Psychological Assessment

Open to Interpretation? Inconsistent Reporting of Lifetime Nonsuicidal Self-Injury Across Two Common Assessments

Kealagh Robinson and Marc S. Wilson

Online First Publication, April 30, 2020. http://dx.doi.org/10.1037/pas0000830

American Psychological Association

Psychological Assessment

ISSN: 1040-3590

http://dx.doi.org/10.1037/pas0000830

Open to Interpretation? Inconsistent Reporting of Lifetime Nonsuicidal Self-Injury Across Two Common Assessments

Kealagh Robinson and Marc S. Wilson
Victoria University of Wellington

Nonsuicidal self-injury (NSSI) is typically assessed using either single-item questionnaires or checklists of common behaviors, but preliminary research suggests that checklists produce higher lifetime prevalence rates. In 2 preregistered studies (combined $n = 1{,}364$), we tested whether memory cueing afforded by behavioral checklists accounts for this discrepancy. Participants reported their lifetime NSSI history using both a single-item and a checklist, with presentation order randomized across participants. Nearly a third of participants reported inconsistent NSSI histories on the 2 assessments, with participants 1.57 times more likely to report an NSSI history on a checklist than on a single-item. Counter to the memory account, this discrepancy was evident even when participants completed the checklist first, suggesting that the increased prevalence estimates captured by checklists are unlikely to simply reflect memory facilitation. Across the 2 samples, 12.5% of participants would have been incorrectly screened out in 2-step assessments; these participants were more likely to have engaged in NSSI historically, less likely to self-injure by cutting, and (in Study 2 only) were more likely to be men. These studies suggest that the inconsistencies across 2 of the most common NSSI assessments arise because people dissimilar to the lay conceptualization of self-injury are less likely to endorse a single-item, even when they have affirmed engaging in self-injury behaviors on a checklist. We argue that single-item and checklist assessments capture different aspects of NSSI, such that future research should distinguish between behaviorally identified NSSI assessed with behavioral checklists and self-identified NSSI assessed with single-item assessments.

Public Significance Statement
Young adults were 1.57 times more likely to report a lifetime history of nonsuicidal self-injury on a behavioral checklist than on a single screening question, with no evidence that this difference is because behavioral checklists better facilitate recall. Instead, people dissimilar to the lay understanding of self-injury were more likely to report inconsistent self-injury histories, suggesting that screening questions rely on subjective interpretation of nonsuicidal self-injury behaviors.

Keywords: nonsuicidal self-injury, self-harm, behavioral assessment, agreement, consistency

Supplemental materials: http://dx.doi.org/10.1037/pas0000830.supp

Nonsuicidal Self-Injury (NSSI), when a person deliberately and directly injures themselves *without* the intention to take their life, represents a significant mental health concern. Approximately 20% of college students report a lifetime NSSI history (Swannell, Martin, Page, Hasking, & St John, 2014), with between 2 and 14% having engaged in the behavior within the past year (Serras, Saules, Cranford, & Eisenberg, 2010; Wilcox et al., 2012). Cross-sectionally, NSSI is associated with poorer psychosocial wellbeing (Giletta, Scholte, Engels, Ciairano, & Prinstein, 2012; Muehlenkamp, Brausch, Quigley, & Whitlock, 2013; Rotolone & Martin, 2012), and greater psychopathology (Nock, Joiner, Gordon, Lloyd-Richardson, & Prinstein, 2006). Longitudinally, NSSI is a risk factor for poorer psychological wellbeing (Andrews, Martin, Hasking, & Page, 2014; Robinson et al., 2019), the onset of new psychiatric disorders (Wilkinson, Qiu, Neufeld, Jones, & Goodyer, 2018), and suicidality (Ribeiro et al., 2016). Given the important associations and corollaries of NSSI, research into the onset, development, recovery, and prevention of the behavior is critical. However, these aims rely on our ability to reliably and robustly assess NSSI.

At present there are no objective measures of NSSI that provide a cost-effective, feasible, and ethical alternative to self-report. NSSI is a stigmatized behavior (Heath, Toste, Sornberger, & Wagner, 2011; Lloyd, Blazely, & Phillips, 2018). The majority of people who self-injure do so privately (Klonsky & Olino, 2008), and are unwilling to disclose the behavior to friends and family

Kealagh Robinson and Marc S. Wilson, School of Psychology, Victoria University of Wellington.

We thank Gina Grimshaw, Jessica Garisch, and Gloria Fraser for their valuable feedback during the preparation of this article.

Correspondence concerning this article should be addressed to Kealagh Robinson, School of Psychology, Victoria University of Wellington, Kelburn Parade, P.O. Box 600, Wellington 6012, New Zealand. E-mail: kealagh.robinson@vuw.ac.nz

(Hasking, Rees, Martin, & Quigley, 2015; Klineberg, Stansfeld, & Bhui, 2013) or seek support from a mental health professional (Hasking et al., 2015; Whitlock, Eckenrode, & Silverman, 2006). As such, we need to not only determine the reliability and validity of self-report NSSI assessments, but also to identify the source(s) of any assessment discrepancies so that we have a clear understanding of what we are assessing and how well.

Self-report NSSI assessments vary widely across studies, with one meta-analysis of 128 unique samples identifying 76 different measurement tools (Swannell et al., 2014). NSSI assessments range from structured interviews (e.g., the Suicide Attempt Self-Injury Interview [SASII]; Linehan, Comtois, Brown, Heard, & Wagner, 2006), to checklists of common NSSI behaviors (e.g., the Deliberate Self-Harm Inventory [DSHI]; Gratz, 2001), to single-item questions (e.g., "Have you ever tried to hurt yourself on purpose, without trying to kill yourself (for example burning, cutting, or scratching yourself)?"; Wilkinson et al., 2018; for an overview of extant NSSI assessment measures, see Klonsky & Lewis, 2014).

This heterogeneity in assessment format is likely due, in part, to the different benefits each format affords. Although interview assessments have received the most psychometric investigation and are typically considered the "gold-standard" for assessing self-injury (Klonsky & Lewis, 2014; Lungu, Wilks, Coyle, & Linehan, 2018), fewer than 25% of studies assess NSSI using interviews (Swannell et al., 2014; You et al., 2018). This low uptake is likely because of the resource-intensive nature of interviews. For example, the SASII takes up to 30 min and is administered one-to-one by trained personnel (Linehan et al., 2006). However, research on other stigmatized topics, such as suicidality, finds that participants are less likely to disclose engaging in these behaviors to an interviewer than on a self-administered questionnaire (Kaplan et al., 1994; Velting, Rathus, & Asnis, 1998; for a review, see Tourangeau & Yan, 2007), raising concerns about interviews underreporting NSSI. In contrast, self-administered questionnaires are more suited to large-scale study designs and to participant comfort disclosing the behavior. Self-administered questionnaires encompass omnibus measures assessing several NSSI domains (e.g., Suicidal Behaviors Questionnaire; Linehan, 1981), behavioral checklists (e.g., DSHI; Gratz, 2001), and single-item measures (e.g., the item assessing NSSI from the Trauma Symptom Inventory; Briere, 1995).

Given their prevalence in the literature (Muehlenkamp, Claes, Havertape, & Plener, 2012; Swannell et al., 2014), here we focus on behavioral checklist and single-item measures of NSSI. Behavioral measures ask participants to indicate whether, and to what extent, they have engaged in common NSSI behaviors (e.g., cutting, scratching, or burning their skin), and typically provide researchers with continuous measures of NSSI methods, severity, and frequency. In particular, meta-analyses demonstrate that the Deliberate Self-Harm Inventory (DSHI; Gratz, 2001) is the most frequently used NSSI measure (Swannell et al., 2014; You et al., 2018) and DSHI scores have shown good internal consistency and adequate test–retest reliability over a 3 week period ($\alpha = .82$, $\varphi = .68$; Gratz, 2001).

In comparison, single-item assessments ask participants to report their NSSI history, usually with a dichotomous response-format reflecting "*NSSI history*" and "*no NSSI history*" (e.g., Alfonso & Dedrick, 2010; Laye-Gindhu & Schonert-Reichl, 2005; Wilkinson et al., 2018). Although some single-item measures are drawn from larger validated scales (e.g., the NSSI item from the Trauma Symptom Inventory; Briere, 1995), others are created in-house by researchers (e.g., Wilkinson et al., 2018).

Single-item assessments are often used as part of a two-step procedure—only those who report an NSSI history on the screening question go on to complete additional items assessing their self-injurious behaviors (e.g., Robinson et al., 2019; Ross & Heath, 2002; Wilkinson et al., 2018). Ethics committees are often concerned that asking about specific behaviors may encourage NSSI thoughts and behaviors (Lloyd-Richardson, Lewis, Whitlock, Rodham, & Schatten, 2015), despite evidence that this is not the case (e.g., Muehlenkamp, Swenson, Batejan, & Jarvi, 2015; Whitlock, Pietrusza, & Purington, 2013). Thus, best-practice guidelines for conducting NSSI research recommend this two-step assessment procedure as one way to minimize potential iatrogenic effects (Hasking, Lewis, Robinson, Heath, & Wilson, 2019). However, the consequences of this two-step procedure for assessment are not yet known.

The choice of assessment is an important one. Meta-analyses of the prevalence (Muehlenkamp et al., 2012; Swannell et al., 2014), associations (Bresin & Schoenleber, 2015; You et al., 2018), and risk factors of NSSI (Fox et al., 2015) all find that assessment method moderates the meta-analytic estimates, suggesting that assessment choice substantively affects the conclusions that may be drawn (and indeed, the scientific consensus of the field). Given their popularity and convenience within the empirical literature, understanding the relative benefits and weaknesses of both single-item assessments and behavioral checklists is important. Two meta-analyses of lifetime NSSI prevalence rates found that point estimates almost doubled when NSSI was assessed using a behavioral checklist compared with a single-item (23.6 vs. 12.5%, and 26.7 vs. 11.8%, respectively; Muehlenkamp et al., 2012; Swannell et al., 2014). These large differences provide preliminary evidence of poor agreement between single-item and behavioral checklist assessments.

