INTERNATIONAL JOURNAL OF -EXERCISE SCIENCE-

Original Research

The Psychometric Performance of State Mindfulness Scales Around Sitting and Walking on Desert Trails: A Pilot Study

Dustin W. Davis†[1], Elias M. Malek*[1], Rob Salatto†[2], Marcus M. Lawrence‡[3], Jacob W. Manning‡[3], Mark DeBeliso‡[3], Merrill R. Landers‡[4], Graham R. McGinnis‡[1], James W. Navalta‡[1]

[1]Department of Kinesiology and Nutrition Sciences, University of Nevada, Las Vegas, Las Vegas, NV, USA; [2]Department of Kinesiology, Vanguard University, Costa Mesa, CA, USA; [3]Department of Kinesiology and Outdoor Recreation, Southern Utah University, Cedar City, UT, USA; [4]Department of Physical Therapy, University of Nevada, Las Vegas, Las Vegas, NV, USA

*Denotes student, †Denotes early-career investigator, ‡Denotes established investigator

ABSTRACT

International Journal of Exercise Science 18(4): 27-42, 2025. https://doi.org/10.70252/PCSB3043 State mindfulness is a dynamic construct reflecting current mindfulness, or purposeful attention to the present moment with openness, non-judgment, non-reactivity, and acceptance. Despite its popularity and research in psychology and medicine, measuring state mindfulness remains a challenge. Validity and reliability of the State Mindfulness Scale (SMS) and State Mindfulness Scale for Physical Activity (SMS-PA) in applied settings like nature require more data to be established. This pilot study aimed to evaluate these scales and introduce a new, one-item scale, the Visual Analog Scale-Mindfulness (VAS-M), for quicker mindfulness assessment. Participants completed the SMS, SMS-PA, and VAS-M upon arrival at a desert trail, after 10 minutes of sitting, and after 10 minutes of walking. A subset repeated the intervention 24 hours later. The study found that the SMS and SMS-PA could be given before, during, and after nature immersion. The SMS-PA showed evidence of concurrent validity with the SMS ($\rho = .89$, $p < .001$). However, the VAS-M did not meet the criteria for concurrent validity with the other scales. None of the scales met the criteria for test-retest reliability. This study underscores the need to validate state mindfulness scales, especially in non-laboratory settings. While the SMS-PA shows promise, further validation and refinement of the VAS-M are needed. The findings enhance our understanding of measuring state mindfulness, particularly in natural environments. While the scales are already in use, our study helps address gaps in evidence and informs future decisions about their application.

Keywords: Outdoors, forest bathing, green exercise, human-nature relationship

Introduction

Mindfulness is both the ability to and the practice of purposefully bringing one's attention to the present, moment-to-moment experience with openness, non-judgment, non-reactivity, and acceptance.[1,2] Mindfulness came from Buddhist philosophy and has been used by Western psychologists and medical doctors since the mid-20th century.[2–4] Since then, researchers and

clinicians have widely adopted mindfulness-based interventions, with systematic reviews reporting on their mental benefits.[5–7] Mindfulness is not just an intervention but also an outcome, measured as either a trait or state construct[8]. Trait mindfulness is thought to be intrinsic and reflect personality traits and a person's disposition toward mindfulness,[8,9] usually assessed by self-report questionnaires such as the Five Facet Mindfulness Questionnaire.[10] In contrast to trait mindfulness, state mindfulness is thought to be dynamic and dependent on context, meaning it fluctuates across time and place because of internal and external factors such as emotions and environments.[1,11,12] Like trait mindfulness, state mindfulness is usually assessed by self-report questionnaires,[11–15] some of which are the focus of the present study.

Assessing changes in state mindfulness across activities is crucial for finding suitable candidates for mindfulness-based interventions. Ensuring that measured changes in state mindfulness are real and meaningful requires scales with strong psychometric properties, including validity and reliability. Establishing such properties enables the confident assessment and interpretation of changes in state mindfulness, aiding the development of effective mindfulness-based interventions. Despite progress in scale development, researchers and clinicians need more evidence to validate and establish the reliability of scales measuring state mindfulness before, during, and after acute interventions. Among these scales, the State Mindfulness Scale (SMS) and State Mindfulness Scale for Physical Activity (SMS-PA) show promise.[11–15] The SMS has 21 Likert-type items that assess total state mindfulness, mindfulness of the mind, and mindfulness of the body.[11,12] The SMS-PA is a shorter, 12-item version of the SMS designed to assess the physical aspect of state mindfulness (total, mind, and body) right after physical activity.[13]

Published research has focused on validating the SMS and SMS-PA for practical research applications. Tanay and Bernstein introduced the SMS in 2013, showing its internal consistency, convergent, discriminant, and incremental validity compared to other mindfulness measures, as well as its context-specific stability and sensitivity to change in adults in Israel.[12] Andrade et al. later reported evidence for the validity of the Portuguese version in Portuguese adults.[16] In 2023, Ruimi et al. compiled the existing research on the SMS, which largely supports the scale's validity.[11] The SMS-PA, introduced by Cox et al. in 2016, showed internal consistency and construct validity in U.S. adults, with significant correlations between SMS-PA items and concepts related to mindfulness.[13] Cox et al. noted significant differences in SMS-PA scores based on two key factors: the quality of the physical activity and participants' proficiency in mindful exercise.[13] Regarding quality, activities inherently conducive to mindfulness, such as yoga, yielded higher SMS-PA scores than activities not explicitly designed for mindfulness, such as sport, aerobic exercise, and strength training. Additionally, participants with higher skill levels in mindful exercise had higher SMS-PA scores than those with lower skill levels.[13] Since this seminal study, authors have reported evidence for the validity of the SMS-PA in U.S. middle-school students,[17] a Spanish version of the SMS-PA used with Spanish youth,[15] and a Brazilian Portuguese version used with Brazilian adults who practice physical activities and sports.[14,18,19]

Though validation efforts are ongoing and definitive evidence of validity and reliability is not yet available, some researchers have used the SMS-PA to measure state mindfulness before, during, and after interventions. Two years after introducing the SMS-PA, Cox et al. used it to

assess U.S. adults' psychological responses to both traditional and mindful treadmill walking.[20] Other researchers have used the scale to examine changes in state mindfulness before and after coastal walking[21] and bouldering.[22] A 2023 study protocol for a randomized controlled trial on bouldering psychotherapy and mental model therapy proposes the use of the SMS-PA.[23] And a 2024 cross-sectional study reported differences in state mindfulness between elite speed skaters and university students of physical education, using the SMS-PA.[24]

