Received: 26 February 2025 | Revised: 19 November 2025 | Accepted: 8 December 2025

DOI: 10.1002/jts.70046

RESEARCH ARTICLE

Measuring resilience using language modeling: A computational approach to observing resilience

Syeda Mahwish[1,2] | **Ryan L. Boyd**[3] | **Vasudha Varadarajan**[1] | **Roman Kotov**[4] | **Benjamin J. Luft**[5] | **H. Andrew Schwartz**[1] | **Sean A. P. Clouston**[2,6]

[1]Department of Computer Science, Stony Brook University, Stony Brook, New York, USA

[2]Program in Public Health, Renaissance School of Medicine at Stony Brook University, Stony Brook, New York, USA

[3]Department of Psychology, School of Behavioral and Brain Sciences, University of Texas at Dallas, Richardson, Texas, USA

[4]Department of Psychiatry, Renaissance School of Medicine at Stony Brook University, Stony Brook, New York, USA

[5]Department of Medicine, Renaissance School of Medicine at Stony Brook University, Stony Brook, New York, USA

[6]Department of Family, Population, and Preventive Medicine, Renaissance School of Medicine at Stony Brook University, Stony Brook, New York, USA

Correspondence
Sean Clouston, Program in Public Health, 101 Nichols Rd., Health Sciences, Center, Renaissance School of Medicine at Stony Brook University, Stony Brook, NY, 11794, USA.
Email: Sean.Clouston@stonybrookmedicine.edu

Funding information
National Institute on Alcohol Abuse and Alcoholism, Grant/Award Number: R01 AA028032; National Institute for Occupational Safety and Health, Grant/Award Number: U01 OH011321; National Institute on Aging, Grant/Award Number: R01 AG049953

Abstract

We developed *resilience using language modeling* (ReLM) to measure resilience in language through a novel natural language processing approach called *archetype analysis*. Our model conceptualizes resilience as a process of maintaining healthy functioning after an adverse event. ReLM is theoretically synthesized through nine facets of resilience reviewed from various sources as reflected in language that captures its dynamic capacity: optimism, sense of social support, emotional maturity, uncertainty tolerance, flexible mindset, coping toolkit, cognitive reappraisal, belief in a higher power, and continued activities of daily living. ReLM uses a language model to embed language in a semantic space, with cosine similarity to each facet's prototype statements calculated to quantify a theoretically derived facet score. We applied ReLM to 1,859 voicemails collected from 211 responders to the September 11, 2001, World Trade Center terrorist attacks. Principal component analysis on training and test sets identified a single latent factor from the facet scores, $\lambda = 5.02$ (56% variance explained), and measurement invariance testing confirmed scalar invariance across training and test subsets, $\Delta\chi^2(8) = 8.89$, $p = .352$, indicating ReLM scores reflected the same underlying construct in both sets. A one-way analysis of variance showed significant differences in posttraumatic stress disorder (PTSD) symptom trajectories across resilience quartiles, $F(3, 169) = 5.18$, $p = .002$, with high resilience showing the largest improvements in PTSD after 4 years ($M = -0.212$). Using an archetype-based language model, ReLM offers a theoretically grounded approach to measuring resilience through natural language, capturing psychological processes in narratives, and enabling dynamic assessment.

Seneca (ca. 37–62/2008) originally suggested that hardships in life help people grow and prepare for future challenges, thereby supporting the view that resilience is a "steeling effect" built through confronting and growing from adverse experiences (Owen, 2023). For decades, resilience scholars have described resilience as the ability to respond effectively to adversity and recover from stress (Denckla et al., 2020). Definitions of resilience have evolved toward a "dynamic systems" approach, shifting from the idea of merely "bouncing back" or recovering to seeing adverse experiences as "perturbations in normal functioning" that stabilize over time (Bonanno, 2004). Therefore, here, we draw on the prevailing understanding of resilience as "a stable trajectory of healthy functioning after a highly adverse event" (Southwick et al., 2014, p. 2), emphasizing the importance of long-term processes and mechanisms in the face of ongoing challenges.

Systematic reviews of resilience measures have identified 15 putative measures but noted substantial methodological weaknesses among them (Mesman et al., 2021) and emphasized the lack of a "gold standard" resilience measure and an urgent need for reliable tools applicable across populations. For example, the Connor–Davidson Resilience Scale (CD-RISC; Connor & Davidson, 2003), Brief Resilience Scale (Smith et al., 2008), and Resilience Scale for Adults (RSA; Friborg et al., 2005) are considered reliable tools for capturing overlapping attributes of resilience, including competence, adaptability, optimism, spirituality, and inner strength. Yet, these measures are inconsistent in definition and item selection, resulting in fragmented definitions of resilience. For example, whereas the RSA stresses personal and social competence, the CD-RISC includes high standards, perseverance, and spirituality. To date, no measures reflect all of the dynamic and temporal aspects of resilience that are deemed essential for a comprehensive understanding of resilience.

Current resilience measures rely solely on self-reported or expert-rated assessments that depend on an individual's self-awareness and interpretation and can miss how resilience varies across individuals and contexts. We sought to address these limitations in existing resilience measures by developing a language-based assessment that allows resilience to be captured as an evolving and adaptive process. Natural language processing (NLP) provides a modern, systematic approach to measuring resilience-related factors in language and has been used to monitor depression and anxiety over time (Mangalik et al., 2024), assess suicide risk (Homan et al., 2022), and identify opioid and alcohol use (Matero et al., 2023; Jose et al., 2022). NLP has also outperformed self-report in predicting PTSD symptom trajectories and outcomes (Son et al., 2023; Oltmanns et al., 2021). Given NLP's potential, researchers have begun to propose that NLP is a better method for assessing psychological constructs than surveys alone (Kjell et al., 2024).

Despite important advancements in the use of NLP applications in mental health assessment, no language-based measure of resilience exists. This study proposes the first NLP-based measure of resilience using language modeling (ReLM). ReLM uses an archetype-based approach, like Varadarajan et al. (2024) used to quantify suicide risk, to capture linguistic expressions of resilience in natural language. Archetype methods go beyond traditional word-count models like linguistic inquiry and word count (LIWC; Pennebaker et al., 2001) by assessing the semantic proximity of spontaneous speech to theoretically grounded prototype statements, thereby capturing contextual meaning. ReLM conceptualizes resilience as a dynamic system that enables stable functioning following adversity. It integrates core elements of existing resilience theories within a linguistic framework, recognizing that language itself is dynamic and context-sensitive (Carling et al., 2023). This perspective allows resilience to be studied as an evolving process rather than a fixed trait, reflected naturally in the way people express themselves. We applied this approach to trauma-exposed responders to the September 11, 2001, terrorist attacks on the World Trade Center (WTC). This cohort has demonstrated diverse posttraumatic symptom trajectories—including persistent (9.7%), remitted (7.9%), and partial (5.9%) PTSD presentations (Bromet et al., 2016)—and has been studied for factors supporting long-term adaptation (Bonanno et al., 2006; Pietrzak et al., 2013). Previous research on resilience has relied almost exclusively on questionnaire-based assessments, leaving largely unexplored how these psychological processes naturally emerge in everyday language. Although some studies have employed resilience dictionary–based linguistic methods (Kang et al., 2022), no comprehensive language-based measure grounded in contemporary resilience theory exists to date. Our study fills this critical gap by providing the first comprehensive, language-based effort to capture how resilience unfolds in real-world contexts, addressing a long-standing limitation in the literature. This provides a new lens for identifying specific factors that may be strengthened to enhance resilience, allowing clinicians, researchers, and organizational leaders in high-stress environments, such as the military, health care, and emergency response, to focus their efforts more precisely rather than intervening blindly.

METHOD

Participants and procedure

Recruitment and approval

Linguistic data came from the longitudinal Personality and Health Study in trauma-exposed WTC responders (Waszczuk et al., 2019), which recruited responders with a history of clinical or subclinical PTSD symptom levels. The study received approval from the Committee on Research Involving Human Subjects (CORIHS) at Stony Brook University. All participants provided written informed consent and completed an in-person baseline interview and three follow-up interviews.

