Personality and Individual Differences 257 (2026) 113785

Contents lists available at ScienceDirect

Personality and Individual Differences

journal homepage: www.elsevier.com/locate/paid

The words can harm scale: Measuring beliefs about harmful speech

Samuel Pratt [a,*], Payton J. Jones [c], Benjamin W. Bellet [b], Richard J. McNally [c], Kurt Gray [d]

[a] Department of Psychology, University of California, Los Angeles, United States of America
[b] Massachusetts Mental Health Center, Boston, MA, United States of America
[c] Department of Psychology, Harvard University, United States of America
[d] Department of Psychology, Ohio State University, United States of America

ARTICLE INFO

Keywords:
Words Can Harm Scale
Harmful speech
Moral psychology
Political correctness

ABSTRACT

People differ in their belief that speech can cause lasting psychological harm. We present the ten-item Words Can Harm Scale (WCHS) as a valid and reliable measure of this belief. Items assess attitudes about harmful speech (e.g., "Vulnerable people should not be exposed to certain kinds of speech, as this might harm them") and written words (e.g., "I could be left emotionally scarred by something I read"). In a representative sample of U.S. adults (N = 956), the WCHS demonstrated strong internal consistency (α = 0.92) and robust two-week test-retest reliability (r = 0.80). People higher in the belief that words can harm tended to be younger, female, non-White, and politically liberal. People with higher WCHS scores rated themselves as higher in intellectual humility, empathy, moral grandstanding, and the belief in the importance of silencing others. They were also more likely to support political correctness and endorse trigger warnings and safe spaces. People who believed that words can harm had worse mental health: they reported being more anxious and depressed, less resilient, and having more difficulties in emotion regulation. The WCHS is a reliable tool for measuring beliefs about the harmfulness of words—a divisive issue within modern cultural discourse.

1. Introduction

In some cases, it seems clear that words can harm. For example, Oliver Wendell Holmes famously observed that "falsely shouting fire in a theatre and causing a panic" could cause harm (Schenck v. United States, 1919). Other common examples of potentially harmful speech include direct threats of violence or calls to enact imminent violence. However, whereas legal boundaries have often focused on speech associated with direct physical harms (e.g., physical violence), speech may also have adverse psychological consequences.

First, consider an extreme case, in which a mother is extremely verbally abusive to a child over a period of years: berating, manipulating, and threatening her child daily. Such abusive speech has strong psychological consequences, even in the absence of physical abuse (Teicher et al., 2006). The distinction between psychological harm and physical harm becomes further blurred when considering that persistent verbal abuse is correlated with physical differences in the brain (Choi et al., 2009) and inflammation in the body (Miller & Chen, 2010). In this sense, the psychological harm caused by extremely threatening speech could be considered a type of physical harm (Feldman Barrett, 2017).

Yet some types of speech are likely generative of stress but may or may not cause long-term psychological harm. How harmful are insults? Stressful confrontations? Racial slurs? Criticisms of one's religion? Lectures about genocide or sexual violence? These questions animate contemporary debates about the limits of free speech, including whether we ought to provide trigger warnings in the classroom, mandate microaggression training in the workplace, or allow controversial speakers a platform on college campuses. At the heart of these debates lies an important belief: the belief that words can cause genuine and lasting harm.

1.1. The belief that words can harm

People likely vary in their belief that words can cause lasting psychological harm, and this belief may have significant cultural implications. For example, efforts to deplatform or otherwise silence controversial speakers on college campuses are sometimes rooted in the argument that the speaker's words are a form of violence against vulnerable groups (Feldman Barrett, 2017; Lukianoff & Haidt, 2018). The argument that words can harm has also been used to justify institutional efforts to prohibit certain language. In 2022, the Stanford University IT department launched a program entitled the "Elimination

* Corresponding author at Department of Psychology, University of California, Los Angeles, 1285 Franz Hall, Los Angeles, CA, 90095, United States of America.
E-mail address: sampratt@g.ucla.edu (S. Pratt).

https://doi.org/10.1016/j.paid.2026.113785
Received 28 September 2025; Received in revised form 9 March 2026; Accepted 16 March 2026

of Harmful Language Initiative," which proposed removing from Stanford's websites a set of words considered to be biased or discriminatory, including "blind review," "handicap parking," "OCD," "victim," "immigrant," and "American" (The Wall Street Journal, 2022). The initiative was quickly paused due to strong pushback, presumably from people who disagreed that these words are harmful. Understanding when and why people view speech as harmful—as opposed to merely rude, distasteful, or incorrect—is important because harmful acts are likely to be moralized, giving rise to outrage and calls for punishment (Gray & Pratt, 2025; Pratt, Rosenfeld, et al., 2025).

Who is most likely to believe that words can harm? Based on past discussions, we might expect this belief to be stronger among those who are younger and more politically liberal (Haslam, 2016; Lukianoff & Haidt, 2018). Likewise, groups that are more often targeted by aggressive or derogatory speech—including women, racial and ethnic minorities—may have stronger beliefs that speech can harm given their past personal experiences (Beth Nielsen, 2000).

Personality and clinical variables may also predict the belief that words can harm. People high in negative emotionality might be more sensitive to minor or unintentional verbal slights (Bleske-Rechek et al., 2023). At the same time, people high in empathy might endorse this belief out of concern for protecting others from prejudice (McGrath et al., 2019). The clinical concept of anxiety sensitivity (Reiss et al., 1986), the belief that the experience of anxiety is harmful or dangerous, may also be important. Anxiety sensitivity may be one reason why individuals might believe that words are dangerous—they may fear even short-term experiences of anxiety.

