale of this kind? We think that the answer is yes because of the pressing need for short, accessible, self-administered, and online-appropriate assessments. Second, is it okay that time intervals are a little different between attention and mood items, with the former mostly highlighting the past month and the latter the last week? We think that this is adequate for a behavioral measure of cognition and mood for two reasons: fluctuations that are significant in mood and attention may not operate on the same time scales (Esterman & Rothlein, 2019; Irrmischer et al., 2018; McConville & Cooper, 1997; Zanesco et al., 2022), and relatedly, a scale of attention and mood needs to offer time intervals for probing in which opportunities to capture fluctuations in attention and mood are available and likely.

A limitation of the work is that validation of the BAMS-7 may be constrained by the fact that the ADHD, Anxiety, and Depression cohorts were defined by self-reports of a clinical diagnosis. Date of diagnosis was not reported, nor was current status; it is worth noting, however, that studies of prevalence of clinical disorders, including ADHD, are typically based on self-report (Barbaresi et al., 2013; Faraone et al., 2006). In addition, because many people in the ADHD, Anxiety, and Depression cohorts may have been receiving medication or other treatment to reduce their symptoms, it may be surprising that the BAMS-7 subscales could successfully discriminate between the three disorders. Further use with more traditional cohorts would be helpful, and it is possible that the classification performance measures reported here underestimate the measures for untreated patients.

Overall, the pattern of results indicate that self-administered, online instruments with brevity and accessibility that measure multiple constructs are possible in the research landscape. Scales such as the BAMS-7 show the promise of utilizing large-scale datasets from online research for improving measurement and understanding of cognition and mood in the wild for social and clinical sciences.

VI. References

Arechar, A. A., & Rand, D. G. (2021). Turking in the time of COVID. *Behavior Research Methods*, *53*(6), 2591-2595.

Armstrong, M. J., & Okun, M. S. (2020). Diagnosis and treatment of Parkinson disease: A review. *JAMA*, *323*(6), 548-560.

Bar, M. (2009). A cognitive neuroscience hypothesis of mood and depression. *Trends in Cognitive Sciences*, *13*(11), 456-463.

Barbaresi, W. J., Colligan, R. C., Weaver, A. L., Voigt, R. G., Killian, J. M., & Katusic, S. K. (2013). Mortality, ADHD, and psychosocial adversity in adults with childhood ADHD: A prospective study. *Pediatrics*, *131*(4), 637-644.

Beavers, A. S., Lounsbury, J. W., Richards, J. K., Huck, S. W., Skolits, G. J., & Esquivel, S. L. (2013). Practical considerations for using exploratory factor analysis in educational research. *Practical Assessment, Research, and Evaluation*, *18*(1), 6.

Bridger, R. S., Johnsen, S. Å. K., & Brasher, K. (2013). Psychometric properties of the cognitive failures questionnaire. *Ergonomics*, *56*(10), 1515-1524.

Broadbent, D. E., Cooper, P. F., FitzGerald, P., & Parkes, K. R. (1982). The cognitive failures questionnaire (CFQ) and its correlates. *British Journal of Clinical Psychology*, *21*(1), 1-16.

Brysbaert, M. (2024). Designing and evaluating tasks to measure individual differences in experimental psychology: A tutorial. *Cognitive Research: Principles and Implications*, *9*, 11.

Carriere, J. S., Cheyne, J. A., & Smilek, D. (2008). Everyday attention lapses and memory failures: The affective consequences of mindlessness. *Consciousness and Cognition*, *17*(3), 835-847.

Chen, C., Hu, Z., Jiang, Z., & Zhou, F. (2018). Prevalence of anxiety in patients with mild cognitive impairment: A systematic review and meta-analysis. *Journal of Affective Disorders*, *236*, 211-221.

Cheyne, J. A., Carriere, J. S., & Smilek, D. (2006). Absent-mindedness: Lapses of conscious awareness and everyday cognitive failures. *Consciousness and Cognition*, *15*(3), 578-592.

Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research, and Evaluation*, *10*(1), 7.

Cyr, A. A., & Anderson, N. D. (2019). Effects of question framing on self-reported memory concerns across the lifespan. *Experimental Aging Research*, *45*(1), 1-9.

D'Agati, E., Curatolo, P., & Mazzone, L. (2019). Comorbidity between ADHD and anxiety disorders across the lifespan. *International Journal of Psychiatry in Clinical Practice*, *23*(4), 238-244.

de Winter, J. C., Dodou, D., & Hancock, P. A. (2015). On the paradoxical decrease of self-reported cognitive failures with age. *Ergonomics*, *58*(9), 1471-1486.

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), 1-13.

Esterman, M., & Rothlein, D. (2019). Models of sustained attention. *Current Opinion in Psychology, 29*, 174-180.

Eyre, H., Baune, B., & Lavretsky, H. (2015). Clinical advances in geriatric psychiatry: A focus on prevention of mood and cognitive disorders. *Psychiatric Clinics*, *38*(3), 495-514.

Faraone, S. V., Biederman, J., & Mick, E. (2006). The age-dependent decline of attention deficit hyperactivity disorder: A meta-analysis of follow-up studies. *Psychological Medicine*, *36*(2), 159-165.

Fast, L., Temuulen, U., Villringer, K., Kufner, A., Ali, H. F., Siebert, E., ... & Ritter, K. (2023). Machine learning-based prediction of clinical outcomes after first-ever ischemic stroke. *Frontiers in Neurology*, *14*, 1114360.

Fernandes, L., & Wang, H. (2018). Mood and cognition in old age. *Frontiers in Aging Neuroscience*, *10*, 286.

Franklin, M. S., Mrazek, M. D., Anderson, C. L., Johnston, C., Smallwood, J., Kingstone, A., & Schooler, J. W. (2017). Tracking distraction: The relationship between mind-wandering, meta-awareness, and ADHD symptomatology. *Journal of Attention Disorders*, *21*(6), 475-486.

Gkintoni, E., & Ortiz, P. S. (2023). Neuropsychology of generalized anxiety disorder in clinical setting: A systematic evaluation. *Healthcare*, *11*(17), 2446.

Goodman, J. K., & Wright, S. (2022). MTurk and online panel research: The impact of COVID-19, bots, TikTok, and other contemporary developments. In C. Lamberton, D. Rucker, & S. A. Spiller (Eds.), *The cambridge handbook of consumer psychology* (pp. 1-34). Cambridge University Press.

Gulpers, B., Ramakers, I., Hamel, R., Köhler, S., Voshaar, R. O., & Verhey, F. (2016). Anxiety as a predictor for cognitive decline and dementia: A systematic review and meta-analysis. *The American Journal of Geriatric Psychiatry*, *24*(10), 823-842.

Guvendir, M. A., & Ozkan, Y. O. (2022). Item removal strategies conducted in exploratory factor analysis: A comparative study. *International Journal of Assessment Tools in Education*, *9*(1), 165-180.

Hardy, J. L., Nelson, R. A., Thomason, M. E., Sternberg, D. A., Katovich, K., Farzin, F., & Scanlon, M. (2015). Enhancing cognitive abilities with comprehensive

training: A large, online, randomized, active-controlled trial. *PLoS One*, *10*(9), e0134467.

Hobbiss, M. H., Fairnie, J., Jafari, K., & Lavie, N. (2019). Attention, mindwandering, and mood. *Consciousness and Cognition*, *72*, 1-18.

Irrmischer, M., van der Wal, C. N., Mansvelder, H. D., & Linkenkaer-Hansen, K. (2018). Negative mood and mind wandering increase long-range temporal correlations in attention fluctuations. *PLoS One*, *13*(5), e0196907.

