



How we ask matters: The impact of question wording in single-item measurement of suicidal thoughts and behaviors

Brooke A. Ammerman^{a,*}, Taylor A. Burke^b, Ross Jacobucci^a, Kenneth McClure^a

^a University of Notre Dame, Department of Psychology, Notre Dame, IN, USA

^b Alpert Medical School of Brown University, Department of Psychiatry and Human Behavior, Providence, RI, USA

A B S T R A C T

The present study aimed to extend prior literature on single-item assessment by examining response consistency (1) between several commonly used single-item assessments of suicidal ideation, planning, and attempts, and (2) across three timeframes (past month, past year, and lifetime) commonly employed in the literature. Participants ($N = 613$) were recruited from an online community, Amazon Mechanical Turk (mTurk). Participants were administered three sets of four distinct single-items assessing suicidal ideation, suicidal planning, and suicide attempt history, respectively. Items were drawn from well-known large-scale studies (e.g., National Comorbidity Survey; World Health Organization Mental Health Survey Initiative, Youth Risk Behavior Survey) and commonly used suicide risk assessments (i.e., Self-Injurious Thoughts and Behaviors Interview). Through examinations of intraclass correlations and confirmatory factor analyses, findings suggested mixed response agreement across most outcomes and timeframes. Response inconsistency among items assessing suicidal ideation and among items assessing suicidal planning were partly attributed to minor, yet important, language differences. Given findings that even minor language changes in suicidal ideation and planning items may inflate or restrict prevalence estimates in a meaningful way, it will be important for researchers and clinicians alike to pay close attention to the wording of single items in designing research studies, interpreting findings, and assessing patient risk.

Suicide is a leading cause of death worldwide (World Health Organization [WHO], 2020) and research examining suicidal thoughts and behaviors (STBs) has proliferated in recent decades (Franklin et al. 2017). As a significant proportion of studies in the field of suicide utilize single items to assess for STBs (e.g., “Have you ever attempted suicide?”) and clinical practices oft rely on similar assessments (e.g., Patient Health Questionnaire-9; Kroenke et al. 2001) for STB risk detection, increased attention has been focused on understanding how method of measurement may impact STB findings. Despite the far-reaching implications of single-item STB assessments, ranging from an accurate depiction of prevalence rates to identifying key STB risk factors, there has been limited empirical investigation aimed at better understanding factors that impact measurement accuracy.

Within this small body of literature, studies have highlighted the potential misclassification associated with utilizing single-item measurement (Ploderl et al., 2011; Millner et al. 2015; Nock and Kessler, 2006). Initial research in this area, utilizing a combination of single-items and follow-up probing questions, found that 25% of individuals endorsing a lifetime suicide attempt were false positives and 10% were false negatives (Ploderl et al., 2011). More recently, this work was extended to examine the validity of single-items through coding

narrative descriptions of suicidal actions and employing numerous follow-up questions with more precise language regarding lifetime suicide attempts and suicidal ideation. A false positive rate of 11.3% was found for lifetime suicide attempts and 8.8% for lifetime suicidal ideation (Millner et al. 2015). Further underscoring the potential impact of such misclassification, authors used simulation analysis to conclude that the severity of misclassification observed in their study would significantly reduce power and decrease the accuracy of statistical tests predicting their occurrence (Millner et al. 2015). In another study, Hom et al. (2019) examined consistency in responding to suicide attempt assessments using both single-item self-report measures and clinician-rated interview measures, finding that 35.4% of United States service members exhibited inconsistent suicide attempt reporting across measure formats.

One key factor that has been highlighted as contributing to the imprecise measurement and misclassification of STBs is the inconsistency in definitions of STB outcomes (Hom et al. 2019; Millner et al. 2015, 2017), which has been a long-standing concern in the field more broadly (e.g., Silverman et al., 2007). For example, in considering suicidal ideation, definitions can range from including wishes to die and thoughts of killing oneself to having intent to kill oneself or thoughts of a

* Corresponding author.

E-mail addresses: bammerm1@nd.edu (B.A. Ammerman), taylor_burke@brown.edu (T.A. Burke), rjacobuc@nd.edu (R. Jacobucci), kmclur5@nd.edu (K. McClure).

