



Development and Validation of the State Moral Elevation Scale: Assessing State-Level Elevation Across Nonclinical and Clinical Samples

Adam P. McGuire^{1,2,3} · Candice L. Hayden¹ · Rawda Tomoum¹ · A. Solomon Kurz^{2,3}

Accepted: 24 March 2022 / Published online: 21 April 2022

This is a U.S. government work and not under copyright protection in the U.S.; foreign copyright protection may apply 2022

Abstract

Research on moral elevation has steadily increased and identified several psychosocial benefits that bear relevance to both the general population and people with psychological distress. However, elevation measurement is inconsistent, and few state-level measures have been created and critically evaluated to date. To address this gap, the State Moral Elevation Scale (SMES) was developed and tested using an online sample ($N=930$) including subsamples of general participants (nonclinical) and those who screened positive for mental health symptoms (clinical). Factor analysis indicated a single factor structure with nine items that demonstrated excellent reliability. Multigroup confirmatory factor analysis indicated good fit statistics and strict measurement invariance across clinical and nonclinical subsamples. Lastly, correlational analyses with related constructs provided evidence of construct validity for both subsamples. Thus, the SMES is a psychometrically valid and reliable assessment tool for state-level elevation which can be used in both general and clinical populations.

Keywords Moral elevation · Psychometric analysis · State measure · Measurement invariance

1 Introduction

Moral elevation is a positive emotional state that is described as feeling inspired or moved after witnessing another person perform a remarkable act of virtue (Haidt, 2003). Research interest on this topic has increased over the past 15 years, which has led to important discoveries about what facilitates moral elevation and the positive correlates of those

✉ Adam P. McGuire
amcguire@uttyler.edu

¹ Department of Psychology and Counseling, The University of Texas at Tyler, 3900 University Blvd, HPR 237, Tyler, TX 75799, USA

² VISN 17 Center of Excellence for Research On Returning War Veterans, Waco, TX, USA

³ Central Texas Veterans Health Care System, Temple, TX, USA

experiences (Pohling & Diessner, 2016; Thomson & Siegel, 2017). However, a significant problem within this area of study is the inconsistency in measurement, particularly when assessing state-level elevation. As researchers strive to expand the elevation literature, a reliable state-level measure will be essential for future studies that need to target important gaps such as (a) assessing specific contexts and elicitors of elevation, (b) determining within-person variability over time or in daily life, and (c) measuring potential benefits to well-being that can be attributed to state elevation induction. In this paper, we will review the distinctive features of state elevation, the utility of measuring this construct, and then describe the development and validation of a novel measurement tool for state-level elevation.

1.1 Distinctive Features

Before considering measurement, it is important to note elevation is a distinct positive emotion that differs from alternative, *other-praising* emotions such as gratitude and admiration. Specifically, elevation is distinguished by the theorized trigger, emotional and physical response, and action tendency or motivations. In a series of studies, Algoe and Haidt (2009) investigated and described differences among these domains across other-praising emotions. First, elevation is elicited by seeing another person engage in a behavior that is perceived as virtuous, whereas gratitude is triggered when someone perceives themselves as the recipient or benefactor of some positive outcome, often initiated by another person (McCullough et al., 2001). Admiration is elicited by witnessing someone demonstrate remarkable skill or talent. Notably, some researchers have described admiration as stemming from either skill or virtue, and there are parallels between admiration of virtue and elevation (Onu et al., 2016). However, consistent with the work of Algoe and Haidt, we will continue describing how elevation is distinctive from the aspect of admiration that refers to skills or talents, which is independent of the virtuous nature of that behavior.

Although admiration and elevation may share some similar subjective feelings, such as awe, elevation also includes subjective feelings of closeness to others (Erickson & Abelson, 2012), self-transcendence (Van Cappellen et al., 2013), and love towards humanity (Diessner et al., 2013). Elevation is also associated with a cluster of physical responses that are not equally shared across admiration and gratitude, which include warmth in chest, lump in throat, piloerection (i.e., goosebumps), and tears in eyes (Algoe & Haidt, 2009; Landis et al., 2009; Oliver et al., 2012).

Lastly, the motivational response to elevation is a strong desire to emulate the witnessed virtue and engage in behavior similar to the moral exemplar (Algoe & Haidt, 2009). Gratitude distinctly elicits a motivation to repay the person who is responsible for the benefactor's positive outcome, or sometimes, to *pay it forward* to another person, whereas admiration is linked with motivations to pursue one's personal goals and strive to succeed. Overall, these unique features support the operationalization of elevation as a separate, positive emotion and highlight aspects of elevation that should be considered in a state-level measure.

1.2 Previous Elevation Measures

During a review of the literature, we identified 50 studies between the years 2006 and 2021 as including some form of elevation measurement. One commonly used measure is the Engagement with Beauty Scale, which includes a moral beauty subscale. This scale has

been validated in previous studies and demonstrates strong psychometric properties (Diessner et al., 2008). However, the moral beauty subscale measures a trait-like disposition or general tendency to feel inspired when exposed to moral beauty, which is distinct from understanding the experience of elevation in a given moment. Past work on measures of discrete, state-level experiences has demonstrated the presence of within-person variability that is not captured by aggregate or trait assessments (e.g., Fleeson, 2004). Researchers have also identified within-person variability in similar positive emotional states like gratitude (e.g., McGuire, et al., 2019a; Nezlek et al., 2019). Thus, dependency on trait-like measures could miss valuable information regarding the experience of elevation. For example, another method of measuring elevation that is often cited is based on a scale originated by Landis and colleagues (2009). Although this measure has strong face validity and is rooted in consultation with founding researchers in this area, it is not a stringent assessment of state elevation. Instead, this measure instructs participants to aggregate experiences and rate the frequency each item typically occurs when witnessing a virtuous act. This method does not account for context and is distinct from the level of elevation one might experience in a given moment following a specific event or exposure to targeted stimuli. To date, one assessment that appears to be the closest measure to state-level elevation is the Chinese Version of the Moral Elevation Scale developed by Ding and colleagues (2014). However, the initial psychometric study was conducted in Chinese, and it is unclear if this measure has been tested in other populations or if an English translation would result in the same reliability and validity characteristics in other samples.

Separate from the distinction between trait and state measures, it is important to note there is no clear, commonly used scale in the field of elevation research. We examined all items used across the 50 studies and identified 83 distinct items (see Supplemental Material Table S1). Assuming there is general agreement on the targeted construct of elevation, one legitimate concern is that the variability in assessments could give one pause when interpreting the validity and reliability of results within this body of literature as a whole. To avoid this potential conflict and to bolster this area of inquiry moving forward, access to a valid measurement tool that can be used widely is imperative.

1.3 Correlates and Benefits of Elevation

The continued study of elevation with proper assessment is important because elevation has been linked with several desirable outcomes. For example, elevation is associated with a wide range of enhanced social functioning that includes higher levels of compassion (Erickson et al., 2018; Siegel et al., 2014), prosocial behavior (Cox, 2010; Sparks et al., 2019; Van de Vyver & Abrams, 2015), increased desire to connect or help other people (Schnall & Roper, 2012; Schnall et al., 2010; Siegel & Thomson, 2017), as well as greater cooperation with and reduced stigma for perceived outgroups (Freeman et al., 2009; Lai et al., 2014; Oliver et al., 2015). Previous work has also found links with elevation and other markers of wellbeing including satisfaction with life, purpose in life, gratitude, hope, and personal growth (Diessner et al., 2008; Landis et al., 2009).

The potential psychosocial benefits of elevation are relevant to the general population at large, but this construct should also be of interest to clinical populations that suffer from psychological distress and their clinicians as the mental health field aims to cultivate these same benefits as treatment outcomes, or mechanisms to outcomes. Although the vast majority of research on elevation has examined the general population or nonclinical samples, select studies have identified specific benefits in clinical populations. For example, daily elevation has

been associated with lower levels of depression, anxiety, and hostility in a clinical sample with moderate depressive symptoms (Erickson & Abelson, 2012). Additionally, early research suggests elevation might benefit those with trauma exposure as indicated by correlations with higher posttraumatic growth and compassion motives (Tingey et al., 2019), along with more treatment engagement and lower avoidance symptoms in veterans (McGuire, Nosen, et al., 2019). Thus, more research is warranted to explore the potential benefits of elevation for those with significant psychological distress; however, any advances in this area of inquiry will be significantly limited and potentially undermined until a reliable measurement tool for state elevation is established.

