

Psychological Assessment

Self-Injurious Thoughts and Behaviors Interview— Revised: Development, Reliability, and Validity

Kathryn R. Fox, Julia A. Harris, Shirley B. Wang, Alexander J. Millner, Charlene A. Deming, and Matthew K. Nock

Online First Publication, April 23, 2020. <http://dx.doi.org/10.1037/pas0000819>

CITATION

Fox, K. R., Harris, J. A., Wang, S. B., Millner, A. J., Deming, C. A., & Nock, M. K. (2020, April 23). Self-Injurious Thoughts and Behaviors Interview—Revised: Development, Reliability, and Validity. *Psychological Assessment*. Advance online publication. <http://dx.doi.org/10.1037/pas0000819>

Self-Injurious Thoughts and Behaviors Interview—Revised: Development, Reliability, and Validity

Kathryn R. Fox
University of Denver

Julia A. Harris
University of Utah

Shirley B. Wang and Alexander J. Millner
Harvard University

Charlene A. Deming
Durham Veterans Affairs Health Care System, Durham,
North Carolina

Matthew K. Nock
Harvard University

The Self-Injurious Thoughts and Behaviors Interview (SITBI) is a widely used measure of the presence, frequency, and characteristics of suicide and self-harming thoughts and behaviors. In response to advances in the conceptualization of these outcomes, and the potential for online data collection, we created a revised version of the SITBI (SITBI-R) and tested its psychometric properties via in-person interview and online self-report formats. Across two studies, the SITBI-R demonstrated strong psychometric properties for both assessment formats. In Study 1, outcomes measured via the SITBI-R showed convergent validity with those assessed with the Columbia Suicide Severity Rating Scale, another interview assessing suicidal thoughts and behaviors. The SITBI-R also showed strong alternate-forms reliability across nearly all outcomes assessed via both assessment formats. In Study 2, the SITBI-R showed strong test–retest reliability via the online assessment format. Across both studies, reliability was strongest for more recent outcomes (e.g., past year vs. lifetime) and for more commonly assessed outcomes of suicidal thoughts, plans, and attempts than for other, less commonly assessed behaviors (e.g., suicide gestures, interrupted suicide attempts, and aborted suicide attempts). The results of these two studies suggest that the SITBI-R provides reliable and valid measurement of key self-injurious outcomes both in person and online.


Public Significance Statement

The present study provides evidence that an updated version of the Self-Injurious Thoughts and Behaviors Interview is a reliable and valid measure of a wide range of self-injurious thoughts and behaviors. Moreover, results indicate concordance between online self-report and in-person interview versions of this measure.

Keywords: assessment, self-injury, suicide, suicide attempts, nonsuicidal self-injury

Self-injurious thoughts and behaviors (SITBs) are a significant global public health challenge. Lifetime prevalence rates for non-lethal SITBs, which cause substantial economic and social-

emotional burden (Centers for Disease Control and Prevention, 2017; Copeland, Goldston, & Costello, 2017), range from 3% to 9% cross-nationally (Nock et al., 2008). Despite decades of research on SITBs, we have a limited understanding of why people engage in and how to best prevent these behaviors. Accurate and thorough measurement of SITBs is critical to increase our understanding of these outcomes. Working toward this goal, Nock, Holmberg, Photos, and Michel (2007) created the Self-Injurious Thoughts and Behaviors Interview (SITBI), a comprehensive and semistructured interview that provides standardized measurement of the presence of several different SITBs and characteristics of these behaviors (e.g., ages of onset, urge/intensity of thoughts, severity of behaviors). The SITBI has strong psychometric properties (Nock et al., 2007) and has been used in hundreds of research studies since its original development (e.g., Barrocas,

 Kathryn R. Fox, Department of Psychology, University of Denver; Julia A. Harris, Department of Psychology, University of Utah; Shirley B. Wang and Alexander J. Millner, Department of Psychology, Harvard University; Charlene A. Deming, Durham Veterans Affairs Health Care System, Durham, North Carolina; Matthew K. Nock, Department of Psychology, Harvard University.

The SITBI-R can be accessed at <https://osf.io/ctqdf/>.

Correspondence concerning this article should be addressed to Kathryn R. Fox, Department of Psychology, University of Denver, 2155 South Race Street, Denver, CO 80210. E-mail: kathryn.fox@du.edu

Hankin, Young, & Abela, 2012; Stanley, Boffa, Hom, Kimbrel, & Joiner, 2017; Stewart et al., 2015; Washburn, Juzwin, Styer, & Aldridge, 2010). Highlighting its popularity, the SITBI has also been translated into multiple languages, including Spanish and German (Fischer et al., 2014; García-Nieto, Blasco-Fontecilla, Paz Yepes, & Baca-García, 2013).

Despite its strengths, the original SITBI and research on this measure have several areas for improvement. First, the original SITBI assesses only a limited range of SITBs. Specifically, it assesses the presence and characteristics of five forms of SITBs: suicidal ideation, suicide plans, suicides gestures, suicide attempts, and nonsuicidal self-injury (NSSI). Other related behaviors not included in the original include aborted suicide attempts (i.e., getting close to killing oneself and at the last minute deciding not to), interrupted suicide attempts (i.e., getting close to killing oneself and at the last minute someone or something else intervening to prevent it from occurring), and dangerous behaviors enacted with ambivalent or no suicidal intent. Second, recent research suggests that some of the original items in the SITBI, in particular single items assessing suicide ideation, plan, and attempt, may lead to unreliable, or inconsistent, responses across people despite similar experiences (e.g., Millner, Lee, & Nock, 2015). Third, data on the reliability and validity of the original SITBI are based on interviewer-assessed SITBs. The reliability and validity of an online format for the SITBI remains unknown. Similarly, data are needed on whether people prefer to answer questions about their SITB histories to someone face to face, or anonymously online. Fourth, as assumptions of suicidal intent as binary may be insufficient and inaccurate in some instances (e.g., Brown, Henriques, Sosdjan, & Beck, 2004; Fox, Millner, & Franklin, 2016), assessing suicidal intentions outside the context of specific behaviors (e.g., suicide attempts) is necessary to provide a more comprehensive understanding of the full spectrum of SITBs.

Accordingly, the current series of studies aimed to develop and evaluate a revised version of the SITBI (SITBI-R) to address each of these limitations. Toward this goal, we first created the SITBI-R, including updated language and additional forms of SITBs. Second, we tested the validity of the SITBI-R compared to another widely used interview. Third, we adapted the SITBI-R to an online self-report format to assess reliability indices of this online format both across time and in comparison to the original interview format. Fourth, we assessed respondents' preferences regarding self-report versus person-to-person interview. We provide further description of each aim below. These tests were conducted across two unique studies.

Aim 1. Update the SITBI to Include Additional Behaviors and to Provide More Precise Assessment Language

We focused on six areas to improve the clarity of the SITBI assessment. These areas involved updating the language used to assess specific SITBs and adding additional behaviors to the assessment; these updates were included in both Study 1 and Study 2.

Dangerous SITB-Related Behaviors

We added an item that assessed engagement in risk-taking behaviors to “tempt fate” (i.e., “Sometimes people do dangerous

things as a way to tempt fate, for example, taking a lot of drugs, driving too fast, volunteering for dangerous missions, walking into traffic, or acting recklessly. Have you ever in your life engaged in a dangerous activity like this as a way to tempt fate?”). This item was added in light of evidence that these behaviors are relatively common, and that engaging in these risky behaviors to “tempt fate” is a significant and unique (statistical) predictor of first suicide attempts (Nock et al., 2018), above and beyond other known risk factors (e.g., NSSI, mental disorders, duration and controllability of suicide ideation). We also added a second question pertaining to whether activities to tempt fate were carried out when the respondent was unsure whether they wanted to live or die. Unlike other SITBs assessed with the SITBI-R, assessing these dangerous behaviors to tempt fate allows for the measurement of behaviors in which the actor is leaving to chance/external factors whether s/he will live or die—a behavior that involves taking action but does not quite meet the threshold of acting to intentionally end one's own life. This represents a category of behavior not well studied in the current literature.