Linguistic data and PTSD symptoms

Responders completed a 2-week voicemail protocol, responding daily to two questions: "What was the worst part of your day?" and "What was the best part of your day?" Transcriptions were processed via TranscribeMe (HIPAA-compliant), resulting in 1,859 voicemails from 211 WTC responders, averaging 75 words per person per day (SD = 47.6). Demographic characteristics and PTSD symptoms, assessed using the PTSD Checklist for *DSM-5* (PCL-5; Weathers et al., 2013), were collected annually over 3 years.

Inclusion and exclusion criteria

As over 99% of responders in the monitoring program are fluent in English, only English-speaking participants were included. To ensure sufficient language data, participants needed at least two eligible voicemails (i.e., 50 or more words each; Kern et al., 2016). For PTSD trajectory analyses, we only included participants with at least two follow-up PCL-5 assessments, yielding a final sample of 174 responders.

ReLM development

The primary study aim was to develop ReLM, a language-based resilience measure grounded in an archetype-based framework that synthesizes insights from existing resilience literature and real-world narratives. To facilitate measuring resilience in a linguistic context, we determined nine facets consistent with stable functioning in the face of adversity.

Several facets, such as optimism, flexible mindset, coping toolkit, belief in a higher power, and uncertainty tolerance, were derived from common resilience measures and frameworks (Connor & Davidson, 2003; Friborg et al., 2005; Masten, 2014; Rutter, 2006; Smith et al., 2008). Though these facets are typically viewed as correlates, our approach conceptualized them as core components that constitute resilience as it manifests linguistically, reflecting their prominence across theoretical models and their role in adaptive processes. We also sought to directly observe and document natural language patterns as individuals described their lives after experiencing stressful events. To this end, we reviewed publicly available narratives from individuals who faced adversity, including veteran interviews archived by the Library of Congress (American Legion Auxiliary Unit, Aagenes, & Larson, 1942; Baby Girl Nay Nay, 2023; Harrison et al., 1991; Julie Dawn Beauty, 2018; Prior et al., 1944). Through these accounts, we observed expressions of emotional maturity, social support, cognitive reappraisal, and engagement in daily activities, themes that both aligned with existing resilience measures and reinforced their theoretical importance across diverse adversity contexts. By combining insights from empirical research and lived experiences from the publicly available narratives we examined, ReLM reflects both what resilience *is theorized to be* and how it *sounds* when expressed in everyday language. The following section defines each of the nine facets we identified as foundational to this measure.

Facets of ReLM

Optimism has been linked to resilience (Forgeard & Seligman, 2012). Thus, we defined optimism as a combination of explanatory and dispositional optimism (Carver et al., 2009; Peterson & Steen, 2009) and measured it as a general expectation that good things will happen in the future or that the future will be favorable. *Sense of social support* (SoS) differs from social support itself and is linked to resilience after adversity and supporting trauma-exposed individuals (Bonanno et al., 2007). Although definitions vary, SoS generally involves feeling cared for, valued, respected, and belonging to a group, and was defined here as the perception and experience of being cared for, valued, and part of a network of mutual assistance and social relationships. *Flexibility mindset* is defined closely as adaptable thinking and has been defined as adaptability in thought patterns and behaviors during stressful or adverse events (Bonanno, 2021; Bonanno et al., 2023; Sassenberg et al., 2022). We defined flexibility mindset as the ability and willingness to adapt one's thinking, behavior, and responses to changing situations and new information. *Continued activities of daily living* (CADL) refers to the ability to maintain daily routines and tasks following a stressful experience, regardless of whether these tasks are essential. We defined CADL as the continuation of any daily activities as a marker of perseverance and adjustment, underscoring the strength needed to maintain normalcy after adversity. *Cognitive reappraisal* is often understood as the ability to alter emotions by changing the way one thinks about something (Clark, 2022); for ReLM, we defined cognitive reappraisal as changing one's thoughts about a situation to modify its emotional impact. *Coping toolkit* focuses specifically on the behaviors or actions used during periods of adversity and was defined as a collection of strategies, techniques, and resources that individuals can use to manage stress and adversity (Folkman & Moskowitz, 2004). This toolkit is metaphorical, representing a personalized set of coping mechanisms that are drawn upon in challenging situations. *Uncertainty tolerance* involves the tolerance of ambiguity. In the literature, this facet, though not extensively explored, has been linked to resilience (Boss, 2013) and was defined here as the capacity to endure uncertainty without experiencing distress. *Belief in higher power*, which extends beyond organized religion, reflects a desire for connection and meaning (Tanyi, 2002) and was defined as the conviction in the existence of a transcendent entity or force, encompassing beliefs in God or gods, spirits, or abstract forces. *Emotional maturity* is not often emphasized in the resilience literature but is sometimes defined as the ability to understand and regulate emotions (Jobson, 2020). We defined emotional maturity as the capacity to identify, understand, and express emotions effectively

Resilience archetype

To process linguistic data, we needed prototype statements linked to each facet. Prototype statements were developed through an iterative process combining insights from the resilience literature and thematic analysis of publicly available narratives (e.g., the previously described interviews with trauma-exposed individuals and veterans). Next, we quantified prototype statements using a transformer-based encoder model—Sentence RoBERTa Large (Liu et al., 2019; Reimers & Gurevych, 2019), pretrained on a large and diverse corpus of text for natural language understanding tasks. Sentence RoBERTa Large is a language model (LM) designed to understand and analyze semantic differences between sentences. This LM was selected due to its superior performance in sentence-level semantic similarity tasks, consistently ranked among the top-performing models on the Semantic Textual Similarity Benchmark (Yang et al., 2020). To assess the generalizability of our approach, we also evaluated the model using three alternative LMs. Each LM quantified the four curated prototypical statements (e.g., "I enjoy taking walks to help destress") to facilitate facet measurement (e.g., optimism, CADL). The iterative process of forming prototype statements involved refining the statements until we achieved high consistency (i.e., Cronbach's alpha) across the sentence embeddings for each facet, providing evidence that the statements within facets clustered conceptually in the same vector space (Figure 1, Supplemental Table S1).

Measures

PTSD symptoms

PTSD symptoms were assessed using the 20-item PCL-5 (Weathers et al., 2013), which is tied to the PTSD criteria outlined in the *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.; *DSM-5*; American Psychiatric Association, 2013). Participants rated each item on a scale of 0–4, with higher scores indicating higher symptom levels (range: 0–80). The PCL-5 has demonstrated high internal consistency (Cronbach's $\alpha > .90$), including in these WTC responders ($\alpha = .95$), good test–retest reliability ($r = .82$), and robust convergent and discriminant validity for assessing PTSD symptoms (Blevins et al., 2015). Although Blevins et al. reported strong test–retest reliability, they observed lower PCL-5 scores at retest. This variability aligns with other findings (Madsen et al., 2014), highlighting that the longitudinal tracking of PTSD symptoms offers better insight into chronic PTSD than single time-point assessments, leading to the calculation of PCL-5 score changes over 4 years in this sample.

Resilience

ReLM uses an archetype-based model to enable a nuanced understanding of sentiment or emotional tone (Figure 2). Prototype statements for each facet were transformed into embeddings, and cosine similarities were computed between these statements. The similarities were averaged to generate Facet Scores, where higher values indicated stronger alignment with a given trait. A single Resilience Score was derived by combining the facet scores using principal component analysis (PCA). To ensure clarity and distinctiveness among facets, we conducted a cosine similarity analysis between the facets themselves and verified that prototype language can be reliably differentiated (cosine similarity < .55) from other facets' prototype statements.

Given the temporal variability of naturalistic language, we aggregated the data across multiple time points. Krippendorff's alpha values over 14 days were low (10–.21; Supplementary Table S2), indicating day-to-day fluctuations. To enhance reliability, we computed weighted averages of daily word counts to create stable participant-level summaries.

Data analysis

Sentence embeddings and similarity computations were implemented in Python (Version 3.9) using the archetype-based model, with code publicly available on GitHub and the Open Science Framework (OSF). Statistical validation and analyses were conducted in RStudio (Version 2024.12.0).

