American Psychological Association

Psychological Assessment

ISSN: 1040-3590

2021, Vol. 33, No. 3, 218–229
https://doi.org/10.1037/pas0000976

Inconsistencies in Self-Reports of Suicidal Ideation and Attempts Across Assessment Methods

Charlene A. Deming[1, 2, 3], Julia A. Harris[4], Franckie Castro-Ramirez[1], Jeffrey J. Glenn[5], Christine B. Cha[6], Alexander J. Millner[1], and Matthew K. Nock[1]

[1] Department of Psychology, Harvard University
[2] Department of Psychiatry and Behavioral Sciences, Duke University Medical Center
[3] Durham VA Health Care System, Durham, NC, United States
[4] Department of Psychology, University of Utah
[5] Deloitte LLP, Arlington, VA, United States
[6] Department of Counseling and Clinical Psychology, Teachers College, Columbia University

Suicide researchers commonly use a variety of assessment methods (e.g., surveys and interviews) to enroll participants into studies and assign them to study conditions. However, prior studies suggest that different assessment methods and items may yield different responses from participants. This study examines potential inconsistencies in participants' reports of suicidal ideation (SI) and suicide attempt (SA) across commonly used assessment methods: phone screen interview, in-person interview, self-report survey, and confidential exit survey. To test the reliability of the effects, we replicated the study across two samples. In both samples, we observed a notable degree of inconsistent reporting. Importantly, the highest endorsement rates for SI/SA were on a confidential exit survey. Follow-up assessments and analyses did not provide strong support for the roles of purposeful inaccuracy, random responding, or differences in participant experiences/conceptualizations of SI. Although the reasons for such inconsistencies remain inconclusive, results suggest that classification of suicidal/control participants that uses multiple items to capture a single construct, that uses a Graded Scale to capture a broad spectrum of thoughts and behaviors, and that takes into account consistency of responding across such items may provide clearer and more homogenous groups and is recommended for future study.

Public Significance Statement

People are inconsistent in their reports of SI and SA across different methods of assessment, with the highest rates of endorsement on more anonymous measures, even in spite of apparent efforts to provide accurate information. These findings have important implications for both researchers and clinicians whose work relies on accurately identifying individuals impacted by suicidal thoughts and behaviors.

Keywords: suicidal ideation, suicide attempt, self-report, communication

Supplemental materials: https://doi.org/10.1037/pas0000976.supp

Suicide is the 10th leading cause of death in the United States (Centers for Disease Control & Prevention, 2017). Despite decades of research, our ability to predict this devastating outcome has not improved over time (Franklin et al., 2017). One possible reason for our continued difficulty around understanding and decreasing suicide is the poor measurement of suicidal outcomes, such as suicidal ideation (SI; i.e., thoughts about purposefully ending one's own life) and suicide attempts (SAs; i.e., action taken to purposefully end one's own life). Misclassification or inconsistent responses of the critical outcomes of interest reduce our ability to identify factors associated with these outcomes, impeding progress, and understanding.

There are multiple problems with how SI and SA are currently assessed. One issue is that people may purposefully choose not to disclose SI and/or SA to others. Prior research suggests that fewer than half of people with SI disclose their thoughts (e.g., Eskin et al., 2015). There are several potential reasons for this type of nondisclosure. First, concerns about negative or unwelcome responses from others, such as stigma or unwanted hospitalization, could decrease people's likelihood of reporting SI/SA (Hom et al., 2017). Second, because SI typically is transient in nature and most often lasts for only minutes to 1 hr at a time (Nock et al., 2009),

This article was published Online First March 11, 2021.
Charlene A. Deming https://orcid.org/0000-0002-3093-9436
Julia A. Harris https://orcid.org/0000-0003-0780-854X
Alexander J. Millner https://orcid.org/0000-0001-6092-2857
This research was funded by the Military Suicide Research Consortium (MSRC), an effort supported by the Office of the Assistant Security of Defense for Health Affairs (Award W81XWH-10-2-0181). Opinions, interpretations, conclusions, and recommendations are those of the authors and not necessarily endorsed by MSRC or the Department of Defense.
Correspondence concerning this article should be addressed to Charlene A. Deming, PhD, Department of Psychology, Harvard University, William James Hall, 33 Kirkland Street, Cambridge, MA 02138. Email: cad347@mail.harvard.edu

if SI is not present during an assessment, the person being assessed may be more likely to deny such thoughts. Third, even in cases in which a person intends to disclose their SI, the assessment methods used may cause confusion. More specifically, the field generally relies on single-item measures to assess for the presence of suicidal outcomes (e.g., "Have you ever had thoughts of suicide?" "Have you ever made a suicide attempt?"). This can lead to reporting errors if the single-item used is vague or unclear as people may vary in their subjective conceptualizations of what it means to have "thoughts of suicide" or have made a "suicide attempt." For instance, one study found that nearly 9% of participants who endorsed SI indicated thoughts that were not consistent with research definitions of this construct (i.e., thoughts of death, depression-related thoughts, passive SI, or no relevant thoughts) (Millner et al., 2015). Such differences can lead to the misclassification of individuals, which in turn decreases statistical power and increases the chances of drawing inaccurate statistical conclusions (Millner et al., 2015). It is critical that researchers and clinicians be able to accurately classify those at risk for SI and SA. Doing so improves researchers' ability to identify risk factors for SI and SA, and allows clinicians to provide appropriate treatment to those in need.

Prior studies comparing different assessment methods find that the *less personal* and the *more anonymous* the assessment is, the higher the rates of suicide endorsement. In terms of being less personal, people are more likely to disclose suicide-relevant material through computer-based (Greist et al., 1973) or questionnaire-based (Kaplan et al., 1994) assessments compared to in-person interviews. With regard to being more anonymous, rates of SI endorsement are higher when suicide assessment is performed anonymously than when the identity of the person being assessed is known (Anestis & Green, 2015; Warner et al., 2011). Suicide researchers commonly use a variety of assessment methods (e.g., phone screen interview) and screening criteria (e.g., presence of lifetime SI) to enroll participants into their studies and to assign participants to study conditions. However, prior studies have not directly tested if and how reporting of SI and SA differ across assessment methods or screening items and laboratory visit measures. This is important to examine because if suicidal status at phone screen and during laboratory visit do not match, this could indicate that either participants are not who they claim to be and/or the research questions posed to participants are unclear (raising concerns about the reliability and validity of our measurements). In either instance of inconsistency, there would be a negative impact for both researchers (e.g., heterogeneous study groups may obscure true findings) and clinicians (e.g., limits the ability to provide appropriate intervention). If, however, there are no inconsistences between reporting of SI and SA between phone screen/initial study screen and other forms of assessment, it would provide more confidence that participants truly fit the intended criteria and the current methods of screening are sufficient.

The primary purpose of this study is to better understand the potential variability in people's responses to different methods of assessing the presence of SI and SA. Specifically, we examined the consistency of people's endorsement of SI and SA across: (a) phone screen interview, (b) in-person interview, (c) self-report questionnaire, and (d) confidential exit survey (sealed in an envelope by the participant not to be viewed by the study staff until the end of all data collection). The secondary purpose of this study was to better understand why inconsistencies may exist through the examination of participants' self-reported effort and perceived accuracy of responses, as well as through participants' subjective experience/conceptualization of "suicidal thoughts." To increase generalizability, we carried this study out two times: First in a community sample ("Sample 1") and second in a veteran sample ("Sample 2"). Because veterans are at increased risk for suicide and may be particularly affected by stigma, suicide assessment may be especially sensitive for this group.