NSSI Specificity

Regarding NSSI, the original SITBI asked, “Have you ever purposely hurt yourself without wanting to die?” This item correctly referred to a lack of suicidal intent during a self-injurious event, however, a range of behaviors (e.g., excessive alcohol/drug use, disordered eating behaviors) could be misclassified as NSSI or misunderstood to be NSSI by participants unaware of researchers' definition of NSSI. As such, drawing from definitions of NSSI proposed by the American Psychiatric Association (2013), the SITBI-R specified certain categories of behaviors that should not be considered NSSI. These excluded behaviors include those that are socially sanctioned (e.g., tattoos, piercings), those that are minor in nature (e.g., picking at wounds), and those that are considered indirect (e.g., binge eating, purging) forms of self-injury (Claes & Muehlenkamp, 2014; Møhl, la Cour, & Skandsen, 2014; St. Germain & Hooley, 2012).

Research has highlighted the importance of gradations of NSSI severity, with moderate-to-severe behaviors showing stronger associations with psychopathology, hospitalization, and suicidal thoughts and behaviors (Lloyd-Richardson, Perrine, Dierker, & Kelley, 2007; Tang et al., 2011). Therefore, in the SITBI-R, we added a new scale to provide a gradient of NSSI severity, as measured by physical injury, for different methods. We designed this scale to provide a more precise assessment of severity than the approach used in prior research, which assumed that method choice can be used as a proxy for severity (e.g., cutting is more severe than hitting).

Range of Suicide Ideation

The original SITBI included a single question to assess suicide ideation (i.e., “Have you ever had thoughts of killing yourself?”). Recent research has shown that the use of single items such as this can result in imprecise measurement (Millner et al., 2015; Nock & Kessler, 2006). For example, among people who reported that their most severe lifetime thought was “I wish I was dead” among several fine-grained suicidal thoughts, roughly 50% endorsed a previously administered single-item suicidal ideation question and

roughly half denied the same item. To address this, in the SITBI-R we added items assessing a range of specific suicidal thoughts (see the full SITBI-R in our repository on the Open Science Framework, <https://osf.io/ctqdf/>) building on evidence that a non-negligible portion (i.e., 10%) of people who deny “seriously” considering suicide continue to endorse having had the thought “Maybe I should kill myself” (Millner et al., 2015). Moreover, assessment of both broader and more specific suicidal thoughts may clarify which suicide ideation items best capture either passive or active suicidal ideation.

Suicide Planning

The original SITBI also contained single-item assessment of suicide plans (i.e., “Have you ever made a plan to kill yourself”). Although useful, there is no consensus for the definition of a “suicide plan” (Chalker, Comtois, & Kerbrat, 2015; Millner et al., 2015), and it is likely that participants possess different ideas regarding what constitutes a suicide plan. Supporting this lack of clarity, a recent study found that among individuals with suicide ideation who denied making a suicide plan, almost 65% endorsed having thought of a suicide method and 38% had thought of a specific place to kill themselves (Millner et al., 2015). In the SITBI-R, we added questions assessing specific aspects of suicidal planning such as thinking of a suicide method, place, and/or time.

Suicide Threats and Gestures Clarifications

In the original SITBI, suicide threats and gestures were assessed with the following question: “Have you ever purposefully led someone to believe that you wanted to kill yourself, when you really had no intention of doing so?” (Nock et al., 2007). In the SITBI-R we sought to build upon this initial item to provide more specific assessment of these types of communications. In particular, we added language to differentiate across use of language and behaviors, asking, “Have you ever said or done something to purposely lead someone to believe that you wanted to kill yourself when you really had no intention of dying?” This question was asked alongside follow-up questions differentiating between using words versus behaviors as forms of communication.

Suicidal Intent of Self-Injurious Behaviors and Intensity of Suicidal Thoughts

Finally, the original SITBI lacked questions regarding the suicidal intent of different self-harm behaviors. Given evidence that suicidal intent may be more continuous than binary (Brown et al., 2004; Fox et al., 2016), this may have obscured important information regarding peoples’ suicidal behavior. Therefore, the SITBI-R included items assessing several components of suicidal intent (i.e., belief in lethality of method, desire to die, confidence in lethality of event) across suicidal preparatory actions and suicidal behaviors. Additionally, the original SITBI included questions about the “intensity” of suicidal thoughts, and participants frequently asked for intensity to be defined. To decrease the likelihood that participants would require additional clarification, we added an additional question assessing the “urge” to act on suicide thoughts and plans. Results are compared across these two new items.

Aim 2. Examine Psychometric Properties of the SITBI-R

In Study 1, we evaluated interrater reliability and convergent validity of the SITBI-R. Interrater reliability was determined by comparing concordance across two interviewers. Convergent validity was determined by comparison with the Columbia Suicide Severity Rating Scale (C-SSRS; Posner et al., 2011). The C-SSRS was selected because it is an empirically validated, widely used measure across a variety settings and situations that has demonstrated convergent and divergent validity in the assessment of SITBs among adolescents and adults (Posner et al., 2011). The C-SSRS partially corresponds to definitions for suicidal ideation and suicidal behavior endorsed by the Centers for Disease Control and Prevention (Crosby, Ortega, & Melanson, 2011; Giddens, Sheehan, & Sheehan, 2014) and the U.S. Food and Drug Administration (Food and Drug Administration, 2012). Of note, although some consider the C-SSRS to be the “gold standard” instrument in the assessment of suicidal thoughts and behaviors, others see shortcomings in the measure (Giddens et al., 2014). Nonetheless, the C-SSRS was used to evaluate the convergent validity of the SITBI.

Aim 3. Examine Psychometric Properties of an Online Version of the SITBI-R

Online methods for research studies in psychology are increasingly common. Recent research demonstrated that online methods produce similar results to in-person studies across a range of tasks and populations (Casler, Bickel, & Hackett, 2013; Crump, McDonnell, & Gureckis, 2013; Hauser & Schwarz, 2016) including clinical populations (Chandler & Shapiro, 2016; Shapiro, Chandler, & Mueller, 2013) and those engaging in NSSI (e.g., Franklin, Puzia, Lee, & Prinstein, 2014). Given that SITBs are highly stigmatized behaviors (Batterham, Calear, & Christensen, 2013; Oliffe et al., 2016), online methods to assess SITBs are particularly useful because stigma and social desirability may increase biased responses during in-person interviews (Fisher, Pearson, Kim, & Reynolds, 2002). Thus, unlike in-person assessments, online methods allow for increased confidentiality and privacy, and in some cases near anonymity, each of which may increase SITB reporting accuracy (Swannell, Martin, Page, Hasking, & St John, 2014). Of note, although some risk issues emerge when assessing psychopathology and SITBs online, research suggests that SITBs can be safely assessed using online data collection (e.g., Michaels, Chu, Silva, Schulman, & Joiner, 2015), and several recent studies suggest that asking people about their SITBs does not increase their risk (e.g., Cha et al., 2016; Gould et al., 2005; Muehlenkamp, Swenson, Batejan, & Jarvi, 2015; Rudd et al., 2006; Whitlock, Pietrusza, & Purington, 2013).

Given the increasing popularity and feasibility of online data collection for SITBs, validation of an online version of the SITB is necessary for online data collection on this topic. Thus, in Study 1, we examined equivalence (i.e., correspondence; Schulenberg & Yutzenka, 1999) between an in-lab interview and an online self-report version of the SITBI-R. Low equivalence between online self-report and in-lab interview versions of the SITBI would suggest that participants do not understand questions or distinctions across SITBs when asked online without an interviewer to clarify

distinctions. High equivalence would indicate that participants respond similarly when asked about these questions online and in person. In Study 2, we examined the test–retest reliability of the SITBI-R when administered entirely online to an anonymous sample. Of note, we did not include a paper-based, self-report form of the SITBI-R.

Aim 4. Explore Participant Preferences and Perceived Ability to Complete SITB Assessments

Finally, we examined participants' preferences and perceived ability to answer questions about their SITB histories online versus in an interview setting. As mentioned earlier, SITBs can be perceived as stigmatized behaviors, and answering questions about these types of behaviors in an interview setting may be both uncomfortable and difficult for people reporting a history of these behaviors. We included several exploratory questions to examine the degree to which participants preferred online versus in-person/in-lab administration and assessed participants' self-reported accuracy and honesty within these different administration contexts. All study procedures were approved by the university's Human Research Protection Office.