Ismail, Z., Elbayoumi, H., Fischer, C. E., Hogan, D. B., Millikin, C. P., Schweizer, T., ... & Fiest, K. M. (2017). Prevalence of depression in patients with mild cognitive impairment: A systematic review and meta-analysis. *JAMA Psychiatry*, *74*(1), 58-67.

Jonkman, L. M., Markus, C. R., Franklin, M. S., & van Dalfsen, J. H. (2017). Mind wandering during attention performance: Effects of ADHD-inattention symptomatology, negative mood, ruminative response style and working memory capacity. *PLoS One*, *12*(7), e0181213.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, *20*(1), 141-151.

Keller, A. S., Leikauf, J. E., Holt-Gosselin, B., Staveland, B. R., & Williams, L. M. (2019). Paying attention to attention in depression. *Translational Psychiatry*, *9*(1), 279.

Keshavan, M. S., Vinogradov, S., Rumsey, J., Sherrill, J., & Wagner, A. (2014). Cognitive training in mental disorders: Update and future directions. *American Journal of Psychiatry*, *171*(5), 510-522.

Kessler, R. C., Adler, L., Ames, M., Demler, O., Faraone, S., Hiripi, E. V. A., ... & Walters, E. E. (2005). The World Health Organization Adult ADHD Self-Report Scale (ASRS): A short screening scale for use in the general population. *Psychological Medicine*, *35*(2), 245-256.

Knekta, E., Runyon, C., & Eddy, S. (2019). One size doesn't fit all: Using factor analysis to gather validity evidence when using surveys in your research. *CBE—Life Sciences Education*, *18*(1), rm1.

Koster, E. H., Hoorelbeke, K., Onraedt, T., Owens, M., & Derakshan, N. (2017). Cognitive control interventions for depression: A systematic review of findings from training studies. *Clinical Psychology Review*, *53*, 79-92.

Kroenke K, Spitzer R L, & Williams J B (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, *16*(9), 606-613.

Lourenco, S. F., & Tasimi, A. (2020). No participant left behind: Conducting science during COVID-19. *Trends in Cognitive Sciences*, *24*(8), 583-584.

Mather, M., & Carstensen, L. L. (2005). Aging and motivated cognition: The positivity effect in attention and memory. *Trends in Cognitive Sciences*, *9*(10), 496-502.

McConville, C., & Cooper, C. (1997). The temporal stability of mood variability. *Personality and Individual Differences*, *23*(1), 161-164.

Ng, N. F., Osman, A. M., Kerlan, K. R., Doraiswamy, P. M., & Schafer, R. J. (2021). Computerized cognitive training by healthy older and younger adults: Age comparisons of overall efficacy and selective effects on cognition. *Frontiers in Neurology*, *11*, 564317.

Rast, P., Zimprich, D., Van Boxtel, M., & Jolles, J. (2009). Factor structure and measurement invariance of the cognitive failures questionnaire across the adult life span. *Assessment*, *16*(2), 145-158.

Retz, W., Stieglitz, R. D., Corbisiero, S., Retz-Junginger, P., & Rösler, M. (2012). Emotional dysregulation in adult ADHD: What is the empirical evidence? *Expert Review of Neurotherapeutics*, *12*(10), 1241-1251.

Reynolds, G. O., Willment, K., & Gale, S. A. (2021). Mindfulness and cognitive training interventions in mild cognitive impairment: Impact on cognition and mood. *The American Journal of Medicine*, *134*(4), 444-455.

Santos, N. C., Costa, P. S., Cunha, P., Portugal-Nunes, C., Amorim, L., Cotter, J., ... & Sousa, N. (2014). Clinical, physical and lifestyle variables and relationship with cognition and mood in aging: A cross-sectional analysis of distinct educational groups. *Frontiers in Aging Neuroscience*, *6*, 21.

Saragih, I. D., Tonapa, S. I., Porta, C. M., & Lee, B. O. (2022). Effects of telehealth intervention for people with dementia and their carers: A systematic review and meta-analysis of randomized controlled studies. *Journal of Nursing Scholarship*, *54*(6), 704-719.