<https://doi.org/10.1016/j.ypmed.2021.106472>

Received 25 October 2020; Received in revised form 18 February 2021; Accepted 20 February 2021

Available online 16 September 2021

0091-7435/© 2021 Elsevier Inc. All rights reserved.

method by which to kill oneself (i.e., [Brown et al. 2008](#); [Posner et al. 2014](#)). The definition adopted by the authors of a single study or risk assessment scale has direct implications on the found associations, which is underscored by [Berman and Silverman \(2014\)](#) who note, “Determining the presence of suicidal ideation depends upon how the question is asked, [and] how the respondent understands the question...” (pg. 34). Not surprisingly, then, this is further exacerbated when utilizing single-item assessments, which typically focus on one or only a subset of the many components of suicidal ideation (or suicidal behavior) measured.

Nevertheless, there is some evidence of the validity of a single-item approach to measuring STBs (e.g., [Desseilles et al. 2012](#)), supporting regular use in STB research and public health screening initiatives (e.g., Youth Risk Behavior Survey; World Health Organization). Thus, given the widespread use, and the implications for our understanding of STB prevalence estimates and risk factors, it is important to further understand the ramifications of using commonly employed single-items. Further, prior research has largely examined single-item measurement for lifetime STB assessment (e.g., [Hom et al. 2019](#); [Millner et al. 2015](#)); however, given the increased focus on near-term prediction of STBs ([National Institute of Mental Health 2017](#)) and continued concerns about retrospective recall bias in self-report measures of lifetime events (e.g., [Schmier and Halpern 2004](#)), single-item measurements across varying timespans (i.e., past month vs. past year vs. lifetime) need to be explored.

The current study aimed to examine responses to multiple single-item assessments of suicidal ideation, planning, and attempts drawn from large-scale studies across three commonly assessed timeframes: past month, past year, and lifetime. This study aimed to extend prior literature on single-item assessment by examining response consistency between (1) several commonly used single-item assessments of suicidal ideation, planning, and attempts, and (2) across three distinct timeframes (past month, past year, and lifetime) commonly employed in the literature. Findings have the potential to improve STB assessment and to shed light on potential low-burden approaches to refine single-item measurement.

1. Methods

1.1. Participants and procedures

Participants were recruited from an online community, Amazon Mechanical Turk (mTurk). Potential participants on mTurk are able to browse available tasks (i.e., marketing campaigns, focus groups, research studies) and self-select to participate in those of interest. The current sample was recruited as part of a larger online study described as a research study on well-being and social connection. Participants were presented with blocks of questions, where each block represented all items per STB outcome within each timeframe (e.g., all past-month suicidal ideation questions); randomization both within and between the nine blocks was implemented. The study was approved by the first-author’s Institutional Review Board for research ethics.

The final sample consisted of 613 participants. The mean age of participants was 36.37 years old ($SD = 11.03$, Range = 19–74). Overall, 56.6% of participants ($n = 347$) identified as a man, 43.1% ($n = 264$) identified as a woman, 0.3% ($n = 2$) preferred not to answer. The majority of the sample identified as White (75.2%, $n = 461$), followed by African American/Black (12.7%, $n = 78$), Asian (8.3%, $n = 51$), more than one race (2.0%, $n = 12$), American Indian or Alaska Native (0.5%, $n = 3$), or other (0.7%, $n = 4$); a subset preferred not to answer (0.5%, $n = 3$). The majority identified as non-Hispanic/Latinx (83.7%, $n = 513$), with 14.0% ($n = 86$) identifying as Hispanic/Latinx and 2.3% ($n = 14$) preferring not to answer.

1.2. Measures

1.2.1. Suicidal ideation

Four single-items assessing suicidal ideation (SI) were drawn from large-scale studies including STB prevalence assessments, well-utilized STB measures, or relevant empirical studies. Included items were: SI1: “Did you have thoughts of killing yourself?” (drawn from the Army Study to Assess Risk and Resilience in Servicemembers [Army STARRS; [Ursano et al., 2020](#)]; Self-Injurious Thoughts and Behaviors Interview [[Nock et al., 2007](#)]); SI2: “Did you think about committing suicide?” (drawn from National Comorbidity Survey [NCS; [Kessler, 2018](#)]; National Comorbidity Survey-Replication [NCS-R; [Kessler, 2015](#)]); SI3: “Did you seriously consider attempting suicide?” (drawn from the National Longitudinal Study of Adolescent to Adult Health [Add Health; [Harris & Udry, 2018](#)]; Youth Risk Behavior Survey [YRBS; [Center for Disease Control and Prevention \[CDC\], n.d.](#)]); SI4: “Have you seriously thought about committing suicide?” (drawn from WHO Mental Health Survey Initiative; [Kessler et al. 2006](#)). All items were asked with regard to participant experiences in the past month, the past year (i.e., 12-months), and across the lifetime.

