Testing the Coddling Hypothesis: Campus Safetyism and Student Resilience

Samuel Pratt[1,2], Benjamin W. Bellet[1,3], Payton J. Jones[1], Richard J. McNally[1]

[1] Department of Psychology, Harvard University

[2] Department of Psychology and Neuroscience, University of North Carolina at Chapel Hill

[3] Massachusetts Mental Health Center, Boston, Massachusetts, United States

Note: This manuscript is under peer review and may not be the final version of this publication.

Author Note:

Correspondence concerning this article may be sent to Samuel Pratt, Department of Psychology and Neuroscience, University of North Carolina at Chapel Hill, 315 Davie Hall, CB 3150, Chapel Hill, NC 27599, United States of America. E-mail address: sampratt@unc.edu.

Abstract

Objective: We tested the "coddling hypothesis" that students' exposure to campus safetyism—efforts to support student well-being by preventing emotional discomfort (e.g., trigger warnings; safe spaces; deplatforming controversial speakers)—is associated with decreases in their psychological resilience.

Participants: 107 first-year college students.

Methods: Students reported their exposure to campus safetyism at four weekly intervals. We tested the relationship between safetyism exposure and various indicators of resilience at endpoint assessment, controlling for baseline resilience. We also assessed students' own perceived need for safetyism practices.

Results: Perceived need for safetyism was positively related to female gender and the belief that words can harm, and negatively related to self-reported resilience. Analysis using Bayes Factors revealed relative evidence for the null hypothesis that self-reported safetyism exposure was unrelated to changes in psychological resilience.

Conclusions: Safetyism practices were not associated with benefits or harm to student resilience. Implications for higher education are discussed.

Keywords: safetyism, resilience, college, mental health

Testing the Coddling Hypothesis: Campus Safetyism and Student Resilience

In *The Coddling of the American Mind*, Lukianoff and Haidt (2018) expressed concern that American institutions of higher learning have been afflicted by a culture of "safetyism" characterized by the belief that students should be shielded from words, ideas, and experiences that might cause emotional discomfort. They argue that emerging practices implemented on campuses based on principles of safetyism are intended to bolster student well-being but may be iatrogenic.

For example, proponents of *trigger warnings* argue that professors should warn students on syllabi or in the classroom about readings and lecture topics that may trigger adverse emotional responses, which may impede their ability to learn (e.g., Stokes, 2014). Such warnings presumably enable students to mobilize their coping skills to inhibit anxiety or signal a professor's sensitivity to the emotional vulnerabilities of their students. However, a meta-analysis of the efficacy of trigger warnings finds that they increase anticipatory anxiety and have no effect on response anxiety (Bridgland et al., 2023). Another study found that trigger warnings counter-therapeutically increase individuals' belief that trauma is central to their identity (Jones et al., 2020).

Another safetyist practice is the provision of *safe spaces* where students are assured that their emotional well-being will not be jeopardized. The use of safe spaces is multifaceted, ranging from verbal assurances that the classroom is a "safe space" to the provision of physical spaces as emotional accommodation (Ho, 2017). For example, when a controversial speaker participated in a debate at Brown University in 2015, the president of the university organized a safe space for upset students replete with "cookies, coloring books, bubbles, Play-Doh, calming music, pillows, blankets, and a video of frolicking puppies" (Shulevitz, 2015).

Other safetyist practices include restricting speech that some might find offensive ("The Stanford Guide to Acceptable Words," 2022), deplatforming controversial campus speakers ("Campus Deplatforming Database," 2024), and providing mental health resources to students who are experiencing non-clinical levels of stress or anxiety (Haslam et al., 2021). Proponents of these practices argue for their necessity amidst rising levels of student mental health problems (Bower, 2023), whereas opponents argue that safetyism stifles free speech and contradicts core values of higher education (e.g., promoting growth through intellectual discomfort; Haidt, 2017). Lukianoff and Haidt (2018) further theorize that efforts to shield students from emotional discomfort in the classroom will only "coddle" students by encouraging them to view themselves as fragile and hindering their resilience to stressors in the classroom and beyond.

Despite widespread debates about the benefits or drawbacks of safetyism, little research has tested how safetyism practices affect students psychologically (though see Bridgland et al., 2023). As the history of psychological interventions shows, well-intentioned efforts to help sometimes turn out to cause inadvertent harm, from recovered memory therapy to DARE programs (Lilienfeld, 2007). The current study tested the hypothesis that exposure to safetyism is associated with decreases in students' psychological resilience.

The Current Study

We examined the correlates and consequences of safetyism at a northeastern university. Participants reported their exposure to a set of safetyism-related practices on campus in four weekly surveys. We assessed resilience at baseline and endpoint, operationalized broadly to include students' self-perceptions of their resilience and general attitudes toward stress, as well as their stress appraisals and anticipatory anxiety when faced with a real stressor (a brief online public speaking task). This allowed us to test two relevant claims advanced by opponents of safetyism: that safetyism leads students to view themselves as vulnerable, and that safetyism

compromises students' ability to competently cope with challenges (Lukianoff & Haidt, 2018). We hypothesized that greater exposure to safetyism during the study period would be associated with negative changes in each component of resilience. However, we computed Bayes factors to evaluate the probative import of any null results (Etz et al., 2018).