Analyzing data from the SMS-PA and its collection context can offer valuable insights to researchers assessing state mindfulness. This is beneficial for researchers working in practical, real-world settings beyond controlled laboratory environments, such as outdoor natural environments. At the onset of the present study (late 2021), the authors knew of no studies to have tested the psychometric performance of the SMS or SMS-PA before, during, and after an acute nature immersion involving physical activity on desert trails. It was also unclear whether anyone had investigated the psychometric performance of the scales through repeated measures in quick succession, how the SMS-PA scores compared to the original SMS scores in such a setting, or if there was evidence of the scales' test-retest reliability across approximately 24 hours. Filling these gaps is crucial as it would help researchers know whether the scales are suitable for measuring state mindfulness around nature immersion, an increasingly popular intervention to improve mindfulness and mental health.[25–27] Given that completing multi-item scales repeatedly can be burdensome, the authors of the present study created the one-item Visual Analog Scale-Mindfulness (VAS-M) as a quicker alternative to the SMS and SMS-PA, aiming to mitigate response burden. The VAS-M introduced here needed to be tested for validity and reliability.

The present study was a pilot study aimed to evaluate the psychometric performance of the SMS, SMS-PA, and VAS-M when administered to participants three times within a 30-minute period on desert trails. Participants completed the SMS and VAS-M upon arrival at the trailhead, after 10 minutes of sitting nearby, and after 10 minutes of walking on the trail. Participants completed the SMS-PA only after walking. The authors hypothesized evidence of concurrent validity between the SMS-PA and SMS post-walk, as well as between the VAS-M and SMS at baseline (pre-sit), post-sit, and post-walk. The authors also hypothesized evidence of concurrent validity between the VAS-M and SMS-PA post-walk. For a subsample completing the same intervention on two consecutive days, 24 hours apart, the authors hypothesized good test-retest reliability for all three scales. The authors hypothesized nothing about changes in state mindfulness resulting from sitting and walking because of the study's uncontrolled nature. Changes in scores over time were analyzed and reported only to understand the psychometric performance of the scales and to see if the VAS-M scores changed in the same direction as the SMS scores across the three assessments (pre-sit, post-sit, and post-walk).

Methods

Participants

Since the primary aim of the present study was not to assess possible significant effects of the intervention on state mindfulness, a traditional power analysis was not conducted to estimate a

sample size based on an α level of .05 and a β of .80. Large samples are usually needed for robust evaluations of psychometric performance, but the present study was part of a doctoral dissertation under time and recruitment constraints. The team recruited as many participants as possible within the constraints, and the sample size met or exceeded that of typical studies on outdoor walking.[21,28,29] No financial incentives were offered.

The present study's convenience sample included two subsamples drawn from the Thunderbird Gardens Trailhead (n = 19) and the Clark County Wetlands Park (n = 23). These sites were chosen to provide nature immersion with physical activity in natural environments, away from built-up landscapes like cityscapes. Both sites meet the criteria for green spaces as defined in the green exercise literature, where exposure to nature is fundamental to the activity.[30] The desert landscapes chosen for the present study qualify as green spaces despite lacking the greenery one might expect from the term "green exercise," because the key requirement is exposure to natural environments, not necessarily those dominated by green foliage.[29] Internal validity was preserved and substantial confounding was avoided because the subsamples shared identical target populations and inclusion criteria, as well as similar baseline age, height, weight, and body mass index (BMI in $kg \cdot m^{-2}$). All participants gave both verbal and written informed consent. The inclusion criterion was an age range of 18 to 64 years,[12] chosen to align with the participants' age range in the paper that introduced the SMS.[12] People of any sex, gender identity, race, ethnicity, height, mass, and BMI were eligible to participate. Exclusion criteria included being under 18 years or over 64 years of age or being recommended to obtain medical clearance before engaging in exercise, as determined by the American College of Sports Medicine Preparticipation Health Screening Questionnaire for Exercise Professionals.[31] Additionally, people who were pregnant, lactating, or had conditions that could affect their ability to walk safely, as well as those with self-reported intellectual disabilities, were excluded from the study.

The present study was carried out fully in accordance with the ethical standards of the *International Journal of Exercise Science*[32] and the Declaration of Helsinki. Approval for the study was obtained from the respective Institutional Review Boards of the University of Nevada, Las Vegas, and Southern Utah University. The study did not collect data on participants' gender identity but supplied options for self-reporting sex as female, male, or intersex. Participants also self-reported race or ethnicity, and the categories (African American or Black, Asian, Caucasian or White, Hispanic or Latino, Mediterranean, Middle Eastern, Multi-Racial, and Polynesian) were based on participants' written responses without a predefined list. Age and height were also self-reported. Mass, measured clothed but without shoes or items in pockets, was measured using the Tanita TBF-521 Bodyfat Monitor/Scale (Tanita, Tokyo, Japan). The sample was 52% female, 48% male, and 0% intersex. Racial distribution included 10% African American or Black, 10% Asian, 45% Caucasian or White, 21% Hispanic or Latino, 2% Mediterranean, 2% Middle Eastern, 7% Multi-Racial, and 2% Polynesian. The mean age, height, mass, and BMI were 26.0 ± 9.0 years, 169.7 ± 8.7 centimeters, 69.6 ± 15.9 kg, and 23.9 ± 4.5 $kg \cdot m^{-2}$, respectively.

Protocol

Data were collected over two-day periods at two study sites (Figure 1). The first period occurred between approximately 0800 and 1700 on October 1–2, 2021, at the Thunderbird Gardens Trailhead in Cedar City, UT, and the second period occurred between approximately 0800 and 1700 on November 6–7, 2021, at the Clark County Wetlands Park in Las Vegas, NV. Two sites were used to increase recruitment and generalizability, allowing for a more diverse sample and for the findings to apply to different desert trails. The air temperature, relative humidity, and wind speed at each site were recorded with the Kestrel 4400 Heat Stress Tracker Wind/Altitude Meter (Kestrel Instruments®, Nielsen-Kellerman Co., Boothwyn, PA, USA). These variables and descriptions of each site's landscape were reported elsewhere.[33]

Figure 1. Photos of the Thunderbird Gardens Trailhead and linked trails (top) and the Clark County Wetlands Park (bottom).