Method

Participants

Sample 1

Participants were 72 adults (M = 35.1 years, SD = 13.9 years; 66.7% female; 54.2% White) from the Boston community. Based on phone screen data, the total sample was broken into two main groups: 35 *suicide ideators* with the lifetime history of SI and 37 *controls* with no lifetime history of SI or suicidal behaviors. Other study inclusion criteria were: The ability to read, write, and speak English and, for suicide ideators and for a subsample of controls (n = 21; 56.8%), a positive screen for current (i.e., past 2 weeks) major depressive disorder (MDD) as determined via a "yes" response to at least one of two screening items on the Mini-International Neuropsychiatric Interview (MINI; Sheehan et al., 1998). We excluded from all analyses one participant who did not meet the criteria for MDD at the phone screen but was inadvertently admitted into the study as a suicide ideator. Exclusion criteria were: moderate-to-severe cognitive impairment (i.e., any score <3 on the Six-Item Screener for Cognitive Impairment; Callahan et al., 2002) or any other factor that may impair an individual's ability to comprehend and effectively participate in the study (e.g., extreme agitation, violent behavior, and high/imminent suicide risk). Exclusion criteria for the nonsuicidal controls also included current uncertainty about wanting to live (i.e., any score <9 out of 10 on the "desire to live" question in the phone screen; "Currently how would you rate your desire to live, with '10' being you really want to be alive and '0' being you very much want to be dead?"). This criterion was included in order to increase the accuracy of group classification and as a second check to confirm the absence of any current suicide-relevant thought, even an ambivalent wish to live, among "nonsuicidal" controls.

Sample 2

Participants were 169 veterans (M = 43.7 years, SD = 14.8 years; 11.8% female; 70.4% White) from the Boston community. Based on phone screen data, the total sample included: 92 *recent suicide ideators* with SI in the past 30 days and 77 *depressed controls* with no history of SI (passive or active) or SA. Other inclusion criteria for both groups were: A positive screen for current MDD (i.e., reporting "yes" to at least one of two MINI screening items assessing MDD symptoms over the past 2 weeks). The exclusion criteria were identical to those for Sample 1. Also consistent with Sample 1, depressed controls were excluded if they reported uncertainty around the desire to live (i.e., any score <9 out of 10 on the "desire to live" question). Excluding controls with uncertainty around the desire to live was especially important in sufficiently distinguishing controls from suicidal participants in

Sample 2, as all control participants in this sample were purposefully recruited to screen positive for MDD *in the absence of SI*. After data collection but before analyses, we excluded eight suicide ideators who, during their initial phone screen interview, reported a lifetime presence of SI but an absence of SI in the past 30 days ($n = 7$), or provided an inconsistent report of lifetime SI ($n = 1$). An additional 23 depressed controls were excluded based on their responses to the initial phone screen interview, reporting the presence of Passive SI ($n = 18$, "Did you ever wish you were dead or would go to sleep and never wake up again?"), lifetime SI ($n = 7$; "Have you ever had thoughts of killing yourself"), and/or desire to live below "9" ($n = 2$). These 31 participants were inadvertently admitted into the study as they did not meet initial study inclusion criteria via phone screen interview; therefore, they were not included in the total sample ($N = 169$).

Protocol

The protocol was identical across Samples 1 and 2, except where indicated. Both Study 1 and Study 2 were reviewed and approved by Harvard University's Committee on the Use of Human Subjects: University-Area Institutional Review Board.

Prestudy Recruitment and Screening

Participants were recruited via online postings and community flyers. Prior to the laboratory visit, a trained research assistant interviewed all potential participants using a standardized phone screen (described below) to assess study eligibility.

Laboratory Visit

During the laboratory visit, participants provided written informed consent and were asked to complete interview questions, self-report questionnaires, mood inductions, computer-based tasks, confidential exit survey, and suicide risk assessment. Sample 1 participants received $30 for their 3-hr study participation; Sample 2 participants received $100 for their 4-hr study participation. The data presented here are from two larger projects focused on identifying cognitive processes associated with SI and SA. This study reports only on the methods and data applicable to the examination of the consistency of SI and SA reporting.

Measures

Phone Screen Interview

The prestudy phone screen interview assessed: Current desire to live and history of SI and SA using items from the Self-Injurious Thoughts and Behaviors Interview (SITBI; Nock et al., 2007). The SITBI assesses various characteristics (e.g., presence and frequency) of suicidal and nonsuicidal self-injury via structured interview. When examined among a sample of young adults ($M = 17.1$ years old, $SD = 1.9$ years), the SITBI demonstrated strong interrater reliability (average *Kappa* = .99, $r = 1.0$) and 6-month test–retest reliability of presence (average *Kappa* = .70) and frequency (average intraclass correlation coefficient = .44) of SI, suicide plan, SA, suicide gesture, and NSSI (Nock et al., 2007). Further, the SITBI has been shown to have strong concurrent validity via concordance with other established measures of SI (i.e., Schedule for Affective Disorders and Schizophrenia for School-Aged Children—Present and Lifetime Version [K-SADS-PL], Beck Scale for Suicide Ideation; average *Kappa* = .54) and of SA (i.e., K-SADS-PL; *Kappa* = .65) (Nock et al., 2007). Suicide risk assessment and safety planning procedures were performed as needed. Because this study is interested in the impact of anonymity on reports of SI and SA, it is important to note that this initial phone screen was not considered to be anonymous. Prior to asking screening questions, phone interviewers described the sensitive nature of upcoming questions, detailed limits of confidentiality, and collected contact information from phone respondents (i.e., name, address, and phone number). The time lag between completing the phone screen interview and laboratory measures varied across participants. In Sample 1, the average lag was 6.30 days ($SD = 4.98$ days) with a range of 0 and 26 days; in Sample 2, the average lag was 9.20 days ($SD = 12.91$ days) with a range of 0–119 days. Given analyses focused on lifetime (vs. recent or current) SI and SA, it was unlikely, though technically possible, that any "inconsistent" responding occurred due to an actual change in lifetime SI/SA status during the intervening period between phone screen and laboratory measures. Consistent with this conclusion, analyses indicated that "inconsistent" responding was not associated with the length of time interval (analyses available by request from the first author).

In-Person Interview

During the laboratory visit, the history of SI and SA was assessed once again using items from the SITBI (described above; Nock et al., 2007).

Self-Report Scales

During the laboratory visit the *Beck Scale for Suicide Ideation* (BSS; Beck et al., 1988) obtained information relevant to desire to live and SA history (see Supplemental Table 1, for precise items). The *BSS* includes 21 items and measures current (i.e., past week) presence/severity of one's desire to live and SI, as well as, lifetime presence of SA. The *BSS* is a self-report version of the clinician-administered *Beck Scale for Suicide Ideation* (SSI; Beck et al., 1979). The *BSS* has demonstrated strong concurrent validity with the SSI, strong internal consistency, and both divergent and convergent validity (Beck et al., 1988; Beck & Steer, 1991).

Confidential Exit Survey

The confidential exit survey was developed specifically for these studies to allow participants to share details about their experience of the laboratory visit—as well as to confidentially address inclusion and exclusion criteria themselves. Among other items, participants were asked to report on the presence and content of SI to confirm accurate classification as a "suicidal" or "control" participant without fear of financial penalty (i.e., participants received payment regardless of responses), stigma (i.e., participants were told that, "Your answers will never be linked with your name, only a *non-identifying* subject number"), or unwanted treatment (i.e., responses were not connected to risk assessment/safety planning). The goal of collecting this information was to both: (a) improve participant experience in future studies and (b) better understand the

appropriateness of participants included within the study based on prior research to show that anonymity increases accurate disclosure of suicidal thoughts and behaviors (STBs) (Anestis & Green, 2015; Warner et al., 2011). This latter goal, to incentivize honest reporting of SI, was important based on anecdotal evidence that some research participants may over-report SI (i.e., report presence of SI, when it is not truly there) in order to participate and receive compensation from a study, whereas other participants may under-report SI (i.e., report absence of SI, when it is truly there) in order to avoid unwanted help resources/hospitalization and/or to minimize opportunities for self-perceived stigma (e.g., embarrassment). Overall, the confidential exit survey was included in an effort to remove or minimize the influence of factors that threaten accurate reporting around SI. To our knowledge, this is the first confidential survey to be compared directly with answers given during a laboratory visit.