In addition to examining the relationship between safetyism exposure and student resilience, we asked several exploratory questions to gain a broader understanding of campus safetyism. First, we explored how often students reported experiencing various safetyism practices on campus. Second, we examined the extent to which students report that safetyism practices are necessary or beneficial to their psychological well-being. We examined demographic differences in the endorsement of this belief and measured its associations with relevant constructs such as self-reported resilience. Finally, we explored whether several covariates might moderate the relationship between safetyism and student resilience: whether students had a prior psychiatric diagnosis, their general anxiety levels, their belief that words can harm, and their perceived need for safetyism.

Method

All data, materials, and code are available online at https://osf.io/q6dyx/?view_only=24149ca7283e451786f80a85ad9509b3.

Participants

First-year undergraduate students were recruited on a rolling basis during the fall semester of 2022. Prospective participants accessed the first of five surveys via a link in the recruitment materials, after which they were emailed a weekly survey to complete at four weekly intervals.

Measures

Safetyism

Campus Safetyism Scale

We created a 16-item scale to assess participants' frequency of exposure to safetyism-related campus practices during the previous week. Scale items consisted of an ad hoc set of campus practices that capture the essence of "safetyism" based on popular discussions and prior research (Lukianoff & Haidt, 2018). Participants indicated how often they experienced each event in the past week on a sliding scale (0 = never, 50 = 3-4 times, 100 = daily or more often). Critical items assessing safetyism (e.g., "I was given a trigger warning or content warning to notify me of potentially distressing material") were intermixed with noncritical filler items (e.g., "I went to office hours to get extra help for a course"), thereby masking the purpose of the study. Participants' scores across each of the four exposure surveys were averaged to generate a composite score representing safetyism exposure during the study period.

Perceived Need for Safetyism Scale (PNSS)

We developed the 8-item Perceived Need for Safetyism Scale (PNSS) to assess the degree to which an individual believes that practices aimed at shielding them from stress and emotional discomfort are beneficial to their emotional well-being and their academic and social functioning on campus. Items included "Receiving a trigger warning can help me mentally prepare for distressing content" and "Campus practices designed to ensure my emotional safety are an invaluable part of my education" as well as reverse-scored items such as "It is my own responsibility to manage and monitor my emotional well-being." Participants indicated their level of agreement with each statement on a sliding scale ranging from 0 (strongly disagree) to 100 (strongly agree). Responses were averaged to generate a composite score, with higher scores indicating a greater perceived need for safetyism. The PNSS displayed acceptable internal consistency in our sample ($\alpha = 0.68$).

Resilience

Brief Resilience Scale (BRS)

The Brief Resilience Scale (BRS; Smith et al., 2008) measures individuals' perceptions of their resilience, or ability to recover from stressful events. The BRS has demonstrated strong convergent and construct validity in other samples, correlating positively with related components of resilience (e.g., optimism, social support) and negatively with impaired mental health (e.g., anxiety, depression; Smith et al., 2008). Participants indicated their agreement with each of six statements (e.g., "I usually come through difficult times with little trouble") on a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree) at two different timepoints: in survey one (Time 1) and survey five (Time 2). Mean scores were calculated at each timepoint, with higher scores indicating a stronger belief in one's own resilience. The BRS demonstrated good internal consistency in our sample ($\alpha = 0.87$ (Time 1), $\alpha = 0.89$ (Time 2)).

Stress Mindset Measure (SMM)

The Stress Mindset Measure (SMM; Crum et al., 2013) is an 8-item scale that assesses the belief that stress has either positive effects (e.g., "the effects of stress are positive and should be utilized") or negative effects (e.g., "experiencing stress inhibits my learning and growth"). Participants responded to four stress-is-enhancing statements and four stress-is-debilitating statements (reverse coded) on a 5-point Likert scale (0 = strongly disagree, 4 = strongly agree) at two different timepoints: in survey one (Time 1) and survey five (Time 2). Mean scores were calculated at each timepoint, with higher scores indicating a more positive mindset towards the effects of stress. In our sample, the Stress Mindset Measure exhibited acceptable internal consistency ($\alpha = 0.81$ (Time 1), $\alpha = 0.77$ (Time 2)).

Acute Stress Appraisal Scale (ASAS)

Mendes and colleagues (2007) developed a scale (here termed the Acute Stress Appraisal Scale, or ASAS) to measure individuals' appraisal of a stressor as a manageable challenge versus an overwhelming threat. Cognitive appraisals of a stressor can powerfully shape the stress

response (Blascovich & Tomaka, 1996; Stress, Appraisal, and Coping, 1984), and challenge appraisals are generally associated with resilient outcomes relative to threat appraisals (Bonanno et al., 2012; Holbrook et al., 2001; Seery et al., 2010).