The three mindfulness scales used in the present study were the State Mindfulness Scale (SMS), State Mindfulness Scale for Physical Activity (SMS-PA), and Visual Analog Scale-Mindfulness (VAS-M).

The SMS is a 21-item Likert-type scale designed to measure total state mindfulness (SMS-Total), State Mindfulness of Mind (SMS-Mind from items 1–7, 10–12, 15–17, and 19–20), and State Mindfulness of Body (SMS-Body from items 8–9, 13–14, 18, and 21).[11,12] The SMS has demonstrated internal consistency, with Cronbach's α values ranging from .85 to .97.[11,12] In the present study, participants rated how well each SMS item described their experiences in the past 10 minutes, using a scale from 1 (not at all) to 5 (very well). All 21 items were summed to obtain

the SMS-Total score, with higher scores indicating greater state mindfulness. The SMS-Mind and SMS-Body subscales were summed separately, with higher scores similarly indicating greater mindfulness in each factor.

Building on Tanay and Bernstein's SMS,[11,12] Cox et al. introduced the SMS-PA, a 12-item Likert-type scale designed to capture mindfulness during physical activity.[13] The SMS-PA provides three scores: SMS-PA-Total, SMS-PA-Mind (items 1–6), and SMS-PA-Body (items 7–12) and has demonstrated internal consistency, with Cronbach's α values ranging from .87 to .93.[13] In the present study, participants rated how well each item described their experiences in the past 10 minutes while walking, using a scale ranging from 1 (not at all) to 4 (very much), a deviation from the original 0 to 4 SMS-PA to match the SMS. All 12 items were summed to obtain the SMS-PA-Total score, with higher scores indicating greater state mindfulness. The SMS-PA-Mind and SMS-PA-Body subscales were summed separately, with higher scores similarly indicating greater mindfulness in each factor.

The third scale used in the present study was the VAS-M, a single-item scale for measuring total state mindfulness. Participants read the statement, "From moment-to-moment, I noticed and accepted my thoughts, feelings, bodily sensations, and environment without judging them," before marking on a 100-millimeter (mm) horizontal, ungraduated line how well the statement described their experience over the last 10 minutes. The line ranged from "not at all like my experience" on the left anchor (0 mm) to "exactly like my experience" on the right anchor (100 mm). Participants marked their response as a vertical dash, and the score was measured with a standard, 12-inch ruler as the distance in mm from the left anchor to the marked dash. Higher scores were considered evidence of a higher level of state mindfulness.

After consenting, undergoing screening, and supplying participant characteristics, each participant was instructed not to use their phone except for starting and stopping 10-minute timers for sitting and walking. Participants were not supplied with any information or instructions about mindfulness. Each participant received a binder holding several scales organized into three sets: pre-sit, post-sit, and post-walk. The pre-sit and post-sit sets included the SMS and VAS-M, while the post-walk set included the SMS, VAS-M, and SMS-PA. The order of the scales within each set for each participant was random.

Upon receiving the binder, participants at both the Thunderbird Gardens Trailhead and Clark County Wetlands Park completed the pre-sit set of scales. Participants then chose a secluded spot to sit alone for 10 minutes near the trailhead, but away from researchers and other participants. After sitting, participants completed the post-sit set before walking at a self-selected pace for 10 minutes along a trail, having been staggered and instructed not to walk together, near other people, or to talk to anyone. Researchers saw no instances of participants violating the protocol. After the walk, participants completed the post-walk set, returned their binders, and left. Participants at the Clark County Wetlands Park returned approximately 24 hours later to repeat the protocol, starting with the pre-sit set. This was done to assess for evidence of test-retest reliability across approximately 24 hours when the scales were administered to the same participants in the same environment at a similar time of day.

Researchers removed all the scales from day one from participants' binders before day two to prevent participants from reviewing their earlier ratings.

Statistical Analysis

Researchers used the correct parametric and non-parametric statistical analyses, along with post hoc tests, based on checking each test's statistical assumptions (IBM SPSS Statistics v28; IBM, Armonk, NY, United States). All analyses used an α level of .05, meaning significant findings had a *p*-value < .05. Each analysis included as many participants as possible, based on the availability of complete data for the dependent variable in that analysis. The smallest subsample sizes occurred in the SMS analyses due to a typographical error that resulted in the loss of SMS data for 11 participants at the Thunderbird Gardens Trailhead (Figure 2). Despite this loss, these participants were included in other analyses where their data were complete. The error was corrected from participant 12 onward at that site and for all participants at the Clark County Wetlands Park. All data are presented as sample arithmetic means ± sample SD, unless otherwise specified.

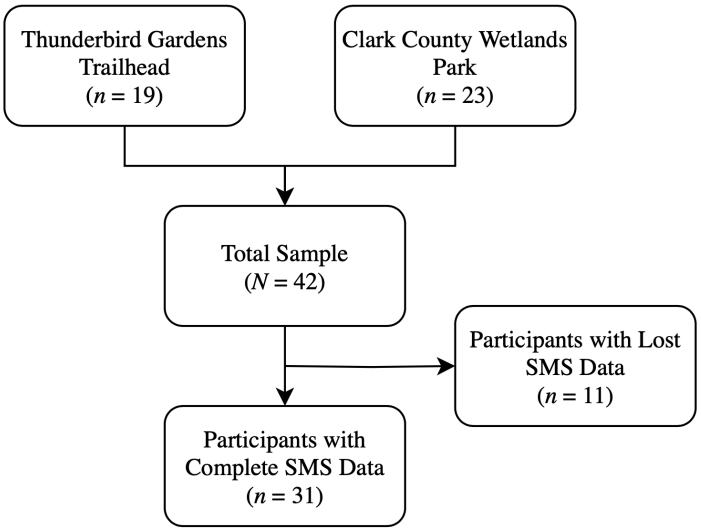

Figure 2. Flowchart of sample composition and loss of SMS data for analysis.