1.2. The Words Can Harm Scale (WCHS)

Understanding the psychology of the belief that words can harm requires confirming that there is a reliable and valid way to measure this belief—which is the goal of our study. The Words Can Harm Scale (WCHS) is a ten-item scale originally developed to capture individual differences in the belief that words can cause lasting psychological harm (Bellet et al., 2018). The scale has been used in several published papers (Bellet et al., 2018; Celniker et al., 2022; Jones et al., 2020; Pratt, Jones, et al., 2025), demonstrating strong internal consistency (Cronbach's α = 0.89–0.92). However, no study to date has provided a full psychometric validation of the measure. This study is the first to formally assess the factor structure, reliability, and validity of the WCHS in a large, nationally representative sample. We examined how the scale correlates with a network of related constructs including demographic variables, personality and individual difference variables (e.g., emotional stability), social and moral belief variables (e.g., political ideology), and clinical variables (e.g., anxiety). We conclude by offering several practical future uses of the WCHS.

2. Method

This study was approved by the University of North Carolina at Chapel Hill Office of Human Research Ethics (IRB23-1911). The preregistered study design and analysis plan along with all data, materials, and code are publicly available on the Open Science Framework (https://osf.io/gf8j5/?eda628f546f24fa09beaa11beeebbcc1). All analyses were conducted in the R software environment (Version 4.4.2; R Core Team, 2024).

2.1. Participants

Participants were adult U.S. residents recruited online via Prolific. We aimed to recruit approximately 1000 participants based on Comrey and Lee's (2013) recommendations for exploratory factor analysis, which classifies samples of this size as "excellent" because they substantially reduce sampling error in the correlation matrix and yield more stable, replicable estimates of factor loadings and latent structure. We received 1061 survey responses, and after applying our preregistered exclusion criteria, the final sample consisted of 956 participants. To assess test-retest reliability, we recontacted participants two weeks later to complete the WCHS a second time; 756 completed the follow-up (79% retention). We used Prolific's nationally representative sample option to stratify participants based on age, sex, and ethnicity. The final sample had a mean age of 46.15 (SD = 15.76; range 18–83) and was evenly split between females (49.8%) and males (48.7%) with 1.4% non-binary. Our sample closely mirrored the U.S. population in terms of age, gender, and race. Full demographic and exclusion details are provided in the Supplementary Materials.

2.2. Procedure and measures

After consenting and passing all screening checks, participants completed the WCHS and a set of personality and individual difference measures, social and moral belief measures, and clinical measures in a randomized order, followed by demographic questions. The full survey, including all scale items and anchors, is available on the OSF page (https://osf.io/gf8j5/?eda628f546f24fa09beaa11beeebbcc1). Participants were compensated $4 for their participation in the 25-minute study and an additional $1 for completing the two-week follow-up.

2.2.1. The Words Can Harm Scale (WCHS)

The WCHS is a ten-item scale measuring the belief that words can cause lasting psychological harm. Items assess attitudes about speech in general (e.g., "Vulnerable people should not be exposed to certain kinds of speech, as this might harm them") as well as several items more specific to reading words (e.g., "I could be left emotionally scarred by something I read"). Participants read the prompt "Please read the following statements, and indicate your level of agreement with each one" and provided responses on a sliding scale (1 = *strongly disagree*, 25 = *somewhat disagree*, 50 = *neither agree nor disagree*, 75 = *somewhat agree*, 100 = *strongly agree*). The full item wordings are displayed in Table 1.

2.2.2. Demographic variables

Participants reported their age, gender, race, highest level of education, religious affiliation, total household income (i.e., SES), and completed the MacArthur Scale of Subjective Social Status (Adler et al., 2000).

2.2.3. Personality and individual differences variables

The Big Five dimensions of personality (extraversion, agreeableness, conscientiousness, emotional stability, and openness) were assessed using the Ten-Item Personality Inventory (TIPI; Gosling et al., 2003). The General Self-Efficacy Scale (GSE; Schwarzer & Jerusalem, 1995) measured participants' perceived ability to problem-solve and overcome challenges. Intellectual humility was measured using the Leary Intellectual Humility Scale (Leary et al., 2017). Trait empathy was measured using the empathic concern subscale of the Interpersonal Reactivity Index (IRI; Davis, 1980). The Tendency for Interpersonal Victimhood scale (TIV; Gabay et al., 2020) measured the disposition to view oneself as a victim across four dimensions: need for recognition, moral elitism, lack of empathy, and rumination.

2.2.4. Social and moral belief variables

Political ideology was measured on a 7-point Likert scale (1 = *very liberal*, 7 = *very conservative*), and political party affiliation was measured using a multiple-choice question. Right-wing authoritarianism was measured using the Very Short Authoritarianism scale (VSA; Bizumic & Duckitt, 2018). The 13-item Left-Wing Authoritarianism Scale (Costello & Patrick, 2023) measured left-wing manifestations of authoritarianism across three dimensions: anti-hierarchical aggression (LWA-13 AHA), anti-conventionalism (LWA-13 AC), and top-down censorship (LWA-13 TDC).

Several measures assessed specific cultural attitudes. The Trigger

Table 1
Factor loadings and descriptive statistics of the WCHS (α = 0.92).

Item	*M* (*SD*)	Factor loading	Corrected item-total correlation
1. I could be left emotionally scarred by something I read.	45.45 (29.44)	0.67	0.67
2. I could be traumatized without ever being touched, just through someone's hurtful words.	59.00 (30.37)	0.73	0.71
3. Reading a book can be emotionally damaging, depending on who is reading it.	51.96 (27.82)	0.67	0.67
4. A person might develop posttraumatic stress disorder or at least some of its symptoms from something they read.	47.02 (27.34)	0.69	0.68
5. I should be careful about what I say, as it could permanently damage someone's emotional health.	66.05 (26.63)	0.82	0.77
6. Vulnerable people should not be exposed to certain kinds of speech, as this might harm them.	53.84 (27.08)	0.72	0.68
7. Even if I try to think about them in a different way, hurtful words could be damaging nonetheless.	68.93 (24.88)	0.77	0.73
8. Exposing someone to a triggering idea can seriously damage their mental health.	62.27 (26.89)	0.83	0.79
9. There is great power in the words we choose, either to heal others or to permanently harm them.	77.19 (22.32)	0.68	0.62
10. Even a simple phrase can be emotionally traumatizing for someone vulnerable.	66.40 (25.84)	0.82	0.77

Warnings Attitudes Assessment (TWAA; Bellet et al., 2018) measured support for trigger warnings. The Safe Spaces Attitudes Assessment (SSAA; Pratt, Jones, et al., 2025) measured support for designating the classroom as a "safe space." The Concern for Political Correctness scale (CPC; Strauts & Blanton, 2015) measured the tendency to call out or be emotionally affected by politically incorrect language. The Belief in the Importance of Silencing Others scale (BISO; Tsfati & Dvir-Gvirsman, 2018) measured support for suppressing harmful viewpoints.[1] The Moral Grandstanding Motivation scale (Grubbs et al., 2019) measured the tendency to share one's moral beliefs as a means of gaining social status.