The ASAS comprises two subscales. The demand evaluation subscale (ASAS-D) includes five items assessing appraisals of task demands (e.g., "the upcoming task is very stressful"). The resource evaluation subscale (ASAS-R) includes five items assessing appraisals of coping resources in the context of the task (e.g., "I have the abilities to perform the upcoming task successfully"). Participants indicated their endorsement of the ten statements on a 7-point Likert scale (1 = strongly disagree, 7 = strongly agree) at two timepoints: after learning about the public speaking task in the first survey (Time 1) and after learning about the public speaking task in the final survey (Time 2). To score this measure, we first averaged demand and resource evaluations separately. Then, average demand evaluations were divided by average resource evaluations to generate a challenge/threat ratio at each timepoint. Scores greater than one represent a threat appraisal, indicating that the demands of the task are perceived as exceeding one's coping resources. Scores less than one represent a challenge appraisal, indicating that one's coping resources are perceived as outweighing the demands of the task. After collecting data, we removed one item from the demand subscale (i.e., "a poor performance on this task would be very distressing for me") and one item from the resource subscale (i.e., "it is very important to me that I perform well on this task") to improve the reliability of the measure. After removing these items, both subscales of the measure displayed acceptable to good internal consistency (ASAS-D: $\alpha = 0.83$ (Time 1), $\alpha = 0.88$ (Time 2); ASAS-R: $\alpha = 0.70$ (Time 1), $\alpha = 0.77$ (Time 2)).

Momentary Anxiety Scale

Both before and after learning that they would be completing a public speaking task, participants rated the degree to which they felt a variety of emotions in the present moment via slider bars (0 = not at all, 100 = extremely). The target emotion was anxiety; the other emotions (e.g., sad, angry, disgusted) masked the target emotion. This measure was completed at four different timepoints: before (Time 1) and after (Time 2) learning about the public speaking task in the first survey, and before (Time 3) and after (Time 4) learning about the public speaking task in the final survey. Change scores were calculated in each of the two surveys to represent the change in anxiety, or anticipatory anxiety, in response to the public speaking tasks administered at baseline and endpoint.

Additional Variables

We explored whether several additional variables might moderate the relationship between safetyism exposure and changes in resilience.

Psychiatric History

Participants reported whether they had ever been diagnosed with a psychiatric or psychological problem. Participants who responded yes were prompted to select all applicable diagnoses from a dropdown menu.

Generalized Anxiety Disorder Scale –7-item – Modified

The Generalized Anxiety Disorder Scale – 7-item (GAD-7; Spitzer et al., 2006) was modified to assess anxiety symptoms in general, rather than over a specified period. Participants reported how often they experienced the symptoms of generalized anxiety disorder (e.g., "not able to control or stop worrying") on a 4-point Likert scale (0 = not at all, 3 = nearly every day). Responses to each item were summed to generate a composite score. The GAD-7 demonstrated good internal consistency in our sample ($\alpha = 0.87$).

Words Can Harm Scale – Modified

The Words Can Harm Scale (WCHS; Bellet et al., 2018) measures the degree to which an individual believes that potentially offensive or distressing words can cause serious harm. Participants indicated their level of agreement with nine statements (e.g., "I could be left emotionally scarred by something I read") on a sliding scale ranging from 0 (strongly disagree) to 100 (strongly agree). Because we were interested individuals' perceptions that words can be harmful to *themselves*, six of the original items were reworded to refer to the self, as opposed to other people. Responses were averaged to obtain a composite score, with higher scores indicating a stronger belief that words can cause harm. The WCHS demonstrated excellent internal consistency in our sample ($\alpha = 0.93$).

Procedure

Survey 1 (Baseline)

Eligibility requirements constituted being (1) a first-year undergraduate student, (2) at least 18 years of age, (3) a fluent English speaker, and (4) willing to complete all five surveys. Eligible participants provided informed consent and then completed the Psychiatric History questionnaire, PNSS, WCHS, GAD-7, SMM, BRS, Momentary Anxiety Scale, and standard demographic questions.

Next, participants were asked to complete a brief online public speaking task. The task was similar in format to in-person tasks used in previous studies to elicit an anxiety response (McNally et al., 2013). Participants were given the following information:

> "In the next section we would like to assess your ability to perform under pressure. You will have 2 minutes to prepare a 4-minute speech on a topic of our choosing. The survey software will record your speech and automatically submit it once 4 minutes have elapsed. Researchers will assess your submitted speech based on your preparation, persuasiveness, and delivery."

Participants were neither recorded nor assessed on the quality of their speeches; these instructions were merely intended to facilitate a motivated performance situation in which individuals' stress appraisals and anxiety responses could be assessed.

Immediately after receiving the instructions for the speech task, participants were informed that "before you complete this task, we're going to ask you a few questions about how you're feeling right now regarding the task you're about to complete." Participants then completed the Momentary Anxiety Scale for a second time as well as the ASAS to assess challenge and threat appraisals. The survey then advanced to a two-minute speech preparation period during which participants received the following prompt (modified from the Trier Social Stress Test; Allen et al., 2017): "Imagine that you are interviewing for your dream job. You will have four minutes to convince your potential employer that you should get the job." The survey then advanced to a four-minute speech delivery period.

Participants accessed all subsequent surveys via a link provided in a weekly email.

Surveys 2, 3, and 4

In surveys 2, 3, and 4, participants completed the Campus Safetyism Scale at weekly intervals. Participants reported how frequently they had experienced a variety of safetyism practices and filler campus events during the prior week.

Survey 5

Participants completed the final survey four weeks after completing the baseline survey. First, participants completed the Campus Safetyism Scale for a fourth and final time before completing the SMM, BRS, and Momentary Anxiety Scale. Next, participants were told that they would perform a brief public speaking task (the same instructions administered at baseline). Next, they completed the Momentary Anxiety Scale and the ASAS to assess their response to the

public speaking task, and were then informed that the study was completed[1]. Participants were then thanked for their participation and fully debriefed about the purpose and hypotheses of the study.