Direct evidence that participants report different lifetime NSSI histories across common assessments comes from a small sample of 260 college students who reported their NSSI on a single-item followed by two commonly used behavioral checklists (Lund et al., 2018). Of the 79 participants who reported engaging in NSSI, 43.0% reported NSSI across all three measures, 25.3% reported NSSI on two measures, and 31.6% reported NSSI on only one measure. NSSI was more commonly reported on the two behavioral checklists than the single-item questionnaire (89.9% of the NSSI subsample screened positive on the Inventory of Statement About Self-Injury [ISAS], Klonsky & Glenn, 2009 and 74.7% on the DSHI, compared with 46.8% on the single-item). Overall, 8.5% of participants reported inconsistent NSSI histories between the single-item and the DSHI, and 13.1% reported inconsistent histories between the single-item and the ISAS. This study provides preliminary evidence of considerable variability in lifetime NSSI history across assessments. However, as yet it is unclear *why* participants report different NSSI histories on behavioral checklists compared with single-item assessments.

One intuitive explanation for why participants respond differently across assessments is because they are carelessly selecting responses regardless of the item content. Common methods for identifying careless responders include instructed response items

(e.g., "Please select strongly agree for this item"), and statistical techniques for identifying unusual patterns in raw data (e.g., long-string index, multivariate outliers; Curran, 2016; Meade & Craig, 2012). Careless responding prevalence estimates range from 3.5 to 12% (Johnson, 2005; Meade & Craig, 2012), with instructed response items largely as good at identifying careless responders as statistical techniques (Curran, 2016).

However, instead of careless (likely unsystematic) responding, assessment discrepancies may result from *systematic* differences in responding. One hypothesis argues that behavioral checklists "outperform" single-item assessments because each item of the checklist acts as a recognition memory cue (Lundh, Karim, & Quilisch, 2007; Swannell et al., 2014). In comparison, responding to a single-item is more like a free-recall task that is arguably *more* cognitively demanding, in that it relies on recollection and may not provide a sufficiently strong memory cue for participants to retrieve instances of NSSI from episodic memory (Lundh et al., 2007; Swannell et al., 2014). If memory processes underlie the discrepancy in NSSI prevalence between single-item and behavioral checklists, then the memory benefits afforded by the behavioral checklist should extend to a single-item presented immediately afterward. At present we do not know if this is the case, because two-step procedures and the limited research on NSSI assessment discrepancy (Lund et al., 2018) always presents the single-item question before the behavioral checklist.

In contrast, we could argue that participants interpret a single-item assessment as asking if they are more, or less, similar to people who self-injure (Lundh et al., 2007; Swannell et al., 2014). That is, single-item assessments may produce different NSSI prevalence rates than behavioral checklists because participants evaluate their own behavioral history in light of their personal understanding of what self-injury is, and who self-injures. Critically, this subjective self-identification may be systematic, rather than producing random error. Within lay communities, NSSI is often conceptualized primarily as a behavior that young women engage in by cutting their skin (Lewis, Mahdy, Michal, & Arbuthnott, 2014). Given the stigma associated with NSSI (Heath et al., 2011; Lloyd et al., 2018) and the stereotype of "who self-injures," men, older people, and people who engage in NSSI methods other than cutting may be more hesitant to label their behavior as NSSI on a single-item.

The severity of the behavior may also play a role in how a person interprets their NSSI status. Studies of suicidality find that discrepant responders report less severe suicidality than do consistent responders (Eikelenboom, Smit, Beekman, Kerkhof, & Penninx, 2014; Hart, Musci, Ialongo, Ballard, & Wilcox, 2013; Hom et al., 2019). In a similar manner, adolescents who reported a lifetime history of self-harm (a broader class of behaviors with, and without, suicidal intent; Hawton, Rodham, Evans, & Weatherall, 2002), at 16 but not at 18, were less likely to report recent self-harm at 16 and to have self-injured with suicidal intent compared with consistent reporters (Mars et al., 2016). Turning specifically to NSSI, young adults who reported inconsistent NSSI histories between a behavioral checklist and a single-item also reported fewer NSSI methods, less frequent NSSI engagement, and less psychological distress compared with consistent responders (Lund et al., 2018). Taken together, research across a range of self-injurious behaviors suggests that inconsistent responders may systematically differ from consistent responders.

The Current Study

Although meta-analyses have highlighted a difference in NSSI prevalence depending on assessment type, and a handful of studies have shown that participants report different self-injury histories across assessments, the reasons underlying this assessment discrepancy are not yet clear. Across two studies, we manipulated the order in which participants complete a behavioral checklist and a single-item assessment of NSSI. We establish the extent of agreement across assessments, before testing: (a) if memory facilitation or, (b) careless responding explains discrepant reports of NSSI histories, and (c) if people are interpreting their behavior through a lens of "what counts as self-injury" and "who self-injures" when responding to a single-item NSSI assessment.

We first aim to replicate the finding that fewer people report an NSSI history on the single-item than the behavioral checklist (Lund et al., 2018) in a larger sample. We then test the following preregistered predictions:

1. If people report discrepant NSSI histories because behavioral checklists better facilitate recall memory than do single-item assessments, then assessment discrepancy should be smallest when participants complete the behavioral checklist first, relative to the single-item first.

2. If people report discrepant NSSI histories because they are carelessly responding, then participants who fail an instructed response item should be more likely to report discrepant NSSI histories across assessments.

3. If single-item measures invite participants to reflect on the extent to which they are similar to people who self-injure, then we would expect that:

    a. Participants who deviate from the prototypical definition of "what counts as self-injury" and "who self-injures" (i.e., men, people who do not self-injure by cutting, and older adults) should be more likely to report discrepant NSSI histories across assessments, and

    b. Participants with less recent NSSI and less psychological distress should be more likely to report discrepancies across assessments.

Given the paucity of empirical investigation into the (dis)agreement across NSSI assessments and the potential mechanisms underpinning these assessment discrepancies, we compare assessment responses in two samples to internally replicate effects. Preregistered hypotheses, predictions, design, and analytical plans for both experiments, as well as the deidentified data, analysis syntax, and materials are available at Open Science Framework (Robinson & Wilson, 2020, https://osf.io/8gwju/).

Method

Participants

We invited all students enrolled in introductory psychology courses across two trimesters (Study 1: $n = 835$; Study 2: $n = 1{,}020$) to participate in these studies as part of a broader survey

counting toward a research participation course requirement, with the majority accepting the invitation (Study 1: 77.4%; Study 2: 79.2%). In Study 2 analyses, we excluded 35 participants who took part in both studies (because of the low number, we did not conduct exploratory analyses of how reliably people report discrepant NSSI histories). Across both samples (Study 1: $n = 626$; Study 2: $n = 738$), participants tended to be young adults (Study 1: $M_{age} = 19.31$, $SD = 3.38$; Study 2: $M_{age} = 18.85$, $SD = 2.79$), and identify as female (Study 1: 76.0% identified as female, 23.0% as male, and 1.0% as gender diverse; Study 2: 76.2% identified as female, 23.2% as male, and <1% as gender diverse).[1] The majority of participants identified as Pākehā/New Zealand European (Study 1: 66.9%; Study 2: 69.5%), with the remainder comprised of Māori (indigenous New Zealanders; Study 1: 11.7%; Study 2: 11.3%), Asian (Study 1: 11.0%; Study 2: 9.8%) and Pasifika (Study 1: 5.6%; and Study 2: 3.6%).

Design

In both experiments, we used a between-subjects design in which the independent variable was assessment order; half of participants completed the single-item assessment followed by the behavioral checklist (single-item first condition), while the remainder completed the behavioral checklist followed by the single-item assessment (behavioral checklist first condition). Participants were randomized to each condition. Given that participants could report an NSSI history on both, one, or neither assessment, for clarity we refer to lifetime NSSI reported on the behavioral checklist as "behaviorally identified NSSI," and lifetime NSSI reported on the single-item as "self-identified NSSI." Age, gender, and in Study 2, careless responding, and psychological distress, were the predictor variables, and lifetime NSSI history and the two most common types of assessment discrepancy (participants who reported no self-identified NSSI history with behaviorally identified NSSI, and participants who reported self-identified NSSI ideation with behaviorally identified NSSI) were the dependent variables.

Measures

Nonsuicidal self-injury. At the beginning of both experiments all participants received the following instructions:

> This part of the survey asks questions about some of the things that people sometimes do to hurt themselves. You might not have told anyone about how you have hurt yourself or may not want to tell anyone. By answering these questions honestly, though, you will help us learn how to help others like you. It is very important that you read all the instructions below carefully. Sometimes people can hurt themselves on purpose without intending or expecting to kill themselves. Please only answer these questions if you meant to hurt yourself (not if it was an accident), but without intending to kill yourself. Do not answer yes if you did something accidentally (e.g., you tripped and banged your head accidentally).

Participants were then presented with the two NSSI assessments.

Single-item assessment. All participants saw the item "Please indicate whether you have had thoughts about hurting yourself on purpose, or whether you have hurt yourself on purpose (e.g., punched yourself or objects like walls, prevented wounds from healing, or cut, burnt, scratched or carved your skin, etc.)," and were invited to respond with "*Yes*," "*No*," or "*Thought about it*." The wording of this single-item assessment is similar to others used in the literature (Wilkinson et al., 2018) and includes the response option "Thought about it" to reflect evidence that NSSI ideation and NSSI actions are related, but often distinct (Martin, Bureau, Cloutier, & Lafontaine, 2011).