We computed descriptive statistics to summarize sample demographic characteristics. To assess ReLM's reliability, we randomly split the data into training ($n = 127$) and test

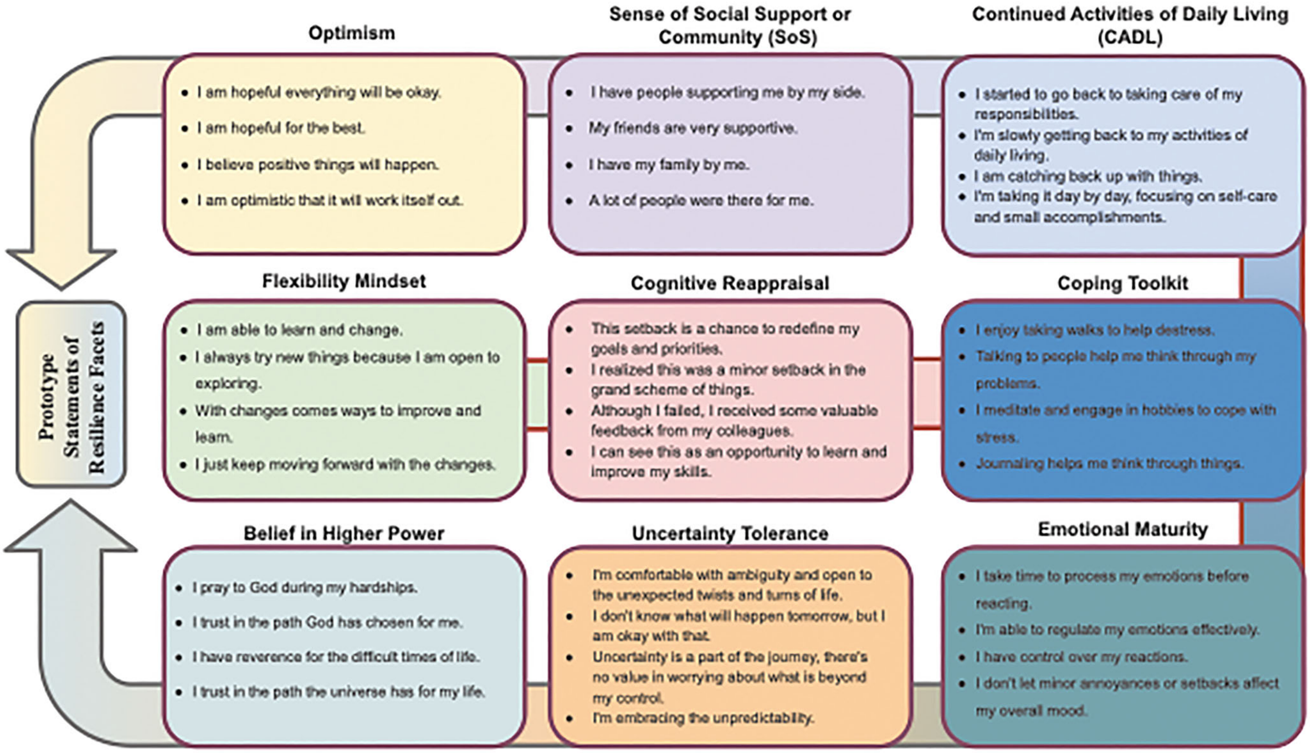

FIGURE 1 Resilience using language modeling (ReLM), showing facets and prototype statements

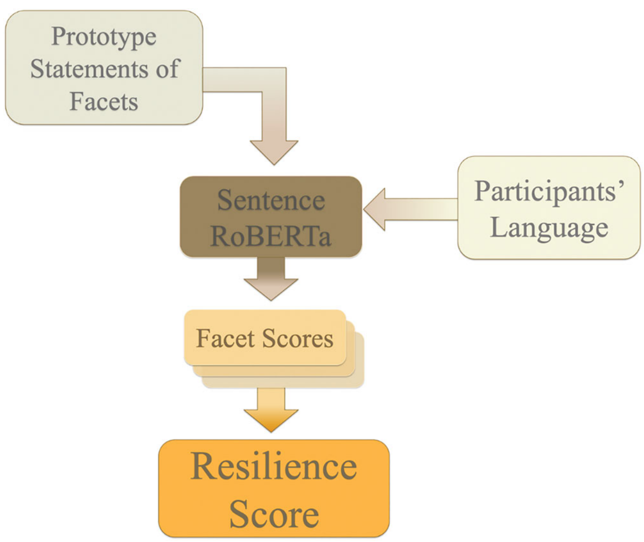

FIGURE 2 Archetype-based measure for resilience using language modeling (ReLM)

($n = 47$) sets (James et al., 2013). A larger training set was preferred over an even (i.e., 50/50) split to support richer language-based pattern detection and enable the manual refinement of prototype statements. The smaller test set was sufficient for preliminary validation and comparison. This approach aligns with common machine learning practices, ensuring robust learning while reserving data for unbiased evaluation (Goodfellow et al., 2016).

We conducted a Kaiser–Meyer–Olkin (KMO) test and Bartlett's test of sphericity to ensure PCA was appropriate. An exploratory PCA was then conducted on the training set for all nine facets to examine their dimensionality and determine underlying constructs. A subsequent PCA on the test set was performed to evaluate whether similar dimensionality and constructs were observed. Pearson correlations were calculated to examine associations among the facets.

Some researchers (e.g., van Dijk et al., 2022) caution against rigid model fit thresholds in confirmatory factor analysis (CFA), advocating for flexible approaches that consider contextual complexity. Following this guidance, we tested ReLM's measurement invariance using a stepwise procedure to ensure consistent construct measurement across the training and test sets. We first assessed configural invariance by allowing factor loadings and intercepts to vary, confirming that the same indicators measured the construct in both sets. Next, we tested metric invariance by constraining loadings and comparing model fit via chi-square differences. Finally, scalar invariance was evaluated by also constraining intercepts, ensuring equivalent measurement across groups. This stepwise testing, alongside PCA, enabled a flexible yet rigorous examination of factor structure, aligning with the view that model fit should reflect research goals and data characteristics. Overall resilience scores for the test set were then calculated using facet weights from both the test and training sets. Pearson's correlations between these scores were calculated to further validate measurement consistency. To address concerns about model specificity and ensure that the ReLM framework was not overly dependent on a single language model, we conducted a series of PCAs using sentence embeddings generated from three additional transformer-based models: MPNet (Song et al., 2020), Mixedbread AI (MXBAI; Lee et al., 2024), and BAAI BGE M3 (Chen et al., 2024).

To make a fair comparison between the ReLM archetype method and traditional approaches, we replicated the archetype procedure using a bag-of-words model via CountVectorizer (CV), computing cosine similarity between CV-derived embeddings of voicemails and facets to assess alignment with resilience constructs. We also extracted 15 LIWC constructs that are comparable to ReLM facets, including constructs of mental and emotional processes (*positive tone*, *positive emotion*, *mental*, *cognitive process*), social orientations (*social references*, *family*, *friend*), daily life activities (*lifestyle*, *leisure*, *home*, *work*), and other relevant constructs (*religion*, *past/present/future-focused*). These served to benchmark ReLM and the included facets of resilience against a traditional sentiment tool to test whether ReLM captures distinct resilience features beyond what simpler comparable linguistic measures already provide. To further examine semantic distinctiveness, we developed a control *positive sentiment* facet using the same ReLM embedding pipeline as our resilience facets. This control condition allowed us to test whether our approach specifically captured resilience-related language or merely general positive sentiment. We conducted bivariate Spearman's correlations among ReLM resilience scores, CV resilience scores, LIWC constructs, and the positive sentiment archetype to assess the degree of overlap between measures. We also included the control facet in a PCA alongside the nine resilience facets to evaluate whether ReLM captures a distinct semantic construct separate from general positive sentiment.

To further validate ReLM, we examined its association with PTSD symptom trajectories among WTC responders over 4 years. Training and test sets were combined for these convergent validity analyses. Individual PCL-5 trajectories were modeled using linear regression, with time (in days) as the predictor and PCL-5 scores as the outcome, yielding slope estimates representing symptom change over time. We employed multiple analytic approaches to capture different aspects of the association between resilience and PTSD symptoms. First, we examined associations using Spearman's correlations due to nonnormal distributions of the PCL-5 slope. Benjamini–Hochberg corrections were applied to control for the false discovery rate in multiple correlation analyses. Second, participants were categorized into resilience quartiles (very low, low, high, very high) to examine potential threshold effects and identify where differences emerged most clearly. Group differences were tested using analyses of variance (ANOVAs) with Tukey's honestly significant difference (HSD) post hoc comparisons and Cohen's d effect sizes.