Analyses

Prior to analyzing the data, we conducted pairwise deletion and mean imputation of scale measures to address missing data. The full details of missing and imputed data are reported in the supplemental materials.

For our primary analyses, we conducted a series of linear regressions to test whether average exposure to safetyism during the study period predicted endpoint scores for each resilience variable, controlling for baseline scores. We included relevant demographic variables as covariates in all models. For each relevant research question, we set up an opposing alternative hypothesis (i.e., the effect does not equal zero) and null hypothesis (i.e., the effect equals zero). We then computed Bayes Factors to compare the relative likelihood of these two hypotheses. Bayes Factors were computed with the lmBF function from the BayesFactor R package with the default 'medium' prior (Morey & Rouder, 2024). Bayes Factors below 0.33 (ratio of 1 to 3) were interpreted as providing substantial evidence for the null hypothesis, and Bayes Factors above 3 (ratio of 3 to 1) were interpreted as providing substantial evidence for the alternative hypothesis. Bayes Factors in between ⅓ and 3 were interpreted as providing insufficient evidence to adjudicate between the two hypotheses.

Results

Seven hundred sixty-two individuals navigated to the first Qualtrics survey via a link in the recruitment materials. Due to a large volume of incomplete survey responses, participants

[1] Because we were only interested in anticipatory anxiety and stress appraisals prior to the delivery of the speech, we did not ask participants to deliver the second speech after they had completed these measures in the final survey. The reason that participants were asked to deliver their speech in the first survey was to give the impression that the task was genuine and would be assessed by researchers.

were only considered enrolled in the study, and therefore emailed the subsequent surveys, if they completed at least one of the dependent variables in the first survey (i.e., completed at least ~27% of the survey). After excluding participants who did not pass the initial eligibility questions, provide informed consent, enter a valid email address, and make sufficient progress on the first survey, we were left with 288 participants who enrolled in the study and were emailed the subsequent surveys.

After completing data collection but prior to examining the data, we excluded 181 participants from analyses for (1) failing to take the final survey and reach at least the first dependent variable (n = 180), and (2) failing to complete at least 80% of the safetyism exposure measure in at least two out of three intermediary surveys (n = 1). The final sample used for analyses included 107 participants who made sufficient progress on the first survey, the final survey, and at least two of three intermediary surveys *(*n = 107; 103; 102; 101; 107 for each respective survey). Table 1 displays the demographic characteristics of our participants.

Table 1
Demographic Characteristics of the Sample (N = 107)

Variable	
Age (*M*, *SD*)	18.3 (0.85)
Gender (*n*, %)	
Female	69 (64.5%)
Male	34 (31.8%)
Other	4 (3.7%)
Ethnicity (*n*, %)	
White	48 (44.9%)
Asian	25 (23.4%)
Hispanic or Latino	11 (10.3%)
Black or African American	9 (8.4%)
Native Hawaiian or Pacific Islander	2 (1.9%)
Other	12 (11.2%)
Sexual Orientation (*n*, %)	
Straight	73 (68.2%)

Bisexual	14 (13.1%)
Gay or lesbian	12 (11.2%)
Other	6 (5.6%)
Psychiatric History (*n*, %)	
Yes	25 (23.4%)
No	82 (76.6%)

Descriptive Statistics and Correlational Analyses

First, we conducted standard descriptive and correlational analyses on our variables to assess relationships between participants' attitudes towards safetyism, their resilience, and other relevant constructs.

Table 2 displays the descriptive statistics for our moderator and outcome variables.

Table 2

Descriptive Statistics of Moderator and Outcome Variables (N = 107)

Variable	Possible Range	Observed Range	Mean	Median	SD
Trait Anxiety (GAD-7)	0-21	0-20	5.87	5.00	4.56
Perceived Need for Safetyism (PNSS)	0-100	9.38-73.00	48.68	49.75	11.89
Belief That Words Can Harm (WCHS)[a]	0-100	0-92	46.35	49.17	20.15
Perceived Resilience (BRS)					
Baseline	1-5	1.67-5	3.38	3.5	0.73
Endpoint	1-5	1.17-5	3.30	3.33	0.75
Stress Mindset (SMM)					
Baseline	0-4	0-2.75	1.67	1.75	0.60
Endpoint	0-4	0-2.75	1.66	1.75	0.55
Acute Stress Appraisal (ASAS)					
Baseline[b]	0.14-7	0.35-4.33	1.31	1.11	0.78
Endpoint[c]	0.14-7	0.15-5.20	1.28	1.06	0.91
Anticipatory Anxiety					
Baseline[d]	-100-100	-43-86	8.17	2.00	22.4920.20
Endpoint[e]	-100-100	-55-66	0.25	0.00	

Note. [a]$n = 106$. [b]$n = 84$. [c]$n = 86$. [d]$n = 84$. [e]$n = 85$.

Table 3 depicts the bivariate correlations between pairs of variables.

