Trigger Warnings and Resilience in College Students: A Preregistered Replication and Extension

Benjamin W. Bellet and Payton J. Jones
Harvard University

Cynthia A. Meyersburg
Foundation for Individual Rights in Education, Philadelphia, Pennsylvania

Miranda M. Brenneman
Coastal Carolina University

Kaitlin E. Morehead
Coastal Carolina University and Foundation for Individual Rights in Education, Philadelphia, Pennsylvania

Richard J. McNally
Harvard University

Trigger warnings notify people that content they are about to engage with may result in adverse emotional consequences. An experiment by Bellet, Jones, and McNally (2018) indicated that trigger warnings increased the extent to which trauma-naïve crowd-sourced participants see themselves and others as emotionally vulnerable to potential future traumas but did not have a significant main effect on anxiety responses to distressing literature passages. However, they did increase anxiety responses for participants who strongly believed that words can harm. In this article, we present a preregistered replication of this study in a college student sample, using Bayesian statistics to estimate the success of each effect's replication. We found strong evidence that none of the previously significant effects replicated. However, we found substantial evidence that trigger warnings' previously nonsignificant main effect of increasing anxiety responses to distressing content was genuine, albeit small. Interpretation of the findings, implications, and future directions are discussed.

Public Significance Statement
This article examines whether giving a trigger warning for distressing content is psychologically helpful or harmful for college students. We found that trigger warnings increase students' anxiety responses to distressing content in the short term, but do not have effects on any other aspects of psychological resilience. Trigger warnings do not appear to be helpful for college students, and, if anything, cause small, temporary increases in anxiety.

Keywords: trigger warning, resilience, college students, preregistered replication

Supplemental materials: http://dx.doi.org/10.1037/xap0000270.supp

Giving a "trigger warning" is the practice of providing notification of the potential adverse emotional consequences of forthcoming content (Boysen, 2017). Trigger warnings originated in online discussion groups for survivors of sexual trauma. Their purpose was to warn participants about posted content capable of triggering reexperiencing symptoms (e.g., flashbacks) characteristic of posttraumatic stress disorder (PTSD), thereby enabling survivors to brace themselves to process such content or to avoid it altogether (Carter, 2015; Wyatt, 2016).

Concerns about sexual trauma among college students (Krebs, Lindquist, Warner, Fisher, & Martin, 2009; Walsh et al., 2012) prompted calls for trigger warnings in the classroom (Stokes, 2014). However, evidence suggests that gradual exposure to trauma cues among those with PTSD allows them to habituate and regain functioning (Institute of Medicine, 2008), whereas avoidance of such cues is associated with higher severity of other PTSD symptoms (Rosenthal, Hall, Palm, Batten, & Follette, 2005). Accordingly, some clinical psychologists have cautioned that trigger warnings may countertherapeutically encourage avoidance of

[ID] Benjamin W. Bellet and [ID] Payton J. Jones, Department of Psychology, Harvard University; [ID] Cynthia A. Meyersburg, Foundation for Individual Rights in Education, Philadelphia, Pennsylvania; [ID] Miranda M. Brenneman, Department of Psychology, Coastal Carolina University; Kaitlin E. Morehead, Department of Psychology, Coastal Carolina University, and Foundation for Individual Rights in Education; [ID] Richard J. McNally, Department of Psychology, Harvard University.

Cynthia A. Meyersburg is now at the Department Psychology, Harvard University. Kaitlin E. Morehead is now at Hilton Grand Vacations, Myrtle Beach, South Carolina.

Correspondence concerning this article should be addressed to Benjamin W. Bellet, Department of Psychology, Harvard University, 33 Kirkland Street, Cambridge, MA 02138. E-mail: bbellet@g.harvard.edu

trauma-related cues in those with PTSD (e.g., McNally, 2014). Regardless of concerns about whether they are helpful for trauma survivors, warnings have been increasingly requested by students, and for a far wider range of content than that related to canonical traumas (e.g., depictions of classism and sexism; Lukianoff & Haidt, 2015; Wilson, 2015).

As their range and frequency of use have expanded beyond concerns about PTSD, trigger warnings have become the center of a spirited debate about whether they are helpful or harmful to students, participants in the workplace, and media consumers in the general population. Proponents argue that they allow members of marginalized groups to participate in online, work, and academic settings without becoming emotionally dysregulated (Carter, 2015; Stokes, 2014). However, others worry that they hamper academic freedom or that they "coddle" students, inflating their sense of vulnerability and creating unreasonable expectations about the world outside the campus gates (Lukianoff & Haidt, 2015). From the standpoint of experimental psychology, an important and empirically tractable question in this debate was whether trigger warnings do in fact strengthen or undermine individual resilience.