Relevant to these studies, in the confidential exit survey, participants were asked to report on their: (a) effort and perceived accuracy in responding to laboratory visit questions (i.e., "... Which best describes the way you responded to the questionnaires?"; see Table 3, for full question prompt and response options); (b) true match to inclusion and exclusion criteria (e.g., "Have you had at least one thought of killing yourself in your entire lifetime?"; see Supplemental Table 1, for additional items); and (c) subjective experience/conceptualization of "suicidal thoughts." With regard to participants' subjective experience and conceptualization of "–+ thoughts," we used the *Graded Scale of Suicide and Related Thoughts* (referred to as "Graded Scale" for the remainder of this article) (Millner et al., 2015) to ask about a graded spectrum of possible thoughts. Specifically, participants were provided with the following prompt, "People have different definitions of exactly what a 'suicidal thought' is. Please check 'Yes' for each thought that you have had/experienced" and were asked to respond to the following items: General thoughts about death (i.e., "What happens to people when they die?", "What will it be like when I die?"), depressive thoughts (i.e., "I wish I could disappear or not exist," "I wish I was never born," "My life is not worth living"), passive SI (i.e., "I wish I was dead [ex: go to sleep and not wake up again"]), and active SI (i.e., "Maybe I should kill myself," "I should kill myself").

Data Analyses

For all analyses, assignment of group membership (i.e., suicide ideator and control) was based on responses to an initial phone screen interview. Analyses used chi-square tests to examine group differences, with Cramer's *V* to indicate effect sizes.[1] As a preliminary step and to confirm that participants classified as suicide ideators based on phone screen generally had higher rates of SI and SA compared to those classified as controls, we tested for group differences in endorsement of SI and SA indicated during laboratory questioning. We next examined the consistency of reporting SI and SA *within* those classified as suicide ideators and controls by comparing rates of endorsement across the various assessment methods (i.e., between phone screen interview and each laboratory measure, as well as between in-person interview, self-report questionnaire, and confidential exit survey laboratory measures), using McNemar's Test for significance to account for nonindependent, repeated-measures observations. Missing data were not counted as inconsistent, meaning that a participant missing response on one assessment method but with otherwise consistent responses was deemed "consistent." This handling of missing data was taken as a conservative approach to identifying inconsistency and one that would, if anything, underestimate the degree of inconsistency overall. Finally, we tested potential reasons for inconsistencies. Specifically, we examined endorsement rates for questions related to potential factors influencing participant responses (i.e., self-reported effort and perceived accuracy, subjective experience/conceptualization of "suicidal thoughts") and tested whether these factors were more commonly reported among those who were inconsistent, versus consistent, in their reports of SI.

Results

Preliminary Analyses: Between-Group Differences in SI and SA

First, we sought to confirm that responses to the phone screen interview generally mapped onto expected reports during the laboratory visit [i.e., groups categorized as "suicidal," compared to "controls" (based on the phone screen), should report greater SI via laboratory measures]. As expected, for both Sample 1 and Sample 2, compared with control participants, participants who endorsed having suicidal thoughts during the phone screen showed significantly higher rates of (a) lifetime SI via all modes of assessment (Table 1; Sample 1: $Vs = .65–.95$; Sample 2: $Vs = .75–.90$), (b) depressive thoughts and all forms of suicide-related thoughts, with effect sizes generally increasing with the severity of the ideation assessed (Sample 1: $V = .51$ for depressive thoughts to $V = .72$ for passive SI; Sample 2: $V = .53$ for depressive thoughts to $V = .71$ for active SI), and (c) lifetime SA via all modes of assessment (Table 2; Sample 1: $Vs = .36–.51$; Sample 2: $Vs = .48–.53$).

How Consistent Were Reports of SI and SA Across Different Assessment Methods?

Regarding SI, across different assessment methods, both control and suicidal participants reported lifetime SI inconsistently (see Table 1). Among Sample 1 control participants, 0.0% reported lifetime SI via phone screen interview, 5.4% reported lifetime SI via in-person interview, and 16.7% reported lifetime SI via confidential exit survey. This pattern also was observed in Sample 2 controls, with 0.0% reporting lifetime SI on the phone screen interview, 11.7% reporting lifetime SI via in-person interview, and 23.7% reporting lifetime SI via confidential exit survey. Suicidal participants were more consistent in their reporting of lifetime SI, with nearly 100% of all such participants in both Sample 1 and Sample 2 reporting SI through phone screen interview, in-person interview, and general SI question on the confidential exit survey (within Sample 2, 96.6% of suicide ideators reported lifetime SI using the general confidential exit survey item). However, and importantly, when looking at the perception/conceptualization of what constitutes SI, among suicide ideators in both Samples 1 and 2, we see that approximately 80% reported lifetime active SI, meaning

[1] The "Results" section includes effect sizes (see Tables, for information on chi-square and *p* values).

Table 1

Reports of Lifetime Suicidal Ideation across Assessment Methods

	Phone screen interview	In-person interview	Confidential exit survey									
			People have different definitions of exactly what a 'suicidal thought' is. Please check "Yes" for each thought that you have had/experienced:									
Groups determined by phone screen data	Sample 1: *Have you ever had thoughts about actually killing yourself?* Sample 2[a]: *Have you ever seriously considered killing yourself?*	*Have you ever had thoughts of killing yourself?*	*1. What happens to people when they die?*	*2. What will it be like when I die?*	*3. I wish I could disappear or not exist*	*4. I wish I was never born*	*5. My life is not worth living*	*6. I wish I was dead (ex: go to sleep and not wake up again)*	*7. Maybe I should kill myself*	*8. I should kill myself*	*Lifetime Presence of Active SI (7. or 8.)*	*Have you had at least one thought of killing yourself in your entire lifetime?*[b]
Sample 1[c]												
Suicide ideators	100.0%	100.0%	77.1%	88.6%	74.3%	54.3%	57.1%	80.0%	80.0%	42.9%	80.0%[e]	100.0%
Controls	0%	5.4%	73.5%	74.3%	20.6%	5.9%	8.8%	8.8%	14.7%	2.9%	14.3%[f]	16.7%
Significance test	$\chi^2_{(1)} = 72.0^{**}$	$\chi^2_{(1)} = 64.4^{**}$	$\chi^2_{(1)} = 0.1$	$\chi^2_{(1)} = 2.4$	$\chi^2_{(1)} = 19.9^{**}$	$\chi^2_{(1)} = 19.1^{**}$	$\chi^2_{(1)} = 18.1^{**}$	$\chi^2_{(1)} = 35.3^{**}$	$\chi^2_{(1)} = 29.5^{**}$	$\chi^2_{(1)} = 15.4^{**}$	$\chi^2_{(1)} = 29.5^{**}$	$\chi^2_{(1)} = 35.9^{**}$
(effect size)	($V = 1.00$)	($V = .95$)	($V = .04$)	($V = .18$)	($V = .54$)	($V = .53$)	($V = .51$)	($V = .72$)	($V = .65$)	($V = .47$)	($V = .65$)	($V = .80$)
Sample 2[d]												
Recent suicide ideators	100.0%	100.0%	73.6%	80.5%	65.5%	52.9%	64.8%	73.9%	77.3%	59.1%	80.8%[g]	96.6%
Depressed controls[b]	0%	11.7%	73.7%	68.0%	13.3%	4.0%	8.0%	9.6%	6.8%	4.1%	6.6%[h]	23.7%
Significance test	$\chi^2_{(1)} = 168.0^{**}$	$\chi^2_{(1)} = 136.0^{**}$	$\chi^2_{(1)} = 0.0$	$\chi^2_{(1)} = 3.3$	$\chi^2_{(1)} = 45.2^{**}$	$\chi^2_{(1)} = 45.6^{**}$	$\chi^2_{(1)} = 55.0^{**}$	$\chi^2_{(1)} = 66.7^{**}$	$\chi^2_{(1)} = 80.7^{**}$	$\chi^2_{(1)} = 54.3^{**}$	$\chi^2_{(1)} = 90.8^{**}$	$\chi^2_{(1)} = 92.8^{**}$
(effect size)	($V = 1.00$)	($V = .90$)	($V = .001$)	($V = .14$)	($V = .53$)	($V = .53$)	($V = .58$)	($V = .64$)	($V = .71$)	($V = .58$)	($V = .75$)	($V = .75$)

[a] Sample 2: Phone Screen Interview asked about lifetime SI in two different ways: (1) *Have you ever seriously considered killing yourself?* (see table) and (2) *Have you ever had thoughts of killing yourself*? Item (2) was asked of $n = 91$ Recent Suicide Ideators and $n = 77$ Depressed Controls; 100% of Recent Suicide Ideators endorsed "yes" to this question and 0% of Depressed Controls endorsed "yes." Significance and effect size are as follows: $\chi^2(1) = 168.0^{**}$, $V = 1.0$.