1.4 Considerations for Measurement in General and Clinical Populations

Despite the potential benefits to both general and clinical populations, it is important to consider group differences that could possibly impact measurement validity when assessing for a positively-valenced construct like elevation. Specifically, symptoms of psychological disorders—such as anxiety, depressive, and posttraumatic stress disorder (PTSD) symptoms—can impact how people perceive various state-level experiences and how they respond to stimuli in their environment. For example, depressive symptoms can be associated with reduced emotional disclosure (Kahn & Garrison, 2009), reduced cognitive processing, and negatively biased memory recall (Claúdio et al., 2012). PTSD is also associated with negative alterations in cognition, which often results in strong negative beliefs about self, others, or the world (American Psychiatric Association, 2013). Additionally, PTSD can lead to emotional numbing and difficulties experiencing positive emotions (e.g., Kashdan et al., 2006). These underlying symptoms within certain clinical populations are clearly relevant to the possible perceptions, experiences, and self-report of state elevation; thus, any scale development will need to assess for measurement invariance across clinical and nonclinical samples prior to future use across both groups.

1.5 Current Study

The purpose of this study is to develop a state-level measure of elevation—the State Moral Elevation Scale (SMES)—and examine the reliability and validity of this measure in both clinical and nonclinical populations. To examine the psychometric properties of this newly developed measure, we conducted exploratory and confirmatory factor analyses (CFA) using three subsamples with a total of 930 participants. Specifically, we aimed to include a minimum of 300 nonclinical participants for exploratory factor analysis (EFA) to determine the factor structure and assess reliability. Next, we aimed to conduct a CFA with two additional subsamples of a minimum of 300 participants each: a clinical and a nonclinical subsample. Participants were assigned to a given subsample based on positive (clinical) or negative (non-clinical) screens from a set of screening questionnaires for mental health symptoms. Lastly, we aimed to conduct a series of analyses to assess construct validity of this measure across both clinical and nonclinical subsamples.

Table 1 Initial 18 candidate items used for state moral elevation scale development

Items	Domain
1. Admiration or respect for the person(s) who did the good deed	Emotion
2. Happy	Emotion
3. Motivated to live in a nobler or virtuous way	Motives
4. "Choked up" (a lump feeling in my throat)	Physical
5. Touched or moved by the goodness of others	Emotion
6. Want to thank or reward the person who did the good deed	Motives
7. Increased heart rate	Physical
8. Warm or glowing feeling in my chest	Physical
9. Somehow "lifted up" or in touch with the better parts of myself	Emotion
10. Want to do something good too	Motives
11. Want to tell the story to other people	Motives
12. Chills, tingles, or goosebumps	Physical
13. Tears in my eyes	Physical
14. Want to be more like the person(s) who did the good deed	Motives
15. Inspired	Emotion
16. Relaxed muscles	Physical
17. More open and loving towards people in general	Emotion
18. Want to become a better person	Motives

2 Method

2.1 Scale Development

First, a pool of 83 items was generated by identifying items used in previous studies that attempted to measure the experience of elevation (see Table S1). Next, the authors categorized each item within one of three primary categories: *Emotional Reaction*, *Physical Reaction*, and *Motivation*. These categories were selected from previous work by Algoe and Haidt (2009) that investigated the features and distinctions of other-praising emotions, as previously described. Any items that did not fit these domains were placed in an *Other* category.

Last, we aimed to simplify this list and reduce participant burden by selecting six items from each primary category for a total of 18 candidate items (see Table 1). Item selection was based on (a) the frequency of use in previous studies (see Table S1) and (b) relevance to the theoretical framework of elevation as determined by the authors. The order of candidate items was randomized. To capture state-level experiences of these items, the following instructions were added: *To what extent did you experience each of the following statements while [insert time/context specifier]*. A Likert-type scale was used with ratings from 0 (*not at all*) to 4 (*extremely*) for each item.

2.2 Participants

Participants were recruited through Amazon Mechanical Turk (MTurk)—an online platform that allows researchers to recruit participants to complete study tasks. Eligibility criteria included (a) at least 18 years or older; (b) must reside in the United States; (c) completed and approved for at least 100 previous MTurk tasks (e.g., online surveys or MTurk Human Intelligence Tasks); (d) at least a 95% approval rating from previous MTurk tasks or surveys. Criteria three and four were included to screen for participants who *successfully* completed substantial MTurk tasks in the past, and to exclude people with a history of rejected payments due to inconsistent responses or poor effort; thus, increasing our confidence in the quality of responses from this pool of participants. All study procedures were reviewed and approved by the local institutional review board. Participants provided assent online rather than written informed consent because they were not asked to provide identifiable information; therefore, participants remained anonymous throughout the study.

Participants were separated into three subsamples at the completion of the study, prior to analysis: two nonclinical subsamples and one clinical subsample. The first nonclinical subsample was used for the preliminary EFA and reliability assessment. The second nonclinical subsample and the clinical subsample, which was determined based on responses to clinical symptom screening measures, were used to assess measurement invariance and validity of the SMES. All participants across all subsamples completed the same study procedures as outlined below. For a summary of demographic characteristics across the entire sample and subsamples, see Table 2.

2.3 Procedures

Participation for this study included two, separate online surveys (see Fig. 1).¹ The first survey (part one) assessed demographic information and characteristic traits, and it included screeners for both clinical symptoms and attention/effort. The second survey (part two) asked participants to watch a 5-minute video clip intended to induce elevation, followed by state-level measurements including the elevation items. The video clip showcased a father who competed in marathons/triathlons with his son who had paraplegia by carrying, pushing, and pulling his son during those races. This video was used in previous studies that examined elevation and was validated as a demonstration of virtuous behavior in one study (Erickson et al., 2018).

Both surveys were advertised as “a survey about personal experiences,” estimated to last approximately 10–15 minutes each, and included a \$1.50 payment for each survey. In the final sample, the median completion time for part one was 8.56 min and 10.03 min for part two, which resulted in a median hourly rate of over \$9.74 per hour across both surveys. This rate is above the median hourly wage for tasks performed on MTurk, which has been estimated at approximately \$2.00 per hour (Hara et al., 2018).

¹ It should be noted all participants were recruited and completed surveys in late-March 2020, during the early phase of the COVID-19 pandemic. We have no reason to believe this affected the results; however, we note in the limitation section that it is possible this historic event could have some unforeseen effect on responses and future research should replicate these findings to rule out potential chronology bias.

Table 2 Demographic characteristics by subsamples

Variable	Overall $n = 930$ $M (SD) / n (%)$	Clinical $n = 321$ $M (SD) / n (%)$	Nonclinical 1 $n = 304$ $M (SD) / n (%)$	Nonclinical 2 $n = 305$ $M (SD) / n (%)$
Age	39.27 (11.98)	36.92 (10.64)	40.13 (12.68)	40.87 (12.26)
Gender				
Male	515 (55.38%)	177 (55.14%)	158 (51.97%)	180 (59.02%)
Female	414 (44.52%)	144 (44.86%)	145 (47.70%)	125 (40.98%)
Female-to-Male Transgender	1 (0.11%)	0 (0%)	1 (0.33%)	0 (0%)
Race				
White	731 (78.60%)	245 (76.32%)	248 (81.58%)	238 (78.03%)
Black or African American	112 (12.04%)	45 (14.02%)	28 (9.21%)	39 (12.79%)
Asian or Asian American	80 (8.60%)	27 (8.41%)	25 (8.22%)	28 (9.18%)
American Indian or Alaska Native	14 (1.51%)	8 (2.49%)	3 (0.99%)	3 (0.98%)
Native Hawaiian or Pacific Islander	6 (0.65%)	4 (1.25%)	0 (0%)	2 (0.66%)
Middle Eastern or North African	3 (0.32%)	3 (0.93%)	0 (0%)	0 (0%)
Hispanic	43 (4.62%)	16 (4.98%)	12 (3.95%)	15 (4.92%)

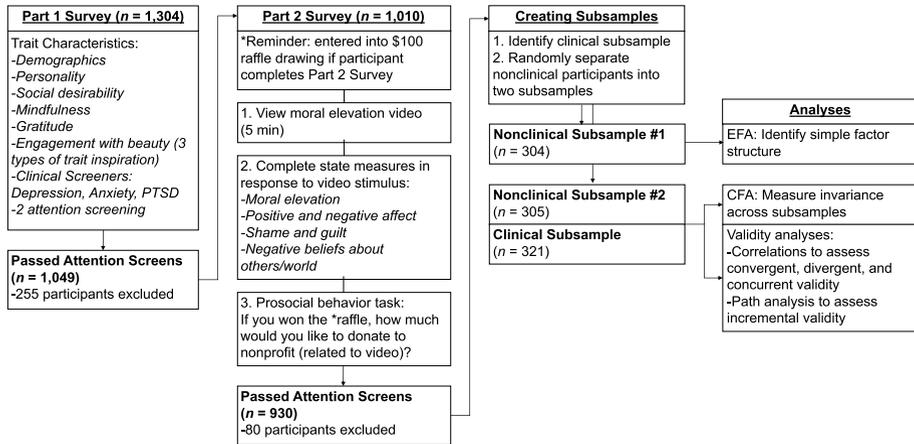


Fig. 1 Study design including screening procedures, descriptions of surveys part 1 and 2, and analyses associated with generated subsamples

2.3.1 Part One Survey

A total of 1,304 participants were recruited to participate in part one of this study. Part one included two attention screeners (see Supplemental Material) placed after approximately one-third and two-thirds of the questionnaires, which assessed whether participants carefully read instructions and responded appropriately. Participants who failed both attention screeners were rejected payment and excluded from the study. To further screen for poor responding, we also excluded participants who reported inappropriate answers to count questions about how many days the participant engaged in specific behaviors during the past week (e.g., in the past *seven* days, I interacted with others on *ten* days). The final sample size for part one was 1,049 after a total of 255 participants were excluded: 66 did not provide consent or complete the study (i.e., did not submit their responses), 43 failed two attention screeners, 3 were excluded for not completing attention screeners, 142 provided inappropriate responses to count questions, and 1 participant was excluded for being under the age of 18.