2.2.5. Clinical variables

The GAD-7 measured symptoms of Generalized Anxiety Disorder (Spitzer et al., 2006) and the Patient Health Questionnaire (PHQ-9; Kroenke et al., 1999) measured symptoms of Major Depressive Disorder, both over the past two weeks. The Difficulties in Emotion Regulation Scale Short Form (DERS-SF; Kaufman et al., 2016)[2] measured common features of emotion dysregulation. The Brief Resilience Scale (BRS; Smith et al., 2008) measured perceived ability to recover from stressful events. The Anxiety Sensitivity Index (ASI; Reiss et al., 1986) measured fear of anxiety-related symptoms. Finally, the Perceived Post-Traumatic Vulnerability Scale (Bellet et al., 2018) measured participants' beliefs about vulnerability to trauma for themselves (PPV-S) and others (PPV-O). Participants read an example trauma scenario and rated the likelihood of experiencing symptoms of posttraumatic stress disorder.

3. Results

3.1. Factor structure and descriptive statistics

We conducted an exploratory factor analysis (EFA) on the WCHS using maximum likelihood extraction and oblimin rotation. A parallel analysis suggested retaining up to three factors; however, the scree plot indicated a clear elbow after the first factor (see Fig. 1). All items loaded strongly on a single factor (range = 0.67–0.83; see Table 1), and a one-factor solution explained 55.1% of the total variance. Internal consistency was high (α = 0.92), and the average corrected item-total correlation was $r = 0.71$.

3.2. Confirmatory factor analysis

To evaluate whether the single factor structure of the WCHS replicated in a second sample, we conducted a confirmatory factor analysis (CFA) using data from the two-week follow up assessment ($n = 756$).[3] A one-factor model was specified in which all ten WCHS items loaded onto a single latent factor. The model was estimated by using maximum likelihood with robust (MLR) standard errors, and missing data were handled using full information maximum likelihood.

All items loaded strongly and significantly onto the latent factor (standardized loadings ranged from 0.66 to 0.84), closely replicating the factor loading pattern observed in the exploratory factor analysis. An evaluation of eigenvalues also clearly indicated a single factor structure. Model fit was acceptable to moderate for CFI and SRMR (CFI = 0.89; SRMR = 0.06), but poor for χ^2 and RMSEA ($\chi^2(35) = 571.27$, $p < .001$, RMSEA = 0.14), likely reflecting some misfit in the item covariances. Taken together, the CFA provides confirmatory evidence that the WCHS is well characterized by a single latent dimension and that this factor structure is stable across time.

3.3. Test-retest reliability

We assessed the two-week test-retest reliability of the WCHS in a subsample of participants ($n = 756$). A two-week interval between measurements is sufficiently long to reduce recall effects while remaining short enough that meaningful attitude change is unlikely (DeVon et al., 2007). The correlation between scores at Time 1 and Time 2 was high, $r = 0.80$, $p < .001$, indicating strong stability over time.

3.4. Construct validity

To examine the construct validity of the WCHS, we assessed its associations with theoretically relevant demographic, personality, social and moral belief, and clinical variables.

3.4.1. Associations between the WCHS and demographic variables

The WCHS was negatively correlated with age ($r = -0.10$), consistent with past research (Celniker et al., 2022). Individuals who scored higher on the WCHS were significantly more likely to be female ($r = 0.19$) and non-White ($r = 0.10$). To explore racial group differences more fully, we conducted a one-way ANOVA comparing mean WCHS scores across self-identified racial groups with at least 30 observations: White/European American ($n = 587$), Black/African American ($n = 122$), Asian ($n = 60$), and Hispanic/Latinx ($n = 65$). The overall effect of race was significant, $F(3, 830) = 11.34$, $p < .001$, $\eta^2 = 0.039$. Tukey's

[1] We reworded one item of the BISO to refer to "others" instead of "Israelis." The adapted scale demonstrated excellent internal consistency in our sample (α = 0.92).

[2] We excluded the three items comprising the Awareness subscale of the DERS-SF due to construct validity concerns documented in prior research (Bhat et al., 2025).

[3] This analysis was conducted at the request of a reviewer and was therefore not preregistered.