[b] The Confidential Feedback Survey included two slightly different versions of the SI content question: Version 1. *Have you ever had a thought of suicide in your entire lifetime?* V2. *Have you had at least one thought of killing yourself in your entire lifetime*? Each participant was asked only one of these two questions and data from Version 1. and Version 2. are combined and presented here.

[c] Sample 1: Suicide Ideator group contains n = 35 participants for each cell, except *Have you had at least one thought of killing yourself in your entire lifetime?*, which was added later in the study and contains $n = 20$ participants. Control group contains $n = 34$ to $n = 37$ participants for each cell.

[d] Sample 2: Recent Suicide Ideator group contains $n = 87$ to $n = 92$ participants for each cell. Depressed Control group contains $n = 73$ to $n = 77$ participants for each cell.

[e] Sample 1: 15/35Suicide Ideators have BOTH 7. and 8.; 42.9%.

[f] Sample 1: 1/34 Controls have BOTH 7. and 8.; 2.9%.

[g] Sample 2: 48/88 Recent Suicide Ideators have BOTH 7. and 8.; 54.5%.

[h] Sample 2: 3/74 Depressed Controls have BOTH 7. and 8.; 4.1%.

* Significant at the .05 level (two-tailed). ** Significant at the .001 level (two-tailed).

that roughly 20% of those recruited and screened as suicide ideators denied lifetime active SI.

Regarding SA, despite consistency across laboratory visit data, inconsistent reporting from phone screen interview to laboratory modes of assessment was observed among suicide ideators and depressed controls (see Table 2). Compared to phone screen interview, Sample 1 participants were significantly more likely to report the presence of lifetime SA via iIn-person interview, self-report questionnaire, and confidential exit survey (McNemar two-sided: All $p = .006$). In fact, Sample 1 suicide ideators nearly doubled their endorsement of SA from the phone screen (22.9%) to laboratory visit (40.0%). The same pattern held true for Sample 2 such that participants were more likely to report SA via a confidential exit survey compared with the phone screen interview (McNemar two-sided: $p = .027$). Among the 122 total participants who said "no" to SA at phone screen interview, 11 of those (9.0%) answered "yes" to SA at the confidential exit survey. The results for Sample 2 were slightly more complicated when comparing phone screen interview data to both in-person interview and self-report data. More specifically, comparisons of the presence of SA reported via phone screen interview and presence of SA reported via in-person interview and self-report questionnaire were not significant (McNemar two-sided: $p = 1.000$ and .181, respectively), though a closer examination revealed intraindividual inconsistencies. That is, among Sample 2, roughly the same number of people said "no" to SA using both modes of assessment (i.e., 127 compared to 128). However, nearly 1 in 10 individuals switched their answers from one mode to another; 8.9% of participants switched answers from phone screen to in-person interview and 8.5% of participants switched answers from phone screen to self-report questionnaire.

When examining consistency across the three laboratory assessment methods of SA for Sample 2 participants, there was some indication that the likelihood of disclosure of SA was associated with the level of anonymity of the assessment method. Among 157 Sample 2 participants who completed all in-person SA questions, 144 participants (91.7%) were consistent across the board with 110 saying "no" and 34 saying "yes" to each question. Among the 13 participants (8.3%) with inconsistencies, the majority of cases indicated potential inconsistency around the in-person interview. Specifically, 9 participants (69.2% of inconsistent cases) occurred when participants answered "yes" at both self-report and confidential exit survey but reported "no" during the in-person interview. Examining pairwise comparison of format types, the proportion of Sample 2 participants reporting the presence of SA via confidential exit survey was different than via in-person interview (McNemar two-sided: $p = .006$). Among the 123 total participants who answered "no" to SA via in-person interview, 11 of those (8.9%) answered "yes" to SA at the confidential exit survey. In contrast, among the 37 total participants who said "yes" to SA via in-person interview, only 1 participant (2.7%) answered "no" to SA at the confidential exit survey. Inconsistencies were less notable and not statistically significantly different when comparing between the in-person interview and self-report (McNemar two-sided: $p = .146$) and between self-report and confidential exit survey (McNemar two-sided: $p = .250$). Among the 126 total participants who answered "no" to SA via in-person interview, 9 of those (7.1%) answered "yes" to SA via self-report questionnaire. Among the 40 total participants who said "yes" to SA via in-person interview, 3 participants (7.5%) answered "no" to SA via self-report questionnaire. Among the 114 total participants who answered "no" to SA via self-report questionnaire, 3 of those (2.6%) answered "yes" to SA at the confidential exit survey. Among the 43 total participants who said "yes" to SA via self-report questionnaire, none answered "no" to SA at the confidential exit survey. The pattern of switching answers from "no" on in-person interview (least anonymous method) to "yes" on the confidential exit survey (most anonymous

Table 2
Reports of Suicide Attempt across Assessment Methods

Groups determined by phone screen data	Phone screen interview	In-person interview	Self-report questionnaire	Confidential exit survey
	Have you ever actually attempted to kill yourself?	*Have you ever made an actual attempt to kill yourself in which you had at least some intent to die?*	*Have you ever attempted suicide?*	*Have you ever made a suicide attempt in your entire lifetime?*
Sample 1[a]				
Suicide ideators	22.9%	40.0%[c]	40.0%[c]	40.0%[c]
Controls	0.0%	0.0%	0.0%	0.0%
Significance test (effect size)	$\chi^2_{(1)} = 9.5^*$ ($V = .36$)	$\chi^2_{(1)} = 18.4^{**}$ ($V = .51$)	$\chi^2_{(1)} = 18.4^{**}$ ($V = .51$)	$\chi^2_{(1)} = 17.9^{**}$ ($V = .50$)
Sample 2[b]				
Recent suicide ideators	45.1%	42.4%	49.4%	52.4%
Depressed controls	0.0%	1.3%	2.6%	3.9%
Significance test (effect size)	$\chi^2_{(1)} = 45.9^{**}$ ($V = .52$)	$\chi^2_{(1)} = 39.2^{**}$ ($V = .48$)	$\chi^2_{(1)} = 45.2^{**}$ ($V = .52$)	$\chi^2_{(1)} = 45.1^{**}$ ($V = .53$)

[a] Sample 1: Suicide Ideator group contains $n = 35$ participants for each cell. Control group contains $n = 36$ to $n = 37$ participants for each cell.
[b] Sample 2: Recent Suicide Ideator group contains $n = 84$ to $n = 92$ participants for each cell. Depressed Control group contains $n = 76$ to $n = 77$ participants for each cell.
[c] Among Sample 1, the 40% of people reporting lifetime SA via In-Person Interview, Self-Report Questionnaire, and Confidential Exit Survey are the same 14 individuals in each cell. In other words, these 14 individuals are consistently reporting lifetime SA via these three distinct assessment methods.
* Significant at the .05 level (two-tailed). ** Significant at the .001 level (two-tailed).

method) provides some indication that disclosure of suicide may be related to the level of anonymity. Regardless of the cause, however, the main finding is that reporting among Sample 2 participants appears inconsistent across numerous methods. An important exception to the pattern of inconsistent reporting was observed for Sample 1 controls. This group consistently reported having made no lifetime SAs across all modes of assessment.

Does Purposeful Inaccuracy, Random Responding, and/or Differences in Participant Experiences/ Conceptualizations of SI Account for Inconsistencies in Reporting?