Two-tailed Pearson's correlation (*r*) was the planned test for assessing concurrent validity, but two-tailed Spearman's rank correlation (ρ) was used for comparisons between scores that did not meet the assumptions of *r*. The strength of correlations was interpreted as follows: ± 1.00 was a perfect correlation, ± 0.70 to ± 0.90 was a strong to very strong correlation, ± 0.50 to ± 0.70 was a moderate to strong correlation, ± 0.30 to ± 0.50 was a weak to moderate correlation, and ± 0.00 to ± 0.30 was no correlation to a weak correlation. The 95% confidence interval (CI) for correlations was estimated using Fisher's *r*-to-*z* transformation. For estimating the 95% CIs for Spearman's ρ, the standard error was estimated using Fieller, Hartley, and Pearson's formula in IBM SPSS Statistics v28.

In addition to the correlations, two 1 × 3 repeated measures ANOVAs were run to see if VAS-M scores changed similarly to SMS scores over time. Five outliers were found among the SMS and VAS-M scores but were kept as they did not seem to result from errors in data measurement or entry. Normality, homogeneity of variance, and sphericity checks were conducted on all scores. Despite some non-normality, the ANOVAs were performed, and the Greenhouse-Geisser method was used to adjust the degrees of freedom because of non-sphericity. Significant ANOVAs were followed by post hoc pairwise comparisons using Bonferroni-corrected paired-samples *t*-tests.

Our team defined evidence of concurrent validity between the SMS-PA and SMS post-walk as a significant positive correlation of at least moderate strength, as shown by the lower bound of the 95% CI, between scores on the two scales. The same criterion was applied when comparing the VAS-M and SMS-PA post-walk and the VAS-M and SMS pre-sit, post-sit, and post-walk. A second criterion was applied when comparing the VAS-M and SMS: Scores on both scales needed to change in the same direction from pre-sit to post-sit to post-walk, as shown by the two 1 × 3 repeated measures ANOVAs.

While the present study does not offer a hypothesis about changes in state mindfulness resulting from the intervention, it reports effect sizes and the minimal detectable change (MDC) based on the 95% CI, called the MDC_{95}. The effect sizes were used to interpret the magnitude of changes seen in the scores, and the MDC_{95} was used as the smallest change on a scale that can be considered a true change beyond the expected measurement error. This approach helped to contextualize the findings and understand the practical significance of any observed changes in state mindfulness scores, albeit understanding that, without a control group, no causal relationship can be determined.

The effect size for the ANOVAs was partial eta squared (η^2_p), which was converted to partial omega squared (ω^2_p) to avoid overestimating the population effect size, particularly with the small sample size.[34] The ω^2_p values were classified according to thresholds by Ferguson[35] and used in a mindfulness study that used the SMS-PA[20]: .04 for a "recommended minimum effect size representing a 'practically' significant effect (RMPE)," .25 for a moderate effect, and .64 for a strong effect. The effect size for post hoc *t*-tests was Cohen's *d*, with thresholds of 0.41 for RMPE, 1.15 for a moderate effect, and 2.70 for a strong effect, based on Ferguson's criteria.[35] Calculating the MDC_{95} for the SMS and VAS-M required each scale's standard error of measurement (SEM), which was derived from the pre-sit intraclass correlation coefficient (ICC) and the pooled pre-sit standard deviation (SD). The primary purpose of calculating ICCs was to obtain the SEM, using pre-sit scores from day one and day two at the Clark County Wetlands Park.

The secondary purpose for calculating ICCs was to evaluate the absolute test-retest reliability of the scales. Relative reliability was assessed using coefficients of variation (CVs). For the SMS and VAS-M, three ICCs and three CVs were calculated: one each for pre-sit, post-sit, and post-walk. Additionally, one ICC and one CV were calculated for the SMS-PA post-walk. The ICC model followed the guidance of Koo and Li[36]: two-way mixed effects, single measurement, with absolute agreement and interpretation based on the 95% CI. Thresholds for absolute reliability

were adapted from Koo and Li's[36] thresholds to avoid overlap: excellent (> .90), good (.76–.90), moderate (.50–.75), and poor (< .50). Thresholds for relative reliability were adapted from Aronhime et al.[37] to avoid overlap: excellent (≤ 10%), good (11–20%), acceptable (21–30%), and poor (> 30%). Our team defined evidence of good test-retest reliability for any of the three scales as having an ICC with the lower bound of the 95% CI at least .76 and a CV no greater than 20.00% at every assessment (pre-sit, post-sit, post-walk).

Results

Table 1 displays the correlations among the SMS, SMS-PA, and VAS-M, and Figure 3 illustrates the changes in SMS and VAS-M scores over time. The correlations supported the hypothesized evidence of concurrent validity between the SMS-PA and SMS post-walk. While the VAS-M scores changed in the same direction as the SMS scores across the three assessments, the correlations did not support the hypothesized evidence of concurrent validity between the VAS-M and SMS pre-sit, post-sit, and post-walk, nor between the VAS-M and SMS-PA post-walk.

Table 1. Correlations among the State Mindfulness Scale, State Mindfulness Scale for Physical Activity, and Visual Analog Scale-Mindfulness.

Comparison	Coefficient	95% CI	*p*-value	Direction	Strength	Met Criterion?
SMS-PA and SMS post-walk	ρ(31) = .89	[.78, .95]	< .001	Positive	Strong to Very Strong	✓
VAS-M and SMS pre-sit	*r*(31) = .58	[.30, .77]	< .001	Positive	Weak to Strong	
VAS-M and SMS post-sit	ρ(33) = .34	[−.003, .61]	.046*	n.s.	n.s.	
VAS-M and SMS post-walk	ρ(32) = .55	[.25, .75]	< .001	Positive	Weak to Strong	
VAS-M and SMS-PA post-walk	ρ(31) = .67	[.42, .83]	< .001	Positive	Weak to Very Strong	

CI: confidence interval; *p*-value: probability value; SMS-PA: State Mindfulness Scale for Physical Activity; SMS: State Mindfulness Scale; VAS-M: Visual Analog Scale-Mindfulness; *r*: Pearson's correlation coefficient; ρ: Spearman's rank correlation coefficient; n.s.: non-significant correlation. "Met Criterion" means that the scale met the study's criterion for concurrent validity, which was a significant positive correlation of at least moderate strength, as determined by the lower bound of the 95% confidence interval, between the scores on the two scales. *Despite having a *p*-value of .046, the correlation between VAS-M and SMS scores post-sit was considered non-significant because the 95% CI crossed the null value of 0.