TABLE 1 Demographic characteristics of World Trade Center responders in the voicemail follow-up study

Variable	*M*	*SD*
Age (years)	55.5	8.3
	n	%
Sex		
Male	153	88.4
Female	20	11.6
Race		
White	157	90.8
Non-White	16	9.2
Ethnicity		
Hispanic	10	5.8
Non-Hispanic	163	94.2
Educational attainment		
High school and some college	105	60.7
College degree and higher	68	39.3

RESULTS

Demographic characteristics

The demographic profile of the WTC responders included in the sample revealed that most participants were male (88.4%), with an average age of 55.5 years ($SD = 8.3$). Most participants identified as White (90.8%) and non-Hispanic (94.2%). Most participants ($n = 105$, 60.7%) completed high school or some college, whereas 68 (39.3%) had a college degree or higher (Table 1).

Participants left a median of 10 voicemails over the 2-week observational period, and voicemails contained an average of 1,015 words ($SD = 678$). Prototype statements within each facet demonstrated strong internal consistency, Cronbach's αs = .72–.87 (Supplementary Table S3). Cosine similarity scores ranged from .27 to .53, remaining below our predetermined threshold of .55 or lower (Supplementary Figure S1).

PCA was determined to be appropriate for both the training and test sets, showing strong sampling adequacy, KMO = .81, and a significant Bartlett's test of sphericity, χ^2 (36, n = 127) = 674.05, $p < .001$ (Table 2). Component loadings were moderately high, indicating the cohesiveness of the resilience construct. In the training set, PCA explained

TABLE 2 Factor weights for resilience facets derived from principal component analysis independently in training and test sets

	Training set (n = 127)	Test set (n = 46)
Facet		
Optimism	.71	.64
Sense of social support	.79	.70
Flexibility mindset	.68	.44
Continued activities of daily living	.73	.82
Cognitive reappraisal	.57	.62
Coping toolkit	.76	.81
Uncertainty tolerance	.78	.66
Belief in higher power	.84	.73
Emotional maturity	.82	.78

56% of the variance (maximal eigenvalue = 5.02), with a sharp drop from the first component to the second and third components, consistent with a one-factor solution (Supplementary Figure S2). Loadings ranged from .57 for cognitive reappraisal to .84 for belief in higher power. To ensure alignment with theoretical constructs, we conducted an exploratory factor analysis (EFA), which yielded similar results (Supplementary Table S3).

Measurement invariance in training and test sets

Next, we established configural invariance by confirming that the factor structure held across the training and test sets (Supplementary Table S4). We tested for metric invariance and found no significant difference between the configural and metric models, $\Delta\chi^2(8) = 9.29$, $p = .318$, indicating equivalent factor loadings across groups (Supplementary Table S5). We then assessed scalar invariance by constraining loadings and intercepts across groups and found no significant difference, $\Delta\chi^2(8) = 8.89$, $p = .352$, thereby supporting full measurement invariance (Supplementary Table S6).

Resilience score correlations between training and test sets

We evaluated the associations between overall resilience scores, calculated using component weights derived from the training set and applied to the test set, and scores calculated directly from the test set data. The analysis revealed an anticipated strong positive correlation between the two variables, $r = .999$, 95% confidence interval (CI) [.997, .999], $p < .001$ (Supplementary Figure S3).

Examination of ReLM using various transformer-based models

Tests of robustness of our ReLM approach across different transformer architectures showed high loadings (i.e., all greater than .57) across the nine resilience facets for all three transformer models (Supplementary Table S7). All four models indicated a one-factor solution, meaning the resilience facets formed a single semantic dimension across architectures, revealing a shared latent structure within the ReLM framework. The variance explained by the first principal component increased with each model (RoBERTa: 54%, MPNet: 65%, MXBAI: 74%, BGE-M3: 80%), suggesting the resilience signal became more cohesive and structured with more powerful models.

Evaluation of ReLM against conventional methods

ReLM demonstrated strong construct validity across multiple domains (Supplementary Table S8). ReLM outperformed CV resilience, the positive sentiment archetype, and LIWC positive tone when analyzing high versus low resilience. Specifically, ReLM-derived resilience scores yielded a significant group difference, $d = 0.35$, $p = .020$, whereas CV resilience scores, $d = -0.13$, $p = .410$, and LIWC positive tone, $d = 0.19$, $p = .233$, did not significantly differ. Not shown in Supplementary Table S6, Spearman's correlation revealed generally low correlations with 15 conceptually related LIWC categories, including emotional (positive tone, positive emotion, mental, cognitive process), social (social references, family, friend), daily life (lifestyle, leisure, home, work), temporal focus (past, present, future), and religious constructs, $r = -.22$–$.41$, with most correlations below .25, indicating that ReLM captures constructs distinct from simple linguistic sentiments. Specifically, the low-to-moderate correlations between ReLM facets and LIWC categories that would be expected be highly correlated, such as optimism and positive tone, $r = .03$, and positive emotions, $r = .21$, or CADL and lifestyle, $r = .12$, leisure, $r = .02$, home, $r = .20$, and work, $r = .05$, demonstrate that facets measure functional resilience capacity rather than the simple frequency of discussing certain topics. Some facets, such as SoS, demonstrated modest associations with social categories, including social references $r = .18$, family, $r = .41$, and friend, $r = .30$, reflecting expected moderately high overlap that was still limited given that the prototypes for the SoS facet were based on those topics. Similarly, belief in higher power was significantly correlated with religion, $r = .24$, though the association was modest, indicating that the belief in higher power facet captures spirituality as a process rather than general religious language.

ReLM versus general positivity

ReLM scores demonstrated no significant correlation with LIWC's positive tone, $r = .12$, or positive emotion scores, $r = .19$, suggesting that the model does not simply mirror the frequency of positive words used. The PCA analysis, which incorporated the control positive sentiment facet with the nine resilience facets derived from ReLM, revealed that the positive sentiment facet aligned strongly with a distinct principal component (.92), whereas all nine resilience facets grouped into the first component (Supplementary Table S9). The orthogonal separation of these two vectors in the PCA biplot (Supplementary Figure S3) significantly emphasizes the semantic distinction between ReLM-based resilience and general positivity.

Correlations between resilience facets

Skewness and kurtosis for all facets were within acceptable ranges. Spearman's correlations (Figure 3) between resilience facets ranged from .25 to .67, $ps < .001$. The strongest correlation was observed between coping toolkit and CADL, $r = .68$, whereas the weakest correlation was between SoS and flexibility mindset, $r = .25$.

PTSD symptom trajectories among participants with high and low resilience

Tukey's post hoc tests revealed that PTSD symptoms among participants in the high resilience quartile (Q3) improved significantly more than those in the very low (Q1), $d = -.65$, $p = .003$, and low (Q2) quartiles, $d = -.58$, $p = .015$, whereas those in the very high resilience quartile (Q4) improved less than those in Q3, $d = -.57$, $p = .015$ (Figure 4). A one-way ANOVA similarly indicated significant differences in PTSD symptom trajectories across resilience quartiles over time, $F(3,169) = 5.18$, $p = .002$. Participants in Q1 showed slight symptom worsening ($M = +0.023$/year), whereas those in Q2 showed mild symptom improvement ($M = -.009$/year). Participants in Q3 showed the most improvement ($M = -0.212$/year), whereas those in Q4 showed no change over time. Accordingly, high (i.e., Q3) resilience showed significant associations with PCL slope, Spearman's $r = -.22$, $p = .003$, even after correcting for multiple testing, whereas the very low, low, and very high (i.e., Q1, Q2, and Q3) groups exhibited different correlation patterns (Supplementary Figure S4). However, associations between ReLM and symptom trajectories were modest.

DISCUSSION

Explorations of resilience through contemporary theoretical frameworks that transcend conventional definitions—as more than just the absence of PTSD—remain elusive. This study utilized the novel ReLM measure to capture a multifaceted view of resilience through language. Incorporating an archetype-based framework alongside advanced computational techniques, ReLM extended language-based measures of psychological phenomena beyond traditional methods such as word count or frequency analyses. By aligning with established measures while advancing contemporary conceptualizations of resilience, this work highlights ReLM's transformative potential to redefine how resilience is measured and studied.