Table 3

Bivariate Correlations Between Variables of Interest at Baseline

Variable	Age	GAD-7	PNSS	WCHS	BRS_1	SMM_1	$ASAS_1$	Anxiety Response$_1$
GAD-7	.08	-						
PNSS	-.21*	.17	-					
WCHS	-.16	.22*	.40***	-				
BRS_1	.11	-.38***	-.43***	-.27***	-			
SMM_1	.00	-.09	-.06	.00	.19*	-		
$ASAS_1$	-.04	.44***	.36***	.27*	-.35***	.02	-	
Anticipatory Anxiety$_1$	-.13	-.05	.16	.13	-.06	.12	.25*	-
N	107	107	107	106	107	107	84	84

Note. GAD-7 = Trait anxiety, PNSS = Perceived need for safetyism, WCHS = Belief that words can harm, BRS_1 = Perceived resilience at baseline, SMM_1 = Stress mindset at baseline, $ASAS_1$ = Acute stress appraisal of the public speaking task at baseline, Anticipatory $Anxiety_1$ = Anxiety change in advance of the public speaking task at baseline.

*$p < .05$, ***$p < .001$.

Perceived Need for Safetyism

Next, we examined which students perceived safetyism as being especially beneficial or necessary for their well-being on campus. We conducted a series of *t*-tests, reported below in Table 4, examining whether the perceived need for safetyism differed based on participants' demographic characteristics and psychiatric history.

Table 4

Differences in Perceived Need for Safetyism by Demographics and Psychiatric History

Variable	N	PNSS	SD	*df*	*t*	*p*	Cohen's

		Mean					*d*
Gender							
Male	34	44.92	11.64	66.52	-2.19	.03*	0.46
Female	69	50.28	11.78				
Sexual Orientation							
Heterosexual	73	48.13	12.23	67.85	-0.58	.56	0.12
Non-heterosexual	33	49.53	11.08				
Ethnicity							
White	48	46.33	12.53	94.49	-1.85	.07	0.36
Non-white	59	50.59	11.02				
Psychiatric History							
Yes	25	49.27	12.19	38.80	0.28	.78	0.06
No	82	48.50	11.83				

Note. PNSS = perceived need for safetyism.

*$p < .05$.

Prevalence of Campus Safetyism

Having established the psychological and demographic correlates of perceiving benefits in safetyism, we next explored students' reported level of exposure to various safetyism practices and whether overall exposure was related to other relevant variables. Average exposure to safetyism practices across all timepoints was not significantly correlated with trait anxiety, $r(105) = .02$, $p = .86$, perceived need for safetyism, $r(105) = .19$, $p = .055$, the belief that words can harm, $r(104) = .04$, $p = .67$, perceived resilience at baseline, $r(105) = -.03$, $p = .79$, or stress mindset at baseline, $r(105) = -.07$, $p = .46$, suggesting that our self-reported measure of safetyism exposure had good discriminant validity and was not merely a function of trait level variables.

Table 5 displays the participants' average exposure to safetyism during the study and average exposure to each individual item on the Campus Safetyism Scale.

Table 5

Average Exposure to Campus Safetyism and to Each Constituent Scale Item Over a Four-Week Period (N = 107)

Variable	Observed Range[a]	Mean	Median	SD
Average Safetyism Exposure	0.33-60.53	21.08	17.30	15.41
Scale Item				
I was informed of the campus mental health resources available to me.	0-72.50	31.54	29.75	20.02
I was given a trigger warning or content warning to notify me of potentially distressing material.	0-69.75	20.35	15.75	17.17
My teacher made certain distressing course content optional to students.	0-76.75	18.26	12.00	19.81
It was emphasized that I should avoid the stress associated with college.	0-67.67	18.72	15.00	18.48
A teacher/administrator arranged a time or place intended to help students deal with distressing course material or campus events (e.g., a "safe space").	0-77.67	19.81	16.00	19.62
I was encouraged to seek out mental health resources in the event that I experience school-related stress.	0-71.25	24.63	21.75	19.28
My teacher emphasized their willingness to accommodate students in order to minimize course-related stress (e.g., make a test open-book, remove assignments from the syllabus, etc.).	0-88.75	24.39	22.50	19.42
A teacher/administrator emphasized the importance of my emotional safety.	0-80.00	23.57	17.33	20.96
A presentation, talk, or class was canceled due to distressing content or offensive viewpoints.	0-73.00	6.46	0.00	13.03
It was emphasized that my emotional well-being is as important as my physical safety.	0-85.00	23.20	18.25	20.46

Note. Each week for four weeks, participants were asked to report how often they experienced each event on a sliding scale (0 = never, 50 = 3-4 times, 100 = daily or more often).

[a]The possible range of each item was 0-100.

Does Exposure to Safetyism Predict Changes in Resilience?

Perceived Resilience

We computed a linear regression to test whether average safetyism exposure over the duration of the study predicted perceived resilience at endpoint assessment, controlling for perceived resilience at baseline and relevant demographic variables. This analysis revealed substantial evidence in favor of the null hypothesis that safetyism exposure was not related to changes in perceived resilience ($\beta = -0.002$, $BF = 0.19$, $n = 106$).

Stress Mindset

We computed a linear regression to test whether average safetyism exposure over the duration of the study predicted stress mindset at endpoint assessment, controlling for stress mindset at baseline and relevant demographic variables. This analysis revealed substantial evidence in favor of the null hypothesis that safetyism exposure was not related to changes in stress mindset ($\beta = 0.001$, $BF = 0.12$, $n = 106$).

Stress Appraisal

We computed a linear regression to test whether average safetyism exposure over the duration of the study predicted participants' stress appraisal of the public speaking task at endpoint assessment as a challenge versus a threat, controlling for their stress appraisal of the task at baseline and relevant demographic variables. This analysis revealed inadequate evidence to adjudicate between the null hypothesis and the alternative hypothesis that exposure to safetyism was related to changes in stress appraisal ($\beta = 0.007$, $BF = 0.50$, $n = 73$). In other words, our sample size was insufficient to inform a conclusion about this hypothesis.