The repeated measures ANOVA on the SMS scores (n = 31) was significant (Figure 3), $F(2, 49) = 54.35$, $p < .001$, $\omega^2_p = .53$ (a moderate effect). The mean SMS score significantly increased from 51.8 ± 19.5 arbitrary units (AU) pre-sit to 81.0 ± 15.1 AU post-sit (mean difference = 29.2 AU, 95% CI [20.1, 38.4], $p < .001$, $d = 1.68$, a moderate effect) and 86.0 ± 16.6 AU post-walk (mean difference = 34.2 AU [23.6, 44.7], $p < .001$, $d = 1.89$, a moderate effect). There was no significant change in the mean SMS score from post-sit to post-walk (mean difference = 4.9 AU [−1.9, 11.8], $p = .232$, $d = 0.31$, not a practically significant effect).

The repeated measures ANOVA on the VAS-M scores (n = 40) was significant (Figure 3), $F(1, 53) = 35.05$, $p < .001$, $\omega^2_p = .36$ (moderate effect). The mean VAS-M score significantly increased from 47.8 ± 28.4 mm pre-sit to 69.9 ± 21.6 mm post-sit (mean difference = 22.1 mm, 95% CI [12.8, 31.4],

$p < .001$, $d = 0.88$, a practically significant effect) and 77.7 ± 20.1 mm post-walk (mean difference = 29.9 mm [18.2, 41.7], $p < .001$, $d = 1.22$, a moderate effect). The mean VAS-M score significantly increased, but only slightly, from post-sit to post-walk (mean difference = 7.9 mm [2.0, 13.8], $p = .006$, $d = .38$, not a practically significant effect).

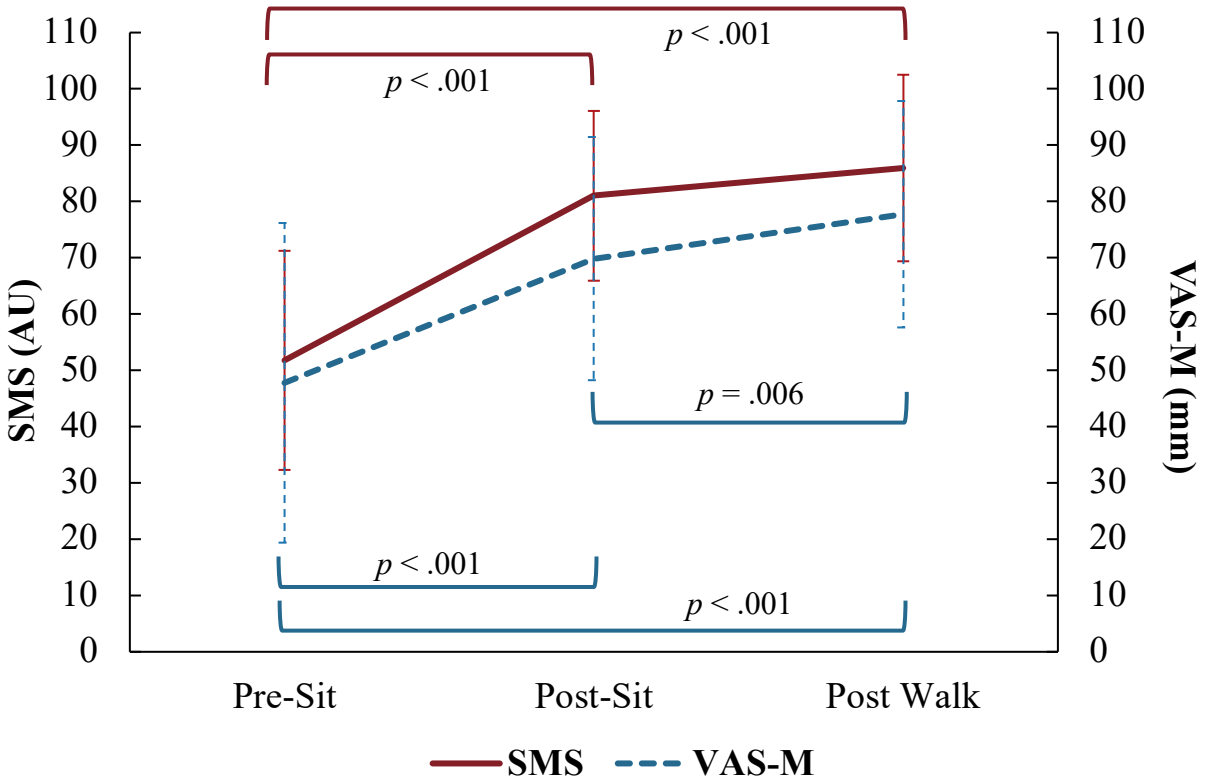

Figure 3. The line graph shows the outcomes of two 1 × 3 repeated measures ANOVAs conducted on the State Mindfulness Scale (SMS) and Visual Analog Scale-Mindfulness (VAS-M) scores, with sample sizes of 31 and 40, respectively. Solid red error bars show the standard deviations of SMS scores, while dashed blue error bars show the standard deviations of VAS-M scores. AU: arbitrary units; mm: millimeters; *p*: probability value.

The SEM and MDC_{95} of the SMS were 10.2 and 29, respectively. Out of 31 participants with scores both pre-sit and post-walk, none had a decrease beyond 29 AU. Thirteen participants (41.9%) showed no change, while 18 (58.1%) had an increase beyond 29 AU.

The results did not support the hypothesized test-retest reliability of the SMS, SMS-PA, and VAS-M (Table 2).

Discussion

Before the present pilot study, several gaps existed in the literature. First, there was a lack of studies testing the psychometric performance of the SMS or SMS-PA before, during, and after nature immersion involving physical activity, which is becoming popular as a mindfulness-based intervention for increasing mindfulness and mental health.[25–27] Second, evidence for the psychometric performance of the scales through repeated measures in quick succession was unclear. Third, there was a lack of comparison between scores on the SMS-PA, adapted from the

SMS, and scores on the original SMS after physical activity in nature. Fourth, uncertainty existed about the scales' test-retest reliability across approximately 24 hours. Finally, no single-item scale to measure state mindfulness was available or known to work as well as the SMS or SMS-PA. The new scale proposed here, the VAS-M, needed to be tested for validity and reliability.

Table 2. Absolute and relative test-retest reliability of the State Mindfulness Scale, State Mindfulness Scale for Physical Activity, and Visual Analog Scale-Mindfulness.