2.3.2 Part Two Survey

Within three days of completing part one, eligible participants received a message notifying them of the opportunity to complete part two of the study. A total of 1,010 participants completed this follow-up survey. The majority of participants (60.43%) completed part two within two days of part one ($M = 3.01$ days, $SD = 2.00$). Part two contained an additional attention screener that asked participants to provide a text response to a question about the elevation video: “What exercise were the people in the video seen doing the most?” Participants who reported inappropriate answers (e.g., “not sure”; “good”) or exercises not demonstrated in the video (e.g., “jumping jacks”; “yoga”) were excluded from data analysis. The final sample size for part two was 930 after a total of 80 participants were excluded: 72 failed the video-specific attention screen and 8 were excluded for having more than 20% missing data (7 participants with 67% missingness, 1 participant with 27% missingness).

2.4 Clinical Sample Screening Measures

The *Patient Health Questionnaire-9* (Kroenke et al., 2001) was used to screen for depressive symptoms in the past two weeks and has demonstrated adequate reliability and validity in past work. Participants rated the nine items on a 0 (*not at all*) to 3 (*nearly every day*) scale. The items were summed, with higher scores indicating greater symptom severity ($\omega_{\text{categorical}}=0.95$ [0.94, 0.95]).

The *Generalized Anxiety Disorder-7* (Spitzer et al., 2006) was used to screen for anxiety symptoms in the past two weeks and has demonstrated validity and reliability in past studies. Participants rated the seven items on a 0 (*not at all*) to 3 (*nearly every day*) scale. The items were summed, with higher scores indicating greater symptom severity ($\omega_{\text{categorical}}=0.94$ [0.93, 0.95]).

To screen for trauma history, participants completed the first item of the *Primary Care PTSD Screener for DSM-5* (Prins et al., 2016), which asks whether you experienced an event that was frightening, horrible, or traumatic. Those who answered *yes* also completed the *PTSD Checklist for the DSM-5* (Weathers et al., 2013) to screen for PTSD symptoms in the past month, which has demonstrated validity and reliability in past work (Blevins et al., 2015). Participants rated the twenty items on a 0 (*not at all*) to 4 (*extremely*) scale. We summed the items, with higher scores indicating greater symptom severity ($\omega_{\text{categorical}}=0.99$ [0.99, 0.99]).

2.5 Measures used in Construct Validity Analyses

2.5.1 Part One Survey Measures

The *Freiburg Mindfulness Inventory-Short Form* (Walach et al., 2006) assessed trait mindfulness. Participants rated the 14 items on a 1 (*rarely*) to 4 (*almost always*) scale, which were summed, with higher scores indicating higher trait mindfulness. The initial study demonstrated validity and reliability in a nonclinical sample, and future studies also found similar psychometric properties with clinical samples (e.g., Kocovski et al., 2013). For this and all other remaining measures, see Table 6 for a summary of reliability estimates.

The *Gratitude Questionnaire-6* (McCullough et al., 2002) assessed trait gratitude. Participants rated the six items on a 0 (*very unlike me*) to 6 (*very much like me*) scale, which were summed, with higher scores indicating higher trait gratitude. The Gratitude Questionnaire-6 was developed and validated in a nonclinical sample with strong psychometric properties. Other studies also found evidence of construct validity and high internal consistency in clinical samples (Ghalesefidi et al., 2018; Ruini & Vescovelli, 2013).

The *Engagement with Beauty Scale* (EBS; Diessner et al., 2008) assessed trait-like tendencies to be inspired by three different kinds of beauty: natural, artistic, and moral beauty. Participants rated a total of 14 items on a 1 (*very unlike me*) to 7 (*very much like me*) scale, summed to create three subscale scores with higher scores indicating greater tendency to be inspired in each domain. The initial psychometric study demonstrated validity and reliability with a nonclinical sample. The authors are not aware of other studies that assessed the psychometric properties of EBS in a clinical population; however, this study found adequate internal consistency for all three subscales (see Table 6 for estimates of composite reliability).

The *Marlowe-Crowne Social Desirability Scale-Short Form* (Strahan & Gerbasi, 1972) assessed social desirability, or a tendency to respond in a way that might be perceived as

socially acceptable. Participants responded true or false to ten statements coded as 0 or 1 with five reversed items. After reverse-scoring, items were summed with higher scores indicating greater social desirability responding.

The *Ten-Item Personality Inventory* (Gosling et al., 2003) assessed the Big Five personality traits with items rated on a 1 (*disagree strongly*) to 7 (*strongly agree*) scale and summed to create five subscale scores: agreeableness, conscientiousness, emotional stability (neuroticism), extraversion, and openness to experience. Higher scores indicated higher levels of each personality domain. Past work has examined the Ten-Item Personal Inventory and found additional support for construct validity (Jonason et al., 2011).

2.5.2 Part Two Survey Measures

The *International Positive and Negative Affect Scale* (Thompson, 2007) assessed state-level positive and negative affect after viewing the elevation stimulus. Participants rated ten items on a 0 (*not at all*) to 4 (*extremely*) scale, which were summed to create subscale scores for positive and negative affect. For this study, the positive affect sum score excluded *inspired* and was limited to the remaining four items because *inspired* was also included as an SMES item. Higher scores indicated higher levels of affect. Initial measurement validation used a nonclinical sample, but previous studies have also demonstrated validity and reliability this measure in clinical samples (e.g., Serafini et al., 2016).

The *State Shame and Guilt Scale* (Marschall et al., 1994) assessed state-level shame and guilt after viewing the elevation stimulus. Participants rated ten items on a 1 (*Not feeling this way at all*) to 5 (*Feeling this way very strongly*) scale, which were summed to create subscale scores for state guilt and shame. Higher scores indicated higher guilt or shame. Previous studies examined the State Shame and Guilt Scale and found evidence of validity and reliability in clinical (e.g., Levinson et al., 2016) and nonclinical samples (e.g., Fedewa et al., 2005).

The *Negative Cognitions about the World* subscale of the *Posttraumatic Cognitions Inventory* (Foa et al., 1999) assessed current negative beliefs about others. Participants rated the seven subscale items on a 1 (*totally disagree*) to 7 (*totally agree*) scale, which was summed to create a total score, with higher scores indicating more severe negative beliefs about others. Although this measure was designed for and tested with clinical samples, select studies have also noted strong reliability when assessing nonclinical or subclinical populations (e.g., Contractor et al., 2020).

2.6 Data Analytic Plan

2.6.1 Identifying Subsamples

All data management and analyses were conducted with R (R Core Team, 2021). First, scores on the mental health screeners were reviewed across the entire sample. Participants who screened positive on the Generalized Anxiety Disorder-7 (score ≥ 10 ; Spitzer et al., 2006), Patient Health Questionnaire-9 (score ≥ 10 ; Levis et al., 2019) or PTSD Checklist for the DSM-5 (score ≥ 31 ; Bovin et al., 2016) were placed in the clinical subsample ($n=321$). The remaining participants were randomly separated into two nonclinical subsamples using the sample function from the base R stats package: one subsample for the EFA ($n=304$) and one for the CFA ($n=305$).

2.6.2 Factor Analysis

First, we used the psych package (Revelle, 2020) to conduct a parallel analysis to help determine how many factors to extract from the SMES items (Horn, 1965). We then used the lavaan package (Rosseel, 2012) to conduct an EFA with the first nonclinical subsample. EFA model fit was assessed using fit indices including comparative fit index (CFI), Tucker-Lewis Index (TLI), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR). A model was indicated as having acceptable fit with $CFI > 0.90$, $TLI > 0.90$, $RMSEA < 0.10$, and $SRMR < 0.08$ (Brown, 2015). Following recommendations from Kenny et al. (2015), who found the RMSEA has an upward bias for structural equation models with small degrees of freedom, we de-emphasized the RMSEA in models with fewer than 100 degrees of freedom. In addition to fit indices, we also examined the magnitudes of the loadings, preferring items with primary loadings above the conventional 0.40 threshold. Standardized factor loadings are reported. We hypothesized the results would indicate a one-factor model because items were selected from three domains that are theorized to collectively distinguish elevation from other positive emotions (Algoe & Haidt, 2009).