Behavioral checklist assessment. All participants completed the simplified version of the Deliberate Self-Harm Inventory (DSHI-s; Gratz, 2001; Lundh et al., 2007) that describes common NSSI behaviors (e.g., cutting skin, punching, or banging the body). For each of the 13 items, participants indicated how frequently they have deliberately engaged in the behavior (in the absence of suicidal intent) on a five-point scale ranging from 0 = *never* to 4 = *many times*, with an additional scale point (1 = *I have thought about it*) included to capture NSSI ideation. We combined the "punched oneself" and "banged head" items into one item (combined item "punched yourself, or banged your head against something, to the extent that caused a bruise to appear") and, given the New Zealand context of these studies, modified two items ("carved words . . ." and "stuck sharp objects . . .") to explicitly exclude tā moko, the body and face marking that is part of Māori culture. As the most widely used behavioral checklist (Swannell et al., 2014; You et al., 2018), the DSHI is well suited to investigate the discrepancy in responding across behavioral and single-item assessments. DSHI scores have demonstrated convergent validity with other self-injury measures, as well as internal consistency ($\alpha = .82$), and adequate test–retest reliability over 4 weeks ($\varphi = .68$; Gratz, 2001). Scores of both the simplified DSHI-s ($\alpha = .90$, Lundh, Wångby-Lundh, & Bjärehed, 2011) and of the DSHI-s modified to capture NSSI ideation and exclude tā moko ($\alpha = .79$; Robinson et al., 2017) have previously shown good internal consistency within adolescent samples. Across both current samples, DSHI-s scores showed good internal reliability (Study 1: $\alpha = .86$, 95% confidence interval, CI [.84, .87], Coefficient H = .89; Study 2: $\alpha = .84$, 95% CI [.83, .86], Coefficient H = .87).

Participants who indicated a lifetime NSSI history on either measure were also asked to report how many times they had engaged in NSSI in the past year: "In the last year, how many times have you deliberately hurt yourself (but without wanting to kill yourself)?" In Study 1 these measures used an open text response format, however this response format resulted in 4.3% unusable data (e.g., "too many to count," "in the 40's or 50's"). So, in Study 2 we measured NSSI frequency with the question "In the past year, on how many occasions have you intentionally hurt yourself?", with the response format: "*never*," "*1–3 times*," "*4–5 times*," "*6–10 times*," "*11–20 times*," "*21–50 times*," and "*>50 times*." Participants who reported no NSSI history on either the single-item or behavioral checklist were assigned an NSSI recency score of never.

Psychological distress. In Study 2, participants completed measures of their depression, anxiety, and stress symptoms in the past week using the Depression, Anxiety, and Stress Scale (DASS-21; Henry & Crawford, 2005). Participants responded to 21 items such as "I felt that I had nothing to look forward to" on a 4-point

[1] Open-text responses coded as per best-practice guidelines (Fraser, Bulbulia, Greaves, Wilson, & Sibley, 2019). Given the low number of participants who identified as gender diverse (Study 1: $n = 6$; Study 2: $n = 5$), we excluded these participants from analyses involving gender to guard against spurious findings.

scale ranging from 0 = *did not apply to me at all* to 3 = *applied to me very much, or most of the time*, with seven items comprising each of the depression, anxiety, and stress subscales. Item scores are totaled and then doubled to create subscales, with higher scores indicating greater depression, anxiety, and stress respectively. Previous research has found that DASS-21 scores have adequate internal consistency and construct validity within young adult samples (Depression: $\alpha = .83$, Anxiety: $\alpha = .78$, Stress: $\alpha = .87$; Norton, 2007). In the current sample, the Depression ($\alpha = .89$, 95% CI [.88, .90], $\omega = .89$, 95% CI [.88, .90]), Anxiety ($\alpha = .80$, 95% CI [.78, .83], $\omega = .81$, 95% CI [.79, .83]), and Stress ($\alpha = .83$, 95% CI [.81, .85], $\omega = .84$, 95% CI [.82, .85]) subscales scores showed good internal consistency.

Careless responding. In Study 2, embedded within a scale approximately 5 to 7 min before the experiment was the item: "This is an attention check. Please select 'agree' for this statement," to which participants could respond on a 7-point scale ranging from 1 = *strongly disagree* to 7 = *strongly agree*.[2] This item is similar to other instructed response items commonly used to identify careless responding (Curran, 2016). Participants who failed this instructed response item (i.e., selected any response option other than 6 = *agree*; $n = 30$, 4.1%) were assigned a careless responding score of 1, while those who responded as instructed were assigned a score of 0.

Procedure

Both experiments were embedded approximately 15 min into an hour-long prescreening survey for a department-wide research pool for students enrolled in introductory psychology courses at a large public university. The prescreening survey was presented online via Surveymonkey and participants took part in their own time and on their own devices. In Study 2, participants completed an attention check approximately 5 to 7 min before the experiment. Following both NSSI assessments, participants reported their past-year NSSI frequency and, in Study 2, their psychological distress. Participants gave informed consent before the experiment and received a list of mental health resources available to them after debriefing. Victoria University of Wellington's Human Ethics Committee provided ethical approval for both studies.

Missing Data

Data from participants who completed the survey several times (Study 1: $n = 85$; Study 2: $n = 66$, identified by identical student ID numbers) were inspected. The most complete survey response for each unique participant was retained, except in cases where the NSSI assessments had been answered in which case the earliest of the responses was retained. Participants who did not complete both assessments (Study 1: $n = 35$, Study 2: $n = 16$) were excluded. Across both studies, participants who were excluded did not differ from those who were included on age (Study 1: $t(659) = 0.78$, $p = .434$; Study 2; $t(787) = 1.37$, $p = .171$), gender (Study1: $\chi^2(1, N = 655) = 0.19$, $p = .659$; Study 2: $\chi^2(1, N = 784) = 0.25$, $p = .621$), or ethnicity (Study 1: $\chi^2(4, N = 643) = 1.92$, $p = .751$; Study 2: $\chi^2(4, N = 775) = 3.77$, $p = .438$).

Within the final Study 1 sample ($n = 626$), Little's Missing Completely at Random (MCAR) test suggested that the pattern of missingness on the DSHI-s was not MCAR, $\chi^2(143, N = 621) = 234.42$, $p < .001$. However, as only 0.34% of values were missing, we deemed this inconsequential, following convention (Schafer, 1999). In contrast, for Study 2 ($n = 738$) Little's MCAR test suggested that the pattern of missingness on the DSHI-s was likely to be MCAR, $\chi^2(95, N = 734) = 85.00$, $p = .759$, with 0.26% of data missing. Using logistic regression, neither gender, age, ethnicity, or order of assessment presentation predicted presence (relative to absence) of DSHI-s missingness (Study 1: *p*s range from .872 to .205; Study 2: *p*s range from .153 to .618).

Analytic Plan

NSSI status was coded as 'No history', 'NSSI ideation', and 'NSSI history' for each of the two assessment types. For the single-item assessment, Yes responses were coded as 'NSSI history', No responses as 'No history', and Thought about it responses were coded as 'NSSI ideation'. For the behavioral checklist, participants who reported engaging in at least one of the 13 behaviors on one or more occasions were coded as 'NSSI history', those who reported thinking about engaging in one or more behaviors (but reported no engagement) as 'NSSI ideation', and those who reported never thinking about engaging in *any* of the 13 behaviors were coded as 'No history'. Participants whose NSSI status differed across the two assessment types were assigned a discrepancy score of 1, while those assigned to the same status on both assessments were assigned a score of 0. Inspection of continuous variables indicated that, although scores for Depression ($M = 9.37$, $SD = 9.10$), Anxiety ($M = 10.97$, $SD = 8.72$), and Stress ($M = 13.18$, $SD = 8.89$) were low (as expected for a community sample), there was no evidence that they differed significantly from normality.

Pearson Chi-Square statistical tests assess for relationships between assessment order (single-item first vs. behavioral checklist first) and NSSI prevalence measured using the single-item and the behavioral checklist, as well as a relationship between careless responding and assessment discrepancy. Logistic regression models test whether age, gender, specific NSSI methods, past year NSSI frequency, and psychological distress are associated with change (vs. stability) across assessment types. Exploratory logistic regression models test whether age, gender, and frequency of past-year NSSI predicts incorrectly screened out participants and those who report NSSI thoughts on the single-item but an NSSI history on the checklist (vs. participants who report an NSSI history on both assessments). All analyses reported here were preregistered unless noted otherwise, and all preregistered predictions are confirmed as supported, or not supported. For preregistered analyses, alpha was set at .05. For exploratory analyses, alpha was set at .01 to better guard against spurious findings. Additional exploratory analyses are reported in the online supplemental materials. All analyses were conducted using SPSS Version 25.

[2] The wording of this instructed response item of this attention check differs from our preregistration ("Please select 'Applied to me very much, or most of the time' for this question to show that you are paying attention") and the item was presented before the experimental manipulation rather than afterwards. These changes were made because this version of the instructed response item was already included in the online survey and having two separate instructed response items was considered unnecessary by the prescreening facilitator.

Results

In Study 1, 60.5% of participants ($n = 379$) reported a lifetime NSSI history on one or both assessments, and 24.1% ($n = 151$) reported NSSI ideation. There were 29.8% of participants ($n = 169$) who reported engaging in NSSI within the past year (44.6% of those who reported lifetime self-injury), with a mean of 8.63 ($SD = 14.31$, *Median* = 3.00) NSSI episodes. In Study 2, 69.4% ($n = 512$) reported a lifetime NSSI history on one or both assessments, and 20.9% ($n = 154$) reported NSSI ideation. There were 32.7% of participants ($n = 393$) who reported engaging in NSSI within the past year (46.4% of those who reported lifetime self-injury); of these, the majority (25.3%) reported engaging in NSSI 1 to 3 times in the past year, 8.7% 4 to 5 times, 4.8% 6 to 10 times, 4.6% 11 to 20 times, 2.6% 21 to 50 times, and 0.8% more than 50 times. Across both studies, the most common NSSI methods were scratching (Study 1: 33.4%, Study 2: 35.2%), cutting (Study 1: 32.0%, Study 2: 34.4%), and punching and/or banging the body (Study 1: 31.6%, Study 2: 41.8%).