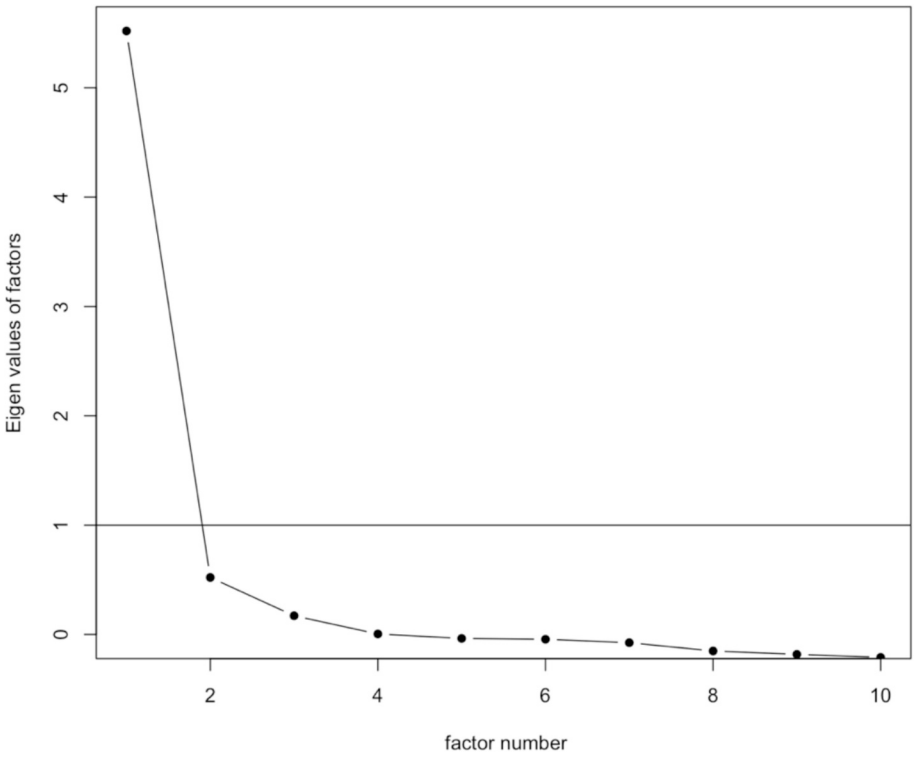

Fig. 1. Scree plot of eigenvalues for the WCHS.
Note. Eigenvalues were obtained from factor analysis. A clear elbow appears after the first factor, supporting a one-factor solution.

post hoc comparisons revealed that Black participants scored significantly higher on the WCHS than both White participants ($M_{diff} = 11.33$, $p < .001$) and Asian participants ($M_{diff} = 10.11$, $p = .008$). The full set of correlations is displayed in Table 2.

3.4.2. *Associations between the WCHS and personality and individual difference variables*

People with higher WCHS scores reported a greater tendency for interpersonal victimhood ($r = 0.24$) and lower emotional stability ($r = -0.18$). The WCHS was positively correlated with self-reported intellectual humility ($r = 0.18$), closely mirroring past work (Celniker et al., 2022). However, this finding should be interpreted cautiously given that intellectual humility questionnaires may track overconfidence in one's knowledge rather than objective metacognitive abilities (Costello et al., 2023).

People with higher WCHS scores rated themselves as more empathic ($r = 0.32$), consistent with prior research showing that people who endorse broader definitions of harm-related concepts (e.g., prejudice; bullying) tend to score higher on the IRI empathic concern subscale (McGrath et al., 2019). Surprisingly, the WCHS also showed a small correlation with the TIV lack of empathy subscale ($r = 0.08$). One possible explanation is that the IRI empathic concern subscale assesses general empathy (e.g., "I often have tender, concerned feelings for people less fortunate than me") whereas the TIV lack of empathy subscale arguably reflects more specific attitudes towards one's adversaries (e.g., "People who claim I have behaved wrongly want me to admit it so they can take advantage of the situation"). The full set of correlations is displayed in Table 3.

3.4.3. *Associations between the WCHS and social and moral belief variables*

The WCHS was consistently associated with progressive political beliefs and support for regulating speech perceived as harmful. WCHS scores were positively correlated with liberal political ideology ($r = 0.29$), Democratic party affiliation ($r = 0.25$), and all three subscales of the Left-Wing Authoritarianism scale (LWA-13). The WCHS was most strongly related to support for top-down censorship (LWA-13 TDC; $r = 0.52$), suggesting a link between viewing speech as harmful and

Table 2
Correlations between the WCHS and demographic variables.

Variable	α	*M*	*SD*	1	2	3	4	5	6	7
1. WCHS	0.92	59.81	20.71							
2. Age	–	46.15	15.76	−0.10**						
3. Gender (female)	–	–	–	0.19**	0.05					
4. Race (White)	–	–	–	−0.10**	0.20**	−0.02				
5. Education	–	4.17	1.37	0.03	0.11**	−0.03	−0.06			
6. Religion	–	–	–	0.00	0.15**	0.09**	−0.12**	0.03		
7. SES	–	3.24	1.54	0.03	−0.02	−0.09**	−0.02	0.38**	−0.00	
8. Subjective social status	–	3.78	1.10	−0.03	0.06	−0.01	−0.04	0.33**	0.11**	0.53**

Note. WCHS = Words Can Harm Scale. SES = Median household income over the past 12 months. Gender was coded 1 = Female, 0 = Male. Race was coded 1 = White, 0 = Non-White. Religion was coded 1 = Religious (any selection), 0 = Nonreligious (atheist or agnostic). Means, standard deviations, and internal consistency coefficients are not shown for dummy coded variables.
** $p < .01$.

Table 3
Correlations between the WCHS and personality and individual difference variables.

Variable	α	*M*	*SD*	1	2	3	4	5	6	7	8	9	10	11	12	13
1. WCHS	0.92	59.81	20.71													
2. Extraversion	–	3.37	1.64	0.03												
3. Agreeableness	–	5.52	1.23	0.14**	0.20**											
4. Conscientiousness	–	5.60	1.32	−0.03	0.08*	0.35**										
5. Emotional stability	–	5.05	1.57	−0.18**	0.25**	0.40**	0.46**									
6. Openness	–	5.28	1.31	0.07*	0.29**	0.28**	0.16**	0.22**								
7. GSE	0.91	3.11	0.53	−0.04	0.25**	0.28**	0.44**	0.54**	0.35**							
8. Intellectual humility	0.89	3.72	0.86	0.18**	0.05	0.13**	0.05	0.10**	0.30**	0.23**						
9. IRI empathic concern	0.87	4.00	0.78	0.32**	0.17**	0.55**	0.13**	0.07*	0.27**	0.14**	0.26**					
10. TIV	0.94	3.78	1.11	0.24**	−0.12**	−0.22**	−0.19**	−0.38**	−0.08*	−0.20**	−0.07*	−0.04				
11. TIV need for recognition	0.89	3.85	1.42	0.27**	−0.08*	−0.18**	−0.21**	−0.37**	−0.03	−0.20**	−0.02	0.00	0.86**			
12. TIV moral elitism	0.84	4.28	1.23	0.22**	−0.11**	−0.02	−0.07*	−0.25**	−0.02	−0.06	0.02	0.09**	0.78**	0.48**		
13. TIV lack of empathy	0.85	3.46	1.23	0.08*	−0.07*	−0.25**	−0.11**	−0.19**	−0.12**	−0.11**	−0.15**	−0.17**	0.84**	0.61**	0.58**	
14. TIV rumination	0.89	3.40	1.54	0.25**	−0.16**	−0.28**	−0.24**	−0.47**	−0.10**	−0.31**	−0.07*	−0.07*	0.84**	0.70**	0.53**	0.61**