There are reasons to suspect that trigger warnings may have unintended adverse consequences in the general population. Trauma survivors who appraise their acute symptoms as harbingers of enduring debilitation are at heightened risk for developing PTSD (Dunmore, Clark, & Ehlers, 2001; Ehring, Ehlers, & Glucksman, 2006). By implying that certain content may trigger intense distress, trigger warnings may encourage people to catastrophically interpret acute symptoms should they encounter serious trauma later in life. Trigger warnings may raise awareness of the struggles of trauma survivors but may also lead people to believe that trauma invariably undermines survivors' ability to cope with everyday stressors. In fact, most people who experience trauma are resilient, with acute posttraumatic symptoms markedly dissipating over time (Breslau & Kessler, 2001).

Trigger warnings may also counterproductively create an expectancy of imminent distress, thereby inducing heightened anxiety and functioning similarly to a nocebo (e.g., Barsky, Saintfort, Rogers, & Borus, 2002). This effect may be especially pronounced in those who believe that triggering content is intrinsically harmful. On the other hand, negative stimuli are less distressing when they are perceived as predictable (Grupe & Nitschke, 2013; Mineka & Kihlstrom, 1978), or do not violate peoples' expectations about how distressing they will be (Telch, Harrington, Smits, & Powers, 2011). Thus, trigger warnings might lower anxiety levels by rendering distressing content predictable. However, people exposed to trigger warnings may also develop the implicit assumption that distressing material will always be predictable, making even relatively innocuous content viewed later without a warning appear more startling and more anxiogenic.

Bellet et al. (2018) were the first to test whether trigger warnings affect the resilience of their recipients. In this experiment, trauma-naïve participants read a series of distressing passages from world literature. Each passage was either preceded or not preceded by a trigger warning (experimental and control conditions, respectively). Participants indicated their anxiety responses to these passages and to subsequently presented mildly distressing passages without warnings and reported on beliefs relevant to psychological resilience. Trigger warnings increased participants' per-

ceptions of their own and other people's vulnerability to developing enduring psychological impairment in the event of experiencing trauma. Overall, trigger warnings did not affect immediate anxiety responses to distressing passages or to the subsequently presented, mildly distressing passages. However, trigger warnings increased immediate anxiety response to distressing passages for those who held a strong belief that words can cause harm.

Other recent studies on the effects of trigger warnings have yielded mixed findings. Bruce, O'Brien, Hoffmann, and Roberts (2019) found that trigger warnings given prior to viewing a film clip produced larger physiological markers of anxiety in participants prior to viewing the film than did PG-13 warnings or no warning. Similarly, Bridgland, Green, Oulton, and Takarangi (2019) found that relative to a control condition, trigger warnings shown prior to the viewing of photographs increased anxiety, negative affect, and negative expectations in participants prior to viewing, but had a trivially small negative effect on arousal levels after the photographs were viewed. Gainsburg and Earl's (2018) set of studies found that participants experienced increased anticipated negative affect and were more likely to avoid both film clips and essays when they had a trigger warning. However, after reading the essays, participants who saw trigger warnings reported slightly less negative affect compared to those who did not receive a warning. Ironically, anticipated negative affect and likelihood of avoidance were increased among participants who believed trigger warnings are protective, consistent with Bellet et al.'s (2018) finding that beliefs can influence trigger warnings' effects. Finally, Sanson, Strange, and Garry (2019) concluded that trigger warnings were overall neither helpful nor harmful. Across six experiments involving college students and Internet users, they found that people experienced similar levels of negative affect and intrusive memories in relation to the content they viewed, regardless of whether they had received a trigger warning.

Given the mixed findings concerning trigger warnings and the small effects in Bellet and colleagues' (2018) study, replication of this original study is warranted. Further, college students were not the focus of the original study, despite being at the center of the debate and presumably more likely than nonstudents to encounter trigger warnings. Thus, the primary aim of the current study was to conduct a preregistered replication and extension of Bellet and colleagues' (2018) original experiment in a college student population. At the suggestion of a reviewer of an earlier draft of this article, we also conducted an analysis that was not preregistered: we assessed college students' attitudes toward trigger warnings to see how they compared with those reported in our original study.