1.2.2. Suicidal planning

Four single-items assessing suicidal planning (SP) were drawn from large-scale studies including STB prevalence assessments, well-utilized STB measures, or relevant empirical studies. Included items were SP1: “Did you make a plan about how you would attempt suicide?” (drawn from Add Health [[Harris & Udry, 2018](#)]; YRBS [[CDC, n.d.](#)]); SP2: “Have you made a plan for committing suicide?” (drawn from NCS [[Kessler, 2018](#)]; NCS-R [[Kessler & Merikangas, 2004](#)]; WW WHO Mental Health Survey Initiative [[Kessler et al., 2009](#)]); SP3: “Have you made a plan to kill yourself?” (drawn from [Millner et al. 2015](#); SITBI [Self-Injurious Thoughts and Behaviors Interview; SITBI; [Nock et al., 2007](#)]); SP4: “Did you think about how you might kill yourself (e.g., taking pills, shooting yourself) or work out a plan of how to kill yourself?” (drawn from Army STARRS [[Ursano et al., 2020](#)]). All items were asked with regard to participant experiences in the past month, the past year (i.e., 12-months), and across the lifetime.

1.2.3. Suicide attempt

Four single-items assessing suicide attempts (SA) were drawn from large-scale studies including STB prevalence assessments, well-utilized STB measures, or relevant empirical studies. Included items were SA1: “Have you attempted to kill yourself?” (drawn from [Millner et al. 2015](#)); SA2: “Have you attempted suicide?” (drawn from NCS [[Kessler, 2018](#)]; NCS-R [[Kessler & Merikangas, 2004](#)]; WHO Mental Health Survey Initiative [[Kessler et al., 2009](#)]); SA3: “Did you make a suicide attempt (i.e., purposefully hurt yourself with at least some intention to die)?” (drawn from Army STARRS [[Ursano et al., 2020](#)]); SA4: “Have you actually attempted suicide?” (drawn from Add Health [[Harris & Udry, 2018](#)]; YRBS [[CDC, n.d.](#)]). All items were asked with regard to participant experiences in the past month, the past year (i.e., 12-months), and across the lifetime.

1.3. Data analysis

To ensure valid and attentive responding, inclusion in the current analysis required: (a) passing an online bot-detection test (i.e., ReCaptcha), (b) passing at least 50% of the 12 included attention check items (i.e., “Please answer ‘3’ on this item”), (c) passing the time completion attention check (i.e., completing the survey within three standard deviations of the mean completion time), and (d) providing responses on the study’s variables of interest. Descriptive analyses (i.e., frequencies) were utilized to present item and item pattern endorsement.

To provide an initial examination of item consistency, we first computed intraclass correlations (ICCs) by outcome within each

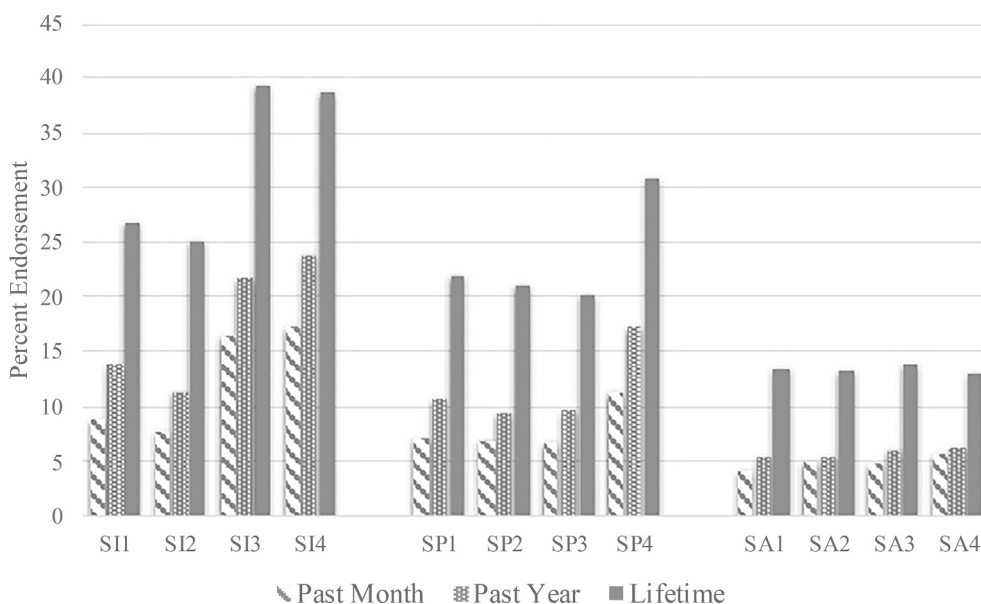


Fig. 1. Single item endorsement by outcome and timeframe.