Study 1 Method

Procedure

Participants were randomized to complete either the in-lab or online component of the study first, set two weeks apart. Within each setting, participants completed self-report questionnaires and then were randomized to complete either the SITBI-R prior to the C-SSRS or vice versa. When in the lab, the SITBI-R and C-SSRS were administered by an interviewer and the interview was recorded. Participants who indicated past-month SITB histories received risk assessment interviews and safety plans (conducted over the phone during the online component). Participants were compensated \$30 for their participation. Data were collected between January 2016 and June 2017.

Participants

Participants were recruited via announcements and advertisements posted across the community, Internet messaging boards, and several university campuses in the Boston and Cambridge metropolitan area. Most announcements specifically requested participants who had engaged in SITBs. Participants who saw the advertisements and were interested in participating ($n = 298$) completed a brief online screening questionnaire that assessed inclusion criteria (i.e., past-year SITB engagement for 90% of participants, 18+ years of age, English fluency, and Internet access). Of the 203 people (68%) who qualified for the study, 114 people (54%) chose to participate. Six participants were excluded from analyses because they completed only one component of the study; therefore, Study 1 included 108 participants (69.2% female; $M_{\text{age}} = 29.13$ years, $SD = 12.80$). Participants primarily reported White race ($n = 83$), although several participants reported being Black ($n = 9$), Asian ($n = 12$), or Other ($n = 4$); 13 participants reported Hispanic ethnicity. Approximately half of participants reported heterosexual orientation ($n = 60$), with remaining partic-

ipants identifying as homosexual ($n = 9$), bisexual ($n = 32$), other ($n = 6$), or prefer not to say ($n = 3$).

Measures

Suicidal Thoughts and Behaviors Interview—Revised (SITBI-R). We assessed SITBs using a revised version of the SITBI (Nock et al., 2007), the SITBI-R. The SITBI-R included new sections and updates to previously existing sections (in accordance with the aforementioned aims), including risk-taking behaviors, NSSI, suicide ideation, suicide planning, preparatory behaviors, suicide threats and gestures, aborted suicide attempts, interrupted suicide attempts, and suicide attempts. In each section, follow-up questions regarding the age of onset, frequency of thought/behavior, duration of thought/behavior, urge/intensity, future likelihood of thought/behavior engagement, and so forth were assessed. In the interview version, laminated cards containing item descriptions were used for many items requiring participants to rate the severity of their thoughts and/or behaviors. Additionally, for items using rating scales, a 5-point (0 = *low/little*, 4 = *very much/severe*) and an 11-point scale (0 = *low/little*, 10 = *very much/severe*) were administered to examine which Likert scale length to use with questions assessing SITB; results from these analyses will be reported in a separate article. The SITBI-R was administered in both interview format with a trained research assistant and self-report via an online Qualtrics survey.

Columbia-Suicide Severity Rating Scale (C-SSRS). The Columbia-Suicide Severity Rating Scale (C-SSRS) is a widely used suicide assessment questionnaire (Posner et al., 2011) and was used to establish convergent validity in the current design. The scale assesses four main constructs: (1) severity of suicide ideation, (2) intensity of suicide ideation, (3) suicidal behaviors, and (4) lethality of suicidal behaviors. Questions range from asking about passive suicide ideation (e.g., "Have you wished you were dead or wished you could go to sleep and not wake up?") to inquiring about suicide attempts (e.g., "Have you made a suicide attempt?"). The C-SSRS was administered both via interview by a trained research assistant, and online via a self-report Qualtrics survey.

Preference and accuracy survey. We created a brief scale to assess participant preferences and self-efficacy in completing questions regarding their SITB histories. In particular, we asked participants to indicate whether they preferred answering questions about their SITB histories in lab, online, or no difference (i.e., preference), whether they felt they were more honest when answering questions about their SITB histories in lab, online, or no difference (i.e., honesty), and whether they felt they were more accurate when providing responses about their SITB histories in lab, online, or no difference (i.e., accuracy).

Interviewers

Interviews were conducted by a master's level doctoral-level student in clinical psychology (KRF) as well as trained undergraduate and postbaccalaureate research assistants who were thoroughly trained and supervised.

Data Analytic Plan

First, we calculated descriptive statistics (frequencies, means, standard deviations) for all items. Second, we examined interrater reliability by calculating the agreement between codes provided by participant interviewers and codes provided by an independent interviewer who listened to and coded randomly selected audio taped interviews. Third, we examined the alternate forms reliability (i.e., in-lab interview vs. online self-report) of the SITBI-R by examining the agreement between responses across these two formats. Of note, alternate-forms reliability often involves a short delay between test administrations (Bolarinwa, 2015). Given the time and potential emotional intensiveness of completing the SITBI-R, we opted for a two-week delay between assessments, resulting in an alternate-form test–retest reliability analysis. Fourth, we examined the convergent validity of items by assessing agreement in SITB reporting across the SITBI-R and the C-SSRS. We also examined agreement between the self-reported frequency of each behavior across these two assessments. Given the different timeframes used to assess frequency of these behaviors across interviews, only lifetime frequency of suicide attempts was compared across these measures. Finally, we examined proportions of participants who reported preferring to answer questions online versus in person, as well as self-reported accuracy in answering questions online versus in person. We then conducted chi-square tests of independence to examine associations among perceived preferences, honesty, and accuracy.

Following Nock and colleagues (2007), we used Kappa statistics (κ) to assess agreement regarding the presence or absence of SITB engagement. κ values range from 0 to 1, with a score of 0 indicating chance-level agreement between raters, values below .4 representing poor agreement, and values greater than .75 representing excellent agreement (Fleiss, Levin, & Paik, 2003). To assess agreement for continuous items, we used two methods: Intraclass Correlation Coefficients (ICCs) and Spearman Correlations. Following the guidelines of Weir (2005), we used two-way, random effects, single score, absolute ICCs. ICCs provide an index of the proportion of variance across assessments that are due to *true variance*, versus error (systematic and random) variance. Higher ICC scores indicate higher agreement; however, results should be interpreted with caution as scores are highly impacted by the homogeneity of responses across participants. In cases where responses are quite similar (e.g., reporting no history of a certain behavior), ICCs tend to be quite low (see Weir, 2005 for a discussion); therefore, we also examined Spearman correlations for continuous outcomes to provide another index of consistency across continuous items. In cases of missing data (e.g., skipped response), listwise deletion was used.

Study 1 Results

Psychometric Properties of the SITBI-R

Descriptive statistics. All descriptive statistics for the SITBI-R in this sample are presented in Table 1. The majority of participants reported lifetime NSSI, suicide ideation, and suicide methods. Nearly half of participants reported a history of suicide preparations, with just over one third reporting lifetime suicide attempts. Thinking of a specific time for a suicide attempt and interrupted suicide attempts were less commonly reported.

Interrater reliability. Twenty audiotaped interviews (18.52%) were randomly selected to be coded for interrater reliability. We found perfect agreement ($\kappa_s = 1$, $\rho_{hos} = 1.0$) for lifetime, past year, and past month presence and frequency of each SITB category except for ever making a suicide threat or gesture, where we found excellent agreement ($\kappa = .77$).

Alternate forms reliability. Reliability indices between responses provided during the in-person interview and responses provided in the online, self-report SITBI-R (administered after a two-week period) are listed in Table 2. κ indices, ranging from .60 to .99, were strongest for the presence of NSSI, suicide ideation, specific suicidal thoughts, seriously considering suicide, aborted suicide attempts, and suicide attempts. κ indices were weakest for the presence of a suicide gesture, endorsement of which was quite uncommon (although similar to the endorsement observed by Nock et al., 2007) in the present sample. ICC indices were stronger for past-year compared to lifetime frequency of nearly all behaviors. Additionally, ICC indices demonstrated primarily strong reliability for the characteristics of SITBs.

Convergent validity. Kappa scores indicated excellent agreement between the SITBI-R and the C-SSRS in assessing the lifetime presence of suicide attempts ($\kappa = .85$), aborted suicide attempts ($\kappa = .78$), interrupted suicide attempts ($\kappa = .75$), and having engaged in suicide preparations ($k = .75$). Indices were weaker, although still good, for ever having thought of a suicide method (SITBI-R: “Have you ever thought of a specific way or method to kill yourself”; C-SSRS: “Have you been thinking about how you might do this”; $\kappa = .73$), or ever thinking about suicide (SITBI-R: “Have you ever had thoughts of killing yourself”; C-SSRS: “Have you actually had any thoughts of killing yourself”; $\kappa = .74$). Examining the lifetime frequencies of suicide attempts reported across the SITBI-R and C-SSRS, ICC and rho demonstrated strong agreement (ICC = .74; $\rho = .79$).