Anticipatory Anxiety

We computed a linear regression to test whether average safetyism exposure over the duration of the study predicted participants' anticipatory anxiety in response to the public speaking task at endpoint assessment, controlling for their anticipatory anxiety response to the task at baseline and relevant demographic variables. Again, we found there was inadequate evidence to adjudicate between the null hypothesis and the alternative hypothesis ($\beta = 0.144$, BF = 0.62, n = 73).

Exploratory Moderation Analyses

We further explored whether four theoretically relevant variables (i.e., psychiatric history; trait anxiety; perceived need for safetyism; and the belief that words can harm) moderated of the relationship between exposure to safetyism and each dependent variable by using regression-based interaction detections. We found substantial evidence (BFs ranging from 0.13 to 0.29) in favor of the null hypothesis; that is, the hypotheses that these variables did not meaningfully moderate the relationship between safetyism exposure and changes in perceived resilience, stress mindset, or stress appraisals. We found insufficient evidence (BFs ranging from 0.38 to 0.98) to adjudicate between the null hypothesis and the alternative hypothesis that these variables moderated the relationship between safetyism exposure and changes in anticipatory anxiety. The full results of these analyses are reported in the online supplemental materials.

Discussion

Debates about safetyism on college campuses—efforts to keep students "safe" from emotionally challenging ideas and experiences via trigger warnings, safe spaces, and related practices—have animated cultural discourse in recent years. Proponents argue that these practices are necessary to support the well-being of vulnerable students (Stokes, 2014), whereas critics argue that they "coddle" students and harm their resilience (Lukianoff & Haidt, 2018). The current study was the first to test the longitudinal impact of campus safetyism on students'

resilience. We find evidence against both claims—greater exposure to safetyism among first-year college students was neither associated with improvements nor reductions in resilience.

Contrary to our hypotheses, we found evidence in favor of the null effects of campus safetyism: greater exposure to safetyism had no relationship with changes in students' perceived resilience or stress mindset. Our sample size was not sufficient to determine any relationship between exposure to safetyism and changes in their overall appraisals of and anxiety response to a stressful public speaking task.

We did not find any relationships between safetyism exposure and adverse consequences. On one hand, this suggests safetyism is unlikely to be harmful to resilience. On the other hand, we found no benefits of these practices either. If safetyism practices make students feel supported and psychologically safe, as proponents argue, we might expect that greater exposure to these practices would bolster students' confidence in their resilience and improve their ability to cope with the public speaking task. Instead, we found no impact, either positive or negative, of safetyism exposure. Strikingly, even those students who believed most strongly in their own need for safetyism practices did not appear to benefit from them in any way. The belief that safetyism was beneficial or necessary was indeed more prevalent among vulnerable students—those who doubted their resilience and believed more strongly that words can harm them—but it was not associated with any positive outcomes of safetyism exposure. Taken together, our findings suggest that safetyism practices neither help nor hurt students.

Why don't safetyism practices seem to make a difference among college students? Though our current results cannot directly answer this question, there are some plausible hypotheses that might be explored in future work. One possibility is that these practices are simply a very small part of the complex and dynamic life of a young person. If they have some kind of effect, it may be trivially small. Another possibility is that safetyism practices do have an

effect, but the effect occurs only upon initial exposure or occurs only in an earlier developmental window. We suspect that the students in our study had already been exposed to many safetyism practices before they entered the university, in high school or earlier.

Some research suggests that beyond safetyism practices, students' safetyism *beliefs* predict a range of outcomes. For example, Lukianoff and Haidt (2018) argue that support for safetyism is rooted in a number of cognitive distortions (e.g., catastrophizing) and beliefs (e.g., the belief that words can harm) that may be maladaptive for adolescents. Supporting this view, one study found that students' endorsement of safetyism-inspired beliefs (e.g., emotional pain or discomfort is dangerous) and practices (e.g., support for trigger warnings) was associated with cognitive distortions (Celniker et al., 2022). However, it is currently unclear whether or not exposure to safetyism practices has a direct impact on safetyism beliefs. The direction of causation may be reversed, bidirectional, or caused by a third variable. Future research could further examine these questions.

Overall, our findings align with experimental research on the (in)efficacy of trigger (content) warnings. Although trigger warnings have been shown to increase anxiety in undergraduates relative to those not receiving them, the anxiogenic effects, when they occur, tend to be relatively small and evanescent (Bellet et al., 2020). Indeed, a meta-analysis of these experiments shows that trigger warnings have virtually no effect on people's response to potentially distressing material aside from increasing anticipatory anxiety (Bridgland et al., 2023). However, a study of trauma survivors with probable posttraumatic stress disorder indicated that trigger warnings iatrogenically increase how central one's trauma is to the survivor's identity (Jones et al., 2020). Trigger warnings do not seem to have any particular clinical benefit, but harmful effects are far less significant than staunch opponents would make it seem—perhaps the same is true of campus safetyism.

Prevalence of Safetyism

Our data provide a window into the prevalence of safetyism on college campuses as reported by students themselves, beyond anecdotal evidence and media reports. In terms of the relative prevalence of safetyism practices, the most controversial and widely publicized instances of safetyism—trigger warnings, safe spaces, and the deplatforming of speakers—were among the rarest.