Note: S11 = Seriously thought about committing suicide; S12 = Seriously considered attempting suicide; S13 = Thought about committing suicide; S14 = Thoughts of killing yourself; SP1 = Made a plan about how you would attempt suicide; SP2 = made a plan for committing suicide; SP3 = Made a plan to kill yourself; SP4 = Thought about how you might kill yourself (e.g., taking pills, shooting yourself) or work out a plan of how to kill yourself; SA1 = Attempted to kill yourself; SA2 = Attempted suicide; SA3 = Made a suicide attempt (i.e., purposefully hurt yourself with at least some intention to die); SA4 = Actually attempted suicide;

timeframe to quantify the agreement in item endorsement for suicidal ideation, suicidal planning, and suicide attempt within the three timeframes, treating the four individual items as fixed, separate raters. ICCs were computed to reflect estimates of single item reliability using two-way fixed calculations within the psych package (Revelle 2019) in R (R Core Team 2020). ICCs less than 0.50 are considered poor, less than 0.75 considered moderate, less than 0.90 considered good, and greater than 0.90 considered excellent (Koo and Li 2016).

Next, to improve our understanding of item consistency, we utilized a series of confirmatory factor analyses (CFA) to evaluate whether each set of four items equally reflect each construct (i.e., ideation, planning, attempts) and are endorsed at similar rates within each timeframe. Within this framework, we fit a sequence of three models: 1) the configural model contained three factors, one for each timeframe; 2) the metric invariance model imposed the equality of factor loadings within each timeframe; 3) the strong invariance model imposed both the equality of factor loadings and item thresholds within each timeframe. The metric invariance model assesses whether each of the four items is equally representative of the same construct, while the strong invariance model assesses whether the endorsement rates per item within each timeframe are similar. Given the interchangeable use of the included items throughout the literature, items were expected to demonstrate metric and strong invariance preserving overall model fit relative to the configural model. For models failing to demonstrate invariance, modified item constraints were considered based on item endorsement frequencies. Similar approaches have been used for multiple group measurement invariance (Svetina et al. 2020) and longitudinal measurement invariance (Liu et al. 2017) for categorical data. This analysis was conducted using weighted least squares to account for each item being dichotomous in the lavaan package (Rosseel 2012) in the R statistical environment. Likelihood ratio tests and fit indices were examined to detect violations of invariance. While ICCs impose equal intercepts and residual variances to assess the ratio of between-person variability divided by overall variability (between-person differences and within-person variance), the sequence of CFA models relax these assumptions to detect the cause of low ICC values.

Finally, the study also conducted a sensitivity analysis to examine if consistency is contingent upon attentive responding. The aforementioned models were further assessed to examine measurement invariance of the suicidal ideation, suicidal planning, and suicide attempt models based on two different participant attentive responding inclusion criteria in a multiple group framework. Specifically, we compared

CFA measurement models for the sample of participants who met the current inclusion criteria but did not pass 100% of attention check items (i.e., passed 50–99% of attention check items; $n = 238$, 38.8%) to the sample of participants who successfully passed all attention check items (i.e., passed 100% of attention check items; $n = 375$, 61.2%). Invariance across attentive responding groups was assessed for the configural, metric, and strong invariance models¹.

2. Results

2.1. Suicidal ideation

Across the four single items, 21.4% ($n = 131$) of participants reported past month suicidal ideation on at least one item, 29.7% ($n = 182$) reported past year suicidal ideation on at least one item, and 44.4% ($n = 272$) reported lifetime suicidal ideation on at least one item. See Fig. 1 for suicidal ideation endorsement by item and Fig. 2 for endorsement patterns across items.