Absolute (ICC)						
Scale	Pre-Sit	Rating	Post-Sit	Rating	Post-Walk	Rating
SMS	.65 [.11, .87] $p < .001$	Poor to Good	.50 [.13, .76] $p = .006$	Poor to Good	.85 [.57, .94] $p < .001$	Moderate to Excellent
SMS-PA	-	-	-	-	.90 [.58, .97] $p < .001$	Moderate to Excellent
VAS-M	.58 [.05, .83] $p < .001$	Poor to Good	.70 [.12, .89] $p < .001$	Poor to Good	.95 [.76, .98] $p < .001$	Good to Excellent
Relative (CV as %)						
Scale	Pre-Sit	Rating	Post-Sit	Rating	Post-Walk	Rating
SMS	16.2	Good	7.2	Excellent	5.3	Excellent
SMS-PA	-	-	-	-	5.5	Excellent
VAS-M	36.9	Poor	18.1	Good	7.8	Excellent

ICC: intraclass correlation; CV: coefficient of variation; SMS: State Mindfulness Scale; SMS-PA: State Mindfulness Scale for Physical Activity; VAS-M: Visual Analog Scale-Mindfulness. The top half of the table presents absolute test-retest reliability as intraclass correlations (ICCs) with 95% confidence intervals (CIs) in brackets for pre-sit, post-sit, and post-walk. The rating corresponding to each assessment is in the column to the left of the ICC for that assessment. The bottom half of the table presents relative test-retest reliability as coefficients of variation (CVs), with the rating corresponding to each assessment in the column to the left of the CV for that assessment. The ICCs were interpreted based on 95% CIs and thresholds adapted from Koo and Li.[36] The CVs were interpreted based on thresholds adapted from Aronhime et al.[37]

The present pilot study aimed to fill the gaps by evaluating the psychometric performance of the SMS, SMS-PA, and VAS-M when administered to participants three times within 30 minutes on a desert trail. Our main finding was that the SMS and SMS-PA can be given to participants in quick succession before, during, and after nature immersion involving physical activity. The SMS-PA showed evidence of concurrent validity with the SMS post-walk through a significant, positive, strong to very strong correlation between the SMS-PA and SMS scores. In contrast, the VAS-M did not show evidence of concurrent validity with the SMS at any assessment (pre-sit, post-sit, or post-walk) or with the SMS-PA post-walk. None of the three scales showed evidence of test-retest reliability across approximately 24 hours.

No adverse effects from the intervention or scale completion were documented. However, the present study's methods were not designed to collect qualitative data, such as participants' thoughts and feelings about using the scales. Therefore, it would be inappropriate to report anecdotal evidence heard during data collection. Future research could address two key areas: first, investigating how participants comprehend the scale items, and second, evaluating the response burden of completing longer scales, particularly in studies with closely spaced repeated measures. Methods such as cognitive interviews or think-aloud protocols could be used to explore participants' understanding of scale items, while assessments of participant

fatigue or adherence could offer insight into the burden of repeated scale completion. These approaches would help refine the use of mindfulness scales in applied settings like nature.

Regarding the SMS-PA, the scale showed evidence of concurrent validity with the SMS post-walk. This result was expected because six of its 12 items were directly borrowed from the SMS.[12,13] The remaining six items on the SMS-PA are original, focusing on thoughts, feelings, and bodily sensations specifically related to physical activity.[13] The present study may be the first direct comparison of the SMS-PA with the SMS in an applied setting. Validation of the SMS-PA has involved exploratory and confirmatory factor analyses on data presumably collected indoors, or at least in an environment more controlled than a desert trail.[13,14,17] Evidence of concurrent validity in a real-world setting is valuable when combined with factor analysis results. Our evidence shows that the SMS-PA measures the intended construct similarly to the trusted measure, the SMS, in a place where the scale will be used. The present study adds to the existing literature by showing that researchers can use the SMS-PA to gather unique data on mindfulness during physical activity while keeping fidelity to the original SMS. The SMS-PA could support research on suitable physical activities for mindfulness-based interventions.

Unlike the SMS-PA, the VAS-M did not show evidence of concurrent validity with the SMS or SMS-PA, likely due to its single-item design, which may be insufficient for capturing mindfulness's multidimensional nature.[11-13] To enhance the VAS-M's clarity and utility, we propose three targeted improvements. First, the VAS-M could include three separate items addressing key mindfulness dimensions—attention, awareness, and acceptance—creating a more comprehensive scale that remains concise compared to other available scales. Second, clearer instructions and visual examples (e.g., showing a correct vertical dash) should be provided, as participants appeared unsure of how to mark responses, leading to imprecise markings such as circles or Xs. Third, researchers could improve the precision of the scale by adopting digital VAS-M formats, which could lower or eliminate the risk of participants making imprecise markings by hand. Until these adjustments are made, we recommend caution in interpreting VAS-M data. Future research should investigate whether the VAS-M shows better evidence of concurrent validity in groups with mindfulness experience or if a total redesign is necessary.

A key strength of the present study is the provision of test-retest reliability data for the SMS over a shorter period, using more robust indices than earlier studies.[11] Earlier studies predominantly relied on Pearson's *r*,[11,12,16] which does not account for variability between measurements. Our results did not meet ICC or CV criteria for test-retest reliability over 24 hours, despite the use of consistent interventions. However, factors such as life stressors or changes in routine between day one and day two were not measured, potentially contributing to inter-day differences in state mindfulness. Additionally, repeated exposure to the mindfulness scales across both days may have influenced participants' responses, a limitation that should be considered when interpreting the test-retest reliability findings. This raises the broader question of whether state mindfulness scales, which measure an inherently dynamic and context-dependent construct, should be expected to show high test-retest reliability.[1,11,12] Future research should explore whether test-retest reliability is a suitable or necessary metric for state mindfulness scales.