Introducing a prototype statement method for quantifying resilience via NLP, the study developed reliable and interrelated resilience facets. A unifactorial structure with strong loadings across all facets underscored the facets' interconnectedness. Measurement invariance testing confirmed consistent associations between the facets and the underlying construct across training and test sets. Scalar invariance further demonstrated robustness and comparable resilience levels across sets. Critically, when situating ReLM within the broader methodological landscape, it demonstrated superior sensitivity and specificity compared to bag-of-words vectorization and lexicon-based affective analysis, indicating a conceptual and computational advancement through its archetype-based approach. These comparisons underscore the importance of contextual cues grounded in theory, as captured by the archetypes approach used to develop ReLM. This approach cannot be reduced to the frequency of individual words or overall positivity; instead, it reflects a broader semantic landscape encompassing the domain in which resilience exists.

We addressed concerns that ReLM tracks only affective positivity by analyzing resilience separately from sentiment. The weak link between ReLM and related LIWC categories, along with the distinct PCA separation of a curated control facet, positive sentiment, supports the idea that ReLM reflects a rich, grounded construct. This distinction is vital for research and clinical applications, where resilience must be evaluated independently of mood.

Though we identified nine distinct facets of resilience, some prototype statements, such as those for cognitive reappraisal and flexibility mindset, exhibited higher similarity compared to others and might logically appear similar within the language space. Flexibility

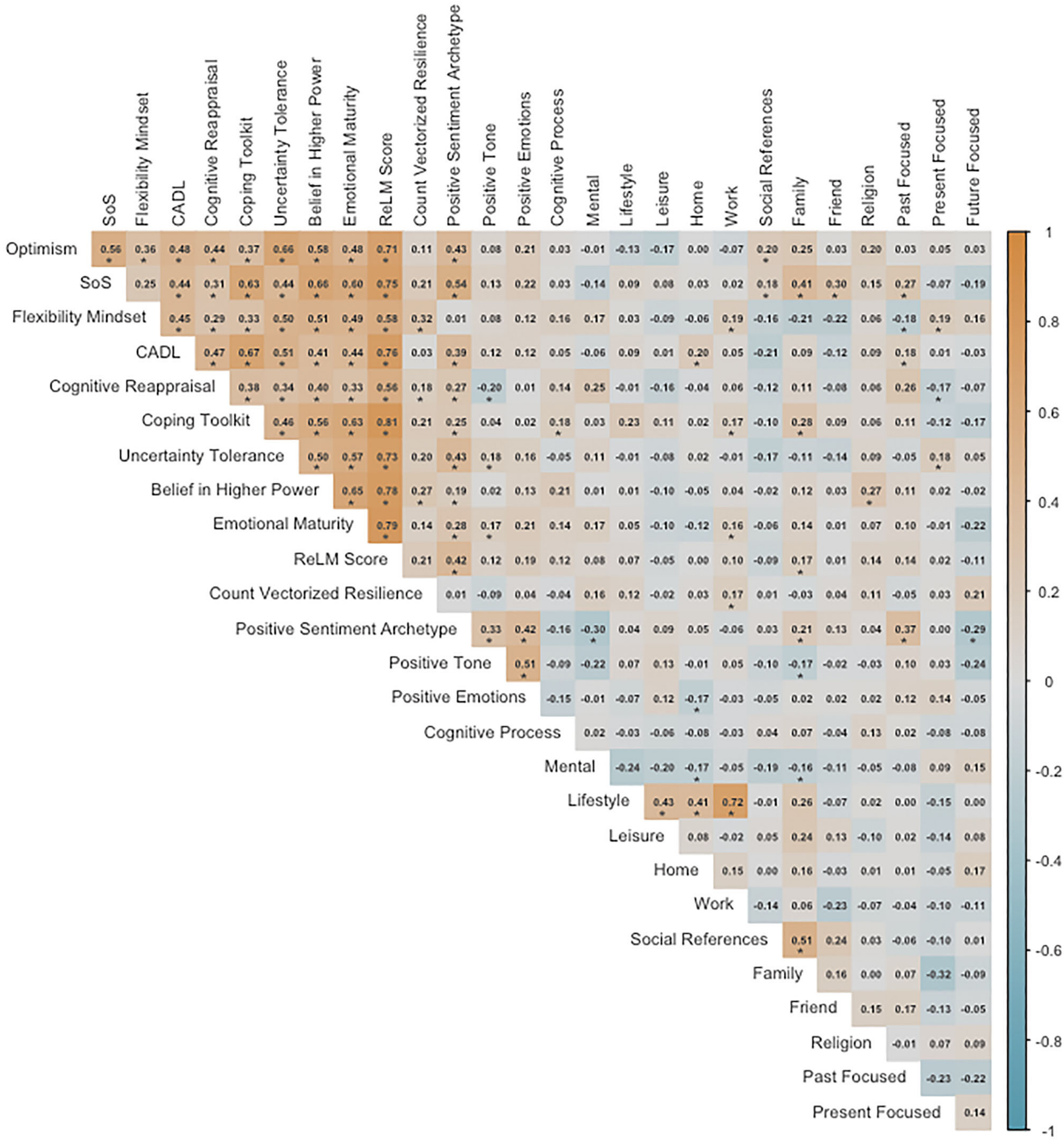

FIGURE 3 Spearman's correlations between resilience using language modeling (ReLM) facets and traditional linguistic measures
Note: SoS = sense of social support or community; CADL = continued activities of daily living.
***$p < .001$.

mindset is characterized by more open and exploratory language, whereas cognitive reappraisal is associated with emotion-focused language centered on reinterpreting a situation. This distinction was evident in this sample of WTC responders, as these two facets were only moderately correlated. These results imply that resilience facets operate in complementary but distinct ways.

The high similarity between the emotional maturity and coping toolkits facets within the archetype-based framework reflects that the ability to recognize and manage emotions is connected to the ability to effectively use coping strategies. These facets remain distinct constructs, as emotional maturity focuses on regulating emotions in a balanced way, whereas coping toolkit emphasizes the practical application of strategies to handle stressors and

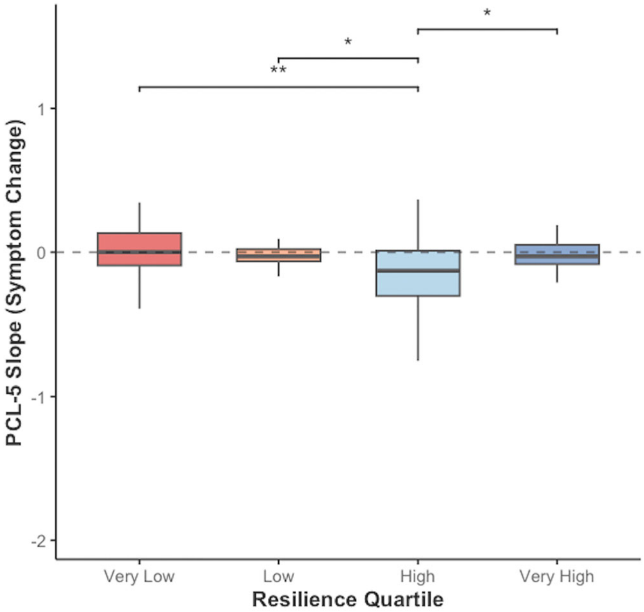

FIGURE 4 Trajectories of posttraumatic stress disorder (PTSD) symptom severity, stratified by quartiles of resilience
Note: Boxplots depict changes in PTSD Checklist for *DSM-5* (PCL-5) scores (slopes) across resilience quartiles, with negative values indicating symptom improvement. The dashed line at $y = 0$ indicates the point at which no symptom change was evident.

challenges. The strongest correlation we observed was between CADL and coping toolkit, reflecting a focus on daily routines and practical coping strategies. High correlations were also observed between SoS and resilience facets, notably coping toolkit, belief in higher power, and emotional maturity. These findings align with research highlighting that higher levels of perceived social support are associated with reduced trajectories of experienced depressive and anxiety symptoms over time (Pijnenburg et al., 2024), whereas the association between the belief in higher power and coping toolkit facets is consistent with studies emphasizing the role of spirituality in supporting coping behaviors (Schuster et al., 2001).