ICCs for suicidal ideation items were mixed. Among participants endorsing presence of suicidal ideation on at least one item, ICCs were poor across past month, past year, and lifetime timeframes (ICC = 0.10, ICC = 0.12, ICC = 0.29). Ont WWhen all participants were included (including participants who did not endorse any item) moderate agreement (ICC = 0.56, ICC = 0.55, ICC = 0.71) was observed. Complete results are shown in Table 1² and Table 2.

The suicidal ideation configural model demonstrated adequate fit (CFI = 0.99, RMSEA = 0.07); however, items failed to demonstrate both metric invariance ($\chi^2(9) = 41.82$, $p < 0.001$) and strong invariance ($\chi^2(9) = 272.73$, $p < 0.001$). Based on item endorsement rates, a modified constraint model was considered, restricting parameters for S11 and S12 as well as S13 and S14 to be equal pairwise. Pairwise constraint of factor loadings within timeframes (i.e. $\lambda_{S11} = \lambda_{S12}$, $\lambda_{S13} = \lambda_{S14}$), however, did not demonstrate significantly worse fit ($\chi^2(6) = 8.06$, $p = 0.23$) relative to the configural model. Pairwise constraint of thresholds reduced model fit relative to the pairwise constrained loading model ($\chi^2(6) = 14.51$, $p = 0.02$). The decrement in fit was much less than complete threshold invariance, suggesting potential discrepancies in thresholds within item pairs; however, no clear result emerges. Pairwise constraint metric invariance model estimates are shown in Table 3.

Pattern					Past Month			Past Year			Lifetime		
	SI1	SI2	SI3	SI4	<i>n</i>	Total % (<i>n</i> = 613)	Within % (<i>n</i> = 131)	<i>n</i>	Total % (<i>n</i> = 613)	Within % (<i>n</i> = 182)	<i>n</i>	Total % (<i>n</i> = 613)	Within % (<i>n</i> = 272)
1					26	4.2%	19.8%	40	6.5%	22.0%	128	20.9%	47.1%
2					4	0.7%	3.1%	2	0.3%	1.1%	5	0.8%	1.8%
3					5	0.8%	3.8%	3	0.5%	1.6%	6	1.0%	2.2%
4					11	1.8%	8.4%	20	3.3%	11.0%	20	3.3%	7.4%
5					2	0.3%	1.5%	6	1.0%	3.3%	2	0.3%	0.7%
6					1	0.2%	0.8%	5	0.8%	2.7%	1	0.2%	0.4%
7					1	0.2%	0.8%	4	0.7%	2.2%	0	0.0%	0.0%
8					6	1.0%	4.6%	2	0.3%	1.1%	3	0.5%	1.1%
9					1	0.2%	0.8%	5	0.8%	2.7%	4	0.7%	1.5%
10					1	0.2%	0.8%	3	0.5%	1.6%	3	0.5%	1.1%
11					45	7.3%	34.4%	50	8.2%	27.5%	61	10.0%	22.4%
12					3	0.5%	2.3%	6	1.0%	3.3%	5	0.8%	1.8%
13					8	1.3%	6.1%	5	0.8%	2.7%	4	0.7%	1.5%
14					7	1.1%	5.3%	9	1.5%	4.9%	16	2.6%	5.9%
15					10	1.6%	7.6%	22	3.6%	12.1%	14	2.3%	5.1%
16					482	78.6%	--	431	70.3%	--	359	58.6%	--

Fig. 2. Suicidal ideation endorsement patterns by item.

Note: SI1 = Seriously thought about committing suicide; SI2 = Seriously considered attempting suicide; SI3 = Thought about committing suicide; SI4 = Thoughts of killing yourself; Grey boxes represent presence (i.e., “yes” to the specific item) and white boxed represent the absence (i.e., “no” to the specific item) of each suicidal ideation experience; Within sample descriptive statistics are based on the sample reporting the presence of the outcome of interest on at least one item within the specified timeframe.

Table 1
Intraclass correlations and 95% Confidence Intervals for all participants.

	Past month	Past year	Lifetime
Suicidal ideation	0.56 [0.52, 0.59]	0.55 [0.52, 0.59]	0.71 [0.68, 0.73]
Suicidal planning	0.64 [0.61, 0.67]	0.64 [0.61, 0.67]	0.72 [0.70, 0.75]
Suicide attempts	0.58 [0.55, 0.61]	0.58 [0.54, 0.61]	0.82 [0.80, 0.84]

Note: *n* = 613.