Despite this generally strong agreement, reliability indices assessing follow-up questions in the C-SSRS demonstrated lower agreement with the SITBI-R. For example, the C-SSRS includes the question “Have you started to work out or worked out the details of how to kill yourself?” In comparison to questions that assessed the presence of a suicide method in the SITBI-R (i.e., “Have you ever thought of a specific way or method to kill yourself?”), kappa was substantially lower ($\kappa = .22$), indicating that these question wordings may actually index unique constructs. Additionally, in the C-SSRS, behaviors are coded as suicide attempts when participants report engaging in dangerous behaviors alongside the belief that they could have died from the behavior. Clinical judgment is recommended for coding of a suicide attempt or dangerous behavior in these instances. Allowing for these different and broader suicide attempt definitions may have resulted in lower agreement between the two measures.

Participant Preferences and Perceived Ability to Completing SITB Assessments

About half of participants (51.52%) reported that they preferred answering questions regarding their SITB histories in the lab, with fewer participants preferring online methods (33.32%), and even fewer having no preference (15.16%). Similarly, nearly half of participants reported that they were more accurate in answering these types of questions in the lab

Table 1
Frequencies, Means, and Standard Deviations of In-Lab and Online SITB Responses

Variable	Study 1 (<i>N</i> = 110)		Study 2 (<i>N</i> = 188)	
	<i>N</i> (%)	<i>M</i> (<i>SD</i>)	<i>N</i> (%)	<i>M</i> (<i>SD</i>)
Lifetime				
NSSI	78 (72.2%)	187.61 (461.46)	121 (64.4%)	337.01 (920.79)
Suicide ideation	95 (88.8%)	804.92 (1881.01)	187 (99.5%)	799.81 (1744.81)
Disappear	84 (78.5%)		119 (61.7%)	
Never born	51 (47.7%)		101 (52.3%)	
Life not worth living	78 (72.9%)		161 (83.4%)	
Sleep forever	76 (71.0%)		119 (61.7%)	
Wish dead	79 (73.8%)		109 (56.5%)	
Maybe kill self	73 (68.2%)		109 (56.5%)	
Should kill self	50 (46.7%)		103 (53.4%)	
Going to kill self	39 (36.4%)		73 (37.8%)	
Suicide method	87 (81.3%)		117 (62.9%)	
Suicide place	56 (52.3%)		79 (42.5%)	
Suicide time	26 (24.3%)		57 (31.0%)	
Seriously considered	71 (66.4%)			
Suicide plan	85 (81.0%)	151.46 (449.40)	113 (58.5%)	650.63 (5834.99)
Threat or gesture	23 (21.5%)		31 (16.7%)	10.79 (110.01)
Suicide preparation	57 (48.7%)		92 (49.5%)	
Aborted attempt	37 (31.6%)	0.38 (2.04)	63 (33.9%)	1.03 (4.80)
Interrupted attempt	14 (13.3%)	0.17 (0.49)	24 (13.0%)	0.51 (1.99)
Suicide attempt	37 (34.9%)	0.69 (1.29)	74 (40.0%)	4.20 (15.27)
Dangerous activity	61 (52.1%)		81 (43.1%)	
Ambivalence to die	53 (45.3%)		70 (37.23%)	
Past Year				
NSSI	54 (50.0%)	12.32 (38.66)	100 (51.8%)	57.58 (260.91)
Suicide ideation	74 (71.8%)	48.05 (83.93)	160 (82.9%)	58.16 (100.19)
Suicide plan	65 (60.7%)	16.79 (52.55)	77 (39.9%)	20.49 (56.88)
Threat OR gesture	10 (0%)	0.31 (1.66)	13 (0.7%)	0.51 (3.79)
Aborted attempt	12 (11.5%)	0.38 (2.04)	30 (15.5%)	1.03 (4.80)
Interrupted attempt	3 (2.6%)	0.03 (0.17)	11 (0.6%)	0.19 (1.00)
Suicide attempt	7 (0.1%)	0.13 (0.60)	23 (11.9%)	0.56 (2.52)

Note. NSSI = nonsuicidal self-injury.

(45.45%), with fewer participants reporting no difference (30.30%), and even fewer reporting increased accuracy online (24.24%). Finally, and in contrast, half of participants reported feeling they could be equally honest answering these questions in the lab and online (49.49%), with fewer reporting that they felt more honest answering these questions online (30.30%) and even fewer in lab (20.20%).

Chi-square tests of independence were conducted to examine relationships among preferences and perceived honesty and accuracy. The relationships between preferences and perceived accuracy ($\chi^2(4, N = 97) = 13.77, p = .008$) and preferences and perceived honesty ($\chi^2(4, N = 97) = 20.33, p < .001$) were each significant. Indeed, participants who preferred completing the SITBI-R in person reported that they were either more accurate in person ($n = 31$) or equally accurate ($n = 11$) compared to online ($n = 7$); similarly, participants who preferred completing the SITBI-R online reported that they were either more accurate online ($n = 12$) or equally accurate ($n = 12$) compared to in person ($n = 9$). Regarding honesty, participants who preferred completing the SITBI-R in person were significantly more likely to report that they were either more honest in person ($n = 20$) or equally honest ($n = 27$) compared to online ($n = 2$). Participants who preferred completing the SITBI-R online were significantly more likely to report that

they were either more honest online ($n = 14$) or equally honest ($n = 12$) compared to in person ($n = 7$).

Study 2 Method

Procedure

To recruit participants for Study 2 (an online examination of the test-retest reliability of the SITBI-R), we posted advertisements targeting people who had engaged in SITBs at some point in their lifetime on online forums related to psychopathology, suicide, and NSSI (e.g., reddit.com/r/depression; reddit.com/r/suicidewatch). Interested participants were given additional information about the study and completed an online screening questionnaire to determine eligibility (i.e., past-year engagement in some form of SITB; English fluency; Internet access). After providing consent, participants were e-mailed a link to the first assessment and two weeks later were e-mailed a link to the second assessment. To help maintain participant anonymity, we asked that participants use e-mail addresses without identifiable information (e.g., last name). For both assessments, participants completed self-report questionnaires before completing the SITBI-R. Participants were compensated with a \$30 Amazon gift card following completion of the