Once an acceptable factor structure was established, we used the lavaan package to conduct a multigroup CFA (MGCFA), which assessed SMES for measurement invariance between the clinical and second nonclinical subsample (i.e., participants not included in EFA). Specifically, we assessed nested models by imposing greater restrictions for successive models in the following order: equivalence of gross factor structure across groups (configural invariance), factor loading equivalence (metric invariance), intercept equivalence (scalar invariance), latent variance equivalence, and latent mean equivalence. Following Cheung and Rensvold (2002), we continued to impose greater restrictions unless the CFI decreased by more than 0.01. Because the SMES items were based on five ordinal response options, many of which had asymmetric distributions in our subsamples, we used the weighted least square mean and variance adjusted (WLSMV) estimator in all EFAs and CFAs (see Rhemtulla et al., 2012).

2.6.3 Reliability

Given the criticisms contemporary methodologists have levied against Cronbach's α (e.g., Kelley & Pornprasertmanit, 2016; McNeish, 2018), we estimated all composite reliabilities using coefficient ω . When using item-level data to compute scale reliability estimates, it is important to consider how the items are distributed, primarily with respect to whether they are grossly asymmetric and whether they are best described as continuous or categorical. Taking cues from Rhemtulla et al. (2012), we treated items as categorical when they contained 4 or fewer categories, or if they had 5 categories and markedly asymmetric distributions. We treated items as continuous if they had more than 6 categories or had 5 categories and roughly symmetric distributions. Accordingly, we estimated composite reliability with $\omega_{\text{categorical}}$ for measurements of clinical symptoms (anxiety, depression, and PTSD), state guilt and shame, state positive and negative affect, mindfulness, social desirability, and state elevation. The items for measurements of engagement with beauty, gratitude, personality and posttraumatic cognitions were sufficiently continuous that we estimated their composite reliabilities with ω . Following recommendations from Kelley and Pornprasertmanit (2016), we used the MBESS package (Kelley, 2020) to accompany all reliability estimates with 95% confidence intervals.

2.6.4 Construct Validity Analyses

Additionally, we examined construct validity across the clinical and second nonclinical subsample, combined. To assess concurrent validity, we examined the correlations between state elevation scores and trait mindfulness, trait gratitude, and state positive affect. Given that state elevation requires an awareness and recognition of another person's virtuous behavior, we hypothesized that greater dispositional awareness of one's surroundings as indicated by trait mindfulness would be positively associated with state elevation. Similarly, trait gratitude involves a greater predisposition to be aware of others' positive or prosocial behaviors (e.g., acknowledging one is the benefactor of another's good deed) and to be thankful for those behaviors; thus, we hypothesized state elevation would be positively correlated with trait gratitude. Lastly, since elevation is a positive emotion, we hypothesized state elevation would be positively correlated with state positive affect (i.e., general positive emotion) experienced after watching the video.

We also assessed for convergent validity by examining the correlations between state elevation scores and the Engagement with Beauty Scale subscales: artistic, natural, and moral beauty. All three subscales represent a predisposition to experience a similar emotional response of feeling inspired and/or moved by a stimulus that is perceived as remarkable; however, the moral beauty subscale specifically measures responses to virtuous behavior (i.e., acts of moral beauty). Thus, we expected state elevation scores to correlate with all subscales scores given that there is likely an underlying predisposition to be moved in general; however, we hypothesized the association would be strongest for moral beauty.

To assess divergent validity, we examined the correlations between state elevation scores and other state-level measures of negative emotional and cognitive responses to the same video stimulus. Specifically, we hypothesized that scores on the positive emotional experience of state elevation would be negatively correlated with general negative affect, guilt, and shame. Similarly, we hypothesized that SMES scores would be negatively correlated with negative thoughts or beliefs about others.

3 Results

3.1 Exploratory Factor Analysis

In the first nonclinical subsample, the SMES items contained no missing data. Across all items, all five response options were endorsed, with some items showing roughly symmetrical distributions (e.g., 8 and 16), whereas others were markedly asymmetrical (e.g., 1 and 5). On the whole, the initial set of 18 items was highly correlated, with a median intra-item polychoric correlation of 0.68 and an interquartile range of 0.55 to 0.77. The sole stand-out was item 16 (*muscles feel relaxed*; see Fig. 2), which demonstrated small to near-zero correlations with the other items. The parallel analysis used the weighted least squares estimator to evaluate a polychoric correlation matrix in 1000 iterations. It indicated we could extract as many as three factors with eigen values greater than would be expected from random data. The eigen values for the first three factors were 11.7, 0.8, and 0.6. Accordingly, all EFA models were estimated in batches of three, to allow us to contrast one-, two-, and three-factor solutions for each combination of items.

The first three EFAs were based on all 18 items, with Model 1.1 containing one factor, Model 1.2 containing two factors, and Model 1.3 containing three factors (see

Table 3 for summary of all model fit statistics). All three models showed poor fit by the model χ^2 , but acceptable fit by the CFI, TLI, and SRMR. The RMSEA was poor for both Model 1.1 and Model 1.2, but acceptable for Model 1.3. In Model 1.1, the single-factor model, all items loaded well with the exception of item 16 ($\lambda=0.37$, 95% confidence interval [0.27, 0.46]), which was foreshadowed by the correlation matrix in Fig. 2. The pattern of loadings in Model 1.2 was not readily interpretable. Though one may have expected the items in Model 1.3 to separate by their theorized categories of *Emotional Reaction*, *Physical Reaction*, and *Motivation*, they did not.

We fit the next batch of EFAs after dropping item 16, which was the one item showing poor performance in Model 1.1. The fit statistics for Model 2.1 (the single-factor model) and Model 2.2 (the two-factor model) were comparable to their counterparts, above—their χ^2 and RMSEA values were poor, but they had acceptable CFI, TLI, and SRMR values. Also like before, the RMSEA was only acceptable for the three-factor model, Model 2.3. As with the first batch of models, Model 2.1 was characterized by strong factor loadings, but the pattern of loadings was difficult to interpret in both Model 2.2 and Model 2.3.

For the third batch of EFAs, we aimed to create a simple structure by eliminating additional items. Using the three categories that guided initial item selection—*Emotional Reaction*, *Physical Reaction*, and *Motivation*—we retained the three items with the highest factor loadings for each category in Model 2.1, resulting in nine items. One exception was made for the motivation items based on theoretical rationale; specifically, we did not retain item 10 (*like I want to do something good too*) because of redundancy with other motivation items. Once again, we fit three models with these items, which included from one (Model 3.1) to three factors (Model 3.3). As before, all models failed the χ^2 test of exact

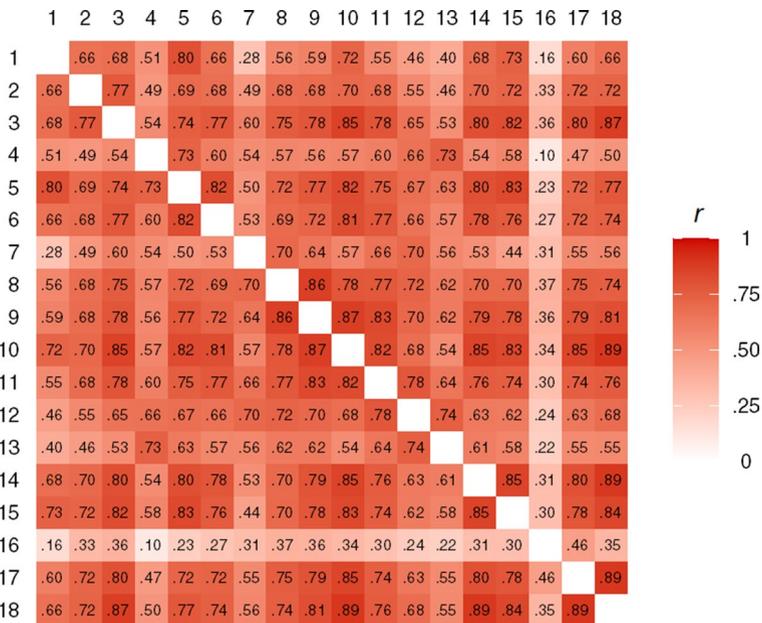


Fig. 2 Polychoric correlation matrix for initial item pool of 18 items for the State Moral Elevation Scale in the first nonclinical subsample (n = 304)