Method

All experimental methods, stimuli, and informed consent procedures for the current study received institutional approval from the Harvard University Committee on the Use of Human Subjects. The design and proposed data analyses were preregistered on the Open Science Framework (OSF; see https://osf.io/egpyz). Once participants entered the online experiment, the procedure was the same as Bellet et al.'s (2018) except that we did not administer an Implicit Association Test (Greenwald, McGhee, & Schwartz, 1998) or the World Assumptions Scale (WAS; Janoff-Bulman, 1989). We omitted these measures because they were unrelated to the key variables in our previous experiment. In our preregistra-

tion, we specified a sequential data collection procedure, in which we planned to collect a large enough sample from each site to reach a decisive level of evidence within a Bayesian statistical framework for either replication or nonreplication of the previously significant effects. This criterion was defined as a Bayes factor (BF) > 10 (decisive evidence for replication) or BF < .10 (decisive evidence for nonreplication). If neither site's sample yielded decisive evidence by the end of the specified recruitment period (the end of the 2018–2019 academic year), we planned to pool the samples from each site for all analyses. For the current study, this procedure indicated that the pooling of samples from both sites was warranted. See S1 in the online supplemental materials for a detailed description of the sequential data collection procedure and an examination of differences between site-specific samples.

Participants

We recruited participants from two sites: a private university in the northeastern United States (Site 1) and a regional state university in the southern United States (Site 2). Participants were recruited via each site's respective study pool website. At Site 1, individuals interested in participating were directed from the study advertisement to the experimental survey, delivered online via Qualtrics. At Site 2, participants completed the same online Qualtrics-based experimental survey in the laboratory. Once they had entered the survey, participants answered a single-item screening question that asked whether they had ever experienced a canonical stressor that would qualify for a Criterion A trauma in a diagnosis of PTSD in the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM–5*; American Psychiatric Association, 2013). Participants endorsing exposure to a traumatic event were excluded from the current study. Next, participants confirmed that they were full-time undergraduate students. A total of 487 trauma-naïve full-time undergraduates were recruited; six participants were excluded from analyses because they had received a diagnosis of PTSD, and 19 participants were excluded from analyses because they answered content-based attention checks incorrectly. This left 426 participants from Site 1 and 36 participants from Site 2 (i.e., *N* = 462).

Materials

Measures and stimuli used in our experiment can be viewed at the OSF page for the original experiment (https://osf.io/rk4yu/). We measured participants' emotional reaction to passages from world literature that would commonly appear in a typical high school or college literature class. Passages were standardized by word length, and their duration of time on the screen was set to a minimum of 20 s before participants could proceed to the next screen. Three types of passages were used. "Neutral" passages contained no disturbing content (e.g., a character description from Herman Melville's *Moby-Dick*). "Mildly distressing" passages contained themes of violence, injury, or death but did not contain graphic details (e.g., a description of a battle from James Bradley's *Flags of our Fathers*). "Markedly distressing" passages contained graphic scenes of violence, injury, or death (e.g., the murder scene from Fyodor Dostoevsky's *Crime and Punishment*). We previously conducted a pilot study on Amazon's Mechanical Turk to

confirm that each passage elicited a level of anxiety consistent with its assigned category (Bellet et al., 2018).

Measures

Perceived Posttraumatic Vulnerability Scale—Self (PPVS-S). The PPVS-S (Bellet et al., 2018) is a 19-item questionnaire that assesses the extent to which participants believe that if they were to experience trauma, they would suffer persistent and debilitating emotional harm. Participants are asked to imagine themselves experiencing a trauma, and to indicate their level of endorsement for each statement concerning its effects (e.g., "I would lose my grip on reality") on a 100-point scale (1 = *disagree*, 100 = *agree*). These responses were averaged for a composite score. The PPVS-S displayed excellent internal consistency in the current study (α = .93).

Perceived Posttraumatic Vulnerability Scale—Other (PPVS-O). The PPVS-O (Bellet et al., 2018) is a 19-item questionnaire that assesses the extent to which participants believe that if any person were to experience trauma, they would experience persistent and debilitating emotional harm. Participants are asked to imagine a hypothetical "average" person experiencing a trauma and to indicate their level of endorsement for each statement concerning its effects (e.g., "He/She would feel isolated and alone") on a 100-point scale (1 = *disagree*, 100 = *agree*). The PPVS-O displayed excellent internal consistency in the current study (α = .93).

Words Can Harm Scale (WCHS). The WCHS (Bellet et al., 2018) is a 10-item scale that assesses the degree to which participants believe that offensive words can cause serious harm to them or to people in general. Participants indicated their level of endorsement for each statement (e.g., "I could be traumatized without ever being touched, just through someone's hurtful words.") on a 100-point scale (1 = *disagree*, 100 = *agree*). Responses were averaged for a composite score. The WCHS has displayed convergent validity in its association with a measure of perception of threat from internal sensations (Jones, 2019), and displayed good internal consistency in the current study (α = .87).