Analyses of multiple embedding models underscored the robustness of the ReLM framework. All tested transformer architectures—RoBERTa, MPNet, MXBAI, and BGE-M3—yielded one-factor solutions with strong loadings across the nine resilience facets, revealing a shared semantic structure not tied to any single embedding architecture. Although Sentence-RoBERTa was selected for its broad adoption and performance, future research can incorporate newer models to enhance sensitivity and semantic richness. This highlights ReLM's potential to evolve alongside the advancing NLP landscape, offering increasing utility for tracking resilience, as newer models explained a larger share of variance in the principal component.

Building on these findings, resilience quartiles demonstrated significant differences in PTSD symptom trajectories over time. As demonstrated in Supplementary Figure S4 and much of the treatment literature on PTSD and other complex disorders, associations between these conditions and their treatments are nonlinear and take on a reiterative process (Gallagher et al., 2020; Krebs et al., 2018; Stein et al., 2012). It is possible that our measure cannot fully capture the highest level of resilience because it captures active engagement with various facets. Alternatively, very high resilience might be a state of maintenance rather than active engagement, which ReLM may not fully capture without further probing related discussions on emotions, coping mechanisms, and stress processing. The restricted range of PTSD symptom change in our sample constrains conclusions about ReLM's predictive validity for symptom trajectories. The overlap in ReLM scores across trajectory groups suggests that although the measure captures differences at the extremes, its discriminative power for intermediate levels of resilience requires further validation.

Our analysis revealed that the participants in the high resilience group (Q3) showed significantly larger symptom reductions than those in the very low group (Q1), though those in the very high group (Q4) did not demonstrate this effect. We emphasize that these differences occurred within a sample with limited overall symptom change and substantial between-group overlap. However, rather than demonstrating strong predictive validity, the findings suggest that ReLM can differentiate resilience levels in ways that relate to symptom patterns, with the clinical utility of these distinctions requiring validation in more dynamic populations.

Though our research does not equate the absence of pathology with resilience, and associations with PCL-5 scores alone should not define the measure's validity, we have demonstrated ReLM's robustness through comprehensive factor analyses showing that resilience facets form a coherent construct. The quartile-based distinctions that map onto meaningful differences in long-term psychological adjustment provide additional construct validity evidence, demonstrating ReLM's ability to capture clinically relevant resilience processes that predict recovery trajectories.

Beyond the nonlinear association between resilience and PCL-5 trajectories, several limitations warrant consideration. First, although this study was unique, it used a small sample, which restricted our analytical power. Therefore, future research using ReLM should employ diverse and sufficiently large samples that allow for stronger conclusions. Additionally, the current measure has only been tested in an English-speaking, predominantly male population, raising questions about its generalizability to other groups. This highlights the need for assessments with international validity and emphasizes conceptualizing resilience in broader social contexts (Denckla et al., 2020). Future work should examine these facets in other populations. Second, the magnitude of symptom change in this sample was modest, which limits our ability to demonstrate strong predictive associations between ReLM scores and symptom trajectories. Future studies with larger and more diverse samples should employ nonlinear modeling techniques to better capture the dynamic interplay between resilience, as expressed in language over time, and psychological outcomes to fully evaluate the measure's predictive utility. In addition, the sample did not include individuals without a history of PTSD symptoms. This does not necessarily limit the applicability of ReLM to other populations; however, as the measure was developed independently and then applied here, the observed Q3 threshold for significant PTSD symptom decline may not generalize to other traumatized populations or contexts. Our cohort was also relatively homogeneous, and participants shared exposure to the WTC event, had similar careers, and made similar occupational choices, which may limit the findings' generalizability to other trauma populations. Based on the differential PCA loadings across language models (Supplementary Table S3), with larger models explaining more variance than other models, future research should examine how model architecture characteristics influence the association between model size and the variance explained in resilience measurement. Furthermore, we did not compare our measure against existing resilience scales. Although we validated the ReLM using PCL-5 symptom trajectories, a common approach in resilience research, this criterion focuses primarily on posttrauma recovery and may not fully capture the multidimensional nature of resilience as we have defined it. Comparisons with a validated subjective measure would offer more comprehensive validation, such as coping or resilience scales.

Bonanno (2021) introduced the "resilience paradox" to highlight the challenge researchers face when trying to predict resilience outcomes. This paradox arises partly due to a lack of uniformity across resilience questionnaires that measure a wide range of behaviors and traits but fail to consolidate them into a cohesive framework. ReLM addresses these limitations by integrating multiple components of resilience into one measure. This innovative approach allows ReLM to accommodate situational variability, providing a more comprehensive and context-sensitive assessment of resilience. Thus, ReLM holds the potential to target the resilience paradox by offering a method that aligns more closely with the complex and dynamic nature of resilience, advancing beyond traditional static questionnaires. Capturing resilience in its dynamic nature over time is an area emphasized by Watson et al. (2011), who highlight the importance of establishing reliable surveillance metrics to track survivors' progress and improve planning for necessary services and resources. ReLM holds significant potential in this regard, as its language-based approach allows for the mapping of resilience over time without requiring individuals to be physically tracked or assessed in a clinical setting.

In this study, we aimed to validate ReLM as a novel method that could enable researchers to naturalistically characterize resilience during routine interactions with trauma-exposed individuals, without the need for structured surveys or lengthy diagnostic assessments. PTSD can become chronic and can even worsen over time2025). This approach, therefore, meets a growing need to identify patients' preventive strengths without interrupting the clinical treatment flow, allowing for the routine adjustment of existing treatment plans based on specific resilience dimensions (e.g., enhancing social support or strengthening cognitive reappraisal) as required. By identifying resilience dimensions, ReLM could empower clinicians to improve areas in which a patient with PTSD may face challenges. These dimensions could be monitored and addressed dynamically, thereby fostering a more human-centered and relationship-driven approach to care.

AUTHOR NOTE

This research was supported by the National Institute for Occupational Safety and Health (U01 OH011321), the National Institute on Alcohol Abuse and Alcoholism (R01 AA028032), and the National Institute on Aging (R01 AG049953).

No generative AI was used while drafting this manus cript. We used ChatGPT and Grammarly to catch grammatical and typographical errors.

OPEN PRACTICES STATEMENT

This study relies on RoBERTa, which is an open-access large language model and is publicly accessible online. RStudio is publicly accessible and free to use. The code and data necessary to replicate this study are available on OSF: osf.io/3wsdq/?view_only=ce7bb3fb428343d1bfad04c234537325

AUTHOR CONTRIBUTIONS

Syeda Mahwish: conceptualization, investigation, writing - original draft, visualization, methodology, formal analysis, project administration, data curation, writing - review and editing. Ryan L. Boyd: conceptualization, investigation, methodology, formal analysis, software, data curation, supervision, resources, writing - review and editing. Vasudha Varadarajan: conceptualization, investigation, methodology, validation, writing - review and editing, formal analysis, software, data curation, supervision. Roman Kotov: conceptualization, funding acquisition, methodology, writing - review and editing, resources, data curation. Benjamin J. Luft: conceptualization, investigation, funding acquisition, writing - review and editing, project administration. H. Andrew Schwartz: conceptualization, funding acquisition, investigation, methodology, validation, writing - review and editing, software, project administration, supervision. Sean A. P. Clouston: conceptualization, investigation, funding acquisition, writing - original draft, writing - review and editing, project administration, supervision.