Overall, across both those with consistent responding and those with inconsistent responding, people did not appear to intentionally respond inaccurately (see Table 3). Most participants (across Samples 1 and 2: 55.3–80.0%) indicated in the confidential exit survey that, "I read every question, answered accurately, and for the most part, found it easy to fill out." Only a small percentage (across Samples 1 and 2: 0.0–2.4%) reported purposely answering "inaccurately" or "randomly." A larger percentage (across Samples 1 and 2: 20.0–42.9%) reported that despite reading every question and trying their best they "found it difficult to give one number that best described me." Further, on average, the control group participants reported more thoroughly reading and accurately responding to assessment questions compared to the suicide ideator group participants, who reported more difficulty in selecting one response that best describes them. This difference was significant for both Sample 1 ($\chi^2_{(1)} = 4.2$, $p < .05$, $V = .25$) and Sample 2 ($\chi^2_{(3)} = 10.1$, $p < .05$, $V = .25$).

Given that a number of participants reported putting forth effort but still found it difficult to report accurately, we examined and found that people had differing subjective experiences/conceptualizations of "suicidal thoughts." Using the Graded Scale to better understand what people have experienced/conceptualize as "suicidal thoughts," we saw that people varied from "no" relevant thoughts to general thoughts about death, to depressive thoughts, to passive SI, to active SI. Figures 1 and 2 depict the highest level of "suicidal thought" endorsed both by suicide ideators and by controls. These figures suggest that the majority of SI participants' highest level of endorsement was within the active SI range of thinking (i.e., "Maybe I should kill myself," "I should kill myself"). More specifically, across the two samples, 80.0–80.8% of SI participants endorsed active SI; meaning that ~20% of suicidal participants denied active SI on the Graded Scale. Conversely, 6.6–14.3% of control participants endorsed active SI whereas 77.1–82.9% of control participants endorsed "no" thoughts or general thoughts about death (i.e., "What happens to people when they die," "What will it be like when I die?"). Incongruously, 5.7–12.4% of suicidal participants reported "no" to all thoughts on the Graded Scale (including both active- and passive-SI) or endorsed only general thoughts about death. Participants whose highest endorsement of thoughts fell in the middle of the severity spectrum—depressive thoughts (i.e., "I wish I could disappear or not exist," "I wish I was never born") and passive SI (i.e., "My life is not worth living," "I wish I was dead")—were a mixture of suicidal and control participants. Specifically, across the two samples, 6.7–14.3% of suicidal participants' highest level of endorsement was depressive thoughts or passive SI; 8.6–10.5% of control participants' highest level of endorsement was depressive thoughts or passive SI. This suggests that people whose strongest thoughts are

Table 3
Participant Effort/Accuracy of Response by Study Groups

	Confidential Exit Survey (Mutually Exclusive Response Options)			
	We asked you to fill out some questionnaires about your past thoughts, feelings or behaviors. Some people find it easy to give a number in response to these questions while others may find it difficult (for example, a person may think, "sometimes I feel like a 5 and other times a 1"). Some people may not want to give out the information and may answer inaccurately. Others might be bored, not read the questions and just answer randomly. Which best describes the way you responded to the questionnaires?			
Groups Determined by Phone Screen Data	*I read every question, answered accurately and, for the most part, found it easy to fill out.*	*I read every question, answered accurately but often found it difficult to give one number that best described me.*	*I read the questions but sometimes I did not want to give out that information so I answered inaccurately.*	*I did not read some questions or most of the questions, and answered randomly.*
Sample 1[a]				
Suicide ideators	57.1%	42.9%	0.0%	0.0%
Controls	80.0%	20.0%	0.0%	0.0%
Significance test (effect size)		$\chi^2_{(1)} = 4.2^*$ ($V = .25$)		
Sample 2[b]				
Recent suicide ideators	55.3%	41.2%	2.4%	1.2%
Depressed controls	78.7%	20.0%	1.3%	0.0%
Significance test (effect size)		$\chi^2_{(3)} = 10.1^*$($V = .25$)		

[a] Sample 1: Suicide Ideator group contains $n = 35$ participants. Control group contains $n = 35$ participants.
[b] Sample 2: Recent Suicide Ideator group contains $n = 85$ participants. Depressed Control group contains $n = 75$ participants.
* Significant at the .05 level (two-tailed). ** Significant at the .001 level (two-tailed).

depressive ideation or passive ideation may be unsure about whether to endorse or deny single-item questions regarding suicidal thoughts.

Next, we examined whether participant-reported effort and accuracy in responding to study questions or experience/conceptualization of what constitutes a "suicidal thought" could help to explain why some participants reported inconsistently across modes of assessment. Participants were defined as "inconsistent" if their reports of lifetime SI presence differed at any point between phone screen interview, in-person interview, and confidential exit survey (excluding Graded Scale responses). Participants were defined as "consistent" if their reports of lifetime SI presence remained the same across phone screen interview, in-person interview, and confidential exit survey. Individuals were counted as "consistent" if they reported the consistent presence of lifetime SI or consistent absence of lifetime SI.

We examined the association between consistency in reporting lifetime SI and reported effort/accuracy in responding to general study questions (see Table 4). Data showed that the people who were inconsistent were not necessarily the same people who found the assessment items difficult to complete and/or reported inaccurate/random responding. Specifically within Sample 1, of those with inconsistent responses to lifetime SI presence across measures, 0 out of 6 participants (0.0%) reported that they "found it difficult" to answer questions; all 6 out of 6 (100.0%) reported that they "read every question, answered accurately . . . found it easy to fill out," and none (0 out of 6; 0.0%) reported purposeful "inaccuracy" or "random" responding. Results generally were similar for Sample 2, with over half of those with inconsistent reporting of lifetime SI across measures (11 out of 20; 55.0%) reporting that they "read every question, answered accurately . . . found it easy to fill out," slightly less than half (9 out of 20; 45.0%) reporting that they "found it difficult" to answer questions, and none (0 out of 20; 0.0%) reporting purposeful "inaccuracy" or responding "randomly." The presence of inconsistent responding did not significantly differ as a function of participant-reported ease/accuracy of answering questions for either Sample 1 ($V = .20$) or Sample 2 ($V = .12$).

Finally, we examined the association between "inconsistent" responders (as defined above) and the location on the Graded Scale where people endorsed their highest "suicidal thought." More specifically, we examined two groups based on their Graded Scale responses: (a) participants who reported their highest "suicidal thought" on the extreme ends of the scale (i.e., clear reporting of "no" relevant thoughts, general thoughts about death, or active SI; the bottom three groups and the top two groups from Figures 1 and 2), and (b) participants who reported their highest suicidal thoughts in the middle range of the scale (i.e., depressive ideation, passive SI; the middle four groups from Figures 1 and 2), which represents those whose thoughts were ambiguous and could not be clearly categorized as absence or presence of SI. We hypothesized that those on the ends of the Graded Scale, with clearly absent or present SI, would report SI more consistently across modes of assessment, whereas those in the middle range of the Graded Scale, with ambiguous thoughts, would report SI less consistently. Contrary to this prediction, data showed that the people who were inconsistent in their responses to SI questions were generally not the same people whose thoughts were ambiguous. Specifically, in Sample 1, of those with the inconsistent response to lifetime SI presence across modes of assessment, 0 out of 6 participants (0.0%) reported ambiguous SI; all 6 out of 6 (100.0%) reported more clearly present or absent SI.