Table 2
Study 1 Equivalence Forms Reliability

Thought/behavior	Lab no; online no	Lab no; online yes	Lab yes; online no	Lab yes; online yes	κ [95% CI]	Year ICC [95% CI]	Year corr.	Life ICC [95% CI]	Life corr.
Dangerous activity tempt fate	35	7	10	48	0.66 [.51, .80]	0.73 [.57, .84]	0.75**	n/a	n/a
Dangerous activity + ambivalent	39	9	11	40	0.60 [.44, .75]	0.75 [.58, .86]	0.75**	n/a	n/a
Nonsuicidal self-injury	23	5	1	70	0.84 [.72, .96]	0.72 [.61, .80]	0.84**	0.52 [.25, .65]	0.88**
Cut severity								0.79 [.67, .87]	0.73**
Scrape severity								0.52 [.07, .80]	0.55*
Hit severity								0.82 [.67, .91]	0.84**
Burn severity								0.65 [.31, .84]	0.81**
Bit severity								0.43 [-.04, .75]	0.46
Ever threat OR gesture	74	4	6	15	0.69 [.51, .87]				
Threats	80	3	6	9	0.61 [.39, .84]	0.37 [.17, .53]	0.61**	0.62 [.47, .73]	0.66**
Gestures	90	6	2	1	0.17 [-.17, .51]	Insuf	Insuf.	0.42 [.24, .57]	0.40**
Suicidal ideation frequency								0.51 [.34, .65]	0.72**
Thoughts of killing yourself?	9	3	2	85	0.75 [.55, .96]	n/a	n/a	n/a	n/a
Seriously considered killing yourself?	29	3	5	61	0.82 [.70, .94]	n/a	n/a	n/a	n/a
I wish I could disappear or not exist	15	8	5	71	0.62 [.43, .81]	n/a	n/a	n/a	n/a
I wish I were never born	49	9	3	42	0.76 [.63, .89]	n/a	n/a	n/a	n/a
My life is not worth living	21	3	7	68	0.74 [.59, .89]	n/a	n/a	n/a	n/a
Sleep and never wake up again	15	12	6	66	0.51 [.31, .71]	n/a	n/a	n/a	n/a
I wish I were dead	22	5	8	64	0.68 [.52, .84]	n/a	n/a	n/a	n/a
Maybe I should kill myself	26	6	9	58	0.66 [.50, .82]	n/a	n/a	n/a	n/a
I should kill myself	41	14	1	43	0.70 [.57, .83]	n/a	n/a	n/a	n/a
I am going to kill myself	53	11	4	31	0.68 [.53, .83]	n/a	n/a	n/a	n/a
Ideation average urge								0.47 [.29, .63]	0.46**
Ideation average intensity								0.40 [.20, .56]	0.41**
Ideation worst point urge								0.71 [.58, .80]	0.69**
Ideation worst point intensity								0.59 [.43, .71]	0.60**
Suicide plans frequency						0.69 [.57, .78]	0.72**	0.64 [.50, .74]	0.73**
Suicide plans: method	13	6	5	75	0.63 [.44, .83]	n/a	n/a	n/a	n/a
Suicide plans: place	38	6	12	42	0.63 [.48, .78]	n/a	n/a	n/a	n/a
Suicide plans: time	64	10	8	17	0.53 [.34, .72]	n/a	n/a	n/a	n/a
Average seriously consider act									
Worst point seriously consider act									
Average confidance lethality								0.56 [.39, .70]	0.56**
Suicide preparatory behaviors	38	7	6	48	0.74 [.60, .87]	n/a	n/a	n/a	n/a
Aborted attempt	60	4	4	30	0.82 [.70, .94]	.52 [.36, .65]	0.60**	0.49 [.32, .62]	0.77**
Interrupted attempt	80	4	6	7	0.53 [.27, .78]	Insuf	Insuf	0.50 [.33, .62]	0.52**
Suicide attempt	60	2	4	32	0.87 [.76, .97]	0.64 [.51, .74]	0.87**	0.72 [.61, .81]	0.82**
Most recent desire to die								0.83 [.58, .93]	0.82**
Most recent confidence in lethality								0.69 [.43, .85]	0.67**
Most harmful desire to die								Insuf.	Insuf.
Most harmful confidence in lethality								Insuf	Insuf

Note. CI = confidence interval; κ = kappa; ICC = intraclass correlation coefficient; corr. = spearman correlation; n/a = not assessed; Insuf. = insufficient cases to assess (i.e., fewer than five total instances). ** $p < .05$. * $p < .01$.

second assessment. Data were collected between March 2016 and June 2016.

Participants

Online forum members who saw the advertisements and were interested in participating ($n = 274$) completed a brief online screening questionnaire that assessed inclusion criteria (i.e., past-year SITB engagement, 18+ years of age, English fluency, and Internet access). Of the 235 people (86%) who qualified for the study, 80% chose to participate. Participants were 188 adults reporting past-year engagement of some form of SITB (58.5% female; $M_{\text{age}} = 26.94$ years, $SD = 6.54$). Of these participants, 46 (24.47%) only completed the first assessment and were therefore excluded from the reliability and validity analyses. Most participants were White ($n = 169$), with a small number of participants reporting other races (Black [$n = 2$], Asian [$n = 6$], and Other [$n = 11$]). Four participants reported Hispanic ethnicity. Most participants identified as heterosexual ($n = 126$), with remaining participants identifying as homosexual ($n = 7$), bisexual ($n = 37$), other ($n = 10$), or prefer not to say ($n = 8$).

Data Analytic Plan

We calculated the test–retest reliability (i.e., dependability; Revelle & Condon, 2018) of the SITBI-R by examining agreement between the reported lifetime presence (κ) and frequency (two-way, random effects, single-score ICCs, Spearman correlations) for each type of SITB reported and the presence and frequency reported two weeks later. As with Study 1, we also examined proportions of participants who reported preferring to answer questions online versus in-person, as well as self-reported accuracy in answering questions about SITB engagement. Chi-square tests of independence were again used to examine associations among perceived preferences, honesty, and accuracy. In cases of missing data (e.g., skipped response), listwise deletion was used.

Study 2 Results

Descriptive Statistics for the SITBI-R

The frequencies, means, and standard deviations of participants' responses to the SITBI-R in this sample are presented in Table 1. As with Study 1, participants reported extensive SITB histories. Almost all participants reported a lifetime history of suicide ideation, with over half reporting a lifetime history of NSSI and suicide planning, and just under half reporting suicide preparations.

Test-Retest Reliability of the SITBI-R

Kappa scores for the lifetime presence of each outcome reported at the two time points ranged from poor to excellent (see Table 3). Kappa scores were weakest for lifetime assessment of suicide ideation; however, it is notable that there were only 4 total disagreements, out of 141 instances. This low kappa value likely relates to the low number of people saying “no” to this item. Kappa values were similar (i.e., overlapping confidence intervals) be-

tween Study 1 and 2 when considering presence/absence of nearly all assessed outcomes.

ICC and Spearman rho indices reflected strong test–retest reliability for past-year and lifetime frequency of SITBs, although these indices tended to be stronger for the assessment of past-year frequency. ICCs were weakest for frequency of suicide threats/gestures and frequency of interrupted and aborted suicide attempts. ICCs were similar (i.e., overlapping confidence intervals) between Study 1 and Study 2 when considering lifetime and past-year frequency of all variables. Finally, ICC indices for characteristics of SITBs were primarily strong. However, ICC indices were weakest in the assessment of NSSI severity characteristics.

Participant Preferences and Perceived Ability to Complete SITB Assessments

Finally, we examined how participants who participated in Study 2 responded to items assessing preference and self-reported accuracy in answering questions about SITB engagement. Most (66.92%) participants in Study 2 said they would prefer to answer questions about their SITB histories online, with only 4.51% reporting they would prefer to answer questions like this in person, and 28.57% indicating no preference. Similarly, most (66.92%) participants in Study 2 reported that they felt they could be more honest online; 28.57% reported no difference; and only 4.51% said in person. Finally, in terms of accuracy, 75.49% of participants reported that they could be more accurate online, 16.54% reported no difference, and 10.53% reported in person.

Chi-square tests of independence were conducted to examine relationships among preferences and hypothesized honesty and accuracy. The relationship between preferences and perceived accuracy ($\chi^2(4, N = 131) = 52.36, p < .001$) and preferences and perceived honesty ($\chi^2(4, N = 97) = 52.07, p < .001$) were each significant. Indeed, participants who said they would prefer completing the SITBI-R online reported that they would be either more accurate online ($n = 88$) or equally accurate ($n = 12$) compared to in person ($n = 8$); participants who said they would prefer completing the SITBI-R in person reported that they would be more accurate in person ($n = 6$) or equally accurate ($n = 2$) compared to online ($n = 1$). Regarding honesty, participants who said they would prefer completing the SITBI-R online were significantly more likely to report that they would be either more honest online ($n = 83$) or equally honest ($n = 23$) compared to in person ($n = 2$). Participants who said they would prefer completing the SITBI-R in person were more likely to report that they would be either more honest in person ($n = 4$) or equally honest ($n = 5$) compared to online ($n = 1$).