2.2. Suicidal planning

Across the four single items, 14.0% (*n* = 86) of participants reported past month suicidal planning on at least one item, 20.7% (*n* = 127) reported past year suicidal planning on at least one item, and 35.1% (*n* = 215) reported lifetime suicidal planning on at least one item. See Fig. 1 for suicidal planning endorsement by item and Fig. 3 for endorsement

Table 2
Intraclass correlations and 95% Confidence Intervals for participants endorsing at least one item of the corresponding construct.

	Past month	Past year	Lifetime
Suicidal ideation	0.10 [0.04, 0.17]	0.12 [0.06, 0.18]	0.29 [0.23, 0.34]
Suicidal planning	0.27 [0.18, 0.37]	0.32 [0.24, 0.40]	0.38 [0.32, 0.44]
Suicide attempts	0.12 [0.02, 0.25]	0.10 [0.02, 0.22]	0.32 [0.24, 0.41]

Note: *n*_{SImonth} = 131, *n*_{SIyear} = 182, *n*_{SIlife} = 272, *n*_{SPmonth} = 86, *n*_{SPyear} = 127, *n*_{SPlife} = 215, *n*_{SAmonth} = 54, *n*_{SAyear} = 63, *n*_{SAlife} = 106.

patterns across items.

ICCs for suicidal planning items demonstrated a similar pattern of agreement as the suicidal ideation items (see Tables 1 and 2). Poor agreement (ICC = 0.27, ICC = 0.32, ICC = 0.38) was observed for those endorsing suicidal planning on at least one item and moderate agreement (ICC = 0.64, ICC = 0.64, ICC = 0.72) was observed for the full

Table 3
Factor loadings, correlations, and threshold estimates for suicidal ideation pairwise constraint metric invariance model.

	Loading (SE)	Threshold (SE)	Variance
Month			
SI1	0.90 (0.02)	1.53 (0.07)	0.18
SI2	0.90 (0.02)	1.42 (0.07)	0.18
SI3	0.98 (0.01)	0.98 (0.06)	0.04
SI4	0.98 (0.01)	0.94 (0.06)	0.04
Year			
SI1	0.90 (0.02)	1.09 (0.06)	0.18
SI2	0.90 (0.02)	1.21 (0.07)	0.18
SI3	0.96 (0.01)	0.78 (0.06)	0.08
SI4	0.96 (0.01)	0.71 (0.06)	0.08
Lifetime			
SI1	0.97 (0.01)	0.62 (0.05)	0.06
SI2	0.97 (0.01)	0.68 (0.06)	0.06
SI3	0.98 (0.01)	0.28 (0.05)	0.05
SI4	0.98 (0.01)	0.28 (0.05)	0.05
Correlations			
	Month	Year	Lifetime
Month	1.00	0.946 (0.02)	0.785 (0.03)
Year	–	1.00	0.830 (0.02)
Lifetime	–	–	1.00

Note: SI1 = Seriously thought about committing suicide; SI2 = Seriously considered attempting suicide; SI3 = Thought about committing suicide; SI4 = Thoughts of killing yourself; factor variances set to 1 and all intercepts to 0 for model identification.

sample.

The suicidal planning configural model demonstrated good fit (CFI = 1.00, RMSEA = 0.02); however, items failed to demonstrate metric invariance ($\chi^2(9) = 40.09, p < 0.001$) as well as strong invariance ($\chi^2(9) = 160.37, p < 0.001$). Based on item endorsement, a modified constraint model was considered, allowing SP4 to vary but restricting the remaining three items to be equal. The model constraining loadings for SP1, SP2, SP3 to be equal but SP4 to vary (i.e. $\lambda_{SP1} = \lambda_{SP2} = \lambda_{SP3}, \lambda_{SI4}$) did not significantly reduce model fit ($\chi^2(6) = 8.16, p = 0.23$); moreover, imposing the same pattern of constraints on thresholds resulting in the same conclusion ($\chi^2(6) = 4.66, p = 0.59$). Table 4 shows modified constraint strong invariance model estimates.

2.3. Suicide attempts

Across the four single items, 8.8% ($n = 54$) of participants reported a pastmonth suicide attempt on at least one item, 10.3% ($n = 63$) reported a past year suicide attempt on at least one item, and 17.3% ($n = 106$) reported a lifetime suicide attempt on at least one item. See Fig. 1 for suicide attempt endorsement by item and Fig. 4 for endorsement patterns across items.