Trigger Warning Attitudes Assessment (TWAA). The TWAA assesses respondents' attitudes toward trigger warnings. Participants are first given a short definition of trigger warnings, and then are asked "Do you think that trigger warnings should be used?" If participants answer "Yes," they are then asked to select all reasons that they believe trigger warnings should be used from a list of potential reasons (e.g., protection of psychologically vulnerable populations, providing for the needs of minority groups) that includes a fillable "other" category.

Demographics questionnaire. This questionnaire asks for participants' gender, self-reported race and ethnicity, religiosity, and political orientation. Religiosity and political orientation were assessed with a 5-point Likert scale (1 = *not religious*, 5 = *extremely religious*; 1 = *very liberal*, 5 = *very conservative*).

Psychiatric history questionnaire. This questionnaire asks participants whether they have "ever been diagnosed with a psychiatric or psychological problem." If participants answer *yes*, they are asked to choose all diagnoses that apply from a list, including an "other" option that allows a free response of any disorders not listed on the questionnaire.

Experimental Procedures

This study was a two-group randomized controlled experiment with a control and experimental group. After providing informed consent and passing all screening questions, participants were randomly assigned to either the no warning (control) or trigger warning (experimental) condition. Participants in both conditions then read three mildly distressing passages in random order. After each passage, they used slider bars ranging from 0 (*not at all*) to 100 (*very much*) to rate the intensity of their response on the following dimensions: sad, happy, afraid, anxious, angry, content, disgusted; degree of unpleasant emotion overall; and degree of anticipated long-term negative emotion. The target emotion was anxiety; the other items were fillers included to diminish demand effects. The average of these anxiety responses served as each participant's baseline anxiety response.

Next, participants read another series of 10 passages in random order. Five were neutral, and the other five were markedly distressing. In the trigger warning condition, each of the markedly distressing passages was preceded by a trigger warning screen which had to be acknowledged by clicking a radio button (*TRIGGER WARNING: The passage you are about to read contains disturbing content and may trigger an anxiety response, especially in those who have a history of trauma.*). The no warning condition participants viewed a screen that indicated they were about to view the next passage, which was also acknowledged by clicking a radio button. Participants rated the intensity of their anxiety responses along with the filler questions after each markedly distressing passage. The difference between the average of these anxiety responses and the baseline anxiety response constituted each participant's immediate anxiety change. After completion of condition-specific passage presentations, participants read three more mildly distressing passages presented in random order and answered the same postpassage questions as before. The difference between the average of these anxiety responses and the baseline anxiety response constituted each participant's follow-up anxiety change. Next, participants completed the PPVS-S, PPVS-O, and WCHS in random order. Participants then responded to the TWAA, the demographics questionnaire, and the psychiatric diagnosis history questionnaire. Finally, participants were provided with a debriefing form. Participants at both sites who completed the entire survey attentively (as assessed by content-based attention checks embedded within the survey) were compensated one course credit each.

Plan of Analysis

The de-identified dataset and R code for all analyses are publicly available at the OSF page for this study (https://osf.io/693gj/). First, we conducted analyses analogous to those of Bellet and colleagues (2018). We examined the descriptive statistics of the demographic and psychiatric characteristics of the sample, as well as participants' attitudes toward trigger warnings. Next, we determined whether the experimental groups (trigger warning and no warning conditions) showed significant differences in any demographic or psychiatric characteristics. If this was the case, we added those variables as covariates in each of the subsequent multiple regression analyses. To determine the effect of trigger warnings on each variable (PPVS-S, PPVS-O, immediate anxiety change, and follow-up anxiety change), we conducted a linear regression with the variable of interest as a dependent variable, and condition (dummy coded as 1 = *trigger warning condition*, 0 = *no warning condition*) as an independent variable. To assess the presence of an interaction between trigger warnings and the belief the words can harm in predicting immediate anxiety change, we conducted a multiple regression with immediate anxiety change as the dependent variable, and condition, WCHS score, and the cross-product of condition and WCHS score as independent variables.