Note. WCHS = Words Can Harm Scale. GSE = Generalized Self-Efficacy Scale. TIV = tendency for interpersonal victimhood.
* $p < .05$.
** $p < .01$.

endorsing the use of institutional power to suppress unacceptable speech (e.g., "The government should shut down right-wing internet sites and blogs that promote nutty, hateful positions"). The WCHS was also associated with the belief in the importance of silencing others (BISO; $r = 0.36$) and moral grandstanding ($r = 0.19$), two constructs that reflect motivations to suppress opposing viewpoints and gain status over one's moral adversaries. These results align with prior work suggesting that ideas are more likely to be censored when perceived as harmful (Kubin et al., 2025). The full set of correlations is displayed in Table 4.

3.4.4. Associations between the WCHS and clinical variables

Examining the relationship between the WCHS and clinically relevant variables revealed a strikingly consistent pattern: in every case, higher WCHS scores were significantly associated with indicators of worse psychological well-being, including depression and anxiety, difficulties in emotion regulation, anxiety sensitivity, perceived post-traumatic vulnerability for the self and others, and lower resilience. The full set of correlations is displayed in Table 5.

To supplement these analyses, we explored differences in WCHS scores across several clinical screening thresholds. Participants who met the cutoff for moderately severe depression (PHQ-9 $\geq$ 15; $n = 110$) scored an average of 6.9 points higher on the WCHS ($M = 65.90$, $SD = 19.72$) than those below the cutoff ($M = 59.02$, $SD = 20.71$), $t(142.13) = 3.42$, $p < .001$, $d = 0.33$. Participants with moderate or severe anxiety (GAD-7 $\geq$ 10, $n = 207$), scored 5.5 points higher ($M = 64.12$, $SD = 19.36$) than those with minimal or mild anxiety ($M = 58.62$, $SD = 20.92$), $t(350.43) = 3.56$, $p < .001$, $d = 0.27$. Finally, participants with low resilience (BRS < 3.00; $n = 295$) scored 4.8 points higher ($M = 63.14$, $SD = 19.70$) than those with normal or high resilience ($M = 58.32$, $SD = 20.99$), $t(598.94) = 3.42$, $p < .001$, $d = 0.23$.

4. Discussion

This study presents the Words Can Harm Scale (WCHS) as a valid and reliable measure of the belief that words can cause lasting psychological harm. The WCHS demonstrated sound factor structure, excellent internal consistency, and strong two-week test-retest reliability. Higher scores were more common among female and non-White participants—two groups that experience higher levels of verbal harassment and identity-based prejudice (Beth Nielsen, 2000). Those endorsing this belief also tended to be younger and politically liberal, which makes sense given that progressives often see language as inextricably tied to power structures (Swetha & Aravind, 2025).

The belief that words can harm was also linked with personality characteristics, beliefs, and well-being. People with higher WCHS scores rated themselves as more agreeable, intellectually humble, and empathic; however, they also scored lower in emotional stability and higher in moral grandstanding and the tendency for interpersonal victimhood. The WCHS also tracked cultural and moral attitudes relating to free speech, consistent with the view that perceptions of harm are central to morality (Gray & Pratt, 2025). People with higher WCHS scores were more likely to support political correctness, endorse trigger warnings and safe spaces, believe in the importance of silencing others, and endorse the use of top-down censorship to suppress problematic speech. Finally, people who scored higher on the WCHS consistently reported worse well-being: they were more anxious and depressed, less resilient, more anxiety-sensitive, had more difficulties regulating their emotions, and viewed themselves and others as especially vulnerable to trauma.

4.1. Implications

These results suggest that the WCHS captures a coherent, unidimensional belief about the psychological harmfulness of words. This belief varies meaningfully across individuals but is largely stable within individuals over a two-week period, meaning that the scale can be used

Table 4
Correlations between the WCHS and social and moral belief variables.

Variable	α	*M*	*SD*	1	2	3	4	5	6	7	8	9	10	11	12
1. WCHS	0.92	59.81	20.71												
2. Political ideology (liberal)	–	4.48	1.78	0.29**											
3. Political party (Democrat)	–	–	–	0.25**	0.81**										
4. TWAA	–	1.85	0.36	0.42**	0.19**	0.21**									
5. SSAA	–	1.65	0.48	0.38**	0.12**	0.15**	0.35**								
6. CPC	0.94	3.46	1.52	0.42**	0.40**	0.36**	0.25**	0.30**							
7. Moral grandstanding	0.86	3.45	1.04	0.19**	0.04	0.03	0.08*	0.13**	0.42**						
8. BISO	0.91	3.47	1.29	0.36**	0.18**	0.15**	0.20**	0.30**	0.53**	0.53**					
9. VSA	0.81	4.34	1.84	−0.05	−0.56**	−0.51**	−0.04	0.11**	−0.18**	0.14**	0.10**				
10. LWA-13	0.88	3.43	1.18	0.44**	0.49**	0.43**	0.30**	0.41**	0.55**	0.34**	0.56**	−0.23**			
11. LWA-13 AHA	0.83	2.86	1.41	0.20**	0.27**	0.21**	0.13**	0.18**	0.35**	0.34**	0.44**	−0.18**	0.79**		
12. LWA-13 AC	0.83	3.30	1.55	0.31**	0.63**	0.56**	0.20**	0.21**	0.52**	0.25**	0.40**	−0.50**	0.84**	0.61**	
13. LWA-13 TDC	0.80	3.99	1.43	0.52**	0.29**	0.28**	0.36**	0.55**	0.46**	0.25**	0.51**	0.07*	0.79**	0.39**	0.46**

Note. WCHS = Words Can Harm Scale. TWAA = Support for trigger warnings. SSAA = Support for safe spaces. CPC = Concern for political correctness. BISO = Belief in the importance of silencing others. VSA = Right-wing authoritarianism. LWA-13 = Left-wing authoritarianism. Political ideology was coded 1 = Very Conservative, 7 = Very Liberal. Political party was coded 1 = Democrat, 0 = Republican. M and SD not shown for dummy coded variables.
* $p < .05$.
** $p < .01$.

to reliably differentiate individuals on this belief. Importantly, individual differences in the WCHS appear to be psychologically meaningful. For example, people who met clinical screening thresholds for anxiety, depression, or low resilience had significantly higher WCHS scores compared to people below these thresholds.