Notably, the two practices most frequently reported by students were (1) being informed of campus mental health resources and (2) being encouraged to seek out mental health resources if they experience school-related stress. This seems relevant to debates about the pervasiveness of mental health culture, mental health awareness, and the "psychiatrization" of everyday experiences of distress (Haslam et al., 2021). Many consider mental health awareness efforts to be positive, but others warn about expanding psychiatric concepts potentially contributing to over-diagnosis and greater psychological vulnerability (Haslam et al., 2021; Jones & McNally, 2021). Our results suggest that among safetyism-related practices, there seems to be an emphasis on heightening student awareness to available mental health resources and encouraging their utilization. Future research might consider evaluating the efficacy of these efforts (Foulkes & Andrews, 2023).

Limitations and Future Directions

Our study was not without limitations. One limitation is the difficulty in reliably assessing resilience as a construct. Some researchers have cautioned against the use of questionnaires to measure resilience—although self-reported measures demonstrate strong cross-sectional correlations with relevant psychological constructs, they tend to be less reliable in predicting future mental health outcomes (Bonanno, 2021). One strength of our study was the use of a behavioral social stressor task to supplement self-reports of resilience. Still, future

researchers might consider exploring a broader range of resilience-related variables and outcomes related to safetyism, such as mental health service utilization, the likelihood of receiving a psychiatric diagnosis following a stressful event, or academic outcomes.

A second limitation was the short duration of the study (i.e., four weeks). It is possible that our study was too short to capture any effects of safetyism that would materialize on a longer timescale. Relatedly, our study design precludes us from capturing or controlling for any effects of safetyism that have already materialized among participants. Safetyism has become a widely discussed topic not just in terms of its prevalence on college campuses but in high schools, parenting styles, and culture writ large. If students had already encountered safetyism upon enrolling in our study, either through the media or direct exposure, perhaps any positive or negative effects had already materialized. Anticipating the potentially limiting effect of prior safetyism exposure, we intentionally selected a sample of first-semester college students to ensure that our participants had no prior direct exposure with college campus safetyism.

Finally, we intentionally cast a wide net in the study of safetyism by examining a broad, ad hoc set of campus practices conceivably aimed at shielding students from potentially distressing words, ideas, and experiences. In doing so, we sought to understand how college students might be affected by an environment in which they are encouraged to avoid, rather than embrace or engage with, emotional discomfort. The scale we developed is by no means an exhaustive list of safetyism practices, and we encourage future researchers to contribute to this new body of work by further investigating the individual practices we have identified, or by identifying new mental health practices ripe for empirical investigation. For example, while there is now a substantial amount of research on trigger warnings (Bridgland et al., 2023), very little research has investigated the efficacy of safe spaces (however, see Gainsburg & Earl, 2022). We encourage future researchers to follow the research agenda paved by scholars of trigger

warnings: rigorously testing the efficacy of mental health practices that have been employed widely in popular culture prior to empirical testing.

Conclusion and Implications for Higher Education

Formal clinical practice requires that the effectiveness of psychological treatments be established through rigorous empirical testing before being employed. In higher education settings, a variety of safetyism practices have proliferated in recent years aimed at bolstering student well-being by shielding them from emotional discomfort. This study was the first to longitudinally measure the relationship between safetyism practices and college students' resilience. We found evidence that self-reported exposure to safetyism had little relationship to student outcomes for better or for worse. Exposure to safetyism did not relate to decreases in psychological resilience within our study. However, to the extent that these practices are aimed at supporting students in their ability to master challenging situations, our findings suggest no positive relationship with this goal.

Acknowledgements

The authors thank the Harvard College Research Program for generous funding that made this study possible.

Declaration of Interests

The authors report there are no competing interests to declare. The authors confirm that the research presented in this article met the ethical guidelines, including adherence to the legal requirements, of United States of America and received approval from the Institutional Review Board of Harvard University.

Author Contributions Statement

SP, BWB, and RJM Contributed to study conception and design. SP conducted data analysis under the guidance of PJJ, BWB, and RJM. SP wrote the first draft of the manuscript and all other authors provided critical revisions and approved the final version for publication. All authors agree to be accountable for all aspects of the work.

Funding Details

This work was supported by a grant from the Harvard College Research Program awarded to the first author.

Data Availability Statement

The data that support the findings of this study are openly available in the Open Science Framework repository at https://osf.io/q6dyx/?view_only=24149ca7283e451786f80a85ad9509b3

References

Allen, A. P., Kennedy, P. J., Dockray, S., Cryan, J. F., Dinan, T. G., & Clarke, G. (2017). The Trier Social Stress Test: Principles and practice. *Neurobiology of Stress*, *6*, 113–126. https://doi.org/10.1016/j.ynstr.2016.11.001

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

Bellet, B. W., Jones, P. J., Meyersburg, C. A., Brenneman, M. M., Morehead, K. E., & McNally, R. J. (2020). Trigger warnings and resilience in college students: A preregistered replication and extension. *Journal of Experimental Psychology: Applied*, *26*(4), 717–723. https://doi.org/10.1037/xap0000270

Blascovich, J., & Tomaka, J. (1996). The Biopsychosocial Model of Arousal Regulation. In *Advances in Experimental Social Psychology* (Vol. 28, pp. 1–51). Elsevier. https://doi.org/10.1016/S0065-2601(08)60235-X