Discussion

In this study, we introduce the SITB-R and describe its psychometric properties across lab interview and online formats. ICC, kappa, and Spearman rho indices were primarily strong for outcomes assessed with the SITBI-R. This was true of both the in-person interview and the online, self-report versions of the SITBI-R. Furthermore, validity indices were good to excellent when comparing responses to the SITBI-R and C-SSRS, suggesting strong convergent validity. Results suggest that, across both

Table 3
Study 2 Test-Retest Reliability

Thought/behavior	T1 no; T2 no	T1 no; T2 yes	T1 Yes; T2 no	T1 Yes; T2 yes	κ [95% CI]	Year ICC [95% CI]	Year corr.	Life ICC [95% CI]	Life corr.
Dangerous activity tempt fate	80	11	13	38	0.63 [.50, .76]	0.56 [.29, .75]	0.65**	n/a	n/a
Dangerous activity + ambivalent	85	11	18	27	0.51 [.36, .66]	0.79 [.59, .90]	0.73**	n/a	n/a
Nonsuicidal self-injury	63	1	3	74	0.94 [.89, .99]	0.52 [.39, .63]	0.84**	0.65 [.49, .78]	0.82**
Cut severity								0.17 [-.31, .57]	0.23
Scrape severity								0.46 [.21, .66]	0.63**
Hit severity								0.59 [.31, .78]	0.76**
Burn severity								0.15 [-.30, .56]	0.31
Bit severity								0.65 [.49, .78]	0.82**
Ever threat OR gesture	115	2	7	14	0.72 [.55, .89]				
Threats	115	1	1	5	0.83 [.59, .99]	0.81 [.75, .86]	0.68**	0.33 [.18, .47]	0.74**
Gestures	120	2	6	13	0.73 [.56, .90]	0.80 [.73, .85]	0.75**	0.41 [.26, .54]	0.78**
Suicidal ideation frequency	1	0	3	137	0.39 [-.15, .93]	0.79 [.72, .85]	0.61**	0.51 [.38, .62]	0.72**
Thoughts of killing yourself?	15	3	10	111	0.64 [.46, .82]	n/a	n/a	n/a	n/a
Seriously considered killing yourself?	64	1	6	72	0.90 [.83, .97]	n/a	n/a	n/a	n/a
I wish I could disappear or not exist	67	8	14	54	0.69 [.57, .81]	n/a	n/a	n/a	n/a
I wish I were never born	10	12	6	115	0.46 [.24, .67]	n/a	n/a	n/a	n/a
My life is not worth living	61	2	10	70	0.83 [.74, .92]	n/a	n/a	n/a	n/a
Sleep and never wake up again	68	4	7	64	0.85 [.76, .94]	n/a	n/a	n/a	n/a
I wish I were dead	68	7	11	57	0.75 [.64, .86]	n/a	n/a	n/a	n/a
Maybe I should kill myself	77	0	13	53	0.81 [.72, .91]	n/a	n/a	n/a	n/a
I should kill myself	93	7	6	37	0.79 [.67, .90]	n/a	n/a	n/a	n/a
I am going to kill myself								0.48 [.33, .60]	0.56**
Ideation average urge								0.66 [.55, .75]	0.68**
Ideation average intensity								0.80 [.74, .86]	0.83**
Ideation worst point urge								0.80 [.72, .85]	0.81**
Ideation worst point intensity								0.46 [.32, .58]	0.86**
Suicide plans frequency	64	1	8	66	0.87 [.79, .95]	0.85 [.79, .89]	0.78**		
Suicide plans: Method	90	2	10	37	0.80 [.69, .91]	n/a	n/a	n/a	n/a
Suicide plans: Place	94	4	12	26	0.69 [.55, .83]	n/a	n/a	n/a	n/a
Suicide plans: Time									
Average seriously consider act									
Worst point seriously consider act									
Average confidence lethality	79	5	6	48	0.81 [.74, .93]	n/a	n/a	0.64 [.48, .77]	0.58**
Suicide preparatory behaviors	98	6	4	30	0.81 [.69, .92]	0.69 [.60, .77]	0.63**	0.27 [.11, .42]	0.74**
Aborted attempt	123	2	2	11	0.83 [.67, .99]	0.41 [.27, .54]	0.43**	0.43 [.29, .56]	0.75**
Interrupted attempt	93	2	12	31	0.75 [.62, .87]	0.67 [.57, .76]	0.84**	0.65 [.54, .73]	0.86**
Suicide attempt								0.38 [.02, .65]	0.37
Most recent desire to die								0.76 [.56, .88]	0.78**
Most recent confidence in lethality								0.68 [.30, .88]	0.61**
Most harmful desire to die								0.68 [.29, .88]	0.66**
Most harmful confidence in lethality									

Note. CI = confidence interval; κ = kappa; ICC = intraclass correlation coefficient; corr. = spearman correlation; n/a = not assessed.
* $p < .05$. ** $p < .01$.

interview and online formats, the SITBI-R provides a reliable and valid measurement of key self-injurious outcomes, including NSSI, suicidal thoughts, suicide plans, and suicidal behaviors, as well as behaviors that have recently shown promise predicting suicide attempts, such as “tempting fate” with dangerous behaviors. Several aspects of the results warrant further discussion.

First, interrater reliability of SITB engagement assessed within the SITBI-R was nearly perfect, and test–retest reliability for ever engaging in most SITBs was also quite strong. This is consistent with test–retest reliability reported for the original SITBI (Nock et al., 2007). Results suggest that the SITBI-R maintains strong agreement across raters. Second, convergent validity between outcomes assessed via the SITBI-R and C-SSRS were also strong. This suggests that the SITBI-R provides valid assessment of SITB engagement. However, a few inconsistencies are worth noting. The SITBI-R and C-SSRS assess suicide planning in different ways. Whereas the SITBI-R defines a suicide plan as thinking of a suicide method, place, or time, and contains items to assess each of these components, the C-SSRS assesses suicide planning via items assessing thinking of a “specific” suicide method or via an item assessing whether participants have been thinking about how they might kill themselves. These discrepancies parallel the lack of clinical and research consensus regarding what constitutes a suicide plan (Chalker et al., 2015; Millner et al., 2015). Researchers and clinicians using these measures should consider that certain items assessing suicide plans are more consistent across these two measures than others. Additionally, although assessment of suicide attempts showed strong reliability across the C-SSRS and SITBI-R, this should be interpreted with caution. In particular, across both assessments, the same suicide attempt definition was used: one requiring that the self-harming behavior was enacted with at least some, nonzero self-reported intent to die. However, the definition typically used in the C-SSRS is broader; in particular, self-harming behaviors are considered suicidal when a participant states suicidal intent or when suicidal intent can be inferred (i.e., from engaging in a highly lethal act even when suicidal intent is denied, from engaging in a behavior that the participant reports could have been lethal; Posner, Oquendo, Gould, Stanley, & Davies, 2007). Allowing for these disparate suicide attempt definitions may have resulted in lower agreement across the two assessments.

Third, results provide at least three lines of evidence that online, self-report versions of the SITBI-R can be administered reliably, at least for people willing and able to understand and answer questions about their SITB histories. In particular, Study 1 demonstrated strong equivalence across in-lab and online administrations of the SITBI-R. Of note, when comparing online versus in-person SITB assessment, we were not able to approximate the influence of anonymity, which may have decreased potential differences that would otherwise be observed between online and in-person assessments. However, results of Study 1 indicate that people respond similarly to an interviewer and to an online self-report questionnaire when answering questions about their SITB histories. Further supporting the use of the SITBI-R online, Study 2 demonstrated strong test–retest reliability for entirely online assessments of suicide and self-injurious thoughts and behaviors, with nearly no significant differences (as determined by overlapping confidence intervals across Kappa and ICC effects) compared to those indices observed in Study 1. Finally, reliability coefficients of the

online SITBI-R were not significantly different from those observed in the assessment of the original SITBI, with confidence intervals overlapping with the effects observed by Nock and colleagues, 2007. These results suggest that online and in-person SITBI-R assessments can likely be used interchangeably, and that the SITBI-R can be reliably used in clinical settings and in research studies capitalizing on both in-person and online methods.

Fourth, across Studies 1 and 2, alternate forms and test–retest reliability were lowest, although still fair, for interrupted suicide attempts and suicide threats/gestures. Lower reliability may relate to several factors. Most importantly, suicide threats/gestures and interrupted suicide attempts were the least commonly endorsed SITBs across both studies. As mentioned earlier, low endorsement tends to disproportionately impact reliability indices.

Alternatively, however, lower reliability indices may relate to construct definitions and/or validity. Regarding interrupted suicide attempts, participants may have experienced difficulty disentangling aborted from interrupted suicide attempts. For example, during in-person interviews, participants often tried to endorse a history of interrupted suicide attempt when an outside event (e.g., an unrelated phone call, text) changed their mindset and they decided after that event not to attempt suicide. However, within our coding schemes, such instances are coded as “aborted suicide attempts” unless someone or something specifically intervened on the suicide attempt (e.g., walked into the room as someone was attempting suicide, physically restrained the individual at risk, took away the method the individual was using, etc.). Given potential difficulty disentangling aborted from interrupted suicide attempts in online and in-person assessments, we recommend that researchers consider additional language to probe these distinctions. Additional research may also be useful to determine if such distinctions are meaningfully related to future SITB risk.