To test for replication, we calculated a replication BF (Verhagen & Wagenmakers, 2014). The *t* test statistic and degrees of freedom for the effect of interest calculated from the replication experiment was combined with analogous statistics from the original experiment to derive a BF indicating the weighted likelihood ratio that the replication data reflect the effect in the original experiment versus the null hypothesis that there is no effect. Replication BFs greater than 1 indicate a stronger likelihood of a successful replication of a given effect; BFs less than 1 indicate a stronger likelihood of an unsuccessful replication. If data collection ceased in accordance with our preregistration's specified sampling timeline before we could derive BFs greater than 10 or less than .10, we considered a BF of greater than or equal to 3 as substantial evidence for the success of replicating a given effect, and a BF of less than or equal to .33 as substantial evidence for the failure of replicating a given effect (Wetzels et al., 2011). In order to compute replication BFs for regression-based *t* statistics, we modified Verhagen and Wagenmaker's (2014) original code to account for changes in degrees of freedom due to intercepts and covariates. See S1 in the online supplemental materials for the R code for all analyses, S2 in the online supplemental materials for the R code for the modified replication BF function, and S3 in the online supplemental materials for the dataset used.

Results

Sample Characteristics

Our pooled sample contained mostly participants who identified as female ($n = 283$, 61.3%), with one participant ($n = 1$, 0.2%) who preferred not to specify a gender. The rest of the participants identified as males ($n = 178$, 38.5%). The mean age of the participants was 20 years old ($SD = 2.33$ years). Most participants racially identified as Caucasian ($n = 277$, 60.0%), with other participants identifying as Asian or Pacific Islander ($n = 110$, 23.8%), Black ($n = 34$, 7.4%), Native American or Alaska Native ($n = 3$, 0.6%), multiracial ($n = 26$, 5.6%), or "other" ($n = 12$, 2.6%). A substantial minority of participants identified their ethnicity as Hispanic ($n = 48$, 10.4%). Most participants identified as at least "somewhat religious" ($n = 269$, 58.2%), and a majority of participants identified as at least "somewhat liberal" ($n = 309$, 66.9%). A minority of participants ($n = 88$, 9.5%) had been diagnosed with at least one lifetime psychiatric disorder. A majority of participants believed that trigger warnings should be used (87.7%, $n = 405$). Of these, 94.3% ($n = 382$) believed that trigger warnings are needed to protect psychologically vulnerable individuals, such as those with PTSD; 64.0% ($n = 259$) believed they are needed to protect members of any minority group; and 53.1% ($n = 215$) believed they are needed to protect people in general.

Analyses

We first conducted group comparisons between conditions on all demographic variables to ensure successful randomization. No group comparisons were significant, so we proceeded with our planned regression analyses without controlling for any demographic characteristics, only including condition (0 = *no warning condition*, 1 = *trigger warning condition*) as a predictor of each outcome of interest. Table 1 shows the regression-based effects on each outcome of interest from the original and replication studies, and the replication BFs indicating the success of the replication attempt for each effect. We found substantial evidence for non-replication of trigger warnings' effect on perceived posttraumatic vulnerability for oneself (BF = .16), and decisive evidence for nonreplication of trigger warnings' effect on perceived posttraumatic vulnerability for others (BF = .08). However, we found substantial evidence that the small increases in immediate anxiety response to distressing content caused by trigger warnings in the first study were replicated in the second study (BF = 4.30), even though this effect was nonsignificant in the first study. We found substantial evidence for nonreplication of the effect of trigger warnings on anxiety change in response to subsequent, mildly distressing content viewed without a warning (BF = .22). Finally, we found decisive evidence for nonreplication of the interaction effect between trigger warnings and the belief that words can harm on immediate anxiety change in response to distressing material (BF = .04).

Discussion

This study provides insight into the replicability of Bellet and colleagues' (2018) study on trigger warnings and examines the effects of trigger warnings in a trauma-naïve college student population. Trigger warnings' effects on beliefs about posttraumatic vulnerability and their interaction with the belief that words can harm did not replicate in a college student sample. However, the previously nonsignificant effect of trigger warnings on anxiety levels, albeit small, is likely genuine. One explanation of these findings is that the previously observed effects were trivial, and that trigger warnings are more or less functionally inert, as others

have suggested (Sanson et al., 2019). However, what we failed to find is just as important as what we did find: in neither our original nor current experiment did trigger warnings work as intended. That is, they failed to reduce anxiety to distressing content. Rather, both studies strongly imply that trigger warnings reliably cause small increases in anxiety in college students without a history of trauma. This finding is consistent with research suggesting that trigger warnings increase anxiety in the short term (Bridgland et al., 2019; Bruce et al., 2019).

Another possible reason for nonreplication is that our college student sample differed in meaningful ways from the that of the original experiment, which used a crowd-sourced online sample. For example, the college student sample used in the second experiment was considerably younger than participants in the previous experiment, had a lower prevalence of lifetime psychiatric disorder diagnoses, and was compensated with course credit instead of money. The fact that this study was an extension into a different population with different characteristics limits the interpretability of the replication attempt to some extent. However, that we were able to provide insight into how trigger warnings function in a population for which they are especially relevant is a strength of the current study.