ORCID

Sean A. P. Clouston https://orcid.org/0000-0002-6124-0329

REFERENCES

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

American Legion Auxiliary Unit (Fertile, Minnesota)., Aagenes, D. F., & Larson, C. (1942). *Personal narrative: Donald Frederick Aagenes collection*. Library of Congress. https://www.loc.gov/item/afc2001001.44002/

Baby Girl Nay Nay. (2023, June 24). *The most emotional moment of my life | Losing my mother* [Video]. YouTube. https://www.youtube.com/watch?v=XKE4_i5DzcQ

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

Bonanno, G. A. (2004). Loss, trauma, and human resilience: Have we underestimated the human capacity to thrive after extremely aversive events? *The American Psychologist*, *59*(1), 20–28. https://doi.org/10.1037/0003-066X.59.1.20

Bonanno, G. A. (2021). The resilience paradox. *European Journal of Psychotraumatology*, *12*(1), Article 1942642. https://doi.org/10.1080/20008198.2021.1942642

Bonanno, G. A., Chen, S., & Galatzer-Levy, I. R. (2023). Resilience to potential trauma and adversity through regulatory flexibility. *Nature Reviews Psychology*, *2*(11), 663–675. https://doi.org/10.1038/s44159-023-00233-5

Bonanno, G. A., Galea, S., Bucciarelli, A., & Vlahov, D. (2006). Psychological resilience after disaster: New York City in the aftermath of the September 11th terrorist attack. *Psychological Science*, *17*(3), 181–186. https://doi.org/10.1111/j.1467-9280.2006.01682.x

Bonanno, G. A., Galea, S., Bucciarelli, A., & Vlahov, D. (2007). What predicts psychological resilience after disaster? The role of demographics, resources, and life stress. *Journal of Consulting and Clinical Psychology*, *75*(5), 671–682. https://doi.org/10.1037/0022-006X.75.5.671

Boss, P. (2013). Resilience as tolerance for ambiguity. In D. S. Becvar (Ed.), *Handbook of family resilience* (pp. 285–297). Springer Science + Business Media. https://doi.org/10.1007/978-1-4614-3917-2_17

Bromet, E. J., Hobbs, M. J., Clouston, S. A., Gonzalez, A., Kotov, R., & Luft, B. J. (2016). *DSM-IV* post-traumatic stress disorder among World Trade Center responders 11–13 years after the disaster of 11 September 2001 (9/11). *Psychological medicine*, *46*(4), 771–783. https://doi.org/10.1017/S0033291715002184

Carling, G., Cronhamn, S., Lundgren, O., Bogren Svensson, V., & Frid, J. (2023). The evolution of lexical semantics dynamics, directionality, and drift. *Frontiers in Communication*, *8*, Article 1126249. https://doi.org/10.3389/fcomm.2023.1126249

Carver, C., Scheier, M., Miller, C. J., & Fulford, D. (2009). Optimism. In C. Snyder & S. Lopez (Eds.), *Oxford handbook of positive psychology* (2nd ed., pp. 303–311). Oxford University Press.

Chen, J., Xiao, S., Zhang, P., Luo, K., Lian, D., & Liu, Z. (2024). *BGE M3-embedding: multi-lingual, multi-functionality, multi-granularity text embeddings through self-knowledge distillation*. ArXiv. https://arxiv.org/abs/2402.03216

Clark, D. A. (2022). Cognitive reappraisal. *Cognitive and Behavioral Practice*, *29*(3), 564–566. https://doi.org/10.1016/j.cbpra.2022.02.018

Connor, K. M., & Davidson, J. R. (2003). Development of a new resilience scale: The Connor–Davidson Resilience Scale (CD-RISC). *Depression and Anxiety*, *18*(2), 76–82. https://doi.org/10.1002/da.10113

Denckla, C. A., Cicchetti, D., Kubzansky, L. D., Seedat, S., Teicher, M. H., Williams, D. R., & Koenen, K. C. (2020). Psychological resilience: an update on definitions, a critical appraisal, and research recommendations. *European Journal of Psychotraumatology*, *11*(1), Article 1822064. https://doi.org/10.1080/20008198.2020.1822064

Folkman, S., & Moskowitz, J. T. (2004). Coping: pitfalls and promise. *Annual Review of Psychology*, *55*, 745–774. https://doi.org/10.1146/annurev.psych.55.090902.141456

Forgeard, M. J. C., & Seligman, M. E. P. (2012). Seeing the glass half full: A review of the causes and consequences of optimism. *Pratiques Psychologiques*, *18*(2), 107–120. https://doi.org/10.1016/j.prps.2012.02.002

Friborg, O., Hjemdal, O., Rosenvinge, J. H., & Martinussen, M. (2005). A new rating scale for adult resilience: What are the central protective resources behind healthy adjustment? *International Journal of Methods in Psychiatric Research*, *14*(1), 65–76. https://doi.org/10.1002/mpr.15

Gallagher, M. W., Long, L. J., & Phillips, C. A. (2020). Hope, optimism, self-efficacy, and posttraumatic stress disorder: A meta-analytic review of the protective effects of positive expectancies. *Journal of Clinical Psychology*, *76*(3), 329–355. https://doi.org/10.1002/jclp.22882

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

Harrison, H. R., Scott, K. M., Women in Military Service For America Memorial Foundation, Inc., & Reborchuk, M. (1991). *Personal narrative: Holly R. Harrison collection*. Library of Congress. https://www.loc.gov/item/afc2001001.46660/

Homan, S., Gabi, M., Klee, N., Bachmann, S., Moser, A. M., Duri', M., Michel, S., Bertram, A. M., Maatz, A., Seiler, G., Stark, E., & Kleim,

B. (2022). Linguistic features of suicidal thoughts and behaviors: A systematic review. *Clinical Psychology Review*, *95*, Article 102161. https://doi.org/10.1016/j.cpr.2022.102161

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112, p. 18). Springer.

Jobson, M. C. (2020). Emotional maturity among adolescents and its importance. *Indian Journal of Mental Health*, *7*(1), 35–41.

Jose, R., Matero, M., Sherman, G., Curtis, B., Giorgi, S., Schwartz, H. A., & Ungar, L. H. (2022). Using Facebook language to predict and describe excessive alcohol use. *Alcoholism*, *Clinical and Experimental Research*, *46*(5), 836–847. https://doi.org/10.1111/acer.14807

Julie Dawn Beauty. (2018, January 18). *How I got over my divorce to find love & happiness* [Video]. YouTube. https://www.youtube.com/watch?v=-mHAK8GDAAo

Kang, Y.-B., McCosker, A., Kamstra, P., & Farmer, J. (2022). Resilience in web-based mental health communities: Building a resilience dictionary with semiautomatic text analysis. *JMIR Formative Research*, *6*(9), Article e39013. https://doi.org/10.2196/39013

Kern, M. L., Park, G., Eichstaedt, J. C., Schwartz, H. A., Sap, M., Smith, L. K., & Ungar, L. H. (2016). Gaining insights from social media language: Methodologies and challenges. *Psychological Methods*, *21*(4), 507–525. https://doi.org/10.1037/met0000091

Kjell, O. N., Kjell, K., & Schwartz, H. A. (2024). Beyond rating scales: With targeted evaluation, large language models are poised for psychological assessment. *Psychiatry Research*, *333*, Article 115667. https://doi.org/10.1016/j.psychres.2023.115667

Krebs, P., Norcross, J. C., Nicholson, J. M., & Prochaska, J. O. (2018). Stages of change and psychotherapy outcomes: A review and meta-analysis. *Journal of Clinical Psychology*, *74*(11), 1964–1979. https://doi.org/10.1002/jclp.22683

Lee, S., Shakir, A., Koenig, D., & Lipp, J. (2024). *Open source strikes bread—new fluffy embeddings model*. Mixedbread AI. https://www.mixedbread.ai/blog/mxbai-embed-large-v1

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). *RoBERTa: A robustly optimized BERT pretraining approach*. ArXiv. https://arxiv.org/abs/1907.11692

Madsen, T., Karstoft, K. I., Bertelsen, M., & Andersen, S. B. (2014). Postdeployment suicidal ideations and trajectories of posttraumatic stress disorder in Danish soldiers: A 3-year follow-up of the USPER study. *Journal of Clinical Psychiatry*, *75*(9), 994–1000. https://doi.org/10.4088/JCP.13m08910

Mangalik, S., Eichstaedt, J. C., Giorgi, S., Mun, J., Ahmed, F., Gill, G., Ganesan, A. V., Clouston, S. A. P., Pinjari, O., Nobles, A. L., Bellisario, G., Fineberg, S., Lee, M., Curtis, B., Ungar, L. H., & Schwartz, H. A. (2024). Robust language-based mental health assessments in time and space through social media. *NPJ Digital Medicine*, *7*(1), Article 109. https://doi.org/10.1038/s41746-024-01100-0

Mann, F. D., Waszczuk, M. A., Clouston, S. A., Feltman, S., Ruggero, C. J., Marx, B. P., Schwartz, J. E., Bromet, E. J., Luft, B. J., & Kotov, R. (2025). A 20-year longitudinal cohort study of post-traumatic stress disorder in World Trade Center responders. *Nature Mental Health*, *3*(7), 789–802. https://doi.org/10.1038/s44220-025-00419-1