ICC patterns for suicide attempt items resembled those of suicidal ideation and planning showing poor agreement for participants endorsing a suicide attempt on at least one item (ICC = 0.12, ICC = 0.10, ICC = 0.32), however demonstrated moderate to good agreement (ICC = 0.58, ICC = 0.58, ICC = 0.82) for the full sample (see Tables 1 and 2).

The suicide attempt configural model fit well (CFI = 1.00, RMSEA < 0.01); moreover, items demonstrated metric invariance ($\chi^2(9) = 4.08, p = 0.91$) and strong invariance ($\chi^2(9) = 6.76, p = 0.66$). Consequently, no subsequent constraint patterns were explored; parameter estimates are shown in Table 5.

2.4. Attentive responding

To assess the potential influence of participant attentive responding inclusion criteria on response consistency, metric and strong measurement invariance was examined across two attentive responding inclusion groups (i.e., 50–99% vs. 100% of attention checks passed) for each of the above CFA models. The suicidal ideation (CFI = 1.00, RMSEA =

0.06) and suicide attempt (CFI = 1.00, RMSEA < 0.01) configural group models demonstrated good fit. The suicidal planning group configural model did not converge; however, the modified constraint metric and strong invariance models fit well (both CFI = 0.99; both RMSEA = 0.03). For suicidal ideation, all models demonstrated both metric and strong invariance across groups. The suicidal planning modified constraint model demonstrated metric and strong invariance across groups. Group metric invariance was observed for the suicidal attempt configural model but strong invariance was not. Parameter estimates are shown in Table 6 and complete attention check invariance results are shown in Table 7.

3. Discussion

The present study aimed to examine endorsement consistency and the measurement structure of single-item measures of suicidal ideation, planning, and attempts across three timeframes. As demonstrated via ICCs, responses on items across nearly all outcomes and timeframes reached only moderate agreement, with the exception of lifetime suicide attempt items, which reached good agreement. When only respondents endorsing one or more items were examined, this dropped to poor agreement for all items. Further, CFA analyses highlighted that these single item measures assessing suicidal ideation and suicidal planning demonstrate differential measurement properties due to minor, yet important, language differences.

Item response patterns (Figs. 2–4) present inconsistent response patterns (i.e., differences between those positively endorsing all items within a specific outcome and timeframe and those positively endorsing only a subset items within a specific outcome and timeframe), ranging from as high as 81.2% (i.e., past month suicidal ideation) to 42.5% (i.e., lifetime suicide attempt). Our findings are in line with Hom et al. (2019) who found that 35.4% of United States service members exhibit inconsistent responding on measures of suicide attempt. Similarly, also in line with prior research (Hom et al. 2019), between 14 and 15 unique patterns across the three outcomes and three timeframes were found, underscoring that inconsistency is not predominantly being driven by one specific faulty single-item measure. Further, it is clear that both false negatives and false positives are likely prevalent across all of the included items, at least to some extent.

Confirmatory factor analysis results suggest that variations in language employed to measure suicidal ideation significantly impacted item endorsement and response consistency. That is, the suicidal ideation items that describe suicidal thoughts as “serious” (SI1 and SI2) were significantly less likely to be endorsed than the suicidal ideation items that did not include a description of ideation severity (SI3 and SI4). Findings indicate that response consistency across the suicidal ideation items would have significantly increased if consistency was examined separately across items that did versus did not use the term “serious.” The impact of the inclusion of the word “serious” is substantial, despite the fact that all four items assess active suicidal ideation (i.e., desire to kill oneself). Indeed, items assessing passive suicidal ideation (i.e., desire to be dead) were explicitly excluded from the current study to reduce variability in the suicidal ideation construct. It is possible that the wording indicates greater suicidal ideation severity, potentially assessing the presence of associated intent to act on these thoughts. However, given the nature of single-item assessments and that “serious” is not defined more explicitly in these items, it is unclear how participants may have interpreted, and thus responded, to this descriptor. Nonetheless, it appears the inclusion of this term may be assessing a more severe, and less prevalent, form of suicidal ideation.