Our results also show that a majority of college students in our sample supported the use of trigger warnings. Further, a considerable proportion of these trigger warning proponents see them as an accommodation measure not just for those with pronounced psychological vulnerabilities, but also for members of any minority group, or for people in general. These findings tally with similar results from the original study by Bellet et al. (2018), and further suggest that many people see trigger warnings as applicable to a far broader range of concerns than just accommodating individuals suffering from PTSD (Boysen, 2017; Lukianoff & Haidt, 2015).

A limitation of both the original and replication studies is the exclusive use of self-report in the measurement of anxiety. It would have been desirable to have multiple measures of anxiety such as skin conductance reactivity as well as self-report on a visual analogue scale. Participants exhibiting increases on both in response to a trigger warning would be more convincingly deemed

Table 1
Statistics for Replication of Effects of Interest

Effect	Outcome variable	Original study statistic[a]	Replication study statistic[a]	Replication Bayes factor[b]
		Study-specific regression statistics		
Condition	PPVS-S	$t(267) = 2.13^*$	$t(460) = .08$	.16
Condition	PPVS-O	$t(267) = 2.19^*$	$t(460) = -.69$	.08
Condition	IAC	$t(267) = .51$	$t(460) = 2.19^*$	4.30
Condition	FAC	$t(267) = .15$	$t(460) = .00$	.22
Condition × WCHS	IAC	$t(265) = 2.79^{**}$	$t(458) = -.29$	.04

Note. The t statistics reflect values corresponding to the effect listed that were derived from regression analyses that included the listed variable as a predictor. PPVS-S = Perceived Posttraumatic Vulnerability Scale—Self; PPVS-O = Perceived Posttraumatic Vulnerability Scale—Other; IAC = immediate anxiety change; FAC = follow-up anxiety change; Condition is a dummy coded variable (0 = *no warning condition*; 1 = *trigger warning condition*); Condition × WCHS = cross-product of condition and Words Can Harm Scale (WCHS) score.
[a] All statistics from original study reflect statistics for effects determined while controlling for self-reported political orientation. No variables were controlled for in the replication study's analyses. [b] Bayes factors ascending higher from 1 indicate stronger relative likelihood that the effect from the original study is present in the replication study. Bayes factors descending lower from 1 indicate stronger relative likelihood that no effect is present in the replication study.
$^* p > .05.$ $^{**} p < .01.$

"anxious" than those whose anxiety occurs only on a self-report measure. However, measuring psychophysiology would be extremely difficult in an online experiment. In any case, other laboratory research indicates that trigger warnings do in fact cause significant increases in psychophysiological signs of anxiety (i.e., heart rate and electrodermal responding) relative to "PG-13" warnings and not receiving any warning (Bruce et al., 2019).

Another limitation of our study is that our sample was primarily composed of college students from a private university in the northeastern United States, with a smaller proportion from a large public university. However, other work has shown similarly trivial effects of trigger warnings in college student samples from other countries (Sanson et al., 2019). Further research should be conducted on trigger warnings with a more nationally representative sample of college students. A further limitation is that we excluded students with a trauma history. However, Sanson et al. (2019) similarly found that trigger warnings trivially increased negative affect in traumatized individuals, even when the triggering material was directly relevant to participants' traumas. Further, other research has found that trigger warnings increase physiological markers of anxiety response in trauma survivors as a function of the severity of their PTSD symptoms (Bruce et al., 2019). Ironically, trigger warnings may (temporarily) worsen the well-being of the very people they are intended to help.

There are additional, substantively distinct concerns about trigger warnings specific to people with PTSD. McNally (2014) suggested that trigger warnings may encourage countertherapeutic avoidance behaviors and reinforce the centrality of trauma to individual's identities, both of which prospectively predict higher levels of PTSD symptoms (Berntsen & Rubin, 2007; Boelen, 2012; Robinaugh & McNally, 2011; Rosenthal et al., 2005). Consistent with this concern, people tend to avoid content accompanied by a trigger warning significantly more frequently than those who do not receive such warnings (Bridgland et al., 2019; Gainsburg & Earl, 2018). In a more direct (albeit cross-sectional) examination of this relationship, Bruce et al. (2019) found a positive association between amount of trigger warning use and severity of PTSD avoidance symptoms. Bruce et al. (2019) also found a positive association between trigger warning use and the extent to which a traumatic event was central to survivors' identities. Future research should examine the generalizability of our findings to a traumatized population and determine the extent to which PTSD severity moderates the effects observed. Additionally, future research should use controlled experimental designs to disambiguate the direction of causality between trigger warning use, avoidance symptoms, and centrality of traumatic events to trauma survivors' identities.