Table 3 Exploratory factor analysis: Model comparisons with single factor model ($n = 304$)

Model	χ^2	df	CFI	TLI	RMSEA	SRMR
<i>Model 1 – 18 items</i>						
M1.1: 1-factor, 18 items	793.13	135	.967	.962	.127	.068
M1.2: 2-factors, 18 items	538.54	118	.979	.972	.108	.047
M1.3: 3-factors, 18 items	255.79	102	.992	.988	.071	.025
<i>Model 2 – excluding item 16</i>						
M2.1: 1-factor, 17 items	751.34	119	.967	.963	.132	.068
M2.2: 2-factors, 17 items	490.99	103	.980	.973	.111	.044
M2.3: 3-factors, 17 items	237.62	88	.992	.988	.075	.024
<i>Model 3 – 9 items with highest loadings</i>						
M3.1: 1-factor, 9 items	249.27	27	.981	.975	.165	.054
M3.2: 2-factors, 9 items	106.43	19	.993	.986	.123	.025
M3.3: 3-factors, 9 items	41.48	12	.998	.993	.090	.016

Note M = Model; χ^2 = model chi-square; df = degrees of freedom; CFI = comparative fit index; TLI = Tucker-Lewis Index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual; the “top 9” items were 3, 4, 8, 9, 12, 14, 15, 17, and 18. In all models, $p < .001$ for the χ^2 test of exact fit

Table 4 Exploratory factor analysis-based factor loadings for the single-factor, 9-item version of the State Moral Elevation Scale ($n=304$)

Items	λ	95% CI
3. Motivated to live in a nobler or virtuous way	.87	[.84, .90]
4. “Choked up” (a lump feeling in my throat)	.62	[.55, .68]
8. Warm or glowing feeling in my chest	.86	[.83, .89]
9. Somehow “lifted up” or in touch with the better parts of myself	.89	[.87, .92]
12. Chills, tingles, or goosebumps	.76	[.70, .81]
14. Want to be more like the person(s) who did the good deed	.91	[.89, .93]
15. Inspired	.89	[.86, .92]
17. More open and loving towards people in general	.89	[.86, .91]
18. Want to become a better person	.96	[.94, .97]

Note λ = standardized factor loading; CI = confidence interval

fit and all showed acceptable CFI, TLI, and SRMR values (see Table 3). Given its positive bias in small- df models (Kenny et al., 2015), we placed less weight on the RMSEA when comparing the models in this batch. Whereas Model 3.2 and Model 3.3 did not contain a readily-interpretable pattern of item loadings, the single-factor model (Model 3.1) contained 9 strong loadings (see Table 4), with an average loading of 0.86 [0.84, 0.88]. The estimated scale reliability was high, $\omega_{\text{categorical}}=0.97$ [0.95, 0.98].

3.2 Multigroup Confirmatory Factor Analysis

Next, we used CFA and MGCFA to replicate the single-factor nine-item model derived from the previous EFA (Model 3.1) and to assess for measurement invariance in the clinical subsample ($n=321$) and the second nonclinical subsample ($n=305$). First, separate single-group CFAs for each subsample indicated acceptable fit for both clinical and non-clinical subsamples across all indices except RMSEA, which we continued to de-emphasize due to its known upward bias in models with small degrees of freedom (Kenny et al., 2015). For a full listing of their fit statistics, see the first two rows in Table 5. The initial MGCFA model—the configural model, which specified gross factor structure equivalence but allowed all parameters to differ by subsample—continued to show adequate fit (see row three in Table 5). We proceeded with more restrictive models and terminated with the most restrictive model, which held the item loadings, item thresholds, latent variances, and latent means constant between the two subsamples. From model to model, the CFI never decreased below the 0.01 threshold and the most restrictive model had acceptable fit statistics for the CFI, TLI, SRMR, and even the RMSEA. Therefore, we concluded the single-factor nine-item SMES fit well in both the clinical subsample and the second non-clinical subsample, and there was evidence of measurement invariance between the two subsamples. In addition, the SMES scale reliability was high in both subsamples (clinical $\omega_{\text{categorical}}=0.93$ [0.90, 0.94]; second nonclinical $\omega_{\text{categorical}}=0.96$ [0.94, 0.97]), as well as in the combination of the two ($\omega_{\text{categorical}}=0.94$ [0.93, 0.95]). The standardized factor loadings in the final MGCFA model were similar to those in the EFA displayed in Table 4 (see Supplemental Material for details).

Table 5 Results from multigroup confirmatory factor analysis of 9-item State Moral Elevation Scale

Model	<i>n</i>	χ^2	<i>df</i>	CFI	TLI	RMSEA	SRMR
<i>Samples</i>							
Clinical subsample	321	148.86	27	.973	.964	.119	.044
Nonclinical sub-sample	305	251.06	27	.981	.974	.165	.045
<i>Equivalence</i>							
Configural	626	403.78	54	.979	.971	.144	.045
Metric	626	307.53	62	.985	.983	.113	.046
Scalar	626	473.29	88	.976	.981	.118	.045
Latent variances	626	373.10	89	.983	.986	.101	.045
Latent means	626	266.31	90	.989	.991	.079	.045

Note Configural = gross factor structure equivalence; Metric = factor loading equivalence; Scalar = item threshold equivalence; Latent variances = latent variance equivalence; Means = latent means equivalence; χ^2 = model chi-square; *df* = degrees of freedom; CFI = comparative fit index; TLI = Tucker-Lewis Index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual; In all models, $p < .001$ for the χ^2 test of exact fit

Table 6 Descriptive statistics for clinical subsample and second nonclinical subsample used for analysis of invariance and validity

Variable	Range	ω [95% CI]	Overall <i>n</i> = 626 <i>M</i> (<i>SD</i>)	Clinical <i>n</i> = 321 (51.3%) <i>M</i> (<i>SD</i>)	Nonclinical <i>n</i> = 305 (48.7%) <i>M</i> (<i>SD</i>)
Clinical Screeners					
Depressive symptoms (PHQ-9)	0–27	.95 [.94, .95]*	7.77 (6.76)	12.99 (5.18)	2.28 (2.59)
Anxiety symptoms (GAD-7)	0–21	.94 [.93, .94]*	7.16 (5.95)	11.63 (4.59)	2.45 (2.69)
PTSD symptoms (PCL-5)	0–80	.99 [.99, .99]*	9.35 (17.70)	15.95 (22.14)	2.39 (5.80)
Personality/Traits					
Extraversion (TIPI)	2–14	.71 [.66, .75]	6.91 (3.43)	6.40 (3.20)	7.46 (3.57)
Agreeableness (TIPI)	2–14	.49 [.40, .56]	10.36 (2.69)	9.35 (2.63)	11.43 (2.32)
Conscientiousness (TIPI)	2–14	.61 [.54, .66]	10.76 (2.73)	9.69 (2.76)	11.90 (2.19)
Openness (TIPI)	2–14	.48 [.39, .56]	9.85 (2.74)	9.35 (2.66)	10.37 (2.72)
Emotional stability (TIPI)	2–14	.78 [.75, .81]	9.46 (3.35)	7.61 (2.93)	11.41 (2.56)
Social desirability (MCSDS-SF)	0–10	.73 [.68, .76]*	4.66 (2.53)	4.13 (2.20)	5.23 (2.73)
Mindfulness (FMI-SF)	14–56	.91 [.90, .92]*	38.02 (8.11)	35.10 (7.21)	41.09 (7.88)
Gratitude (GQ-6)	6–42	.85 [.83, .88]	31.38 (7.70)	28.21 (7.41)	34.72 (6.49)
Engagement with natural beauty (EBS)	4–28	.81 [.78, .84]	20.41 (4.89)	20.40 (4.65)	20.42 (5.13)
Engagement with artistic beauty (EBS)	4–28	.86 [.83, .88]	19.03 (5.40)	19.43 (5.09)	18.61 (5.69)
Engagement with moral beauty (EBS)	6–42	.89 [.87, .91]	30.06 (7.45)	30.07 (6.84)	30.06 (8.04)
State Measures (post-video)					
Moral elevation (SMES)	0–40	.93 [.92, .93]*	22.96 (8.64)	23.17 (8.09)	22.74 (9.19)
Positive affect (I-PANAS)	0–16	.77 [.73, .81]*	9.77 (3.67)	9.41 (3.71)	10.14 (3.59)
Negative affect (I-PANAS)	0–19	.93 [.91, .94]*	2.05 (3.98)	3.59 (5.02)	0.43 (0.95)
Guilt (SSGS)	5–25	.94 [.92, .95]*	8.17 (4.91)	10.31 (5.73)	5.91 (2.25)
Shame (SSGS)	5–25	.95 [.93, .95]*	8.09 (4.65)	10.26 (5.37)	5.80 (1.97)
Negative views of others (PTCI)	7–49	.94 [.93, .95]	26.95 (10.15)	30.12 (9.43)	23.62 (9.83)