Sarris, J., Thomson, R., Hargraves, F., Eaton, M., de Manincor, M., Veronese, N., ... & Firth, J. (2020). Multiple lifestyle factors and depressed mood: A cross-sectional and longitudinal analysis of the UK Biobank (N= 84,860). *BMC Medicine*, *18*, 1-10.

Skirrow, C., McLoughlin, G., Kuntsi, J., & Asherson, P. (2009). Behavioral, neurocognitive and treatment overlap between attention-deficit/hyperactivity disorder and mood instability. *Expert Review of Neurotherapeutics*, *9*(4), 489-503.

Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). Generalized Anxiety Disorder 7 (GAD-7). APA PsycTests.

Tassoni, M. B., Drabick, D. A., & Giovannetti, T. (2022). The frequency of self-reported memory failures is influenced by everyday context across the lifespan: Implications for neuropsychology research and clinical practice. *The Clinical Neuropsychologist*, *37*(6), 1115-1135.

Tomaszewski Farias, S., De Leon, F. S., Gavett, B. E., Fletcher, E., Meyer, O. L., Whitmer, R. A., ... & Mungas, D. (2024). Associations between personality and psychological characteristics and cognitive outcomes among older adults. *Psychology and Aging*. Advance online publication.

van Gool, C. H., Kempen, G. I., Bosma, H., van Boxtel, M. P., Jolles, J., & van Eijk, J. T. (2007). Associations between lifestyle and depressed mood: Longitudinal results from the Maastricht Aging Study. American *Journal of Public Health*, *97*(5), 887-894.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063.

Williams, L. M. (2016). Precision psychiatry: A neural circuit taxonomy for depression and anxiety. *The Lancet Psychiatry*, *3*(5), 472-480.

Wright, D. B., London, K., & Field, A. P. (2011). Using bootstrap estimation and the plug-in principle for clinical psychology data. *Journal of Experimental Psychopathology*, *2*(2), 252-270.

Yapici-Eser, H., Yalcinay-Inan, M., Kucuker, M. U., Kilciksiz, C. M., Yilmaz, S., Dincer, N., ... & Aydemir, O. (2021). Subjective cognitive assessments and N-back are not correlated, and they are differentially affected by anxiety and depression. *Applied Neuropsychology: Adult*, 1-11.

Yates, J. A., Clare, L., & Woods, R. T. (2013). Mild cognitive impairment and mood: A systematic review. *Reviews in Clinical Gerontology*, *23*(4), 317-356.

Yurgelun-Todd, D. (2007). Emotional and cognitive changes during adolescence. *Current Opinion in Neurobiology, 17*(2), 251-257.

Zanesco, A. P., Denkova, E., & Jha, A. P. (2022). Examining long-range temporal dependence in experience sampling reports of mind wandering. *Computational Brain & Behavior*, *5*(2), 217-233.

VII. Additional Notes

Data and code availability: The de-identified data and code will be made available.

Acknowledgments: Lumos Labs, Inc. developed the cognitive training platform (Lumosity) and measures (BAMS-7) used in this study, as well as funded the study through the employment of KPM, AMO, KRK, and RJS. Other members of the company contributed suggestions and ideas during the design of the study and preparation of the manuscript.

Author contributions: KPM: Conceptualization, Data Curation, Methodology, Formal Analysis, Writing – Original Draft Preparation, Writing – Review and Editing; AMO: Conceptualization, Methodology, Writing – Review and Editing; KRK: Conceptualization, Methodology, Visualization, Writing – Review and Editing; RJS: Conceptualization, Data Curation, Methodology, Formal Analysis, Writing – Original Draft Preparation, Writing – Review and Editing.

Competing interests: This study was supported entirely by Lumos Labs, Inc. At the time of the study, KPM, AMO, KRK, and RJS were paid employees of the company, and all hold stock in the company.