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

Bonanno, G. A., Kennedy, P., Galatzer-Levy, I. R., Lude, P., & Elfström, M. L. (2012). Trajectories of resilience, depression, and anxiety following spinal cord injury. *Rehabilitation Psychology*, *57*(3), 236–247. https://doi.org/10.1037/a0029256

Bower, L. J. (2023). The Woman in Black: A Defense of Trigger Warnings in Creating Inclusive Academic Spaces for Trauma-Affected Students through a Feminist Disability Studies Pedagogy. *Journal of Criminal Justice Education*, 1–18. https://doi.org/10.1080/10511253.2023.2264370

Bridgland, V. M. E., Jones, P. J., & Bellet, B. W. (2023). A Meta-Analysis of the Efficacy of Trigger Warnings, Content Warnings, and Content Notes. *Clinical Psychological Science*, 21677026231186625. https://doi.org/10.1177/21677026231186625

Campus Deplatforming Database. (2024). *Foundation for Individual Rights in Education*. https://www.thefire.org/research-learn/campus-deplatforming-database

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*, 111243. https://doi.org/10.1016/j.paid.2021.111243

Crum, A. J., Salovey, P., & Achor, S. (2013). Rethinking stress: The role of mindsets in determining the stress response. *Journal of Personality and Social Psychology*, *104*(4), 716–733. https://doi.org/10.1037/a0031201

Etz, A., Haaf, J. M., Rouder, J. N., & Vandekerckhove, J. (2018). Bayesian Inference and Testing Any Hypothesis You Can Specify. *Advances in Methods and Practices in Psychological Science*, *1*(2), 281–295. https://doi.org/10.1177/2515245918773087

Foulkes, L., & Andrews, J. L. (2023). Are mental health awareness efforts contributing to the rise in reported mental health problems? A call to test the prevalence inflation hypothesis. *New Ideas in Psychology*, *69*, 101010. https://doi.org/10.1016/j.newideapsych.2023.101010

Gainsburg, I., & Earl, A. (2022). Safe here, but unsafe there? Institutional signals of identity safety also signal prejudice in the broader environment. *Journal of Experimental Social Psychology*, *98*, 104232. https://doi.org/10.1016/j.jesp.2021.104232

Haidt, J. (2017, March 2). *Van Jones' Excellent Metaphors About the Dangers of Ideological Safety*. https://heterodoxacademy.org/blog/van-jones-excellent-metaphors/

Haslam, N., Tse, J. S. Y., & De Deyne, S. (2021). Concept Creep and Psychiatrization. *Frontiers in Sociology*, *6*, 806147. https://doi.org/10.3389/fsoc.2021.806147

Ho, K. (2017, January 30). Tackling the Term: What is a Safe Space? *Harvard Political Review*. https://harvardpolitics.com/what-is-a-safe-space/#google_vignette

Holbrook, T. L., Hoyt, D. B., Stein, M. B., & Sieber, W. J. (2001). Perceived Threat to Life Predicts Posttraumatic Stress Disorder after Major Trauma: Risk Factors and Functional Outcome: *The Journal of Trauma: Injury, Infection, and Critical Care*, *51*(2), 287–293. https://doi.org/10.1097/00005373-200108000-00010

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*. https://doi.org/10.1037/tra0001063

Lilienfeld, S. O. (2007). Psychological Treatments That Cause Harm. *Perspectives on Psychological Science*, *2*(1), 53–70. https://doi.org/10.1111/j.1745-6916.2007.00029.x

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.

McNally, R. J., Enock, P. M., Tsai, C., & Tousian, M. (2013). Attention bias modification for reducing speech anxiety. *Behaviour Research and Therapy*, *51*(12), 882–888. https://doi.org/10.1016/j.brat.2013.10.001

Mendes, B. W., Gray, H. M., Mendoza-Denton, R., Major, B., & Epel, E. S. (2007). Why Egalitarianism Might Be Good for Your Health: Physiological Thriving During Stressful

Intergroup Encounters. *Psychological Science*, *18*(11), 991–998. https://doi.org/10.1111/j.1467-9280.2007.02014.x

Morey, R., & Rouder, J. (2024). *BayesFactor: Computation of Bayes Factors for Common Designs* (R package version 0.9.12-4.4) [Computer software]. https://cran.r-project.org/web//packages//BayesFactor/BayesFactor.pdf

Seery, M. D., Weisbuch, M., Hetenyi, M. A., & Blascovich, J. (2010). Cardiovascular measures independently predict performance in a university course. *Psychophysiology*, *47*(3), 535–539. https://doi.org/10.1111/j.1469-8986.2009.00945.x

Shulevitz, J. (2015, March 21). In College and Hiding From Scary Ideas. *The New York Times*. https://www.nytimes.com/2015/03/22/opinion/sunday/judith-shulevitz-hiding-from-scary-ideas.html

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

Stokes, M. (2014, May 29). In Defense of Trigger Warnings. *The Conversation*. https://www.chronicle.com/blogs/conversation/in-defense-of-trigger-warnings

Stress, appraisal, and coping. (1984). Springer Publishing Company.

The Stanford Guide to Acceptable Words. (2022, December 19). *The Wall Street Journal*. https://www.wsj.com/articles/the-stanford-guide-to-acceptable-words-elimination-of-harmful-language-initiative-11671489552