Poorer reliability indices for suicide threats/gestures supports the decision by some (e.g., the Center for Disease Control) to exclude these actions from suicide-related nomenclature (Crosby et al., 2011; Heilbron, Compton, Daniel, & Goldston, 2010). For example, Heilbron and colleagues (2010) argue that the term suicide gesture is often used to describe a wide range of behaviors, decreasing the meaning and utility of such terms. Moreover, they argue that lower reliability in the assessment of these behaviors raises concerns about their conceptual clarity. Although we agree that suicide threats/gestures are a heterogeneous category that showed relatively poor reliability in the present studies, future study and refinement may advance understanding in this important area of research. For example, in the SITBI-R, the question assessing suicide threats/gestures asked whether a participant had done something to make someone think they were going to kill themselves, whether they did this on purpose, and whether they had no intention of dying at the time all, within one item. Assessment of suicide threats/gestures that separates out these different aspects of suicide threats/gestures may lead to more reliable responses and better understanding among participants. Alternatively, more research, including the collection of qualitative data, could lead to improved operationalization and delineation of these types of behaviors, which may, in turn lead to improved reliability. Regardless of these two approaches, collecting data on larger samples where there are more instances of suicide threats/gestures reported may provide better information regarding the reliability of such behaviors.

Finally, participant preferences and perceived honesty and ability varied across studies. In Study 1, participants largely preferred and felt they were more accurate when answering questions about their SITB histories in the lab, to an interviewer (with no difference in terms of honesty). In contrast, in Study 2, participants said that they preferred answering these kinds of questions online, and that they felt they could be more honest and accurate when answering questions about their SITB histories online. These preferences are likely related to both self-selection and consistency biases. Future studies randomizing participants to complete the SITBI-R online versus in person in an interview format may be better suited to testing differences between these methods.

The present study, and the SITBI-R, should be viewed in light of several key limitations. First, the SITBI-R was designed to overcome limitations of the earlier interview to increase clarity and to encompass more forms of self-injurious thoughts and behaviors. However, numerous thoughts and behaviors remain unassessed by the current interview (e.g., indirectly self-harming behaviors including risky/dangerous relationships, excessive substance use.). Moreover, SITBs are not ontologically objective behaviors; they are culturally defined and bound behaviors. As a result, definitions of these behaviors vary across people, clinicians, and researchers, inherently limiting assessment. Second, updated language to assess suicide threats and gestures was meant to increase clarity around the types of behavior, yet the additional language may have inadvertently decreased participant endorsement of these items limiting assessment of these behaviors and power to accurately assess psychometric properties. Third, participants in both studies were samples of convenience, primarily including Caucasians from the United States, and a large number of people identifying as homosexual, lesbian, gay, bisexual, and other. Our sample recruitment technique ensured that participants had experienced a wide range of SITBs, to help validate the assessment. However, the degree to which results would generalize to more representative samples needs to be tested in future studies. Fourth, classic studies on interrater reliability involve participants completing interviews with two independent interviewers (e.g., Ruskin et al., 1998). This method decreases the likelihood of interviewers' follow-up comments and questions influencing the independent coders' responses. In the present study, this stringent method for interrater reliability was not possible as participants were already completing two separate measures of SITB engagement (i.e., SITBI-R, C-SSRS), and adding another interview would have added substantial time and emotional burden for participants. Future studies should consider more stringent interrater reliability methods.

Finally, all participants were paid to participate in the present research. Future research should ensure that psychometric properties remain strong in samples of unpaid participants (e.g., clinical setting, school setting) who may be less motivated to read through all items thoroughly or respond in a thoughtful manner.

Conclusion

Findings from the current studies indicate that the SITBI-R demonstrates primarily strong psychometric properties in both interview and online formats. In particular, the SITBI-R demonstrates strong convergent external validity with the C-SSRS, and strong structural reliability in terms of test-retest, alternative forms, and interrater reports. Use of the SITBI-R in both research

and clinical contexts is thus supported, although additional research seeking to improve the assessment of interrupted suicide attempts and suicide threats and gestures is needed. We encourage use of the SITBI-R to more accurately assess suicide in research and clinical contexts, and to further validate this comprehensive suicide assessment across different populations and age cohorts.

References

- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.
- Barrocas, A. L., Hankin, B. L., Young, J. F., & Abela, J. R. (2012). Rates of nonsuicidal self-injury in youth: Age, sex, and behavioral methods in a community sample. *Pediatrics*, *130*, 39–45. <http://dx.doi.org/10.1542/peds.2011-2094>
- Batterham, P. J., Calear, A. L., & Christensen, H. (2013). Correlates of suicide stigma and suicide literacy in the community. *Suicide and Life-Threatening Behavior*, *43*, 406–417. <http://dx.doi.org/10.1111/sltb.12026>
- Bolarinwa, O. A. (2015). Principles and methods of validity and reliability testing of questionnaires used in social and health science researches. *Nigerian Postgraduate Medical Journal*, *22*, 195–201.
- Brown, G. K., Henriques, G. R., Sosdjan, D., & Beck, A. T. (2004). Suicide intent and accurate expectations of lethality: Predictors of medical lethality of suicide attempts. *Journal of Consulting and Clinical Psychology*, *72*, 1170–1174. <http://dx.doi.org/10.1037/0022-006X.72.6.1170>
- Casler, K., Bickel, L., & Hackett, E. (2013). Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. *Computers in Human Behavior*, *29*, 2156–2160. <http://dx.doi.org/10.1016/j.chb.2013.05.009>
- Centers for Disease Control and Prevention. (2017). *Welcome to WISQARS*. Retrieved from <http://www.cdc.gov/injury/wisqars/index.html>
- Cha, C. B., Glenn, J. J., Deming, C. A., D'Angelo, E. J., Hooley, J. M., Teachman, B. A., & Nock, M. K. (2016). Examining potential iatrogenic effects of viewing suicide and self-injury stimuli. *Psychological Assessment*, *28*, 1510–1515.
- Chalker, S. A., Comtois, K. A., & Kerbrat, A. H. (2015). Impulsivity and suicidal behavior: How you define it matters. *International Journal of Cognitive Therapy*, *8*, 172–192. <http://dx.doi.org/10.1521/ijct.2015.8.2.172>
- Chandler, J., & Shapiro, D. (2016). Conducting clinical research using crowdsourced convenience samples. *Annual Review of Clinical Psychology*, *12*, 53–81. <http://dx.doi.org/10.1146/annurev-clinpsy-021815-093623>
- Claes, L., & Muehlenkamp, J. J. (Eds.). (2014). *Non-suicidal self-injury in eating disorders: Advancements in etiology and treatment*. New York, NY: Springer. <http://dx.doi.org/10.1007/978-3-642-40107-7>
- Copeland, W. E., Goldston, D. B., & Costello, E. J. (2017). Adult associations of childhood suicidal thoughts and behaviors: A prospective, longitudinal analysis. *Journal of the American Academy of Child & Adolescent Psychiatry*, *56*, 958–965.e4. <http://dx.doi.org/10.1016/j.jaac.2017.08.015>
- Crosby, A., Ortega, L., & Melanson, C. (2011). *Self-directed violence surveillance*. Atlanta, GA: Centers for Disease Control and Prevention, National Center for Injury Prevention and Control.
- Crumpp, M. J. C., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PLoS ONE*, *8*, e57410. <http://dx.doi.org/10.1371/journal.pone.0057410>
- Fischer, G., Ameis, N., Parzer, P., Plener, P. L., Groschwitz, R., Vonderlin, E., . . . Kaess, M. (2014). The German version of the self-injurious thoughts and behaviors interview (SITBI-G): A tool to assess non-