One obvious question that arises from the finding that trigger warnings are broadly functionally inert (and if anything, increase anxiety levels) for college students is whether or not trigger warnings should be used in college classrooms. Our results and those of other studies indicate that if teachers are aiming to broadly reduce student anxiety, providing trigger warnings will not accomplish this aim. One possible alternate practice is for teachers to review the syllabus and course readings with students at the outset of each semester. Indeed, this common practice provides an informative overview of the course without insinuating that anyone might become emotionally dysregulated, let alone harmed, by enrolling in it. There are many other potential reasons why teach-

ers might choose to use trigger warnings that are not addressed by this study; these results cannot inform decisions on whether "to warn or not to warn" wholesale and are limited to elucidating acute effects on emotional reactions and trauma-relevant beliefs. However, this study and other recent research efforts are an important step forward in allowing policymakers, educators, and clinicians to make empirically informed decisions about the use of trigger warnings.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

Barsky, A. J., Saintfort, R., Rogers, M. P., & Borus, J. F. (2002). Nonspecific medication side effects and the nocebo phenomenon. *Journal of the American Medical Association, 287,* 622–627. http://dx.doi.org/10.1001/jama.287.5.622

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. http://dx.doi.org/10.1016/j.jbtep.2018.07.002

Berntsen, D., & Rubin, D. C. (2007). When a trauma becomes a key to identity: Enhanced integration of trauma memories predicts posttraumatic stress disorder symptoms. *Applied Cognitive Psychology, 21,* 417–431. http://dx.doi.org/10.1002/acp.1290

Boelen, P. A. (2012). A prospective examination of the association between the centrality of a loss and post-loss psychopathology. *Journal of Affective Disorders, 137,* 117–124. http://dx.doi.org/10.1016/j.jad.2011.12.004

Boysen, G. A. (2017). Evidence-based answers to questions about trigger warnings for clinically-based distress: A review for teachers. *Scholarship of Teaching and Learning in Psychology, 3,* 163–177. http://dx.doi.org/10.1037/stl0000084

Breslau, N., & Kessler, R. C. (2001). The stressor criterion in *DSM–IV* posttraumatic stress disorder: An empirical investigation. *Biological Psychiatry, 50,* 699–704. http://dx.doi.org/10.1016/S0006-3223(01)01167-2

Bridgland, V. M. E., Green, D. M., Oulton, J. M., & Takarangi, M. K. T. (2019). Expecting the worst: Investigating the effects of trigger warnings on reactions to ambiguously themed photos. *Journal of Experimental Psychology: Applied, 25,* 602–617. http://dx.doi.org/10.1037/xap0000215

Bruce, M. J., O'Brien, S. M., Hoffmann, H. L., & Roberts, D. C. (2019, November). The trauma centrality profile and psychophysiological reactivity of students given trigger warnings. In R. J. McNally (Chair), *Trigger warnings! Helpful, harmful, or neither?* Symposium conducted at the meeting of the Association for Behavioral and Cognitive Therapies, Atlanta, GA.

Carter, A. M. (2015). Teaching with trauma: Disability pedagogy, feminism, and the trigger warnings debate. *Disability Studies Quarterly, 35,* 9. http://dx.doi.org/10.18061/dsq.v35i2.4652

Dunmore, E., Clark, D. M., & Ehlers, A. (2001). A prospective investigation of the role of cognitive factors in persistent posttraumatic stress disorder (PTSD) after physical or sexual assault. *Behaviour Research and Therapy, 39,* 1063–1084. http://dx.doi.org/10.1016/S0005-7967(00)00088-7

Ehring, T., Ehlers, A., & Glucksman, E. (2006). Contribution of cognitive factors to the prediction of post-traumatic stress disorder, phobia and depression after motor vehicle accidents. *Behaviour Research and Therapy, 44,* 1699–1716. http://dx.doi.org/10.1016/j.brat.2005.11.013

Gainsburg, I., & Earl, A. (2018). Trigger warnings as an interpersonal emotion-regulation tool: Avoidance, attention, and affect depend on beliefs. *Journal of Experimental Social Psychology, 79,* 252–263. http://dx.doi.org/10.1016/j.jesp.2018.08.006

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology, 74,* 1464–1480. http://dx.doi.org/10.1037/0022-3514.74.6.1464

Grupe, D. W., & Nitschke, J. B. (2013). Uncertainty and anticipation in anxiety: An integrated neurobiological and psychological perspective. *Nature Reviews Neuroscience, 14,* 488–501. http://dx.doi.org/10.1038/nrn3524

Institute of Medicine of the National Academies. (2008). *Treatment of posttraumatic stress disorder: An assessment of the evidence.* Washington, DC: The National Academies Press.