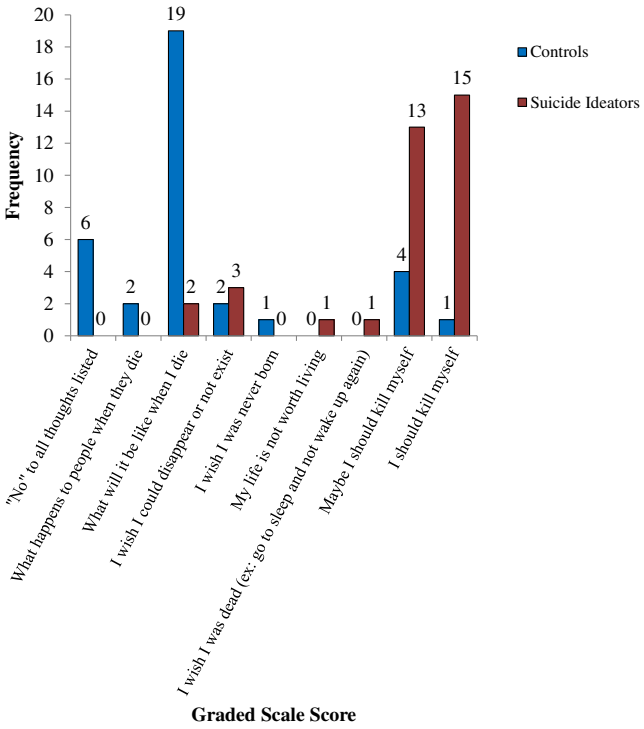
Table 4
Participant Effort/Accuracy of Response by Consistency of Response

	Confidential Exit Survey (Mutually Exclusive Response Options)			
	We asked you to fill out some questionnaires about your past thoughts, feelings or behaviors. Some people find it easy to give a number in response to these questions while others may find it's difficult (for example, a person may think, "sometimes I feel like a 5 and other times a 1"). Some people may not want to give out the information and may answer inaccurately. Others might be bored, not read the questions and just answer randomly. Which best describes the way you responded to the questionnaires?			
Groups determined based on consistency of reported lifetime si across modes of assessment[a]	*I read every question, answered accurately and, for the most part, found it easy to fill out.*	*I read every question, answered accurately but often found it difficult to give one number that best described me.*	*I read the questions but sometimes I did not want to give out that information so I answered inaccurately.*	*I did not read some questions or most of the questions, and answered randomly.*
Sample 1[b]				
Consistent responders	66.2%	33.8%	0.0%	0.0%
Inconsistent responders	100.0%	0.0%	0.0%	0.0%
Significance test (effect size)		$\chi^2_{(1)} = 2.9$ ($V = .20$)		
Sample 2[c]				
Consistent responders	67.9%	29.3%	2.1%	0.7%
Inconsistent responders	55.0%	45.0%	0.0%	0.0%
Significance test (effect size)		$\chi^2_{(3)} = 2.4$ ($V = .12$)		

[a] Sample Participants were defined as "inconsistent" if their reports of lifetime SI presence differed at any point between phone screen interview, in-person interview, and confidential exit survey (not inclusive of response to the Graded Scale). Participants were defined as "consistent" if their reports of lifetime SI presence remained the same across phone screen interview, in-person interview, and confidential exit survey; missing data were not counted as inconsistent. Individuals were counted as "consistent" if they reported consistent presence of lifetime SI or consistent absence of lifetime SI.
[b] Sample 1: Consistent Responders group contains $n = 65$ participants. Inconsistent Responders contains $n = 6$ participants.
[c] Sample 2: Consistent Responders group contains $n = 140$ participants. Inconsistent Responders group contains $n = 20$ participants.
* Significant at the .05 level (two-tailed). ** Significant at the .001 level (two-tailed).

Figure 1

Highest Item Endorsed on the Graded Scale[b] by Group (Sample 1[a])



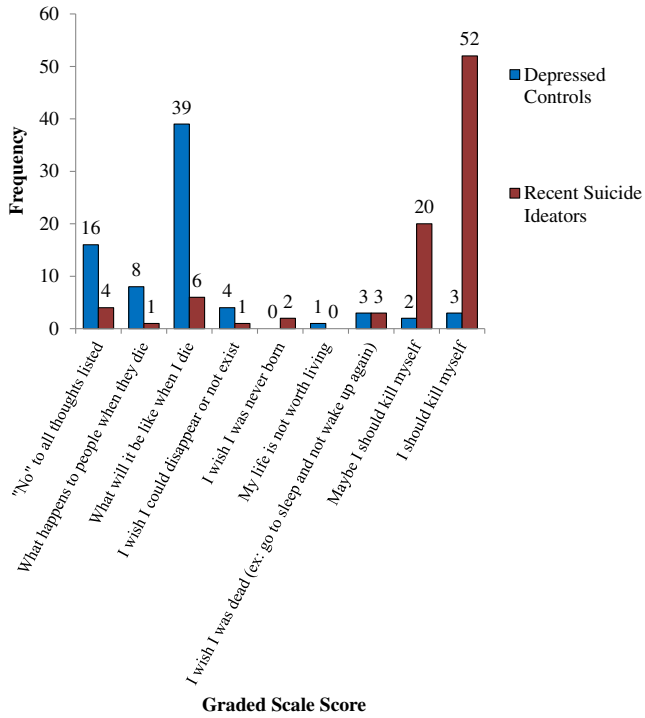
[a] Suicide ideator group contains $n = 35$ participants; $n = 0$ participants had missing data. Control group contains $n = 35$ participants; $n = 2$ participants had missing data.

[b] Graded Scale item asked during confidential exit survey. Question stem reads, *People have different definitions of exactly what a "suicidal thought" is. Please check "Yes" for each thought that you have had/experienced.*

Note. See the online article for the color version of this figure.

Figure 2

Highest Item Endorsed on the Graded Scale[b] by Group (Sample 2[a])



[a] Recent cuicide ideator group contains $n = 89$ participants; $n = 3$ participants had missing data. Control group contains $n = 76$ participants; $n = 1$ participant had missing data.

[b] Graded Scale item asked during confidential exit survey. Question stem reads, *People have different definitions of exactly what a "suicidal thought" is. Please check "Yes" for each thought that you have had/experienced.*

Note. See the online article for the color version of this figure.

Results generally were similar for Sample 2; less than a quarter of those with inconsistent reporting of lifetime SI across measures (4 out of 21; 19.0%) reported ambiguous SI, and the majority of inconsistent responders (17 out of 21; 81.0%) reported clearly absent or present SI (see Table 5). Overall, people with ambiguous SI (in the middle range of the Graded Scale) were not significantly more likely to be inconsistent in their reports of SI across modes of assessment than people with clearly absent or present SI (on the ends of the Graded Scale). This was true for both Sample 1 ($V = .20$) and for Sample 2 ($V = .15$).

Discussion

There were three key findings in this study. First, there was a notable degree of inconsistent reporting, with generally greater endorsement of STBs via laboratory visit compared to phone screen and the *highest* endorsement rates for STBs via confidential exit survey. Second, inconsistencies in reporting SI did not appear to be due to purposeful dishonesty, with no inconsistent respondents reporting purposeful inaccuracy or random responding. Third, participants reported varying subjective experiences/conceptualizations of suicidal thoughts. Participants whose thoughts about suicide were ambiguous and could not be clearly categorized as absence or presence of SI were no more likely to be inconsistent in their reports of SI across modes of assessment than were participants with clearly absent or present SI.

The finding that the highest rates of endorsement of SI and SA were via confidential exit survey is consistent with prior findings showing that increased anonymity, including the inability to link a person to additional help resources was related to increased rates of SI disclosure (e.g., Anestis & Green, 2015; Warner et al., 2011). What is new in this study is the fact that individuals inconsistently report SI and SA across commonly used methods of assessment. For example, up to 23.7% of people recruited as "controls" (i.e., without lifetime SI at phone screen interview) subsequently reported lifetime SI via confidential exit survey, and rates of SA among suicide ideators in one sample almost doubled from phone screen interview (22.9%) to laboratory visit (40.0%). Further, 8.9% of individuals in one sample who answered "no" to lifetime SA during the in-person interview answered "yes" to the same question on the confidential exit survey during the same laboratory session. Inconsistencies also existed for suicide ideators with regard to their perception/conceptualization of SI. Using a Graded Scale inclusive of differing levels of thoughts (i.e., ranging from general thoughts about death to depressive thoughts, to passive SI to active SI), we found the roughly 20% of those recruited as

Table 5
Subjective Experience/Conceptualization of "Suicidal Thoughts" by Consistency of Response