Another key strength of the present study is its use of a diverse sample, unlike previous SMS and SMS-PA studies, which often used homogeneous samples.[11,12,14,15,17–19] Our sample was balanced between females and males and racially diverse, with over half identifying as Black, Indigenous, and People of Color. Additionally, the sample included undergraduate and graduate students and their faculty from Western U.S. universities (mean age 26.0 ± 9.0 years), adding to the evidence that the scales can be used for research in U.S. college samples.[13,20] Although the study did not focus on mental health, its sample avoided the chronic lack of diversity noted in research on the mental health effects of nature.[38]

Besides the present study's findings and strengths, the limitations must be noted. First, participants may have inflated their scores post-sit and post-walk because of recall or response bias, which could have undermined the validity of the data. Researchers took steps to minimize these biases, such as ordering the scales randomly and withholding study hypotheses. Future studies could do better by including scales about other constructs to keep participants from discerning the primary variable under investigation (e.g., state mindfulness). Also, researchers should keep a consistent scale order to avoid an order effect. Second, participants' baseline levels of trait mindfulness were not measured, which may have contributed to the observed changes—or lack thereof—in state mindfulness among some individuals. Future research should incorporate assessments of trait mindfulness to account for its potential influence, at a minimum using baseline levels to inform group allocation and mitigate potential confounding and variability in responses associated with it. Third, the study solely used Likert-type and visual analog scales. Using mixed methods research designs could offer a more comprehensive understanding of participants' experiences and state mindfulness. Third, we only tested the scales on desert trails with a specific nature immersion. To enhance generalizability, future research should investigate various natural environments and include diverse types, frequencies, durations, and intensities of physical activity. The final limitation to note is the loss of 11 SMS values on the first day at the first study site. The first author takes full responsibility for this error and encourages researchers to show great care when editing scales.

In summary, this pilot study addressed key gaps in the literature about the psychometric performance of state mindfulness scales around nature immersion with physical activity. Despite minor data loss, the study successfully used the SMS, SMS-PA, and VAS-M in a short, repeated-measures design. While the SMS-PA shows promise for assessing state mindfulness in a U.S. college sample after walking on desert trails, the single-item VAS-M did not prove to be a viable alternative to the SMS or SMS-PA. Researchers should test the VAS-M in samples with mindfulness experience and improve or rebuild it based on mindfulness theory and literature. The discussion raised important questions about the necessity of test-retest reliability in state mindfulness scales, a topic that deserves further discussion and investigation among mindfulness researchers. As interest in nature-based mindfulness interventions grows, the field needs valid and reliable scales to support related research. It is crucial that we develop, refine, and use psychometrically robust scales responsibly, aligning our choices with the current scientific literature.

Acknowledgements

The authors thank Dr. Jennifer Pharr at the University of Nevada, Las Vegas, who gave valuable feedback on drafts of the manuscript.

The University of Nevada, Las Vegas is situated on the traditional homelands of Indigenous groups, including the Nuwu or Nuwuvi, Southern Paiute People, descendants of the Tudinu, or Desert People.

Southern Utah University wishes to acknowledge and honor the Indigenous communities of this region as original possessors, stewards, and inhabitants of this Too'veep (land), and recognize that the University is situated on the traditional homelands of the Nung'wu (Southern Paiute People).

We acknowledge, honor, and express our gratitude to the original inhabitants and stewards of the land on which Vanguard University sits, the Gabrieleño (or Gabrielino) and Juaneño peoples.

References

1. Bishop SR, Lau M, Shapiro S, et al. Mindfulness: A proposed operational definition. *Clin Psychol (New York)*. 2004;11(3):230-241. https://doi.org/10.1093/clipsy.bph077

2. Kabat-Zinn J. Mindfulness. *Mindfulness*. 2015;6:1481-1483. https://doi.org/10.1007/s12671-015-0456-x

3. Baer RA. Mindfulness training as a clinical intervention: A conceptual and empirical review. *Clin Psychol Sci Pract*. 2003;10(2):125-143. https://doi.org/10.1093/clipsy.bpg015

4. Kabat-Zinn J. Mindfulness-based interventions in context: Past, present, and future. *Clin Psychol (New York)*. 2003;10(2):144-156. https://doi.org/10.1093/clipsy.bpg016

5. Davis DW, Carrier B, Cruz K, Barrios B, Landers MR, Navalta JW. A systematic review of the effects of meditative and mindful walking on mental and cardiovascular health. *Int J Exerc Sci*. 2022;15(2):1692-1734. https://doi.org/10.70252/LTNA9604

6. Kaisti I, Kulmala P, Hintsanen M, et al. The effects of mindfulness-based interventions in medical students: A systematic review. *Adv in Health Sci Educ*. 2024;29(1):245-271. https://doi.org/10.1007/s10459-023-10231-0

7. Li J, Li C, Puts M, et al. Effectiveness of mindfulness-based interventions on anxiety, depression, and fatigue in people with lung cancer: A systematic review and meta-analysis. *Int J Nurs Stud*. 2023;140:104447. https://doi.org/10.1016/j.ijnurstu.2023.104447

8. Tang YY, Hölzel BK, Posner MI. Traits and states in mindfulness meditation. *Nat Rev Neurosci*. 2016;17(1):59. https://doi.org/10.1038/nrn.2015.7

9. Wheeler MS, Arnkoff DB, Glass CR. What is being studied as mindfulness meditation? *Nat Rev Neurosci*. 2016;17(1):59-59. https://doi.org/10.1038/nrn.2015.6

10. Baer RA, Smith GT, Lykins E, et al. Construct validity of the Five Facet Mindfulness Questionnaire in meditating and nonmeditating samples. *Assessment*. 2008;15(3):329-342. https://doi.org/10.1177/1073191107313003

11. Ruimi L, Hadash Y, Tanay G, Bernstein A. State Mindfulness Scale (SMS). In: Medvedev ON, Krägeloh CU, Siegert RJ, Singh NN, eds. *Handbook of Assessment in Mindfulness Research*. Springer, Cham; 2022. https://doi.org/10.1007/978-3-030-77644-2_25-1

12. Tanay G, Bernstein A. State Mindfulness Scale (SMS): Development and initial validation. *Psychol Assess*. 2013;25(4):1286-1299. https://doi.org/10.1037/a0034044

13. Cox AE, Ullrich-French S, French BF. Validity evidence for the State Mindfulness Scale for Physical Activity. *Meas Phys Educ Exerc Sci*. 2016;20(1):38-49. https://doi.org/10.1080/1091367X.2015.1089404

14. Peixoto EM, Palma BP, de França Torres VC, Silva KN de O, Farias RP, Monteiro VTL. Cross-cultural adaptation and validity evidence of the Brazilian version of the State Mindfulness Scale for Physical Activity (SMS-PA). *J Phys Educ Sport*. 2019;19(1):594-602. https://doi.org/10.7752/jpes.2019.01087