Janoff-Bulman, R. (1989). Assumptive worlds and the stress of traumatic events: Applications of the schema construct. *Social Cognition, 7,* 113–136. http://dx.doi.org/10.1521/soco.1989.7.2.113

Jones, P. J. (2019, October). *Sticks and stones: Investigating the belief that words can harm.* Paper presented at the annual meeting of the Foundation for Individual Rights in Education, Boston, MA.

Krebs, C. P., Lindquist, C. H., Warner, T. D., Fisher, B. S., & Martin, S. L. (2009). College women's experiences with physically forced, alcohol- or other drug-enabled, and drug-facilitated sexual assault before and since entering college. *Journal of American College Health, 57,* 639–649. http://dx.doi.org/10.3200/JACH.57.6.639-649

Lukianoff, G., & Haidt, J. (2015, September). The coddling of the American mind. *The Atlantic.* Retrieved from https://www.theatlantic.com/magazine/archive/2015/09/the-coddling-of-the-american-mind/399356/

McNally, R. J. (2014, May 20). Hazards ahead: The problem with trigger warnings, according to the research. *Pacific Standard.* Retrieved from https://psmag.com/education/hazards-ahead-problem-trigger-warnings-according-research-81946

Mineka, S., & Kihlstrom, J. F. (1978). Unpredictable and uncontrollable events: A new perspective on experimental neurosis. *Journal of Abnormal Psychology, 87,* 256–271. http://dx.doi.org/10.1037/0021-843X.87.2.256

Robinaugh, D. J., & McNally, R. J. (2011). Trauma centrality and PTSD symptom severity in adult survivors of childhood sexual abuse. *Journal of Traumatic Stress, 24,* 483–486. http://dx.doi.org/10.1002/jts.20656

Rosenthal, M. Z., Hall, M. L. R., Palm, K. M., Batten, S. V., & Follette, V. M. (2005). Chronic avoidance helps explain the relationship between severity of childhood sexual abuse and psychological distress in adulthood. *Journal of Child Sexual Abuse, 14,* 25–41. http://dx.doi.org/10.1300/J070v14n04_02

Sanson, M., Strange, D., & Garry, M. (2019). Trigger warnings are trivially helpful at reducing negative affect, intrusive thoughts, and avoidance. *Clinical Psychological Science, 7,* 778–793. http://dx.doi.org/10.1177/2167702619827018

Stokes, M. (2014). In defense of trigger warnings. *Chronicle of Higher Education.* Retrieved from http://chronicle.com/blogs/conversation/2014/05/29/in-defense-of-trigger-warnings/

Telch, M. J., Harrington, P. J., Smits, J. A., & Powers, M. B. (2011). Unexpected arousal, anxiety sensitivity, and their interaction on CO_2-induced panic: Further evidence for the context-sensitivity vulnerability model. *Journal of Anxiety Disorders, 25,* 645–653. http://dx.doi.org/10.1016/j.janxdis.2011.02.005

Verhagen, J., & Wagenmakers, E. J. (2014). Bayesian tests to quantify the result of a replication attempt. *Journal of Experimental Psychology: General, 143,* 1457–1475. http://dx.doi.org/10.1037/a0036731

Walsh, K., Danielson, C. K., McCauley, J. L., Saunders, B. E., Kilpatrick, D. G., & Resnick, H. S. (2012). National prevalence of posttraumatic stress disorder among sexually revictimized adolescent, college, and adult household-residing women. *JAMA Psychiatry, 69,* 935–942. http://dx.doi.org/10.1001/archgenpsychiatry.2012.132

Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E. J. (2011). Statistical evidence in experimental psychology: An empirical comparison using 855 *t* tests. *Perspectives on Psychological Science, 6,* 291–298. http://dx.doi.org/10.1177/1745691611406923

Wilson, R. (2015, September 14). Students' requests for trigger warnings grow more varied. *Chronicle of Higher Education.* Retrieved from http://chronicle.com/article/Students-Requests-for/233043

Wyatt, W. (2016). The ethics of trigger warnings. *Teaching Ethics, 16,* 17–35. http://dx.doi.org/10.5840/tej201632427