Note * = reliability estimated by treating items as categorical ($\omega_{\text{categorical}}$); all other reliability estimates are calculated by treating items as continuous (ω). CI = confidence interval; *M* = mean; *SD* = standard deviation; PHQ-9 = Patient Health Questionnaire-9; GAD-7 = Generalized Anxiety Disorder-7; PCL-5 = PTSD Checklist for DSM-5; TIPI = Ten-Item Personality Inventory; MCSDS = Marlowe-Crowne Social Desirability Scale-Short Form; FMI-SF = Freiburg Mindfulness Inventory-Short Form; GQ-6 = Gratitude Questionnaire-6; EBS = Engagement with Beauty Scale; SMES = State Moral Elevation Scale; I-PANAS = International Positive Affect and Negative Affect Schedule, SSGS = State Shame and Guilt Scale; PTCI = Posttraumatic Cognitions Inventory

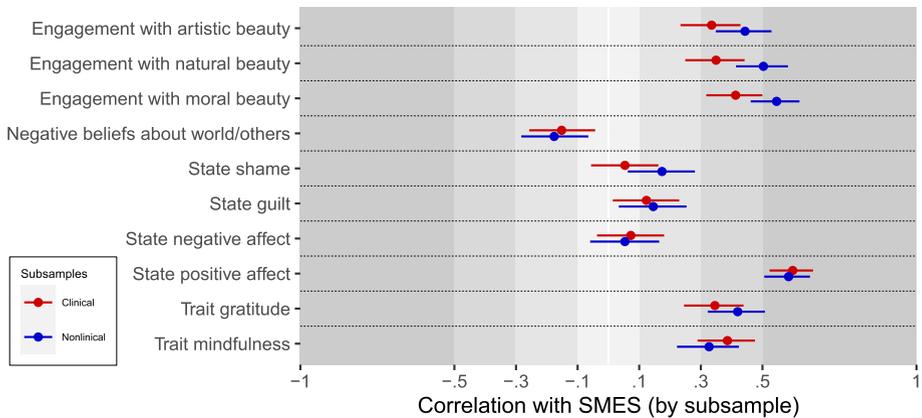


Fig. 3 Results from correlation analyses between State Moral Elevation Scale (SMES) scores and variables assessing construct validity. Dots and lines represent Pearson correlations and their 95% confidence intervals for each subsample, separately. The background is shaded according to conventional effect-size regions of *small* (.1 to .3), *medium* (.3 to .5), and *large* (.5 and up)

3.3 Construct Validity Analyses

Validity analyses were conducted with the same subsamples used for the MGCFA using Pearson correlations. Descriptive statistics for all validity measures are reported across both subsamples in Table 6. Results from correlational analyses are presented in Fig. 3 and are reported for both clinical and nonclinical subsamples. Exact numeric summaries are available in Supplemental Materials (see Table S2).

3.3.1 Concurrent Validity

As hypothesized, there were moderate positive associations between SMES scores and trait mindfulness and trait gratitude, and a strong positive association with positive affect experienced immediately after viewing the same video clip.

3.3.2 Convergent Validity

Results indicated moderate positive associations between SMES scores and all three subscales of trait-like engagement with beauty in both subsamples. As hypothesized, the effects were strongest for the association between state elevation and engagement with moral beauty.

3.3.3 Divergent Validity

As hypothesized, SMES scores were negatively associated with negative views about the world or other people after viewing the video clip in both subsamples; however, the effect

size was surprisingly small. Unexpectedly, state elevation did not demonstrate strong associations with guilt, shame, or negative affect. Results indicated correlations were in the opposite direction, but the effect sizes were near zero, ranging from 0.06 to 0.11, with some confidence intervals that contain or approach zero. For example, the correlation between the SMES and negative affect in the overall sample was 0.06 [$-0.02, 0.14$].

4 Discussion

The purpose of this study was to develop a measure of state-level elevation and assess psychometric properties, including measurement invariance across clinical and non-clinical subsamples.

First, results from the EFA supported the expected single factor structure with nine items. Additionally, findings indicated strong internal consistency and provided preliminary evidence of validity with adequate fit statistics. The nine retained items represent an equal distribution of elevation-related motives, as well as emotional and physical responses. These domains and items are consistent with theory regarding elevation's core and distinctive features compared to other positive emotions. Many of these retained items were also among the most commonly used items in the elevation literature to date. For example, *inspired* was the most frequently used among previous studies reviewed prior to data collection (53.1%; see Table S1). *Warm or glowing feeling in my chest* (36.7%) and *want to become a better person* (32.7%), were also in the top five most frequently used items.

CFA results for each group demonstrated acceptable fit statistics, but more importantly, results from the MGCFA suggested measurement invariance across clinical and nonclinical subsamples at the most restricted levels. Model fit did not degrade at any step of invariance testing and the final model—which included constrained factor loadings, item intercepts, latent variances, and latent means—indicated adequate fit with regard to CFI, TLI, and SRMR, and RMSEA. In sum, these results suggest total scores on the SMES have the same meaning in both clinical and nonclinical subsamples. This finding is particularly important for any future research efforts that aim to assess whether the experience of state elevation can benefit those suffering from psychological distress.

Consistent with hypotheses, findings from correlation analyses in both subsamples provided evidence of concurrent validity with state positive affect, trait gratitude, and trait mindfulness. Additionally, SMES scores were significantly related to all subscales of the EBS, but the highest effects were found with self-reported tendencies to be moved by moral beauty (i.e., trait elevation), as expected. This finding provides initial support for this measure as a state-level assessment of the same trait-like construct captured by the established engagement with moral beauty subscale. Overall, results from correlation analyses with relevant positive constructs, measured at both the state and trait levels across both subsamples, were consistent with hypotheses and provided support for concurrent and convergent validity.

In support of divergent validity, SMES scores were negatively correlated with negative views of the others after watching the video, as expected; however, hypotheses regarding negative correlations with negative emotions (i.e., negative affect, guilt, shame) were not supported. One possible explanation could be restricted variance around negative emotions at the state-level following a positive, inspirational video.

This study design was not intended to induce any emotions prior to watching the video, but perhaps participants would report a more diverse set of negative emotions if they were primed to be in a negative mood prior to the video. Nevertheless, a negative correlation with negative views of others is consistent with the theorized cognitive appraisal of elevation and provides initial support for divergent validity.

4.1 Limitations and Future Directions

Results should be considered in light of several limitations. First, this study used one stimulus to elicit state elevation for all participants; however, it is unknown whether the selected video could be considered a universal stimulus, or if there is significant variability in the type of content or stimulus that elicits elevation across people. Addressing this gap is an important area for future research that will expand our understanding of how elevation is experienced and improve study designs moving forward.

Second, this data was collected with an online survey using a crowdsourcing platform, which involves uncertainty regarding participants' effort and attention during study tasks. Nonetheless, we aimed to increase confidence in the quality of results by including a fair compensation, screening for participants with a history of successful completions, and using multiple attention screeners. Related, there are also limitations to relying on screener tools for mental health symptoms and it is possible that the identified clinical subsample might not be representative of a clinical population with significant distress. Therefore, future research is needed to replicate these results using alternative recruitment methods, including clinical samples with more comprehensive assessment procedures.

Lastly, this data was collected at the early stages of the COVID-19 pandemic in late March 2020; therefore, it is possible responses to survey questions could have been affected by changes in participants' life circumstances at the time, or the subsequent impact on their psychosocial wellbeing. Thus, attempts to replicate these results in future studies could also address potential chronology bias or any unexpected influence associated with the pandemic.

5 Conclusion

Although past work has identified several psychosocial benefits to elevation and provided reason for further exploration, a significant limitation to date is inconsistency in the measurement of state-level elevation experiences. To our knowledge, no state measure has been developed and thoroughly tested for reliability and validity in the English language; particularly with respect to measurement invariance with clinical populations, despite increased interest in studying the potential therapeutic benefits. Overall, results from this study indicate that the SMES is a psychometrically valid and reliable tool for assessing state elevation in both clinical and nonclinical populations. This development is important for critical future directions in the field of elevation research, such as studies that aim to assess specific contexts and elicitors of elevation (i.e., experimental designs), determine within-person variability over time or in daily life (i.e., longitudinal designs), and measure potential benefits to wellbeing that can be attributed to state elevation induction (i.e., intervention studies). It is our hope that this measure facilitates robust research on elevation

moving forward and contributes to our understanding of its full potential for psychosocial wellbeing.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10902-022-00533-2>.