- suicidal self-injury and suicidal behavior disorder. *BMC Psychiatry*, *14*, 265. <http://dx.doi.org/10.1186/s12888-014-0265-0>
- Fisher, C. B., Pearson, J. L., Kim, S., & Reynolds, C. F. (2002). Ethical issues in including suicidal individuals in clinical research. *IRB: Ethics & Human Research*, *24*, 9–14.
- Fleiss, J., Levin, B., & Paik, M. (2003). *Statistical methods for rates and proportions* (3rd ed.) Hoboken, NJ: John Wiley & Sons Inc.
- Food and Drug Administration. (2012, August). *Guidance for industry: Suicidal ideation and behavior: Prospective assessment of occurrence in clinical trials*. Silver Spring, MD: Author.
- Fox, K. R., Millner, A. J., & Franklin, J. C. (2016). Classifying nonsuicidal overdoses: Nonsuicidal self-injury, suicide attempts, or neither? *Psychiatry Research*, *244*, 235–242. <http://dx.doi.org/10.1016/j.psychres.2016.07.052>
- Franklin, J. C., Puzia, M. E., Lee, K. M., & Prinstein, M. J. (2014). Low implicit and explicit aversion toward self-cutting stimuli longitudinally predict nonsuicidal self-injury. *Journal of Abnormal Psychology*, *123*, 463–469.
- García-Nieto, R., Blasco-Fontecilla, H., Paz Yepes, M., & Baca-García, E. (2013). Translation and validation of the “Self-Injurious Thoughts and Behaviors Interview” in a Spanish population with suicidal behaviour. *Revista de Psiquiatría y Salud Mental*, *6*, 101–108.
- Giddens, J. M., Sheehan, K. H., & Sheehan, D. V. (2014). The Columbia-Suicide Severity Rating Scale (C-SSRS): Has the “gold standard” become a liability? *Innovations in Clinical Neuroscience*, *11*, 66–80.
- Gould, M. S., Marrocco, F. A., Kleinman, M., Thomas, J. G., Mostkoff, K., Cote, J., & Davies, M. (2005). Evaluating iatrogenic risk of youth suicide screening programs: a randomized controlled trial. *Journal of the American Medical Association*, *293*, 1635–1643.
- Hauser, D. J., & Schwarz, N. (2016). Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods*, *48*, 400–407. <http://dx.doi.org/10.3758/s13428-015-0578-z>
- Heilbron, N., Compton, J. S., Daniel, S. S., & Goldston, D. B. (2010). The Problematic label of suicide gesture: Alternatives for clinical research and practice. *Professional Psychology: Research and Practice*, *41*, 221–227. <http://dx.doi.org/10.1037/a0018712>
- Lloyd-Richardson, E. E., Perrine, N., Dierker, L., & Kelley, M. L. (2007). Characteristics and functions of non-suicidal self-injury in a community sample of adolescents. *Psychological Medicine*, *37*, 1183–1192. <http://dx.doi.org/10.1017/S003329170700027X>
- Michaels, M. S., Chu, C., Silva, C., Schulman, B. E., & Joiner, T. (2015). Considerations regarding online methods for suicide-related research and suicide risk assessment. *Suicide and Life-Threatening Behavior*, *45*, 10–17. <http://dx.doi.org/10.1111/sltb.12105>
- Millner, A. J., Lee, M. D., & Nock, M. K. (2015). Single-item measurement of suicidal behaviors: Validity and consequences of misclassification. *PLoS ONE*, *10*, e0141606. <http://dx.doi.org/10.1371/journal.pone.0141606>
- Møhl, B., la Cour, P., & Skandsen, A. (2014). Non-suicidal self-injury and indirect self-harm among Danish high school students. *Scandinavian Journal of Child and Adolescent Psychiatry and Psychology*, *2*, 11–18. <http://dx.doi.org/10.21307/sjcap-2014-003>
- 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.
- Nock, M. K., Borges, G., Bromet, E. J., Cha, C. B., Kessler, R. C., & Lee, S. (2008). Suicide and suicidal behavior. *Epidemiologic Reviews*, *30*, 133–154. <http://dx.doi.org/10.1093/epirev/mxn002>
- 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., & Kessler, R. C. (2006). Prevalence of and risk factors for suicide attempts versus suicide gestures: Analysis of the National Comorbidity Survey. *Journal of Abnormal Psychology*, *115*, 616–623. <http://dx.doi.org/10.1037/0021-843X.115.3.616>
- Nock, M. K., Millner, A. J., Joiner, T. E., Gutierrez, P. M., Han, G., Hwang, I., . . . Kessler, R. C. (2018). Risk factors for the transition from suicide ideation to suicide attempt: Results from the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). *Journal of Abnormal Psychology*, *127*, 139–149. <http://dx.doi.org/10.1037/abn0000317>
- Oliffe, J. L., Ogrodniczuk, J. S., Gordon, S. J., Creighton, G., Kelly, M. T., Black, N., & Mackenzie, C. (2016). Stigma in male depression and suicide: A Canadian sex comparison study. *Community Mental Health Journal*, *52*, 302–310. <http://dx.doi.org/10.1007/s10597-015-9986-x>
- Posner, K., Brown, G. K., Stanley, B., Brent, D. A., Yershova, K. V., Oquendo, M. A., . . . Mann, J. J. (2011). The Columbia-Suicide Severity Rating Scale: Initial validity and internal consistency findings from three multisite studies with adolescents and adults. *The American Journal of Psychiatry*, *168*, 1266–1277. <http://dx.doi.org/10.1176/appi.ajp.2011.10111704>
- Posner, K., Oquendo, M. A., Gould, M., Stanley, B., & Davies, M. (2007). Columbia Classification Algorithm of Suicide Assessment (C-CASA): Classification of suicidal events in the FDA’s pediatric suicidal risk analysis of antidepressants. *The American Journal of Psychiatry*, *164*, 1035–1043. <http://dx.doi.org/10.1176/ajp.2007.164.7.1035>
- Rudd, M. D., Mandrusiak, M., Joiner Jr, T. E., Berman, A. L., Van Orden, K. A., & Hollar, D. (2006). The emotional impact and ease of recall of warning signs for suicide: A controlled study. *Suicide and Life-Threatening Behavior*, *36*, 288–295.
- Ruskin, P. E., Reed, S., Kumar, R., Kling, M. A., Siegel, E., Rosen, M., & Hauser, P. (1998). Reliability and acceptability of psychiatric diagnosis via telecommunication and audiovisual technology. *Psychiatric Services*, *49*, 1086–1088.
- Revelle, W., & Condon, D. M. (2018). Reliability from alpha to omega: A tutorial. *PsyArXiv*. Oxford, England: Oxford University Press. <http://dx.doi.org/10.31234/osf.io/2y3w9>
- Schulenberg, S. E., & Yutzenka, B. A. (1999). The equivalence of computerized and paper-and-pencil psychological instruments: Implications for measures of negative affect. *Behavior Research Methods, Instruments, & Computers*, *31*, 315–321.
- Shapiro, D. N., Chandler, J., & Mueller, P. (2013). Using Mechanical Turk to study clinical populations. *Clinical Psychological Science*, *1*, 213–220. <http://dx.doi.org/10.1177/2167702612469015>
- Stanley, I. H., Boffa, J. W., Hom, M. A., Kimbrel, N. A., & Joiner, T. E. (2017). Differences in psychiatric symptoms and barriers to mental health care between volunteer and career firefighters. *Psychiatry Research*, *247*, 236–242. <http://dx.doi.org/10.1016/j.psychres.2016.11.037>
- Stewart, J. G., Kim, J. C., Esposito, E. C., Gold, J., Nock, M. K., & Auerbach, R. P. (2015). Predicting suicide attempts in depressed adolescents: Clarifying the role of disinhibition and childhood sexual abuse. *Journal of Affective Disorders*, *187*, 27–34. <http://dx.doi.org/10.1016/j.jad.2015.08.034>
- St. Germain, S. A., & Hooley, J. M. (2012). Direct and indirect forms of non-suicidal self-injury: Evidence for a distinction. *Psychiatry Research*, *197*, 78–84. <http://dx.doi.org/10.1016/j.psychres.2011.12.050>
- 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>
- Tang, J., Yu, Y., Wu, Y., Du, Y., Ma, Y., Zhu, H., . . . Liu, Z. (2011). Association between non-suicidal self-injuries and suicide attempts in Chinese adolescents and college students: A cross-section study. *PLoS ONE*, *6*, e17977. <http://dx.doi.org/10.1371/journal.pone.0017977>

- Washburn, J. J., Juzwin, K. R., Styer, D. M., & Aldridge, D. (2010). Measuring the urge to self-injure: Preliminary data from a clinical sample. *Psychiatry Research, 178*, 540–544.
- Weir, J. P. (2005). Quantifying test–retest reliability using the intraclass correlation coefficient and the SEM. *Journal of Strength and Conditioning Research, 19*, 231–240.
- 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.

Received May 9, 2019

Revision received January 7, 2020

Accepted February 27, 2020 ■