Groups determined based on consistency of reported lifetime SI across modes of assessment[a]	Confidential Exit Survey—Graded Scale (highest level of "suicidal thought" endorsed)		
	Extremes (Low or High) on SI Range[b] (i.e., no SI, general thoughts about death, active SI)		Middle on SI Range[b] (i.e., general thoughts about death, depressive thoughts, passive SI)
Sample 1[c]			
Consistent responders	87.7%		12.3%
Inconsistent responders	100.0%		0.0%
Significance test (effect size)		$\chi^2_{(1)} = 2.94$ ($V = .20$)	
Sample 2[d]			
Consistent responders	93.1%		6.9%
Inconsistent responders	81.0%		19.0%
Significance test(effect size)		$\chi^2_{(1)} = 3.46$ ($V = .15$)	

[a] Participants were defined as "inconsistent" if their reports of lifetime SI presence differed at any point between phone screen interview, in-person interview, and confidential exit survey. Participants were defined as "consistent" if their reports of lifetime SI presence remained the same across phone screen interview, in-person interview, and confidential exit survey; missing data were not counted as inconsistent. Individuals were counted as "consistent" if they reported consistent presence of lifetime SI or consistent absence of lifetime SI.
[b] Highest participant-reported "suicidal thoughts" on the Graded Scale were divided by: "extremes" (low or high) and "middle. Those classified into the "extremes" category included all those who reported their highest "suicidal thought" on the ends of the scale (i.e., reporting "no" relevant thoughts, general thoughts about death only, and active SI), representing more clearly absent or present thoughts. Those classified into the "middle" category included all those who reported their highest suicidal thoughts were in the middle range of the scale (i.e., depressive ideation and passive SI), representing more ambiguous thoughts.
[c] Sample 1: Consistent Responders group contains $n = 65$ participants. Inconsistent Responders contains $n = 6$ participants.
[d] Sample 2: Consistent Responders group contains $n = 144$ participants. Inconsistent Responders group contains $n = 21$ participants.
* Significant at the .05 level (two-tailed). ** Significant at the .001 level (two-tailed).

suicide ideators via phone screen interview denied lifetime active SI during confidential laboratory-based survey. As previous studies have shown, imprecise group classification (e.g., inadvertently including those with passive SI and active SI under the same "SI" group label) can undermine statistical tests and lead to faulty conclusions (Millner et al., 2015). From a research perspective, this may be problematic when trying to identify specific risk and protective factors relevant to STBs (e.g., risk factors for passive SI vs. risk factors for active SI). From a clinical perspective, inconsistent reports of SI and SA may be problematic when trying to provide the appropriate level of intervention to suicidal patients. Accurate classification is then critical to increasing our ability to better understand and target STBs through both research and clinical interventions.

There are many ways to interpret and understand inconsistencies in reporting. It is possible that participants with true SI initially underreported STBs at phone screen interview so as to avoid perceived stigma or unwanted help, and participants who truly did not have SI initially over-reported severity of STBs via phone screen interview in order to qualify for study participation. It is also possible that participants' lack of understanding questions contributed to inconsistent responses. Examination of self-reported effort and perceived accuracy, as well as subjective experience/conceptualization of SI, helps to shed some light on these possible explanations. Overall, participants reported that they either answered study questions with ease/accuracy or tried their hardest but had difficulty answering. Interestingly, those individuals who reported difficulty answering questions generally were *not* the same as those who inconsistently reported SI across modes of assessment. In fact, only 0.0–45.0% of those with inconsistent SI responses reported difficulty answering questions, with the majority, 55.0–100.0%, of those with inconsistent responses reporting careful reading and accurate/easy answering of questions. Importantly, zero participants (from both studies) with inconsistent reporting indicated purposeful inaccuracy or random responding. These data suggest that even individuals who believed they had read carefully and accurately/easily answered questions were inconsistent in their reports. Given that some participants reported difficulty choosing one response over another, the onus falls on researchers and clinicians to provide clearer and more effective communications of what is meant by suicide-related questions. The majority of participants assessed in this study reported that they attempted to answer accurately. Therefore, if researchers and clinicians can communicate more effectively the intended meaning of suicide-related questions, we may be able to more accurately differentiate those with SI and SA histories from those without. Overall, these findings suggest that participants are answering our assessment items to the best of their ability, but perhaps the intent and definitions of these items are not being clearly communicated to participants.

Speaking further to the possibility of confusion around terms, when using the Graded Scale with thoughts ranging in severity from no death- or suicide-related thoughts up to active SI, we saw that the two extreme ends of this scale (i.e., no thoughts/general thoughts about death and active SI) provided the most homogeneous groups of people, with control participants more commonly on the lower severity and suicide ideators predominately reporting thoughts in the active SI range. The middle levels of the severity scale—inclusive of depressive ideation and passive SI—captured a more heterogeneous mixture of controls and suicidal individuals. We hypothesized that these kinds of thoughts may be more ambiguous, whereas general death thoughts and active SI cleanly separate people who are truly nonsuicidal from those who are truly suicidal. As an analogy, just as thoughts about suicide can range in severity, from general thoughts about death to active desire to kill oneself, eyes can range in color,

from dark brown to blue. If a person has hazel colored eyes, depending on factors such as the lighting in a given room, their eye color may be sometimes described as brown and other times described as green. However, if a person has blue eyes, regardless of factors such as lighting, it seems less likely that their eyes would ever be described as brown (and vice versa). Similarly, "suicidal thoughts" that fall in the middle range of the Graded Scale are ambiguous and, depending on various factors, participants may have more difficulty classifying them as the clear and consistent absence or presence of SI. However, analyses did not support this hypothesis; people whose "suicidal thoughts" were ambiguous were no more likely to be inconsistent in their reports of SI across modes of assessment as compared to individuals with more clearly absent or present SI.

The present findings rule out a number of potential explanations for inconsistent reporting of SI and SA across assessment modes, including purposeful inaccuracy, random responding, and differing experiences/conceptualizations of SI. Future research is needed to explore additional possible explanations for why inconsistencies exist. For example, even if people are motivated to respond accurately and sufficiently understand the definition of SI, they may be limited in their ability to accurately match the definition to their own past experiences due to the influence of cognitive or affective biases. Based on prior research showing that individuals differ in their recall of past events depending on their current mood at the time of retrieval (Laird et al., 1989), it may be that individuals construe their past STBs differently based on current mood, leading to inconsistencies in reporting SI. Along these same lines, it may be that people have simply forgotten or subconsciously suppressed memories of ever having been suicidal, and the memories may resurface later only after repeated questioning.

The most impactful takeaway from these findings is that the specific assessment format and items used for determining group membership matter. Talking with people by phone versus in a laboratory setting results in differing reports of SI and SA. Even more granular than phone versus laboratory settings, the method of questioning (e.g., interview vs. questionnaire) and context of questioning (e.g., part of phone screen/laboratory study linking responses to people vs. confidential/anonymous exit survey not linked to the person, study eligibility, and/or risk intervention) affect answers. Based on the inconsistent reporting of SI and SA, we recommend using multiple items (rather than a single item) and evaluating consistency of responses across items to assess these constructs. This is particularly critical when SI and/or SA represent study eligibility criteria. Based on these data, we also recommend the use of a Graded Scale, such as the Graded Scale of Suicide and Related Thoughts, to define and capture a range of phenomena from general thoughts about death to depressive thoughts to passive SI to active SI. Specifically, endorsement of active SI questions from the Graded Scale (i.e., "Maybe I should kill myself," "I should kill myself") appear to reliably distinguish a "suicidal" group from those who were nonsuicidal or ambiguously suicidal. Both the lack of endorsement of death/depression/suicide-related items and endorsement of general thoughts about death items (i.e., "What happens to people when they die?", "What will it be like when I die?") appear to have reliably distinguished nonsuicidal controls from those who were suicidal or ambiguously suicidal. If the goal of researchers is to obtain clearly distinguishable groups of people with SI versus people without SI, we also propose excluding people who fall in the middle range with the highest endorsement of depressive thoughts and/or passive SI (i.e., "I wish I could disappear or not exist," "I wish I was never born," "My life is not worth living," "I wish I was dead [ex: go to sleep and not wake up again]"). Notably, excluding this middle group will allow researchers to more reliably obtain active suicide ideators; suicide researchers frequently choose to specifically recruit for participants with active SI (i.e., desire to die with at least some interest in taking action to purposefully end one's own life) given this may be more directly related to suicidal behaviors than passive SI (i.e., some desire to die with no interest in purposefully ending one's own life). Using the Graded Scale allows one to gain a better sense of where possible miscommunication of terms could exist between researchers and participants. Finally, for research studies, these data highlight the importance of using a confidential exit survey that is not linked to study eligibility, stigma, or treatment services.