Agreement Across Assessments

Our first aim was to replicate the finding that fewer people report an NSSI history on a single-item assessment than a behavioral checklist. Consistent with previous research, agreement between the single-item and behavioral checklist assessments was low (Study 1: $\kappa = .50$; Study 2: $\kappa = .48$), with nearly a third (Study 1: 31.0%, Study 2: 31.2%) of participants reporting different NSSI histories. Reflecting meta-analytic research, participants were more likely to report a lifetime NSSI history on the behavioral checklist (Study 1: 59.3%; Study 2: 68.6%) than the single-item (Study 1: 37.2%; Study 2: 44.0%; Study 1: $\chi^2(4, N = 626) = 339.20$, $p < .001$, Cramer's $V = .52$; Study 2: $\chi^2(4, N = 738) = 362.83$, $p < .001$, Cramer's $V = .50$). There was a high degree of asymmetry across all discrepant responses; 10.9% ($n = 68$) of Study 1 participants and 14.0% ($n = 103$) of Study 2 participants reported no NSSI history on the single-item but reported engaging in NSSI behavior on the checklist, whereas zero Study 1 participants and 0.7% ($n = 2$) of Study 2 participants reported an NSSI history on the single-item, but no engagement on the checklist. Table 1 presents the breakdown of NSSI status by assessment type and presentation order for both studies.

Memory Facilitation

Next, we considered potential explanations for this high rate of assessment discrepancy. If higher rates of lifetime NSSI on checklists arise because each checklist item cues memory for engaging in NSSI, then we would expect that respondents would be more likely to report an NSSI history on the single-item if they completed the single-item *after* the checklist compared with *before*. Counter to this hypothesis, presentation order did not affect NSSI prevalence as assessed with either the single-item (Study 1: $\chi^2(2, N = 626) = 0.66$, $p = .718$, Cramer's $V = .03$; Study 2: $\chi^2(2, N = 738) = 4.99$, $p = .082$, Cramer's $V = .08$) or the behavioral checklist (Study 1: $\chi^2(2, N = 626) = 3.15$, $p = .207$, Cramer's $V = .07$; Study 2: $\chi^2(2, N = 738) = 0.03$, $p = .985$, Cramer's $V \leq .01$).[3]

Careless Responding

In Study 2 we wanted to test the hypothesis that careless responding can explain the high rates of assessment discrepancy. First, at a descriptive level the proportion of participants who failed the attention-check (4.1%) was substantively lower than the percentage of participants who reported inconsistent NSSI histories between the single-item assessment and the behavioral checklist (31.2%), suggesting that carelessness could not fully account for discrepancies in responding. Moreover, participants who passed the instructed response item were just as likely to show discrepant responding as those who failed, $\chi^2(1, N = 723) = 1.23$, $p = .267$, Cramer's $V = .04$. Thus, no evidence was found for the hypothesis that the high degree of discrepancy found across assessments was because of careless responding.

Individual Differences in Assessment Discrepancy

Our analyses so far show that approximately a third of participants reported inconsistent NSSI histories between a single-item and a behavioral checklist. Two specific forms of discrepant responses made up the majority (Study 1: 74.8%, Study 2: 81.6%) of all inconsistent responses; participants who reported engaging in NSSI behaviors on the checklist, but reported either no history or NSSI ideation on the single-item.[4] We next consider the hypothesis that these types of discrepant responses arise because participants see themselves as dissimilar to the lay understanding of who self-injures and what counts as self-injury.

Typically "Screened Out" Participants

First, we focused on the participants who reported no NSSI history on the single-item and reported engaging in one or more NSSI behaviors on the checklist. These participants would have been screened out in any study using the standard two-step procedure common in the literature. Notably, typically screened out participants reported engaging in an average of two NSSI methods (Study 1: $M = 2.15$, $SD = 1.93$; Study 2: $M = 2.43$, $SD = 2.11$), although this was fewer NSSI methods than the group who consistently reported a NSSI history across both assessments (Study 1: $M = 4.12$, $SD = 2.35$; Study 2: $M = 3.90$, $SD = 2.24$; Study 1: $U = 3413.00$, $p < .001$, $\eta^2 = .16$; Study 2: $U = 6818.50$, $p < .001$, $\eta^2 = .12$).

If participants who engaged in NSSI behaviors but reported no NSSI history on the single-item did so because they saw themselves as dissimilar to the lay understanding of who self-injures, then certain people should be more (or less) likely to be screened out in two step NSSI assessment procedures. Across both studies, the logistic regression model with age, gender, and past-year NSSI frequency as predictor variables was statistically significant, Study 1: $\chi^2(3, N = 620) = 42.93$, $p < .001$, Nagelkerke $R^2 = .28$; Study

[3] Given that presentation order was unrelated to assessment discrepancy, we deviated from our preregistration and collapsed across presentation order (instead of analyzing only the single-item first condition) to maximize sample size.

[4] Given that these specific forms of assessment discrepancy comprise the majority of discrepant responses, we deviated from our preregistration and chose to focus on these specific forms rather than any form of discrepant response.

Table 1
Lifetime NSSI Prevalence Rates as Measured by the Single-Item and Behavioral Checklist Assessments Across Both Studies

	Behavioral checklist					
Single-item first	Study 1 (n = 310)			Study 2 (n = 334)		
Single-item	No history	NSSI ideation	NSSI history	No history	NSSI ideation	NSSI history
No history	81 (26.1%)	22 (7.1%)	37 (11.9%)	72 (21.6%)	12 (3.6%)	36 (10.8%)
NSSI ideation	6 (1.9%)	11 (3.5%)	34 (11.0%)	4 (1.2%)	15 (4.5%)	45 (13.5%)
NSSI history	—	5 (1.6%)	114 (36.8%)	—	1 (0.3%)	149 (44.6%)
Checklist first	Study 1 (n = 316)			Study 2 (n = 404)		
Single-item	No history	NSSI ideation	NSSI history	No history	NSSI ideation	NSSI history
No history	101 (32.0%)	11 (3.5%)	31 (9.8%)	88 (21.8%)	17 (4.2%)	67 (16.6%)
NSSI ideation	2 (0.6%)	13 (4.1%)	44 (13.9%)	3 (0.7%)	15 (3.7%)	39 (9.7%)
NSSI history	—	3 (0.9%)	111 (35.1%)	2 (0.5%)	3 (0.7%)	170 (42.1%)

Note. NSSI = nonsuicidal self-injury. Study 1: n = 195, and in Study 2: n = 229 reported different NSSI histories across the two assessment measures.

2: $\chi^2(3, N = 733) = 113.23$, $p < .001$, Nagelkerke $R^2 = .36$. Consistent with our hypothesis, compared with participants who consistently reported an NSSI history, typically screened out participants had lower past year NSSI frequencies (Study 1: $b = -0.97$, $p = .001$, $SE = 0.30$, odds ratio, $OR = 0.38$, 95% CI [0.21, 0.69]; Study 2: $b = -2.14$, $p < .001$, $SE = 0.34$, $OR = 0.12$, 95% CI [0.06, 0.23]), and, in Study 2 only, were more likely to be men (Study 1: $b = -0.70$, $p = .095$, $SE = 0.42$, $OR = 0.50$, 95% CI [0.22, 1.13]; Study 2: $b = -0.90$, $p = .005$, $SE = 0.32$, $OR = 0.40$, 95% CI [0.22, 0.77]). Across both studies, typically screened out participants did not differ in age (Study 1: $b = -0.03$, $p = .722$, $SE = 0.08$, $OR = 0.97$, 95% CI [0.83, 1.14]; Study 2: $b = -0.09$, $p = .115$, $SE = 0.06$, $OR = 0.91$, 95% CI [0.82, 1.02]).

To test the hypothesis that people who engage in self-injurious behaviors that deviate from the prototypical definition of what counts as self-injury are more likely to report discrepant NSSI histories, we conducted an exploratory binomial logistic regression in which each of the 13 behaviors of the checklist were entered as predictors for whether or not a participant would typically be screened out (behaviorally identified NSSI without self-identified NSSI, compared with both behaviorally *and* self-identified NSSI). NSSI past-year frequency, and (in Study 2 only) gender were added to the first step of the model (Model 1), followed by each of the 13 behaviors (Model 2). Across both studies the logistic regression model was statistically significant, Study 1: $\chi^2(13, N = 626) = 94.91$, $p < .001$, Nagelkerke $R^2 = .73$; Study 2: $\chi^2(13, N = 733) = 168.92$, $p < .001$, Nagelkerke $R^2 = .74$. Participants who reported engaging in cutting behavior were more likely to report consistent NSSI histories (Study 1: $b = -2.16$, $p < .001$, $SE = 0.41$, $OR = 0.12$, 95% CI [0.05, 0.26]; Study 2: $b = -2.02$, $p < .001$, $SE = 0.27$, $OR = 0.13$, 95% CI [0.08, 0.23]). In Study 1 only, participants who reported engaging in burning behavior were more likely to report consistent NSSI histories (Study 1: $b = -1.03$, $p = .007$, $SE = 0.38$, $OR = 0.36$, 95% CI [0.17, 0.75]), while in Study 2 those who used bleach or oven cleaner on skin were more likely to demonstrate inconsistent NSSI histories ($b = 1.71$, $p = .002$, $SE = 0.55$, $OR = 5.51$, 95% CI [1.87, 16.19]). No other NSSI method distinguished between participants who typically are screened out and those who report an NSSI history on both the single-item and the checklist (Study 1: ps range from .998 for using bleach or oven cleaner on skin, to .048 for punching or banging the body; Study 2: ps range from .969 for using acid on skin, to .104 for scratching skin).