In practice, the WCHS provides a valuable psychometric tool for studying cultural issues related to free speech or beliefs about personal emotional vulnerability. Recent work has begun to use the scale to this effect (Bellet et al., 2018; Celniker et al., 2022; Jones et al., 2020; Pratt et al., 2024; Pratt, Jones, et al., 2025). In the present study, WCHS scores correlated more strongly than political ideology with several speech-related beliefs, including endorsement of trigger warnings, safe spaces, political correctness, and silencing opposing viewpoints. This suggests that the WCHS may provide deeper insight into contemporary culture war divides than course proxies like political affiliation alone.

4.2. Limitations and future directions

This study has several limitations that should guide future research. First, as with all self-report measures, responses on the WCHS may be influenced by faking or socially desirable responding (Röhner et al., 2025; Ziegler et al., 2012). Future research should more directly evaluate the extent to which WCHS responses are affected by impression-management processes, for example by using experimental paradigms that manipulate response incentives. Future research could also more intentionally establish the discriminant validity of the scale by including constructs that the WCHS should not correlate with.

This study was cross-sectional, which means that we cannot draw conclusions about the causes and consequences of the belief that words can harm. However, the correlates of the WCHS raise several interesting questions for future research. For example, why is the WCHS consistently associated with worse psychological well-being? One possibility is that people who experience more abusive or derogatory speech, such as women and racial minorities, develop a belief that words can harm through their negative first-hand experiences. Indeed, people from marginalized groups report more psychological distress in response to aggressive language (Sanchez et al., 2018).

A second possibility is that the belief that words can harm is psychologically maladaptive. Prior work finds that the WCHS is associated with cognitive distortions (Celniker et al., 2022), which are known to exacerbate depression and anxiety. Similarly, research on "concept creep" suggests that adopting broader definitions of harm may render individuals more vulnerable to minor stressors (Haslam et al., 2020; Jones & McNally, 2021). Mirroring this, perhaps individuals who believe that words can harm become more emotionally vulnerable, whereas believing that "words will never hurt me" is psychologically protective. Importantly, the two aforementioned hypotheses are compatible: it is possible that people who are negatively impacted by speech develop a stronger belief that words can harm, which then leads to negative psychological outcomes.

A third possibility is that people with pre-existing psychological traits—such as neuroticism or low resilience—are more likely to believe that words can harm and also more likely to have poor mental health. If so, the WCHS may reflect a broader psychological profile rather than a causal belief system. Future experimental work could distinguish between these possibilities by manipulating the belief that words can harm (e.g., Bleske-Rechek et al., 2023) and examining its effects on related variables (e.g., censorship attitudes; anxiety).

Future research could also investigate how beliefs about the harmfulness of words have changed over time. Some scholars argue that this is a relatively new belief—emerging primarily in younger generations as a cohort effect, possibly due to cultural trends like the emergence of social media, rising political polarization, or changes in parenting styles (Lukianoff & Haidt, 2018). Our finding of a small negative correlation between age and WCHS scores ($r = -0.10$) is consistent with this view. However, another possibility is an age effect: perhaps younger people

Table 5
Correlations between the WCHS and clinical variables.

Variable	α	*M*	*SD*	1	2	3	4	5	6	7
1. WCHS	0.92	59.81	20.71							
2. GAD-7	0.93	5.29	5.40	0.21**						
3. PHQ-9	0.91	6.08	6.09	0.16**	0.81**					
4. DERS-SF	0.94	2.02	0.81	0.21**	0.71**	0.70**				
5. BRS	0.91	3.34	0.94	−0.17**	−0.52**	−0.49**	−0.56**			
6. ASI	0.91	19.76	12.32	0.32**	0.54**	0.53**	0.60**	−0.40**		
7. PPV-S	0.95	48.12	20.48	0.40**	0.48**	0.43**	0.50**	−0.48**	0.46**	
8. PPV-O	0.95	57.72	18.22	0.34**	0.34**	0.30**	0.32**	−0.25**	0.32**	0.69**

Note. WCHS = Words Can Harm Scale. GAD-7 = Generalized anxiety disorder symptoms. PHQ-9 = Depressive symptoms. DERS-SF = Difficulties in emotion regulation. BRS = Perceived resilience. ASI = Anxiety sensitivity. PPV-S = Perceived posttraumatic vulnerability of the self. PPV-O = Perceived posttraumatic vulnerability of others.
** $p < .01$.

are more likely to endorse the belief that words can harm, but this belief may decline with age. A third possibility is a period effect, where broader cultural changes—such as increased attention to trauma and mental health awareness—have made everyone, regardless of age, more likely to endorse this belief, reflecting a general "rising tide" of sensitivity to harm (Haslam et al., 2020). Some view this trend as raising important awareness of previously neglected forms of suffering (Cikara, 2016), whereas others see it as a key reason why today's youth seem less resilient and less willing to discuss contentious topics (Lukianoff & Haidt, 2018). Whether the expansion of definitions of harm is good or bad—or likely both, as with many psychological trends—future research using the WCHS will be well equipped to study how and why beliefs about the harmfulness of words have evolved within cultural discourse.

5. Conclusion

The WCHS is a valid and reliable measure of beliefs about harmful speech, a contentious and defining feature of contemporary cultural discourse. People routinely disagree about the necessity of trigger warnings for trauma survivors, whether controversial speakers should be deplatformed, and whether free speech protections should extend to hateful ideas. The WCHS provides a new window into these beliefs and those who hold them, which can contribute to theory development and empirical research on free speech attitudes, debates about "safetyism" and fragility, and the causes of moral disagreement.