15. Ullrich-French S, González Hernández J, Hidalgo Montesinos MD. Validity evidence for the adaptation of the State Mindfulness Scale for Physical Activity (SMS-PA) in Spanish youth. *Psicothema*. 2017;29(1):119-125. https://doi.org/10.7334/psicothema2016.204

16. Andrade C, Arriaga P, Carvalho M. The psychometric properties of the Portuguese version of the State Mindfulness Scale. *Mindfulness*. 2019;10:2661-2672. https://doi.org/10.1007/s12671-019-01241-4

17. Ullrich-French S, Cox A, Cole A, Rhoades Cooper B, Gotch C. Initial validity evidence for the State Mindfulness Scale for Physical Activity with youth. *Meas Phys Educ Exerc Sci*. 2017;21(4):177-189. https://doi.org/10.1080/1091367X.2017.1321543

18. Silva KNDO, Romano AR, Silva MPPD, Peixoto EM. Escala de Estado de Mindfulness para Atividade Física: Novas evidências de validade. *Psipesq*. 2021;15(3):1-18. https://doi.org/10.34019/1982-1247.2021.v15.30302

19. Peixoto EM, Pallini AC, Palma BP, Cox AE. The State Mindfulness Scale for Physical Activity: Further psychometrics properties. *Psicol Cienc Prof*. 2023;43:e257372. https://doi.org/10.1590/1982-3703003257372

20. Cox AE, Roberts MA, Cates HL, McMahon AK. Mindfulness and affective responses to treadmill walking in individuals with low intrinsic motivation to exercise. *Int J Exerc Sci*. 2018;11(5):609-624. https://doi.org/10.70252/AEVJ8865

21. Greco G, Centrone C, Poli L, et al. Impact of coastal walking outdoors and virtual reality indoor walking on heart rate, enjoyment levels and mindfulness experiences in healthy adults. *J Funct Morphol Kinesiol*. 2024;9(1):11. https://doi.org/10.3390/jfmk9010011

22. Wheatley KA. Exploring the relationship between mindfulness and rock-climbing: A controlled study. *Curr Psychol*. 2023;42(4):2680-2692. https://doi.org/10.1007/s12144-021-01593-y

23. Kind L, Luttenberger K, Leßmann V, et al. New ways to cope with depression—study protocol for a randomized controlled mixed methods trial of bouldering psychotherapy (BPT) and mental model therapy (MMT). *Trials*. 2023;24:602. https://doi.org/10.1186/s13063-023-07629-x

24. Rogowska AM, Tataruch R. The relationship between mindfulness and athletes' mental skills may be explained by emotion regulation and self-regulation. *BMC Sports Sci Med Rehabil*. 2024;16:68. https://doi.org/10.1186/s13102-024-00863-z

25. Djernis D, Lerstrup I, Poulsen D, Stigsdotter U, Dahlgaard J, O'Toole M. A systematic review and meta-analysis of nature-based mindfulness: Effects of moving mindfulness training into an outdoor natural setting. *Int J Environ Res Public Health*. 2019;16(17):3202. https://doi.org/10.3390/ijerph16173202

26. Kotera Y, Richardson M, Sheffield D. Effects of shinrin-yoku (forest bathing) and nature therapy on mental health: A systematic review and meta-analysis. *Int J Ment Health Addiction*. 2022;20:337-361. https://doi.org/10.1007/s11469-020-00363-4

27. Vitagliano LA, Wester KL, Jones CT, Wyrick DL, Vermeesch AL. Group nature-based mindfulness interventions: Nature-based mindfulness training for college students with anxiety. *Int J Environ Res Public Health*. 2023;20(2):1451. https://doi.org/10.3390/ijerph20021451

28. Salatto RW. *The Effects of Green Exercise on Perception and Connectedness to Nature*. Ph.D. University of Nevada, Las Vegas; 2021. Accessed November 17, 2022. https://www.proquest.com/docview/2581197573/abstract/B507021D58534C77PQ/1

29. Salatto RW, McGinnis GR, Davis DW, et al. Effects of acute beta-alanine ingestion and immersion-plus-exercise on connectedness to nature and perceived pain. *Int J Environ Res Public Health*. 2021;18(15):8134. https://doi.org/10.3390/ijerph18158134

30. Pretty J, Griffin M, Sellens M, Pretty C. *Green Exercise: Complementary Roles of Nature, Exercise and Diet in Physical and Emotional Well-Being and Implications for Public Health Policy*. University of Essex; 2003:46-49. Accessed January 26, 2023. https://www.semanticscholar.org/paper/Green-Exercise%3A-Complementary-Roles-of-Nature%2C-and-Pretty-Griffin/30cc02a4bdd261586c3bddb1ec930689256055b9

31. American College of Sports Medicine. Chapter 2: Exercise preparticipation health screening. In: *ACSM's Guidelines for Exercise Testing and Prescription*. 10th ed. Lippincott Williams & Wilkins; 2017.

32. Navalta JW, Stone WJ, Lyons TS. Ethical issues relating to scientific discovery in exercise science. *Int J Exerc Sci*. 2020;12(1):1-8. https://doi.org/10.70252/EYCD6235

33. Davis DW, Malek EM, Salatto R, et al. A pilot study on the measurement of connectedness to nature around nature immersion with green exercise on desert trails. *Top Exerc Sci Kinesiol*. 2024;5(1):5.

34. Levine TR, Hullett CR. Eta squared, partial eta squared, and misreporting of effect size in communication research. *Human Comm Res*. 2002;28(4):612-625. https://doi.org/10.1111/j.1468-2958.2002.tb00828.x

35. Ferguson CJ. An effect size primer: A guide for clinicians and researchers. *Prof Psychol Res Pr*. 2009;40(5):532-538. https://doi.org/10.1037/a0015808

36. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15(2):155-163. https://doi.org/10.1016/j.jcm.2016.02.012

37. Aronhime S, Calcagno C, Jajamovich GH, et al. DCE-MRI of the liver: Effect of linear and nonlinear conversions on hepatic perfusion quantification and reproducibility. *J Magn Reso*. 2014;40:90-98. https://doi.org/10.1002/jmri.24341

38. Gallegos-Riofrío CA, Arab H, Carrasco-Torrontegui A, Gould RK. Chronic deficiency of diversity and pluralism in research on nature's mental health effects: A planetary health problem. *Curr Res Environ Sustain*. 2022;4:100148. https://doi.org/10.1016/j.crsust.2022.100148