Given that individual differences in psychological distress have previously been associated with discrepant self-injury reports, within the Study 2 sample we conducted a binomial logistic regression to examine whether depression, anxiety, and stress symptoms negatively predicted behaviorally identified NSSI without self-identified NSSI (compared with both behaviorally identified and self-identified NSSI), controlling for gender and NSSI past-year frequency. The logistic regression model was not statistically significant, $\chi^2(3, n = 733) = 6.60$, $p = .086$, Nagelkerke $R^2 = .38$, providing no evidence that depression ($b = -0.02$, $p = .399$, $SE = 0.02$, $OR = 0.98$, 95% CI [0.94, 1.02]), anxiety ($b = -0.04$, $p = .083$, $SE = 0.02$, $OR = 0.96$, 95% CI [0.92, 1.01]) or stress symptoms ($b = 0.01$, $p = .768$, $SE = 0.03$, $OR = 1.01$, 95% CI [0.96, 1.06]) predicted whether people would have been screened out in a two-step procedure.

Self-Identified NSSI Ideation With Behaviorally Identified NSSI

Finally, we consider the group of discrepant responders who reported engaging in one or more NSSI behaviors on the checklist, and report NSSI ideation (but no action) on the single-item. Although most single-item measures do not provide an ideation only response option, in studies that use a two-step procedure and which do include this response option (e.g., Robinson et al., 2017) these participants typically would go on to answer additional NSSI items. Participants who reported self-identified NSSI ideation with behaviorally identified NSSI reported engaging in an average of two NSSI methods (Study 1: $M = 2.19$, $SD = 1.22$; Study 2: $M = 2.13$, $SD = 1.37$), significantly fewer methods than participants who consistently reported a NSSI history across both assessments (Study 1: $M = 4.12$, $SD = 2.35$; Study 2: $M = 3.90$, $SD = 2.24$; Study 1: $U = 4371.50$, $p < .001$, $\eta^2 = .14$; Study 2: $U = 9976.00$, $p < .001$, $\eta^2 = .11$).

As with the typically screened out participants, if this group showed discrepant NSSI histories because they see themselves as dissimilar to lay understandings of who self-injures, then demographic and NSSI characteristics should predict group membership (relative to people who report an NSSI history on both assessments). We entered age, gender, and past year NSSI frequency as predictor variables within a binomial logistic regression. In Study 1, the logistic regression model was not statistically significant, $\chi^2(3, N = 620) = 2.68$, $p = .443$, Nagelkerke $R^2 = .02$. Participants with behaviorally identified NSSI with self-identified NSSI ideation did not differ by past year NSSI frequency ($b = -0.03$, $p = .177$, $SE = 0.02$, $OR = 0.94$, 95% CI [0.94, 1.01]), gender ($b = -0.07$, $p = .866$, $SE = 0.39$, $OR = 0.94$, 95% CI [0.44, 2.00]), or age ($b = 0.01$, $p = .839$, $SE = 0.05$, $OR = 1.01$, 95% CI [0.92, 1.11]). In comparison, in Study 2, the logistic regression model *was* statistically significant, $\chi^2(3, N = 733) = 50.65$, $p < .001$, Nagelkerke $R^2 = .19$. People with behaviorally identified NSSI and self-identified NSSI ideation had engaged in NSSI less frequently in the past year ($b = -0.90$, $p < .001$, $SE = 0.18$, $OR = 0.41$, 95% CI [0.29, 0.58]), and were more likely to be men ($b = -1.05$, $p = .001$, $SE = 0.32$, $OR = 0.35$, 95% CI [0.19, 0.65]) and younger ($b = -0.018$, $p = .030$, $SE = 0.09$, $OR = 0.83$, 95% CI [0.71, 0.98]).

Next, we considered whether the lay understanding of which behaviors count as self-injury may explain why people report behaviorally identified NSSI with self-identified NSSI ideation. We conducted an exploratory binomial logistic regression to test whether specific NSSI methods reported on the behavioral checklist predicted behaviorally identified NSSI with self-identified ideation (vs. behaviorally identified and self-identified NSSI), while (in Study 2 only) controlling for past-year NSSI frequency, gender, and age. Across both studies the logistic regression model was statistically significant, Study 1: $\chi^2(13, N = 626) = 96.52$, $p < .001$, Nagelkerke $R^2 = .41$; Study 2: $\chi^2(13, N = 733) = 93.13$, $p < .001$, Nagelkerke $R^2 = .48$. Consistent with the account that some NSSI behaviors count more as self-injury, across both studies participants who reported greater engagement in cutting were less likely to show discrepant NSSI histories across assessments (Study 1: $b = -1.02$, $p < .001$, $SE = 0.18$, $OR = 0.36$, 95% CI [0.25, 0.51]; Study 2: $b = -1.30$, $p < .001$, $SE = 0.19$, $OR = 0.27$, 95% CI [0.19, 0.39]). In Study 2 only, participants who reported greater engagement in punching and banging were also less likely to report discrepant NSSI histories across assessments ($b = -0.36$, $p = .006$, $SE = 0.13$, $OR = 0.70$, 95% CI [0.54, 0.90]). No other NSSI method distinguished participants who report behaviorally identified NSSI with self-identified ideation (Study 1: *p*s range from .686 for sticking sharp objects into skin, to .074 for punching or banging; Study 2: *p*s range from .736 for biting skin, to .027 for burning skin).

Finally, we assessed whether recent depression, anxiety, and stress symptoms could help to explain why some participants reported behaviorally identified NSSI with self-identified NSSI ideation. Within the Study 2 sample, we conducted a hierarchical logistic regression with past-year NSSI frequency, gender, and age entered into the first step of the model (Model 1), followed by the measures of psychological distress (Model 2). Counter to predictions, Model 2 was not statistically significant, $\chi^2(3, N = 733) = 1.82$, $p = .612$, Nagelkerke $R^2 = .12$, such that depression ($b < .01$, $p = .608$, $SE = 0.02$, $OR = 1.01$, 95% CI [0.97, 1.05]), anxiety ($b < .01$, $p = .768$, $SE = 0.02$, $OR = 1.00$, 95% CI [0.97, 1.05]), and stress symptoms ($b < .01$, $p = .703$, $SE = 0.02$, $OR = 1.00$, 95% CI [0.96, 1.06]) were unrelated to assessment discrepancy.

Discussion

Agreement across two of the most commonly used NSSI assessments (Swannell et al., 2014; You et al., 2018) was strikingly low in both studies. Participants were 1.57 times more likely to report a lifetime NSSI history when assessed using a behavioral checklist than with a single-item, with 31% of participants in both studies reporting different lifetime NSSI histories across assessments. These studies add to growing evidence from young adult, adolescent, and military samples that between 9 to 35% of participants report inconsistent self-injury histories across assessments (Fliege et al., 2006; Hom, Joiner, & Bernert, 2016; Hom et al., 2019; Mars et al., 2016), raising substantial psychometric concerns for the field of self-injury research.

One intuitively obvious explanation for this discrepancy is that behavioral checklists provide retrieval cues for incidents of self-injury that may have been forgotten. However, we found that assessment order did not influence the rate of discrepancy; 9.8% of participants in Study 1 and 16.6% of participants in Study 2 endorsed items on the checklist, and then went on to report no NSSI history on the single-item. Careless responding also failed to account for this discrepancy.

Two specific types of inconsistent responding made up the majority of discrepant responses; participants who reported engaging in NSSI behaviors on the checklist, but on the single-item reported no NSSI history, or only NSSI ideation. In both of these instances, participants who reported having engaged in cutting behaviors were less likely to report a discrepant response on the single-item, suggesting that this behavior may be seen unequivocally as, and defining of, self-injury, while other behaviors are open to greater interpretation. Within lay communities, cutting is often seen as the prototypical NSSI method (Lewis et al., 2014). Although cutting is one of the most common NSSI methods (Klonsky, 2011; Plener, Libal, Keller, Fegert, & Muehlenkamp, 2009), many people who self-injure do not cut themselves (Garisch & Wilson, 2015), and men are less likely than women to engage in cutting behaviors (Andover, Primack, Gibb, & Pepper, 2010). Compared with participants who both self-identified *and* were behaviorally identified as having a lifetime NSSI history, participants who reported behaviorally identified NSSI and self-identified no NSSI history or NSSI ideation engaged in fewer NSSI methods, and did so less frequently. In Study 2 only, men were also more likely to report discrepant responses. Counter to research on the consistency of self-harm self-reports across a 2-year period (Mars et al., 2016), psychological distress was unrelated to either type of discrepancy.

Given the low agreement across these two self-report assessments, it is reasonable to ask which assessment is more accurate in capturing true NSSI. In general, continuous scores of psychopathologies (i.e., behavioral checklists scores) are more reliable than discrete scores (i.e., single-item scores; Markon, Chmielewski, & Miller, 2011). However, in the absence of a comprehensive objective measure of NSSI neither assessment type can be compared with true NSSI. Given this difficulty, one path forward is to

identify *who* each assessment type captures. Our results suggest that the common practices of measuring NSSI with single-item assessments or two-step procedures are more consistently capturing people who self-injure by cutting, but is likely missing people who engage in less prototypical behaviors or those who do not identify as someone who self-injures.

We now turn to consider participants who are typically screened out of two-step assessment procedures—those who report NSSI behaviors on a checklist but no NSSI history on a single-item. Compared with participants who report an NSSI history on both the behavioral checklist and the single-item, typically screened out participants reported less frequent NSSI, were less likely to self-injure by cutting, and (in Study 2 only) were more likely to be men. Excluding these participants creates three key problems for a literature base built heavily on two-step assessments. First, capturing only a more severe NSSI population leads to misrepresentation (most likely underestimation; Hunter & Schmidt, 1990) of the true effect size of relationships between NSSI and other variables of interest. Second, systematically missing a proportion of the population who self-injures introduces error (and variance) for longitudinal work testing how NSSI affects subsequent well-being, curtailing our ability to understand the long-term repercussions of self-injury. Third, sampling only those who self-identify their self-injurious behaviors as NSSI greatly reduces our ability to understand the resilience of people who self-injure, as the most resilient people may be self-selecting out of such a sample. An evidence base that underrepresents people who engage in NSSI behaviors, but do not label their behavior as such, limits our ability to advance scientific understanding of NSSI. Clinical translations of this research may then result in intervention and prevention strategies that are not effective for a substantial proportion of people who self-injure.