VIII. Supplemental Materials

Experiment 2

MTurk Design for Concordance

We intended to recruit up to 200 individuals ages 18 and older who resided in the United States for a 7 min online Amazon MTurk HIT (human intelligence task) of "Cognition and Emotion" that paid $1.00. Various attention checks in the task reduced the sample from 200 to 150 participants for analysis. The attention checks involved identifying response inconsistencies, random clicking, too little or too long spent on HIT, and bots. One of the attention checks involved having participants re-rate five items from across the different questionnaires; participants were excluded if their average response mismatch was 1-point or more. After demographics, questionnaires included the ASRS, ARCES, Hardy survey (with emphasis on BAMS-7 items), PHQ-9, PANAS, and GAD-7.

Questionnaire Descriptions for Concordance

Adult ADHD Self-Report Scale (ASRS; Kessler et al., 2005). The ASRS symptom checklist from the World Health Organization measures for probable ADHD in adults as well as ADHD symptoms. The checklist asks respondents to indicate how they have felt and conducted themselves over the past 6 months in terms of frequency of inattention and hyperactivity/impulsivity symptoms.

Attention-Related Cognitive Errors Scale (ARCES; Cheyne et al., 2006). The ARCES measures the frequency of cognitive errors in everyday situations that are attributed to attention lapsing. It is well-validated and reliable for remote assessment in populations across the lifespan (Carriere et al., 2008, 2013; Cheyne et al., 2006). Score on the ARCES is related to self-reported cognitive and clinical outcomes from independent questionnaires, including memory failures, boredom, fidgeting, mind wandering, daydreaming, media multitasking, lack of attentional control, and symptoms of depression and ADHD (Franklin et al., 2017; Jonkman et al., 2017; Ralph et al., 2014). Score on the ARCES is related to task-based continuous performance test (CPT) commission errors (Rosenberg et al., 2013), and psychometric work has shown that it is also separable from that of the CFQ (Smilek et al., 2010).

Patient Health Questionnaire (PHQ-9; Kroenke et al., 2001). The PHQ-9 measures for probable major depressive episodes as well as depressive symptom severity. Each item represents one of the diagnostic criteria for major depressive episodes. The PHQ-9 asks participants to report the presence of each symptom within the last 2 weeks.

Positive and Negative Affect Schedule (PANAS; Watson et al., 1988). The PANAS measures positive and negative affect along various emotion and mood dimensions over various time-sensitive intervals, including over the past few weeks. The positive affect score and negative affect score are separable. It is well-validated and reliable for clinical and social sciences.

Generalized Anxiety Disorder (GAD-7; Spitzer et al., 2006). The GAD-7 measures

for probable generalized anxiety disorder as well as anxiety symptom severity. The GAD-7 asks participants to report the presence of each symptom within the last 2 weeks.

MTurk Results for Concordance

***Supplementary Table 1**. Questionnaire correlation matrix of the BAMS-7 Attention and Mood subscales with the 5 existing instruments (ASRS, ARCES, PHQ-9, PANAS, and GAD-7) in Experiment 2.*

	BAMS-7 Attention subscale	**BAMS-7 Mood subscale**	**ASRS Part A**	**ASRS Part B**	**ASRS Total (A+B)**	**ARCES**	**PHQ-9**	**PANAS Pos Aff**	**PANAS Neg Aff**	**GAD-7**
BAMS-7 Attention subscale	-									
BAMS-7 Mood subscale	0.685***	-								
ASRS Part A	-0.750***	-0.659***	-							
ASRS Part B	-0.609***	-0.570***	0.777***	-						
ASRS Total (A+B)	-0.720***	-0.652***	0.942***	0.943***	-					
ARCES	-0.788***	-0.644***	0.834***	0.715***	0.821***	-				
PHQ-9	-0.749***	-0.770***	0.708***	0.684***	0.738***	0.718***	-			
PANAS Pos Aff	0.265**	0.320***	-0.118	0.144△	0.014	-0.069	-0.178*	-		
PANAS Neg Aff	-0.685***	-0.780***	0.652***	0.662***	0.697***	0.623***	0.847***	-0.190*	-	
GAD-7	-0.755***	-0.765***	0.702***	0.674***	0.730***	0.710***	0.916***	-0.158△	0.833***	-