Acknowledgements This material is the result of work with resources and the use of facilities at the the University of Texas at Tyler and supported by the VISN 17 Center of Excellence for Research on Returning War Veterans and the Central Texas Veterans Health Care System. The views expressed herein are those of the authors and do not necessarily reflect the position or policy of the Department of Veterans Affairs or the United States Government. AM and CH collected the data. AM and ASK analyzed the data. AM, CH, and RT drafted the manuscript. CH and RT provided administrative, technical, and material support. All authors provided critical revision of the manuscript for important intellectual content. All authors read, discussed, and approved the final version of the manuscript.

Declarations

Conflict of interest The authors declare that they donot have any conflict of interest to declare.

Data Availability All materials and data for this study are publicly available: <https://osf.io/4g629/>.

References

- Algoe, S. B., & Haidt, J. (2009). Witnessing excellence in action: The 'other-praising' emotions of elevation, gratitude, and admiration. *The Journal of Positive Psychology, 4*(2), 105–127. <https://doi.org/10.1080/17439760802650519>
- American Psychiatric Association. (2013). DSM-5 diagnostic classification. *Diagnostic and Statistical Manual of Mental Disorders (5th Ed.)*, 10. <https://doi.org/10.1176/appi.books.9780890425596>
- Blevins, C. A., Weathers, F. W., Davis, M. T., Witte, T. K., & Domino, J. L. (2015). The posttraumatic stress disorder checklist for DSM-5 (PCL-5): Development and initial psychometric evaluation. *Journal of Traumatic Stress, 28*(6), 489–498. <https://doi.org/10.1002/jts.22059>
- Bovin, M. J., Marx, B. P., Weathers, F. W., Gallagher, M. W., Rodriguez, P., Schnurr, P. P., & Keane, T. M. (2016). Psychometric properties of the PTSD checklist for diagnostic and statistical manual of mental disorders—fifth edition (PCL-5) in veterans. *Psychological Assessment, 28*(11), 1379. <https://doi.org/10.1037/pas0000254>
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). Guilford Publications.
- Cappellen, P. V., Saroglou, V., Iweins, C., Piovesana, M., & Fredrickson, B. L. (2013). Self-transcendent positive emotions increase spirituality through basic world assumptions. *Cognition and Emotion, 27*(8), 1378–1394. <https://doi.org/10.1080/02699931.2013.787395>
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5
- Claúdio, V., Aurélio, J. G., & Machado, P. P. P. (2012). Autobiographical memories in major depressive disorder. *Clinical Psychology & Psychotherapy, 19*(5), 375–389. <https://doi.org/10.1002/cpp.751>
- Contractor, A. A., Banducci, A. N., Jin, L., Keegan, F. S., & Weiss, N. H. (2020). Effects of processing positive memories on posttrauma mental health: A preliminary study in a non-clinical student sample. *Journal of Behavior Therapy and Experimental Psychiatry, 66*, 101516. <https://doi.org/10.1016/j.jbtep.2019.101516>
- Cox, K. S. (2010). Elevation predicts domain-specific volunteerism 3 months later. *The Journal of Positive Psychology, 5*(5), 333–341. <https://doi.org/10.1080/17439760.2010.507468>
- Diessner, R., Iyer, R., Smith, M. M., & Haidt, J. (2013). Who engages with moral beauty? *Journal of Moral Education, 42*(2), 139–163. <https://doi.org/10.1080/03057240.2013.785941>
- Diessner, R., Solom, R. D., Frost, N. K., Parsons, L., & Davidson, J. (2008). Engagement with beauty: Appreciating natural, artistic, and moral beauty. *The Journal of Psychology, 142*(3), 303–332. <https://doi.org/10.3200/jrlp.142.3.303-332>

- Ding, W., Wang, X., Sun, B., & Li, W. (2014). The structure and measurement of the moral elevation. *Advances in Psychology*, 4(6), 777–787. <https://doi.org/10.12677/ap.2014.46102>
- Erickson, T. M., & Abelson, J. L. (2012). Even the downhearted may be uplifted: Moral elevation in the daily life of clinically depressed and anxious adults. *Journal of Social and Clinical Psychology*, 31(7), 707–728. <https://doi.org/10.1521/jscp.2012.31.7.707>
- Erickson, T. M., McGuire, A. P., Scarsella, G. M., Crouch, T. A., Lewis, J. A., Eisenlohr, A. P., & Muresan, T. J. (2018). Viral videos and virtue: Moral elevation inductions shift affect and interpersonal goals in daily life. *The Journal of Positive Psychology*, 13(6), 643–654. <https://doi.org/10.1080/17439760.2017.1365163>
- Fedewa, B. A., Burns, L. R., & Gomez, A. A. (2005). Positive and negative perfectionism and the shame/guilt distinction: Adaptive and maladaptive characteristics. *Personality and Individual Differences*, 38(7), 1609–1619. <https://doi.org/10.1016/j.paid.2004.09.026>
- Fleeson, W. (2004). Moving personality beyond the person-situation debate: The challenge and the opportunity of within-person variability. *Current Directions in Psychological Science*, 13(2), 83–87. <https://doi.org/10.1111/j.0963-7214.2004.00280.x>
- Foa, E. B., Ehlers, A., Clark, D. M., Tolin, D. F., & Orsillo, S. M. (1999). The posttraumatic cognitions inventory (PTCI): Development and validation. *Psychological Assessment*, 11(3), 303. <https://doi.org/10.1037//1040-3590.11.3.303>
- Freeman, D., Aquino, K., & McFerran, B. (2009). Overcoming beneficiary race as an impediment to charitable donations: Social dominance orientation, the experience of moral elevation, and donation behavior. *Personality and Social Psychology Bulletin*, 35(1), 72–84. <https://doi.org/10.1177/0146167208325415>
- Ghalesefidi, M. J., Maghsoudi, J., & Pouragha, B. (2018). Effectiveness of gratitude on psychological wellbeing and quality of life among hospitalized substance abuse patients. *Electronic Journal of General Medicine*. <https://doi.org/10.29333/ejgm/94091>
- Gosling, S. D., Rentfrow, P. J., & Swann, W. B., Jr. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, 37(6), 504–528. [https://doi.org/10.1016/s0092-6566\(03\)00046-1](https://doi.org/10.1016/s0092-6566(03)00046-1)
- Haidt, J. (2003). Elevation and the positive psychology of morality. In *Flourishing: Positive psychology and the life well-lived* (pp. 275–289). American Psychological Association. <https://doi.org/10.1037/10594-012>
- Hara, K., Adams, A., Milland, K., Savage, S., Callison-Burch, C., & Bigman, J. P. (2018). A data-driven analysis of workers' earnings on Amazon Mechanical Turk. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–14. <https://doi.org/10.1145/3173574.3174023>
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30(2), 179–185. <https://doi.org/10.1007/BF02289447>
- Jonason, P., Teicher, E. A., & Schmitt, D. (2011). The TIPI's validity confirmed: Associations with sociosexuality and self-esteem. *Individual Differences Research*, 9, 52–60.
- Kahn, J. H., & Garrison, A. M. (2009). Emotional self-disclosure and emotional avoidance: Relations with symptoms of depression and anxiety. *Journal of Counseling Psychology*, 56(4), 573. <https://doi.org/10.1037/a0016574>
- Kashdan, T. B., Elhai, J. D., & Frueh, B. C. (2006). Anhedonia and emotional numbing in combat veterans with PTSD. *Behaviour Research and Therapy*, 44(3), 457–467. <https://doi.org/10.1016/j.brat.2005.03.001>
- Kelley, K. (2020). MBESS: The MBESS R Package (Version 4.8.0) [computer software and manual]. <https://cran.r-project.org/package=MBESS>
- Kelley, K., & Pornprasertmanit, S. (2016). Confidence intervals for population reliability coefficients: Evaluation of methods, recommendations, and software for composite measures. *Psychological Methods*, 21(1), 69–92. <https://doi.org/10.1037/a0040086>
- Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2015). The performance of RMSEA in models with small degrees of freedom. *Sociological Methods & Research*, 44(3), 486–507. <https://doi.org/10.1177/0049124114543236>
- Kocovski, N. L., Fleming, J. E., Hawley, L. L., Huta, V., & Antony, M. M. (2013). Mindfulness and acceptance-based group therapy versus traditional cognitive behavioral group therapy for social anxiety disorder: A randomized controlled trial. *Behaviour Research and Therapy*, 51(12), 889–898. <https://doi.org/10.1016/j.brat.2013.10.007>
- 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. <https://doi.org/10.1046/j.1525-1497.2001.016009606.x>