Taken together, this research necessitates changes to how NSSI is measured and reported. The current research adds to growing evidence of poor agreement between NSSI self-report assessments (Lund et al., 2018; Lungu et al., 2018), and shows that two of the most common NSSI assessment types may be capturing overlapping, but different populations. One option to address this issue of poor agreement is to move away from self-reported questionnaires assessments altogether in favor of semistructured interviews. Among community adolescents (Ross & Heath, 2002), and women engaged in clinical treatment for borderline personality disorder (Lungu et al., 2018), more people were classified as having a lifetime NSSI history when assessed with a self-report questionnaire compared with a semistructured interview. This discrepancy has been interpreted as evidence that self-report measures overestimate NSSI (Lungu et al., 2018). However, in the absence of an objective measure of NSSI it is difficult to determine the accuracy of either self-report measures or semistructured interviews. Indeed, many of the common interview instruments *begin with* a screening question assessing lifetime NSSI (e.g., the Self-Injurious Thoughts And Behaviors Interview, which begins with "Have you ever had thoughts of purposely hurting yourself without wanting to die?"), followed by branching dependent on a no (skip to next section) or yes response (questions about behavior and frequency; Nock, Holmberg, Photos, & Michel, 2007). That is, semistructured interview assessments typically also follow a two-step assessment procedure, and it remains to be seen if similar patterns of discrepancy as with self-report questionnaires might be evident.

A second option for addressing this measurement problem is to distinguish between NSSI operationalized as "engaging in self-injurious behavior(s)" captured by behavioral checklists and "self-identification as a person who self-injures" captured by single-item and two-step assessments. Distinguishing between *behaviorally identified NSSI* and *self-identified NSSI* may provide greater conceptual precision within both the empirical literature and in client assessment notes. The choice of which assessment type to use is, therefore, dependent on the research question under investigation, as the two operationalizations of NSSI provide different information. For example, the Interpersonal–Psychological Theory of Suicidal Behavior, a prominent theory of suicide, argues that NSSI is a risk factor for subsequent suicide attempts because when a person engages in NSSI behaviors over time they become habituated to the psychophysiological aversiveness of self-injury which, in turn, increases capability for subsequent suicidal behavior (Joiner, Brown, & Wingate, 2005; Joiner, Ribeiro, & Silva, 2012). In this instance, engaging in NSSI *behaviors* is the mechanism by which NSSI confers risk of subsequent suicide, rather than the self-identification as a person who engages in NSSI. Therefore, in this instance capturing behaviorally identified NSSI rather than self-identified NSSI is warranted. We encourage researchers to consider this distinction and report their decision-making process.

Limitations and Future Directions

Across two studies, we found evidence that behavioral checklists and single-item assessments capture different aspects of NSSI. However, this conclusion comes with at least two caveats. First, the single-item assessment has a greater reading complexity than the behavioral checklist, requiring participants to hold in working memory four Boolean alternatives (i.e., "or"). Experimental manipulations have found that greater reading complexity results in longer response times and more midpoint responses (interpreted as a "no-opinion" response; Lenzner, Kaczmirek, & Lenzner, 2010; Velez & Ashworth, 2007), suggesting that working memory load can influence how participants respond to items. Although all participants in the current studies have met the literacy requirements for university enrolment, greater working memory load of the single-item relative to the behavioral checklist may be an alternative explanation for the discrepancy across NSSI assessments. This hypothesis could be tested in future systematic reviews by calculating the reading complexity of each assessment using well-established techniques (Peter, Whelan, Pfund, & Meyers, 2018) before conducting a metaregression to assess whether reading complexity is negatively associated with lifetime NSSI prevalence rates across studies.

A second caveat is the use of college samples. Although the majority of NSSI research has been conducted with college samples (Swannell et al., 2014; and, thus, a strength of the current work is that it tests the agreement of common assessments within commonly used samples), college samples are only able to provide a truncated understanding of human psychology, including mental health (Auerbach et al., 2016; Kovess-Masfety et al., 2016). In particular, 90.7% of participants were aged 17–20 and so the current studies do not contain an adequate age range to stringently test the hypothesis that older participants are more likely to report discrepant NSSI histories across assessments, and so these analyses should be considered exploratory. Given that NSSI peaks in

mid-adolescence (Plener, Schumacher, Munz, & Groschwitz, 2015), and that sample characteristics are associated with assessment type (Swannell et al., 2014), replication across different ages and developmental stages is needed to establish the generalizability of the current studies and to more stringently test whether age moderates assessment discrepancy. It is also worth highlighting that lifetime NSSI prevalence in the current studies, although similar to other New Zealand community adolescent and young adult samples (38%, Fitzgerald & Curtis, 2017; 49%, Garisch & Wilson, 2015), was approximately three-times higher than the international pooled estimate of NSSI prevalence among young adults (Swannell et al., 2014). Given that NSSI varies across cultures (Gholamrezaei, De Stefano, & Heath, 2017), replication in other countries and cultures is warranted.

The current studies are unable to shed light on *why* participants show discrepancy across assessments, only who is *more likely to* report discrepancies. When participants who had reported discrepant suicide attempt histories were asked to explain the discrepancy, many reported changing their operational definition of suicidal behavior across assessments, or misunderstanding the instructions (Velting et al., 1998). Future research could establish whether these explanations extend to NSSI assessments by using a mixed design where participants who report discrepant responses across a single-item and a behavioral checklist are later interviewed and asked to explain why they reported different NSSI histories. Understanding the explanations that participants give for their discrepant responses may help identify the mechanism(s) that distinguish behaviorally identified NSSI from self-identified NSSI for further investigation. Given that people who self-injure often feel ashamed about their self-injury (Rosenrot & Lewis, 2018), one possibility is that reporting a NSSI history on a single-item (i.e., self-identification as a person who self-injures) creates more shame and judgment for the respondent than does a behavioral checklist, and so fewer people report their NSSI history on a single-item. Future research could test this potential mechanism by manipulating shame (e.g., by inviting participants to recall a previous shameful experience) before completing both single-item and behavioral checklist assessments, or by assessing whether people who demonstrate greater socially desirable responding are more likely to report discrepant responses.

Distinguishing between behaviorally identified and self-identified operationalizations of NSSI also necessitates future research comparing the two. We found that people who engage in NSSI behaviors, but do not self-identify as a person who self-injures, engage in fewer NSSI methods, and have done so less recently. Given the cross-sectional nature of these studies, the direction of this relationship remains unclear. Perhaps people with less severe NSSI are less likely to subsequently identify with the behavior. Or perhaps people who identify as someone who self-injures are more likely to subsequently self-injure with greater severity. Longitudinal research tracking the relationship between behaviorally identified NSSI and self-identified NSSI over time may provide valuable insight into who self-injures and who identifies as someone who self-injures.

Conclusion

Across two studies we found that approximately a third of participants report different NSSI histories between a behavioral checklist and a single-item measure. Counter to predictions, neither memory nor careless responding explained this assessment discrepancy. Instead, we found that participants who had not engaged in cutting—the "prototypical" NSSI method within lay communities—were more likely to report discrepancies across assessments. Critically, this suggests that the poor agreement across two of the most common NSSI assessment types may be systematic, rather than the result of random error. The subjective interpretation of what counts as NSSI within single-item and two-step screening measures creates the need for greater conceptual clarity in how NSSI is defined and measured. Decisions regarding operationalization are at the heart of many psychological questions and we are obliged to get these decisions as right as we are able. Failure to do so impedes our scientific understanding and, especially when we conduct research, we hope will inform prevention and intervention, risks resulting in recommendations that will, at best, make a less positive impact than we hope and, at worst, harm our communities of interest.

References

Alfonso, M., & Dedrick, R. E. (2010). Self-injury among early adolescents. *American Journal of Health Education, 41,* 74–84. http://dx.doi.org/10.1080/19325037.2010.10599131

Andover, M. S., Primack, J. M., Gibb, B. E., & Pepper, C. M. (2010). An examination of non-suicidal self-injury in men: Do men differ from women in basic NSSI characteristics? *Archives of Suicide Research, 14,* 79–88. http://dx.doi.org/10.1080/13811110903479086

Andrews, T., Martin, G., Hasking, P., & Page, A. (2014). Predictors of onset for non-suicidal self-injury within a school-based sample of adolescents. *Prevention Science, 15,* 850–859. http://dx.doi.org/10.1007/s11121-013-0412-8

Auerbach, R. P., Alonso, J., Axinn, W. G., Cuijpers, P., Ebert, D. D., Green, J. G., . . . Bruffaerts, R. (2016). Mental disorders among college students in the World Health Organization World Mental Health Surveys. *Psychological Medicine, 46,* 2955–2970. http://dx.doi.org/10.1017/S0033291716001665

Bresin, K., & Schoenleber, M. (2015). Gender differences in the prevalence of nonsuicidal self-injury: A meta-analysis. *Clinical Psychology Review, 38,* 55–64. http://dx.doi.org/10.1016/j.cpr.2015.02.009

Briere, J. (1995). *Trauma Symptom Inventory (TSI) professional manual*. Odessa, FL: Psychological Assessment Resources.