*Note. *p<.05, **p<.01, ***p<.001, △p<.10.*

***Supplementary Table 2**. Item-level concordance analysis with correlation matrix of each BAMS-7 item and related items from the 5 existing questionnaires (ASRS, ARCES, PHQ-9, PANAS, and GAD-7) in Experiment 2.*

BAMS-7	ASRS	ARCES	PANAS	PHQ-9	GAD-7
During the last month, how often have you…	**Over the last 6 months…**	**How frequently have these sorts of things happened to you…**	**Over the past few weeks…**	**Over the last few weeks…**	**Over the last few weeks…**
Lost track of details as you were reading and needed to go back and reread sections?	-	When reading I find that I have read several paragraphs without being able to recall what I read. *r* = -0.623	-	-	-
Misplaced items (e.g., reading glasses, keys) around the house?	How often do you misplace or have difficulty finding things at home or at work? *r* = -0.575	I have absent-mindedly misplaced frequently used objects, such as keys, pens, glasses, etc. *r* = -0.570	-	-	-
Found yourself losing concentration during a conversation?	How often do you have difficulty concentrating on what people say to you, even when they are speaking to you directly? *r* = -0.558	I have lost track of a conversation because I zoned out when someone else was talking. *r* = -0.676	-	-	-
Rate your experience over the last week…					
My ability to concentrate was good.	-	-	Attentive r = 0.598 Alert r = 0.454	Trouble concentrating on things, such as reading the newspaper or watching television. *r* = -0.403	-

I felt anxious.	-	-	Nervous $r = 0.711$	-	Feeling nervous, anxious, or on edge. $r = -0.642$
I was in a bad mood.	-	-	Irritable $r = -0.578$	Feeling bad about yourself - or that you are a failure or have let yourself or your family down. $r = -0.470$	-
I felt sad for no obvious reason.	-	-	Upset $r = -0.627$	Feeling down, depressed, or hopeless. $r = -0.665$	-

Note. All p's<.0001.

Experiment 4

ANCOVA Results with Covariates of Baseline Score and Age

In line with the original ANCOVA analysis (Hardy et al. , 2015), the covariate of baseline BAMS-7 on intervention effects was significant. Participants who had lower pre-intervention Attention and Mood scores exhibited greater post-intervention improvements on the Attention subscale in both Lumosity and Crosswords ($F(1,3485)=1277.94$, $p<.001$) and Mood subscale ($F(1,3485)=1474.06$, $p<.001$). There was also a significant effect of the covariate of age on the Mood subscale ($F(1,3485)=26.57$, $p<.001$) but not the Attention subscale $F(1,3485)=0.27$, $p=.606$), in Lumosity and Crosswords. Participants showed greater improvements on the Mood subscale across interventions with increasing age.

Additional References in Supplemental Materials

Carriere, J. S., Seli, P., & Smilek, D. (2013). Wandering in both mind and body: Individual differences in mind wandering and inattention predict fidgeting. *Canadian Journal of Experimental Psychology*, *67*(1), 19-31.

Ralph, B. C., Thomson, D. R., Cheyne, J. A., & Smilek, D. (2014). Media multitasking and failures of attention in everyday life. *Psychological Research*, *78*(5), 661-669.

Rosenberg, M., Noonan, S., DeGutis, J., & Esterman, M. (2013). Sustaining visual attention in the face of distraction: A novel gradual-onset continuous performance task. *Attention, Perception, & Psychophysics*, *75*(3), 426-439.

Smilek, D., Carriere, J. S., & Cheyne, J. A. (2010). Failures of sustained attention in life, lab, and brain: Ecological validity of the SART. *Neuropsychologia*, *48*(9), 2564-2570.