CRediT authorship contribution statement

Samuel Pratt: Writing – review & editing, Writing – original draft, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Payton J. Jones:** Writing – review & editing, Writing – original draft, Conceptualization. **Benjamin W. Bellet:** Writing – review & editing, Conceptualization. **Richard J. McNally:** Writing – review & editing. **Kurt Gray:** Writing – review & editing, Funding acquisition.

Funding

This research was funded by StandTogether via the Center for the Science of Moral Understanding.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.paid.2026.113785.

Data availability

All data, materials, and code are available on the Open Science Framework (https://osf.io/gf8j5/?eda628f546f24fa09beaa11beeebbcc1).

References

Adler, N. E., Epel, E. S., Castellazzo, G., & Ickovics, J. R. (2000). Relationship of subjective and objective social status with psychological and physiological functioning: Preliminary data in healthy, White women. *Health Psychology, 19*(6), 586–592. https://doi.org/10.1037/0278-6133.19.6.586

Bellet, B. W., Jones, P. J., & McNally, R. J. (2018). Trigger warning: Empirical evidence ahead. *Journal of Behavior Therapy and Experimental Psychiatry, 61*, 134–141. https://doi.org/10.1016/j.jbtep.2018.07.002

Beth Nielsen, L. (2000). Situating legal consciousness: Experiences and attitudes of ordinary citizens about law and street harassment. *Law and Society Review, 34*(4), 1055–1090. https://doi.org/10.2307/3115131

Bhat, N. A., Roopesh, B. N., Bhaskarapillai, B., Chokkanathan, S., & Benegal, V. (2025). Difficulties in Emotion Regulation Scale-Short Form (DERS-SF): Psychometric validation and measurement invariance testing in a sample of urban Indian adolescents. *Indian Journal of Psychological Medicine, 47*(2), 119–126. https://doi.org/10.1177/02537176241232936

Bizumic, B., & Duckitt, J. (2018). Investigating right wing authoritarianism with a very short authoritarianism scale. *Journal of Social and Political Psychology, 6*(1), 129–150. https://doi.org/10.5964/jspp.v6i1.835

Bleske-Rechek, A., Deaner, R. O., Paulich, K. N., Axelrod, M., Badenhorst, S., Nguyen, K., … Lay, P. S. (2023). In the eye of the beholder: Situational and dispositional predictors of perceiving harm in others' words. *Personality and Individual Differences, 200*, Article 111902. https://doi.org/10.1016/j.paid.2022.111902

Celniker, J. B., Ringel, M. M., Nelson, K., & Ditto, P. H. (2022). Correlates of "Coddling": Cognitive distortions predict safetyism-inspired beliefs, belief that words can harm, and trigger warning endorsement in college students. *Personality and Individual Differences, 185*, Article 111243. https://doi.org/10.1016/j.paid.2021.111243

Choi, J., Jeong, B., Rohan, M. L., Polcari, A. M., & Teicher, M. H. (2009). Preliminary evidence for white matter tract abnormalities in young adults exposed to parental verbal abuse. *Biological Psychiatry, 65*(3), 227–234. https://doi.org/10.1016/j.biopsych.2008.06.022

Cikara, M. (2016). Concept expansion as a source of empowerment. *Psychological Inquiry, 27*(1), 29–33. https://doi.org/10.1080/1047840X.2016.1111830

Comrey, A. L., & Lee, H. B. (2013). *A first course in factor analysis* (2nd ed.). Taylor and Francis.

Costello, T. H., Newton, C., Lin, H., & Pennycook, G. (2023, August 6). Intellectual humility questionnaires fail to predict metacognitive skill: Implications for theory and measurement. *PsyArXiv*. https://doi.org/10.31234/osf.io/gux95

Costello, T. H., & Patrick, C. J. (2023). Development and initial validation of two brief measures of left-wing authoritarianism: A machine learning approach. *Journal of Personality Assessment, 105*(2), 187–202. https://doi.org/10.1080/00223891.2022.2081809

Davis, M. H. (1980). *Interpersonal reactivity index*. APA PsycTests. https://doi.org/10.1037/t01093-000

DeVon, H. A., Block, M. E., Moyle-Wright, P., Ernst, D. M., Hayden, S. J., Lazzara, D. J., … Kostas-Polston, E. (2007). A psychometric toolbox for testing validity and reliability. *Journal of Nursing Scholarship, 39*(2), 155–164. https://doi.org/10.1111/j.1547-5069.2007.00161.x

Feldman Barrett, L. (2017, July 14). When is speech violence?. The New York Times. https://www.nytimes.com/2017/07/14/opinion/sunday/when-is-speech-violence.html.

Gabay, R., Hameiri, B., Rubel-Lifschitz, T., & Nadler, A. (2020). The tendency for interpersonal victimhood: The personality construct and its consequences.

Personality and Individual Differences, 165, Article 110134. https://doi.org/10.1016/j.paid.2020.110134

Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (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

Gray, K., & Pratt, S. (2025). Morality in our mind and across cultures and politics. *Annual Review of Psychology, 76*, 663–691. https://doi.org/10.1146/annurev-psych-020924-124236

Grubbs, J. B., Warmke, B., Tosi, J., James, A. S., & Campbell, W. K. (2019). Moral grandstanding in public discourse: Status-seeking motives as a potential explanatory mechanism in predicting conflict. *PLoS One, 14*(10), Article e0223749. https://doi.org/10.1371/journal.pone.0223749

Haslam, N. (2016). Concept creep: Psychology's expanding concepts of harm and pathology. *Psychological Inquiry, 27*(1), 1–17. https://doi.org/10.1080/1047840X.2016.1082418

Haslam, N., Dakin, B. C., Fabiano, F., McGrath, M. J., Rhee, J., Vylomova, E., … Wheeler, M. A. (2020). Harm inflation: Making sense of concept creep. *European Review of Social Psychology, 31*(1), 254–286. https://doi.org/10.1080/10463283.2020.1796080