Cameron, J. E. (2004). A three-factor model of social identity. *Self and Identity, 3,* 239–262. http://dx.doi.org/10.1080/13576500444000047

Curran, P. G. (2016). Methods for the detection of carelessly invalid responses in survey data. *Journal of Experimental Social Psychology, 66,* 4–19. http://dx.doi.org/10.1016/j.jesp.2015.07.006

Eikelenboom, M., Smit, J. H., Beekman, A. T. F., Kerkhof, A. J. F. M., & Penninx, B. W. J. H. (2014). Reporting suicide attempts: Consistency and its determinants in a large mental health study. *International Journal of Methods in Psychiatric Research, 23,* 257–266. http://dx.doi.org/10.1002/mpr.1423

Fitzgerald, J., & Curtis, C. (2017). Non-suicidal self-injury in a New Zealand student population: Demographic and self-harm characteristics. *New Zealand Journal of Psychology, 46,* 156–163. Retrieved from https://www.psychology.org.nz/

Fliege, H., Kocalevent, R. D., Walter, O. B., Beck, S., Gratz, K. L., Gutierrez, P. M., & Klapp, B. F. (2006). Three assessment tools for deliberate self-harm and suicide behavior: Evaluation and psychopathological correlates. *Journal of Psychosomatic Research, 61,* 113–121. http://dx.doi.org/10.1016/j.jpsychores.2005.10.006

Fox, K. R., Franklin, J. C., Ribeiro, J. D., Kleiman, E. M., Bentley, K. H., & Nock, M. K. (2015). Meta-analysis of risk factors for nonsuicidal

self-injury. *Clinical Psychology Review, 42,* 156–167. http://dx.doi.org/10.1016/j.cpr.2015.09.002

Fraser, G., Bulbulia, J., Greaves, L. M., Wilson, M. S., & Sibley, C. G. (2019). Coding open-ended gender identity data in a New Zealand national sample. *Journal of Sex Research*. Advance online publication. http://dx.doi.org/10.1080/00224499.2019.1687640

Garisch, J. A., & Wilson, M. S. (2015). Prevalence, correlates, and prospective predictors of non-suicidal self-injury among New Zealand adolescents: Cross-sectional and longitudinal survey data. *Child and Adolescent Psychiatry and Mental Health, 9,* 28. http://dx.doi.org/10.1186/s13034-015-0055-6

Gholamrezaei, M., De Stefano, J., & Heath, N. L. (2017). Nonsuicidal self-injury across cultures and ethnic and racial minorities: A review. *International Journal of Psychology, 52,* 316–326. http://dx.doi.org/10.1002/ijop.12230

Giletta, M., Scholte, R. H. J., Engels, R. C. M. E., Ciairano, S., & Prinstein, M. J. (2012). Adolescent non-suicidal self-injury: A cross-national study of community samples from Italy, the Netherlands and the United States. *Psychiatry Research, 197,* 66–72. http://dx.doi.org/10.1016/j.psychres.2012.02.009

Gratz, K. L. (2001). Measurement of deliberate self-harm: Preliminary data on the Deliberate Self-Harm Inventory. *Journal of Psychopathology and Behavioral Assessment, 23,* 253–263. http://dx.doi.org/10.1023/A:1012779403943

Hart, S. R., Musci, R. J., Ialongo, N., Ballard, E. D., & Wilcox, H. C. (2013). Demographic and clinical characteristics of consistent and inconsistent longitudinal reporters of lifetime suicide attempts in adolescence through young adulthood. *Depression and Anxiety, 30,* 997–1004. http://dx.doi.org/10.1002/da.22135

Hasking, P. A., Lewis, S. P., Robinson, K., Heath, N. L., & Wilson, M. S. (2019). Conducting research on nonsuicidal self-injury in schools: Ethical considerations and recommendations. *School Psychology International, 40,* 217–234. http://dx.doi.org/10.1177/0143034319827056

Hasking, P., Rees, C. S., Martin, G., & Quigley, J. (2015). What happens when you tell someone you self-injure? The effects of disclosing NSSI to adults and peers. *BMC Public Health, 15,* 1039. http://dx.doi.org/10.1186/s12889-015-2383-0

Hawton, K., Rodham, K., Evans, E., & Weatherall, R. (2002). Deliberate self harm in adolescents: Self report survey in schools in England. *British Medical Journal, 325,* 1207–1211. http://dx.doi.org/10.1136/bmj.325.7374.1207

Heath, N. L., Toste, J. R., Sornberger, M. J., & Wagner, C. (2011). Teachers' Perceptions of Non-Suicidal Self-Injury in the Schools. *School Mental Health, 3,* 35–43. http://dx.doi.org/10.1007/s12310-010-9043-4

Henry, J. D., & Crawford, J. R. (2005). The short-form version of the Depression Anxiety Stress Scales (DASS-21): Construct validity and normative data in a large non-clinical sample. *British Journal of Clinical Psychology, 44,* 227–239. http://dx.doi.org/10.1348/014466505X29657

Hom, M. A., Joiner, T. E., & Bernert, R. A. (2016). Limitations of a single-item assessment of suicide attempt history: Implications for standardized suicide risk assessment. *Psychological Assessment, 28,* 1026–1030. http://dx.doi.org/10.1037/pas0000241

Hom, M. A., Stanley, I. H., Duffy, M. E., Rogers, M. L., Hanson, J. E., Gutierrez, P. M., & Joiner, T. E. (2019). Investigating the reliability of suicide attempt history reporting across five measures: A study of U.S. military service members at risk of suicide. *Journal of Clinical Psychology, 75,* 1332–1349. http://dx.doi.org/10.1002/jclp.22776

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Thousand Oaks, CA: Sage

Johnson, J. A. (2005). Ascertaining the validity of individual protocols from web-based personality inventories. *Journal of Research in Personality, 39,* 103–129. http://dx.doi.org/10.1016/j.jrp.2004.09.009

Joiner, T. E., Jr., Brown, J. S., & Wingate, L. R. (2005). The psychology and neurobiology of suicidal behavior. *Annual Review of Psychology, 56,* 287–314. http://dx.doi.org/10.1146/annurev.psych.56.091103.070320

Joiner, T. E., Ribeiro, J. D., & Silva, C. (2012). Nonsuicidal self-injury, suicidal behavior, and their co-occurrence as viewed through the lens of the interpersonal theory of suicide. *Current Directions in Psychological Science, 21,* 342–347. http://dx.doi.org/10.1177/0963721412454873

Kaplan, M. L., Asnis, G. M., Sanderson, W. C., Keswani, L., De Lecuona, J. M., & Joseph, S. (1994). Suicide assessment: Clinical interview vs. self-report. *Journal of Clinical Psychology, 50,* 294–298. Retrieved from https://onlinelibrary.wiley.com/; http://dx.doi.org/10.1002/1097-4679(199403)50:2<294::AID-JCLP2270500224>3.0.CO;2-R

Klineberg, E., Stansfeld, S. A., & Bhui, K. S. (2013). How do adolescents talk about self-harm? Findings from a qualitative school-based study in England. *The Journal of Adolescent Health, 52,* S78–S79. http://dx.doi.org/10.1016/j.jadohealth.2012.10.183

Klonsky, E. D. (2011). Non-suicidal self-injury in United States adults: Prevalence, sociodemographics, topography and functions. *Psychological Medicine, 41,* 1981–1986. http://dx.doi.org/10.1017/S0033291710002497

Klonsky, E. D., & Glenn, C. R. (2009). Assessing the functions of non-suicidal self-injury: Psychometric properties of the Inventory of Statements About Self-injury (ISAS). *Journal of Psychopathology and Behavioral Assessment, 31,* 215–219. http://dx.doi.org/10.1007/s10862-008-9107-z

Klonsky, E. D., & Lewis, S. P. (2014). Assessment of nonsuicidal self-injury. In M. K. Nock (Ed.), *The Oxford handbook of suicide and self-injury* (pp. 337–351). New York, NY: Oxford University Press.

Klonsky, E. D., & Olino, T. M. (2008). Identifying clinically distinct subgroups of self-injurers among young adults: A latent class analysis. *Journal of Consulting and Clinical Psychology, 76,* 22–27. http://dx.doi.org/10.1037/0022-006X.76.1.22

Kovess-Masfety, V., Leray, E., Denis, L., Husky, M., Pitrou, I., & Bodeau-Livinec, F. (2016). Mental health of college students and their non-college-attending peers: Results from a large French cross-sectional survey. *BMC Psychology, 4,* 20. http://dx.doi.org/10.1186/s40359-016-0124-5

Laye-Gindhu, A., & Schonert-Reichl, K. (2005). Nonsuicidal self-harm among community adolescents: Understanding the "whats" and "whys" of self-harm. *Journal of Youth and Adolescence, 34,* 447–457. http://dx.doi.org/10.1007/s10964-005-7262-z

Lenzner, T., Kaczmirek, L. K., & Lenzner, A. (2010). Cognitive burden of survey questions and response times: A psycholinguistic experiment. *Applied Cognitive Psychology, 24,* 1003–1020. http://dx.doi.org/10.1002/acp.1602

Lewis, S. P., Mahdy, J. C., Michal, N. J., & Arbuthnott, A. E. (2014). Googling Self-injury: The state of health information obtained through online searches for self-injury. *Journal of the American Medical Association Pediatrics, 168,* 443–449. http://dx.doi.org/10.1001/jamapediatrics.2014.187

Linehan, M. M. (1981). *Suicide behaviors questionnaire*. Seattle, WA: University of Washington.