Masten, A. S. (2014). Global perspectives on resilience in children and youth. *Child Development*, *85*(1), 6–20. https://doi.org/10.1111/cdev.12205

Matero, M., Giorgi, S., Curtis, B., Ungar, L. H., & Schwartz, H. A. (2023). Opioid death projections with AI-based forecasts using social media language. *NPJ Digital Medicine*, *6*(1), Article 35. https://doi.org/10.1038/s41746-023-00776-0

Mesman, E., Vreeker, A., & Hillegers, M. (2021). Resilience and mental health in children and adolescents: An update of the recent literature and future directions. *Current Opinion in Psychiatry*, *34*(6), 586–592. https://doi.org/10.1097/YCO.0000000000000741

Owen, J. (2023). Psychological resilience: Connecting contemporary psychology to ancient practical philosophy. *Theory & Psychology*, *33*(3), 366–385. https://doi.org/10.1177/09593543231153820

Oltmanns, J. R., Schwartz, H. A., Ruggero, C., Son, Y., Miao, J., Waszczuk, M., Clouston, S. A. P., Bromet, E. J., Luft, B. J., & Kotov, R. (2021). Artificial intelligence language predictors of two-year trauma-related outcomes. *Journal of Psychiatric Research*, *143*, 239–245. https://doi.org/10.1016/j.jpsychires.2021.09.015

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). *Linguistic inquiry and word count: LIWC 2001*. Lawrence Erlbaum Associates.

Peterson, C., & Steen, T. (2009). Optimistic explanatory style. In C. Snyder & S. Lopez (Eds.), *Oxford handbook of positive psychology* (2nd ed., pp. 313–321). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199396511.013.70

Pietrzak, R. H., Feder, A., Singh, R., Schechter, C. B., Bromet, E. J., Katz, C. L., Reissman, D. B., Ozbay, F., Sharma, V., Crane, M., Harrison, D., Herbert, R., Levin, S. M., Luft, B. J., Moline, J. M., Stellman, J. M., Udasin, I. G., Landrigan, P. J., & Southwick, S. M. (2013). Trajectories of PTSD risk and resilience in World Trade Center responders: An 8-year prospective cohort study. *Psychological Medicine*, *44*(1), 205–219. https://doi.org/10.1017/s0033291713000597

Pijnenburg, L. J., Velikonja, T., Pietrzak, R. H., DePierro, J., de Haan, L., Todd, A. C., Dasaro, C. R., Feder, A., & Velthorst, E. (2024). Perceived social support and longitudinal trajectories of depression and anxiety in World Trade Center responders. *Social Psychiatry and Psychiatric Epidemiology*, *59*(8), 1413–1424. https://doi.org/10.1007/s00127-023-02569-y

Prior, D. F., Dent, Jamie S., Stories of Service–San Diego., Smith, C., T, A., & Unknown. (1944). *Personal narrative: Donald F. Prior collection*. Library of Congress. https://www.loc.gov/item/afc2001001.47925

Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP*; pp. 3980–3990). https://doi.org/10.18653/v1/d19-1410

Rutter, M. (2006). Implications of resilience concepts for scientific understanding. *Annals of the New York Academy of Sciences*, *1094*(1), 1–12. https://doi.org/10.1196/annals.1376.002

Sassenberg, K., Winter, K., Becker, D., Ditrich, L., Scholl, A., & Moskowitz, G. B. (2022). Flexibility mindsets: Reducing biases that result from spontaneous processing. *European Review of Social Psychology*, *33*(1), 171–213. https://doi.org/10.1080/10463283.2021.1959124

Seneca. (2008). *Dialogues and essays* (J. Davie, Trans.). Oxford University Press. (Original work published ca. 37–62).

Schuster, M. A., Stein, B. D., Jaycox, L. H., Collins, R. L., Marshall, G. N., Elliott, M. N., Zhou, A. J., Kanouse, D. E., Morrison,

J. L., & Berry, S. H. (2001). A national survey of stress reactions after the September 11, 2001, terrorist attacks. *New England Journal of Medicine*, *345*(20), 1507–1512. https://doi.org/10.1056/nejm200111153452024

Smith, B. W., Dalen, J., Wiggins, K., Tooley, E., Christopher, P., & Bernard, J. (2008). The Brief Resilience Scale: Assessing the ability to bounce back. *International Journal of Behavioral Medicine*, *15*(3), 194–200. https://doi.org/10.1080/10705500802222972

Son, Y., Clouston, S. A. P., Kotov, R., Eichstaedt, J. C., Bromet, E. J., Luft, B. J., & Schwartz, H. A. (2023). World Trade Center responders in their own words: Predicting PTSD symptom trajectories with AI-based language analyses of interviews. *Psychological Medicine*, *53*(3), 918–926. https://doi.org/10.1017/S0033291721002294

Song, K., Tan, X., Qin, T., Lu, J., & Liu, T.-Y. (2020). MPNet: Masked and permuted pre-training for language understanding. *Advances in Neural Information Processing Systems*, *33*, 16857–16867.

Southwick, S. M., Bonanno, G. A., Masten, A. S., Panter-Brick, C., & Yehuda, R. (2014). Resilience definitions, theory, and challenges: Interdisciplinary perspectives. *European Journal of Psychotraumatology*, *5*(1), Article 25338. https://doi.org/10.3402/ejpt.v5.25338

Stein, N. R., Dickstein, B. D., Schuster, J., Litz, B. T., & Resick, P. A. (2012). Trajectories of response to treatment for posttraumatic stress disorder. *Behavior Therapy*, *43*(4), 790–800. https://doi.org/10.1016/j.beth.2012.04.003

Tanyi, R. A. (2002). Towards clarification of the meaning of spirituality. *Journal of Advanced Nursing*, *39*(5), 500–509. https://doi.org/10.1046/j.1365-2648.2002.02315.x

Varadarajan, V., Lahnala, A., Ganesan, A. V., Dey, G., Mangalik, S., Bucur, A. M., Soni, N., Rao, R., Lanning, K., Vallejo, I., Flek, L., Schwartz, H. A., Welch, C., & Boyd, R. (2024). Archetypes and entropy: Theory-driven extraction of evidence for suicide risk. In A. Yates, B. Desmet, E. Prud'hommeaux, A. Zirikly, S. Bedrick, S. McAvaney, K. Bar, M. Ireland, & Y. Ophir (Eds.), *Proceedings of the 9th Workshop on Computational Linguistics and Clinical Psychology (CLPsych 2024*; pp. 278–291). Association for Computational Linguistics.

van Dijk, W., Schatschneider, C., Al Otaiba, S., & Hart, S. A. (2022). Assessing measurement invariance across multiple groups: When is fit good enough? *Educational and Psychological Measurement*, *82*(3), 482–505. https://doi.org/10.1177/00131644211023567

Watson, P. J., Brymer, M. J., & Bonanno, G. A. (2011). Postdisaster psychological intervention since 9/11. *American Psychologist*, *66*(6), 482–494. https://doi.org/10.1037/a0024806

Waszczuk, M. A., Ruggero, C., Li, K., Luft, B. J., & Kotov, R. (2019). The role of modifiable health-related behaviors in the association between PTSD and respiratory illness. *Behaviour Research and Therapy*, *115*, 64–72. https://doi.org/10.1016/j.brat.2018.10.018

Weathers, F. W., Litz, B. T., Keane, T. M., Palmier, P. A., Marx, B. P., & Schnurr, P. P. (2013). *The PTSD Checklist for DSM-5 (PCL-5)*. https://www.ptsd.va.gov/professional/assessment/adult-sr/ptsd-checklist.asp

Yang, X., He, X., Zhang, H., Ma, Y., Bian, J., & Wu, Y. (2020). Measurement of semantic textual similarity in clinical texts: Comparison of transformer-based models. *JMIR Medical Informatics*, *8*(11), Article e19735. https://doi.org/10.2196/19735

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Mahwish, S., Boyd, R. L., Varadarajan, V., Kotov, R., Luft, B. J., Schwartz, H. A., & Clouston, S. A. P. (2026). Measuring resilience using language modeling: A computational approach to observing resilience. *Journal of Traumatic Stress*, 1–14. https://doi.org/10.1002/jts.70046