Jones, P. J., Bellet, B. W., & McNally, R. J. (2020). Helping or harming? The effect of trigger warnings on individuals with trauma histories. *Clinical Psychological Science, 8*(5), 905–917. https://doi.org/10.1177/2167702620921341

Jones, P. J., & McNally, R. J. (2021). Does broadening one's concept of trauma undermine resilience? *Psychological Trauma Theory Research Practice and Policy, 14*(S1), S131–S139. https://doi.org/10.1037/tra0001063

Kaufman, E. A., Xia, M., Fosco, G., Yaptangco, M., Skidmore, C. R., & Crowell, S. E. (2016). The Difficulties in Emotion Regulation Scale Short Form (DERS-SF): Validation and replication in adolescent and adult samples. *Journal of Psychopathology and Behavioral Assessment, 38*(3), 443–455. https://doi.org/10.1007/s10862-015-9529-3

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (1999). *Patient Health Questionnaire-9*. APA PsycTests. https://doi.org/10.1037/t06165-000

Kubin, E., Von Sikorski, C., & Gray, K. (2025). Political censorship feels acceptable when ideas seem harmful and false. *Political Psychology, 46*(2), 279–299. https://doi.org/10.1111/pops.13011

Leary, M. R., Diebels, K. J., Davisson, E. K., Jongman-Sereno, K. P., Isherwood, J. C., Raimi, K. T., … Hoyle, R. H. (2017). Cognitive and interpersonal features of intellectual humility. *Personality and Social Psychology Bulletin, 43*(6), 793–813. https://doi.org/10.1177/0146167217697695

Lukianoff, G., & Haidt, J. (2018). *The coddling of the American mind: How good intentions and bad ideas are setting up a generation for failure*. Penguin Random House.

McGrath, M. J., Randall-Dzerdz, K., Wheeler, M. A., Murphy, S., & Haslam, N. (2019). Concept creepers: Individual differences in harm-related concepts and their correlates. *Personality and Individual Differences, 147*, 79–84. https://doi.org/10.1016/j.paid.2019.04.015

Miller, G. E., & Chen, E. (2010). Harsh family climate in early life presages the emergence of a proinflammatory phenotype in adolescence. *Psychological Science, 21*(6), 848–856. https://doi.org/10.1177/0956797610370161

Pratt, S., Bellet, B. W., Jones, P. J., & McNally, R. J. (2024, July 3). Testing the coddling hypothesis: Campus safetyism and student resilience. *PsyArXiv*. https://doi.org/10.31234/osf.io/zav5g

Pratt, S., Jones, P. J., Bridgland, V. M. E., Bellet, B. W., & McNally, R. J. (2025). Sending signals: Trigger warnings and safe space notifications. *Journal of Experimental Psychology. Applied*. https://doi.org/10.1037/xap0000541

Pratt, S., Rosenfeld, D. L., Goranson, A., Tomiyama, A. J., Sheeran, P., & Gray, K. (2025). Health behaviors are moralized when perceived to cause harm. *Personality and Social Psychology Bulletin, 0*(0). https://doi.org/10.1177/01461672251372823

R Core Team. (2024). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*. https://www.R-project.org/.

Reiss, S., Peterson, R. A., Gursky, D. M., & McNally, R. J. (1986). Anxiety sensitivity, anxiety frequency and the prediction of fearfulness. *Behaviour Research and Therapy, 24*(1), 1–8. https://doi.org/10.1016/0005-7967(86)90143-9

Röhner, J., Schütz, A., & Ziegler, M. (2025). Faking in self-report personality scales: A qualitative analysis and taxonomy of the behaviors that constitute faking strategies. *International Journal of Selection and Assessment, 33*(1). https://doi.org/10.1111/ijsa.12513

Sanchez, D., Adams, W. N., Arango, S. C., & Flannigan, A. E. (2018). Racial-ethnic microaggressions, coping strategies, and mental health in Asian American and Latinx American college students: A mediation model. *Journal of Counseling Psychology, 65*(2), 214–225. https://doi.org/10.1037/cou0000249

Schenck v. United States, 249 U.S. 47.(1919).

Schwarzer, R., & Jerusalem, M. (1995). *General Self-Efficacy Scale*. APA PsycTests. https://doi.org/10.1037/t00393-000

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

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. https://doi.org/10.1001/archinte.166.10.1092

Strauts, E., & Blanton, H. (2015). That's not funny: Instrument validation of the concern for political correctness scale. *Personality and Individual Differences, 80*, 32–40. https://doi.org/10.1016/j.paid.2015.02.012

Swetha, M., & Aravind, B. R. (2025). Language as power: Analyzing the intersection of linguistics and politics in Ijeoma Oluo's work. *Social Sciences & Humanities Open, 11*, Article 101405. https://doi.org/10.1016/j.ssaho.2025.101405

Teicher, M. H., Samson, J. A., Polcari, A., & McGreenery, C. E. (2006). Sticks, stones, and hurtful words: Relative effects of various forms of childhood maltreatment. *American Journal of Psychiatry, 163*(6), 993–1000. https://doi.org/10.1176/ajp.2006.163.6.993

The Editorial Board. (2022, December 19). The Stanford guide to acceptable words: Behold the school's elimination of harmful language initiative. The Wall Street Journal. https://www.wsj.com/articles/the-stanford-guide-to-acceptable-words-elimination-of-harmful-language-initiative-11671489552?reflink=desktopwebshare_permalink.

Tsfati, Y., & Dvir-Gvirsman, S. (2018). Silencing fellow citizens: Conceptualization, measurement, and validation of a scale for measuring the belief in the importance of actively silencing others. *International Journal of Public Opinion Research, 30*(3), 391–419. https://doi.org/10.1093/ijpor/edw038

Ziegler, M., MacCann, C., & Roberts, R. D. (2012). Faking: Knowns, unknowns, and points of contention. In M. Ziegler, C. MacCann, & R. D. Roberts (Eds.), *New perspectives on faking in personality assessment* (pp. 3–16). Oxford University Press.