- Lai, C. K., Haidt, J., & Nosek, B. A. (2014). Moral elevation reduces prejudice against gay men. *Cognition and Emotion*, 28(5), 781–794. <https://doi.org/10.1080/02699931.2013.861342>
- Landis, S. K., Sherman, M. F., Piedmont, R. L., Kirkhart, M. W., Rapp, E. M., & Bike, D. H. (2009). The relation between elevation and self-reported prosocial behavior: Incremental validity over the five-factor model of personality. *The Journal of Positive Psychology*, 4(1), 71–84. <https://doi.org/10.1080/17439760802399208>
- Levinson, C. A., Byrne, M., & Rodebaugh, T. L. (2016). Shame and guilt as shared vulnerability factors: Shame, but not guilt, prospectively predicts both social anxiety and bulimic symptoms. *Eating Behaviors*, 22, 188–193. <https://doi.org/10.1016/j.eatbeh.2016.06.016>
- Levis, B., Benedetti, A., & Thombs, B. D. (2019). Accuracy of patient health questionnaire-9 (PHQ-9) for screening to detect major depression: Individual participant data meta-analysis. *BMJ*, 365, 11476. <https://doi.org/10.1136/bmj.11476>
- Marschall, D., Sanftner, J., & Tangney, J. P. (1994). *The state shame and guilt scale*. George Mason University.
- McCullough, M. E., Emmons, R. A., & Tsang, J.-A. (2002). The grateful disposition: A conceptual and empirical topography. *Journal of Personality and Social Psychology*, 82(1), 112. <https://doi.org/10.1037/0022-3514.82.1.112>
- McCullough, M. E., Kilpatrick, S. D., Emmons, R. A., & Larson, D. B. (2001). Is gratitude a moral affect? *Psychological Bulletin*, 127(2), 249–266. <https://doi.org/10.1037/0033-2909.127.2.249>
- McGuire, A. P., Erickson, T. M., Quach, C. M., & Willey, B. (2019). Gratitude for better or worse: Differential predictors and affective outcomes of state gratitude in positive and negative contexts. *Journal of Positive School Psychology*, 3(2), 99–111.
- McGuire, A. P., Nosen, E., & Lyons, J. A. (2019). Benefits of moral elevation in veterans with PTSD and moral injury: A proposed theoretical framework and pilot study. *Military Behavioral Health*, 7(3), 315–326. <https://doi.org/10.1080/21635781.2018.1540316>
- McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods*, 23(3), 412–433. <https://doi.org/10.1037/met0000144>
- Nezlek, J. B., Krejtz, I., Rusanowska, M., & Holas, P. (2019). Within-person relationships among daily gratitude, well-being, stress, and positive experiences. *Journal of Happiness Studies*, 20(3), 883–898. <https://doi.org/10.1007/s10902-018-9979-x>
- Oliver, M. B., Hartmann, T., & Woolley, J. K. (2012). Elevation in response to entertainment portrayals of moral virtue. *Human Communication Research*, 38(3), 360–378. <https://doi.org/10.1111/j.1468-2958.2012.01427.x>
- Oliver, M. B., Kim, K., Hoewe, J., Chung, M.-Y., Ash, E., Woolley, J. K., & Shade, D. D. (2015). Media-induced elevation as a means of enhancing feelings of intergroup connectedness. *Journal of Social Issues*, 71(1), 106–122. <https://doi.org/10.1111/josi.12099>
- Onu, D., Kessler, T., & Smith, J. R. (2016). Admiration: A conceptual review. *Emotion Review*, 8(3), 218–230. <https://doi.org/10.1177/1754073915610438>
- Pohling, R., & Diessner, R. (2016). Moral elevation and moral beauty: A review of the empirical literature. *Review of General Psychology*, 20(4), 412–425. <https://doi.org/10.1037/gpr0000089>
- Prins, A., Bovin, M. J., Smolenski, D. J., Marx, B. P., Kimerling, R., Jenkins-Guarnieri, M. A., Kaloupek, D. G., Schnurr, P. P., Kaiser, A. P., Leyva, Y. E., & Tiet, Q. Q. (2016). The primary care PTSD screen for DSM-5 (PC-PTSD-5): Development and evaluation within a veteran primary care sample. *Journal of General Internal Medicine*, 31(10), 1206–1211. <https://doi.org/10.1007/s11606-016-3703-5>
- R Core Team. (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Revelle, W. (2020). psych: Procedures for Psychological, Psychometric, and Personality Research (Version 1.9.12.31) [computer software and manual]. <https://cran.r-project.org/web/packages/psych/index.html>
- Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under sub-optimal conditions. *Psychological Methods*, 17(3), 354–373. <https://doi.org/10.1037/a0029315>
- Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48, 1–36. <https://doi.org/10.18637/jss.v048.i02>
- Ruini, C., & Vescovelli, F. (2013). The role of gratitude in breast cancer: Its relationships with post-traumatic growth, psychological well-being and distress. *Journal of Happiness Studies*, 14(1), 263–274. <https://doi.org/10.1007/s10902-012-9330-x>
- Schnall, S., & Roper, J. (2012). Elevation puts moral values into action. *Social Psychological and Personality Science*, 3(3), 373–378. <https://doi.org/10.1177/1948550611423595>
- Schnall, S., Roper, J., & Fessler, D. M. T. (2010). Elevation leads to altruistic behavior. *Psychological Science*, 21(3), 315–320. <https://doi.org/10.1177/0956797609359882>

- Serafini, K., Malin-Mayor, B., Nich, C., Hunkele, K., & Carroll, K. M. (2016). Psychometric properties of the positive and negative affect schedule (PANAS) in a heterogeneous sample of substance users. *The American Journal of Drug and Alcohol Abuse*, *42*(2), 203–212. <https://doi.org/10.3109/00952990.2015.1133632>
- Siegel, J. T., & Thomson, A. L. (2017). Positive emotion infusions of elevation and gratitude: Increasing help-seeking intentions among people with heightened levels of depressive symptomatology. *The Journal of Positive Psychology*, *12*(6), 509–524. <https://doi.org/10.1080/17439760.2016.1221125>
- Siegel, J. T., Thomson, A. L., & Navarro, M. A. (2014). Experimentally distinguishing elevation from gratitude: Oh, the morality. *The Journal of Positive Psychology*, *9*(5), 414–427. <https://doi.org/10.1080/17439760.2014.910825>
- Sparks, A. M., Fessler, D. M. T., & Holbrook, C. (2019). Elevation, an emotion for prosocial contagion, is experienced more strongly by those with greater expectations of the cooperativeness of others. *PLoS ONE*, *14*(12), e0226071. <https://doi.org/10.1371/journal.pone.0226071>
- Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine*, *166*(10), 1092–1097. <https://doi.org/10.1001/archinte.166.10.1092>
- Strahan, R., & Gerbasi, K. C. (1972). Short, homogeneous versions of the Marlow-Crowne social desirability scale. *Journal of Clinical Psychology*, *28*(2), 191–193. [https://doi.org/10.1002/1097-4679\(197204\)28:2%3c191::AID-JCLP2270280220%3e3.0.CO;2-G](https://doi.org/10.1002/1097-4679(197204)28:2%3c191::AID-JCLP2270280220%3e3.0.CO;2-G)
- Thompson, E. R. (2007). Development and validation of an internationally reliable short-form of the positive and negative affect schedule (PANAS). *Journal of Cross-Cultural Psychology*, *38*(2), 227–242. <https://doi.org/10.1177/0022022106297301>
- Thomson, A. L., & Siegel, J. T. (2017). Elevation: A review of scholarship on a moral and other-praising emotion. *The Journal of Positive Psychology*, *12*(6), 628–638. <https://doi.org/10.1080/17439760.2016.1269184>
- Tingey, J. L., McGuire, A. P., Stebbins, O. L., & Erickson, T. M. (2019). Moral elevation and compassionate goals predict posttraumatic growth in the context of a college shooting. *The Journal of Positive Psychology*, *14*(3), 261–270. <https://doi.org/10.1080/17439760.2017.1402077>
- Van de Vyver, J., & Abrams, D. (2015). Testing the prosocial effectiveness of the prototypical moral emotions: Elevation increases benevolent behaviors and outrage increases justice behaviors. *Journal of Experimental Social Psychology*, *58*, 23–33. <https://doi.org/10.1016/j.jesp.2014.12.005>
- Walach, H., Buchheld, N., Büttenmüller, V., Kleinknecht, N., & Schmidt, S. (2006). Measuring mindfulness—The Freiburg Mindfulness Inventory (FMI). *Personality and Individual Differences*, *40*(8), 1543–1555. <https://doi.org/10.1016/j.paid.2005.11.025>
- Weathers, F. W., Litz, B. T., Keane, T. M., Palmieri, P. A., Marx, B. P., & Schnurr, P. P. (2013). The PTSD Checklist for DSM-5 (PCL-5). *Scale Available from the National Center for PTSD at www.ptsd.va.gov.*

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.