Linehan, M. M., Comtois, K. A., Brown, M. Z., Heard, H. L., & Wagner, A. (2006). Suicide Attempt Self-Injury Interview (SASII): Development, reliability, and validity of a scale to assess suicide attempts and intentional self-injury. *Psychological Assessment, 18,* 303–312. http://dx.doi.org/10.1037/1040-3590.18.3.303

Lloyd, B., Blazely, A., & Phillips, L. (2018). Stigma towards individuals who self harm: Impact of gender and disclosure. *Journal of Public Mental Health, 17,* 184–194. http://dx.doi.org/10.1108/JPMH-02-2018-0016

Lloyd-Richardson, E. E., Lewis, S. P., Whitlock, J. L., Rodham, K., & Schatten, H. T. (2015). Research with adolescents who engage in non-suicidal self-injury: Ethical considerations and challenges. *Child and*

Adolescent Psychiatry and Mental Health, 9, 37. http://dx.doi.org/10.1186/s13034-015-0071-6

Lund, E. M., Thomas, K. B., Nadorff, D. R., Bouchard, L. M., Galbraith, K., & Bradley, A. R. (2018). A comparative analysis of three screening methods for nonsuicidal self-injury in college students. *Journal of College Counseling, 21,* 224–238. http://dx.doi.org/10.1002/jocc.12105

Lundh, L. G., Karim, J., & Quilisch, E. (2007). Deliberate self-harm in 15-year-old adolescents: A pilot study with a modified version of the Deliberate Self-Harm Inventory. *Scandinavian Journal of Psychology, 48,* 33–41. http://dx.doi.org/10.1111/j.1467-9450.2007.00567.x

Lundh, L. G., Wångby-Lundh, M., & Bjärehed, J. (2011). Deliberate self-harm and psychological problems in young adolescents: Evidence of a bidirectional relationship in girls. *Scandinavian Journal of Psychology, 52,* 476–483. http://dx.doi.org/10.1111/j.1467-9450.2011.00894.x

Lungu, A., Wilks, C. R., Coyle, T. N., & Linehan, M. M. (2018). Assessing suicidal and nonsuicidal self-injury via in-depth interview or self-report: Balancing assessment effort and results. *Suicide & Life-Threatening Behavior, 49,* 1347–1359. http://dx.doi.org/10.1111/sltb.12526

Markon, K. E., Chmielewski, M., & Miller, C. J. (2011). The reliability and validity of discrete and continuous measures of psychopathology: A quantitative review. *Psychological Bulletin, 137,* 856–879. http://dx.doi.org/10.1037/a0023678

Mars, B., Cornish, R., Heron, J., Boyd, A., Crane, C., Hawton, K., . . . Gunnell, D. (2016). Using data linkage to investigate inconsistent reporting of self-harm and questionnaire non-response. *Archives of Suicide Research, 20,* 113–141. http://dx.doi.org/10.1080/13811118.2015.1033121

Martin, J., Bureau, J. F., Cloutier, P., & Lafontaine, M. F. (2011). A comparison of invalidating family environment characteristics between university students engaging in self-injurious thoughts & actions and non-self-injuring university students. *Journal of Youth and Adolescence, 40,* 1477–1488. http://dx.doi.org/10.1007/s10964-011-9643-9

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods, 17,* 437–455. http://dx.doi.org/10.1037/a0028085

Muehlenkamp, J., Brausch, A., Quigley, K., & Whitlock, J. (2013). Interpersonal features and functions of nonsuicidal self-injury. *Suicide and Life-Threatening Behavior, 43,* 67–80. http://dx.doi.org/10.1111/j.1943-278X.2012.00128.x

Muehlenkamp, J. J., Claes, L., Havertape, L., & Plener, P. L. (2012). International prevalence of adolescent non-suicidal self-injury and deliberate self-harm. *Child and Adolescent Psychiatry and Mental Health, 6,* 10. http://dx.doi.org/10.1186/1753-2000-6-10

Muehlenkamp, J. J., Swenson, L. P., Batejan, K. L., & Jarvi, S. M. (2015). Emotional and behavioral effects of participating in an online study of nonsuicidal self-injury: An experimental analysis. *Clinical Psychological Science, 3,* 26–37. http://dx.doi.org/10.1177/2167702614531579

Nock, M. K., Holmberg, E. B., Photos, V. I., & Michel, B. D. (2007). Self-injurious thoughts and behaviors interview: Development, reliability, and validity in an adolescent sample. *Psychological Assessment, 19,* 309–317. http://dx.doi.org/10.1037/1040-3590.19.3.309

Nock, M. K., Joiner, T. E., Jr., Gordon, K. H., Lloyd-Richardson, E., & Prinstein, M. J. (2006). Non-suicidal self-injury among adolescents: Diagnostic correlates and relation to suicide attempts. *Psychiatry Research, 144,* 65–72. http://dx.doi.org/10.1016/j.psychres.2006.05.010

Norton, P. J. (2007). Depression Anxiety and Stress Scales (DASS-21): Psychometric analysis across four racial groups. *Anxiety, Stress, and Coping, 20,* 253–265. http://dx.doi.org/10.1080/10615800701309279

Peter, S. C., Whelan, J. P., Pfund, R. A., & Meyers, A. W. (2018). A text comprehension approach to questionnaire readability: An example using gambling disorder measures. *Psychological Assessment, 30,* 1567–1580. http://dx.doi.org/10.1037/pas0000610

Plener, P. L., Libal, G., Keller, F., Fegert, J. M., & Muehlenkamp, J. J. (2009). An international comparison of adolescent non-suicidal self-injury (NSSI) and suicide attempts: Germany and the USA. *Psychological Medicine, 39,* 1549–1558. http://dx.doi.org/10.1017/S0033291708005114

Plener, P. L., Schumacher, T. S., Munz, L. M., & Groschwitz, R. C. (2015). The longitudinal course of non-suicidal self-injury and deliberate self-harm: A systematic review of the literature. *Borderline Personality Disorder and Emotion Dysregulation, 2,* 2. http://dx.doi.org/10.1186/s40479-014-0024-3

Ribeiro, J. D., Franklin, J. C., Fox, K. R., Bentley, K. H., Kleiman, E. M., Chang, B. P., & Nock, M. K. (2016). Self-injurious thoughts and behaviors as risk factors for future suicide ideation, attempts, and death: A meta-analysis of longitudinal studies. *Psychological Medicine, 46,* 225–236. http://dx.doi.org/10.1017/S0033291715001804

Robinson, K., Brocklesby, M., Garisch, J. A., Connell, A. O., Langlands, R., Russell, L., . . . Wilson, M. S. (2017). Socioeconomic deprivation and non-suicidal self-injury in New Zealand adolescents: The mediating role of depression and anxiety. *New Zealand Journal of Psychology, 46,* 126–136. Retrieved from https://www.psychology.org.nz/

Robinson, K., Garisch, J. A., Kingi, T., Brocklesby, M., O'Connell, A., Langlands, R. L., . . . Wilson, M. S. (2019). Reciprocal risk: The longitudinal relationship between emotion regulation and non-suicidal self-injury in adolescents. *Journal of Abnormal Child Psychology, 47,* 325–332. http://dx.doi.org/10.1007/s10802-018-0450-6

Robinson, K., & Wilson, M. S. (2020). *Reports of lifetime nonsuicidal self-injury across a behavioural checklist and a single-item assessment*. Retrieved from Open Science Framework https://osf.io/bq9y2/

Rosenrot, S. A., & Lewis, S. P. (2018). Barriers and responses to the disclosure of non-suicidal self-injury: A thematic analysis. *Counselling Psychology Quarterly*. Advance online publication. http://dx.doi.org/10.1080/09515070.2018.1489220

Ross, S., & Heath, N. (2002). A study of the frequency of self-mutilation in a community sample of adolescents. *Journal of Youth and Adolescence, 31,* 67–77. http://dx.doi.org/10.1023/A:1014089117419

Rotolone, C., & Martin, G. (2012). Giving up self-injury: A comparison of everyday social and personal resources in past versus current self-injurers. *Archives of Suicide Research, 16,* 147–158. http://dx.doi.org/10.1080/13811118.2012.667333

Schafer, J. L. (1999). Multiple imputation: A primer. *Statistical Methods in Medical Research, 8,* 3–15. http://dx.doi.org/10.1177/096228029900800102

Serras, A., Saules, K. K., Cranford, J. A., & Eisenberg, D. (2010). Self-injury, substance use, and associated risk factors in a multi-campus probability sample of college students. *Psychology of Addictive Behaviors, 24,* 119–128. http://dx.doi.org/10.1037/a0017210

Swannell, S. V., Martin, G. E., Page, A., Hasking, P., & St John, N. J. (2014). Prevalence of nonsuicidal self-injury in nonclinical samples: Systematic review, meta-analysis and meta-regression. *Suicide and Life-Threatening Behavior, 44,* 273–303. http://dx.doi.org/10.1111/sltb.12070

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin, 133,* 859–883. http://dx.doi.org/10.1037/0033-2909.133.5.859

Velez, P., & Ashworth, S. D. (2007). The impact of item readability on the endorsement of the midpoint response in surveys. *Survey Research Methods, 1,* 69–74.

Velting, D. M., Rathus, J. H., & Asnis, G. M. (1998). Asking adolescents to explain discrepancies in self-reported suicidality. *Suicide and Life-Threatening Behavior, 28,* 187–196.

Whitlock, J., Eckenrode, J., & Silverman, D. (2006). Self-injurious behaviors in a college population. *Pediatrics, 117,* 1939–1948. http://dx.doi.org/10.1542/peds.2005-2543

Whitlock, J., Pietrusza, C., & Purington, A. (2013). Young adult respondent experiences of disclosing self-injury, suicide-related behavior, and

psychological distress in a web-based survey. *Archives of Suicide Research, 17,* 20–32. http://dx.doi.org/10.1080/13811118.2013.748405

Wilcox, H. C., Arria, A. M., Caldeira, K. M., Vincent, K. B., Pinchevsky, G. M., & O'Grady, K. E. (2012). Longitudinal predictors of past-year non-suicidal self-injury and motives among college students. *Psychological Medicine, 42,* 717–726. http://dx.doi.org/10.1017/S0033291711001814

Wilkinson, P. O., Qiu, T., Neufeld, S., Jones, P. B., & Goodyer, I. M. (2018). Sporadic and recurrent non-suicidal self-injury before age 14 and incident onset of psychiatric disorders by 17 years: Prospective cohort study. *The British Journal of Psychiatry, 212,* 222–226. http://dx.doi.org/10.1192/bjp.2017.45

You, J., Ren, Y., Zhang, X., Wu, Z., Xu, S., & Lin, M. P. (2018). Emotional dysregulation and nonsuicidal self-injury: A meta-analytic review. *Neuropsychiatry, 8,* 733–748. http://dx.doi.org/10.4172/Neuropsychiatry.1000399

Received August 19, 2019
Revision received March 30, 2020
Accepted April 2, 2020 ■