There are several important limitations to keep in mind when interpreting the results from these two samples. First, all modes of assessment used here were reliant on participant self-report and therefore are potentially limited by response bias, introspective ability, participant understanding, and participant insight. Second, and related, there is no objective measure of SI or SA, meaning that we cannot know for certain which participants have or have not experienced SI and/or SA. Future studies should work to incorporate a ground truth (e.g., behavioral measure, validated documentation of SI and/or SA such as via medical record) to which to compare self-reported SI and SA. Third, the precise wording used across modes of SI and SA assessment varied slightly, which may explain some of the differences observed in endorsement rates (see Supplemental Table 1, for exact wording). Wording differences were minor, though we cannot rule out the possibility that participants were understanding questions about similar content areas to mean different things. Future studies of consistent responding should seek to standardize wording from phone screen to various modes of assessment used during laboratory visit. Fourth, there was a time lag between the phone screen interview and the laboratory visit during which a participant's answers to SI and SA questions may have genuinely changed. For example, someone who, based on phone screen data, was recruited as a control participant (with no lifetime SI) could have experienced their first suicidal thought in the time lag between phone screen interview and laboratory visit. This is not likely, given this study focused on the *lifetime* presence of SI and SA and the fact that inconsistency was not associated with length of time lag, and does not explain the discrepancies present between laboratory visit data, which were all collected on the same day. Fifth, some items were double- or triple-barreled with multiple questions embedded within one item (e.g., "I read every question, answered accurately, and for the most part, found it easy to fill out") and it is not clear if respondents were saying "yes" to all (which is what this study assumes) or only some parts of the item. Future studies should ask a single question per item for increased clarity. Finally, the most anonymous measure, the confidential exit survey (on which participants reported the highest rates of SI and SA), was consistently administered after the self-report and interview measures, which was necessary to serve the two-fold purpose of allowing participants to confidentially address inclusion/exclusion criteria and to provide feedback about their laboratory visit (i.e., we could not randomize the order because the confidential exit survey asked about a participant's experience in the study). It is possible that being asked about SI and SA repeatedly throughout the study may have jogged participant

memory or impacted participant understanding of the questions over time—thus influencing participant responses. This means that instead of the confidential nature of this survey increasing reports of SI and SA, rates of reporting may have increased due to priming. In the future, these two outcomes could be separated into different surveys, and the in-laboratory measures assessing SI/SA could be presented in a randomized order.

These limitations notwithstanding, this research reveals inconsistencies in reporting SI and SA that exist across different modes of assessment, in spite of taking a relatively more lenient approach to deeming participants as consistent. Moreover, these results suggest that inconsistencies are not due to purposeful inaccuracy, random responding, or differences in participant experience/conceptualization of SI, and that people may have difficulty answering some questions. These findings are important for researchers and clinicians alike. Future efforts should focus on developing clearer guidelines and practices for the classification of people as suicidal or nonsuicidal by using multiple items to capture a single construct, using a Graded Scale to capture a broad spectrum of thoughts and behaviors, and taking into account consistency of responding across such items. Through such efforts, we may be able to improve not only the assessment of suicide outcomes, but ultimately the prediction and prevention of suicide as well.

References

Anestis, M. D., & Green, B. A. (2015). The impact of varying levels of confidentiality on disclosure of suicidal thoughts in a sample of United States National Guard personnel. *Journal of Clinical Psychology*, *71*(10), 1023–1030. https://doi.org/10.1002/jclp.22198

Beck, A. T., Kovacs, M., & Weissman, A. (1979). Assessment of suicidal intention: The Scale for Suicide Ideation. *Journal of Consulting and Clinical Psychology*, *47*(2), 343–352. https://doi.org/10.1037/0022-006X.47.2.343

Beck, A. T., & Steer, R. A. (1991). *Manual for the Beck Scale for Suicide Ideation*. The Psychological Corporation.

Beck, A. T., Steer, R. A., & Ranieri, W. F. (1988). Scale for Suicide Ideation: Psychometric properties of a self-report version. *Journal of Clinical Psychology*, *44*, 499–505. https://doi.org/10.1002/1097-4679(198807)44:4<499:AID-JCLP2270440404>3.0.CO;2-6

Callahan, C. M., Unverzagt, F. W., Hui, S. L., Perkins, A. J., & Hendrie, H. C. (2002). Six-item screener to identify cognitive impairment among potential subjects for clinical research. *Medical Care*, *40*(9), 771–781. https://doi.org/10.1097/00005650-200209000-00007

Centers for Disease Control and Prevention. (2017). Data & Statistics Fatal Injury Report for 2016.

Eskin, M., Schild, A., Oncu, B., Stieger, S., & Voracek, M. (2015). A cross-cultural investigation of suicidal disclosures and attitudes in Austrian and Turkish university students. *Death Studies*, *39*(10), 584–591. https://doi.org/10.1080/07481187.2015.1037971

Franklin, J. C., Ribeiro, J. D., Fox, K. R., Bentley, K. H., Kleiman, E. M., Huang, X., Musacchio, K. M., Jaroszewski, A. C., Chang, B. P., & Nock, M. K. (2017). Risk factors for suicidal thoughts and behaviors: A meta-analysis of 50 years of research. *Psychological Bulletin*, *143*(2), 187–232. https://doi.org/10.1037/bul0000084

Greist, J. H., Laughren, T. P., Gustafson, D. H., Stauss, F. F., Rowse, G. L., & Chiles, J. A. (1973). A computer interview for suicide-risk prediction. *The American Journal of Psychiatry*, *130*(12), 1327–1332. https://doi.org/10.1176/ajp.130.12.1327

Hom, M. A., Stanley, I. H., Podlogar, M. C., & Joiner, T. E. Jr., (2017). "Are you having thoughts of suicide?" Examining experiences with disclosing and denying suicidal ideation. *Journal of Clinical Psychology*, *73*, 1382–1392. https://doi.org/10.1002/jclp.22440

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*(2), 294–298. https://doi.org/10.1002/1097-4679(199403)50:2<294:AID-JCLP2270500224>3.0.CO;2-R

Laird, J. D., Cuniff, M., Sheehan, K., Shulman, D., & Strum, G. (1989). Emotion specific effects of facial expressions on memory for life events. *Journal of Social Behavior & Personality*, *4*(2), 87–98.

Millner, A. J., Lee, M. D., & Nock, M. K. (2015). Single-item measurement of suicidal behaviors: Validity and consequences of misclassification. *PLOS ONE*, *10*(10), Article e0141606. https://doi.org/10.1371/journal.pone.0141606

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*(3), 309–317. https://doi.org/10.1037/1040-3590.19.3.309

Nock, M. K., Prinstein, M. J., & Sterba, S. K. (2009). Revealing the form and function of self-injurious thoughts and behaviors: A real-time ecological assessment study among adolescents and young adults. *Journal of Abnormal Psychology*, *118*(4), 816–827. https://doi.org/10.1037/a0016948

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., & Dunbar, G. C. (1998). The M.I.N.I. International Neuropsychiatric Interview (M.I.N.I.): The Development and Validation of a Structured Diagnostic Psychiatric Interview for DSM-IV and ICD-10. *Journal of Clinical Psychiatry*, *59*(suppl 20), 22–33.

Warner, C. H., Appenzeller, G. N., Grieger, T., Belenkiy, S., Breitbach, J., Parker, J., Warner, C. M., &Hoge, C. (2011). Importance of anonymity to encourage honest reporting in mental health screening after combat deployment. *Archives of General Psychiatry*, *68*(10), 1065–1071. https://doi.org/10.1001/archgenpsychiatry.2011.112

Received January 23, 2020
Revision received August 2, 2020
Accepted October 6, 2020 ■