With regard to the suicidal planning items, three of the items (SP1, SP2, SP3) had similar endorsement rates and demonstrated greater response consistency, likely due to their similar presentation of the concept of making a plan to kill oneself. However, these three items all significantly differed from the fourth item (SP4), which conflated thinking about a method and making a plan. This conflation of method

Pattern					Past Month			Past Year			Lifetime		
	SP1	SP2	SP3	SP4	n	Total % (n = 613)	Within % (n = 86)	n	Total % (n = 613)	Within % (n = 127)	n	Total % (n = 613)	Within % (n = 215)
1					24	3.9%	27.9%	39	6.4%	30.7%	95	15.5%	44.2%
2					3	0.5%	3.5%	3	0.5%	2.4%	3	0.5%	1.4%
3					1	0.2%	1.2%	3	0.5%	2.4%	8	1.3%	3.7%
4					4	0.7%	4.7%	6	1.0%	4.7%	8	1.3%	3.7%
5					5	0.8%	5.8%	2	0.3%	1.6%	9	1.5%	4.2%
6					1	0.2%	1.2%	2	0.3%	1.6%	4	0.7%	1.9%
7					1	0.2%	1.2%	2	0.3%	1.6%	4	0.8%	1.9%
8					3	0.5%	3.5%	1	0.2%	0.8%	5	0.3%	2.3%
9					2	0.3%	2.3%	2	0.3%	1.6%	2	0.3%	0.9%
10					1	0.2%	1.2%	6	1.0%	4.7%	4	0.7%	1.9%
11					4	0.7%	4.7%	3	0.5%	2.4%	1	0.2%	0.5%
12					2	0.3%	2.3%	10	1.6%	7.9%	6	1.0%	2.8%
13					5	0.8%	5.8%	1	0.2%	0.8%	4	0.7%	1.9%
14					2	0.3%	2.3%	1	0.2%	0.8%	3	0.5%	1.4%
15					28	4.6%	32.6%	46	7.5%	36.2%	59	9.6%	27.4%
16					527	85.9%	--	486	79.3%	--	398	64.9%	--

Fig. 3. Suicidal planning endorsement patterns by item. Note: SP1 = Made a plan about how you would attempt suicide; SP2 = made a plan for committing suicide; SP3 = Made a plan to kill yourself; SP4 = Thought about how you might kill yourself (e.g., taking pills, shooting yourself) or work out a plan of how to kill yourself; Grey boxes represent presence (i.e., “yes” to the specific item) and white boxed represent the absence (i.e., “no” to the specific item) of each suicidal planning experience; Within sample descriptive statistics are based on the sample reporting the presence of the outcome of interest on at least one item within the specified timeframe.

and plan may be responsible for the significantly increased proportion of respondents who endorsed this item. Such conflation may be problematic as it categorizes an individual who has only had a brief passing thought about a potential suicide method to be invalidly classified as having made a plan, thus inaccurately inflating risk. To improve accuracy of risk assessment protocols, it will be important to disentangle these aspects of suicidal ideation and planning in future research.

Contrary to suicidal ideation and planning, suicide attempt response consistency did not significantly differ across items, and agreement ranged from moderate to good agreement when considering responses from the full sample. This finding may help to contextualize Hom et al.’ (2019) study who focused specifically on inconsistent reporting of suicide attempts across both self-report and clinician interview formats. Results from the current study suggest that response inconsistency is likely not due to the wording of the suicide attempt items themselves, but rather that the moderate agreement found across suicide attempt items may be facilitated by using single-item self-report assessments. It

is possible that the inconsistency attributed to the use of single-items in their own right may have inflated inconsistencies found across assessment formats in prior research; this will be an important area to examine more directly moving forward. Our results also indicate that item wording with regards to suicidal ideation and planning is more influential on endorsement than it is for suicide attempt, likely attributed to the greater variability in accepted construct definitions for suicidal ideation and planning (e.g., Millner et al. 2017; Posner et al. 2014). It is generally agreed upon that a suicide attempt is a self-directed, potentially lethal behavior carried out with at least some intention to die/to kill oneself (CDC, 2020), whereas suicidal ideation and planning may include a much larger continuum of experiences, with the delineation between experiences in the hand of the researcher and open to greater interpretability by the respondent. Consequently, both the wording and interpretation of single items assessing suicide attempts may be more consistent, at least as it pertains to the specific suicide attempt items evaluated in the present study.

Table 4
Factor loadings, correlations, and threshold estimates for suicidal planning modified constraint strong invariance model.

	Loading (SE)	Threshold (SE)	Variance
Month			
SP 1–3	0.97 (0.01)	1.49 (0.07)	0.06
SP4	0.94 (0.02)	1.21 (0.07)	0.13
Year			
SP 1–3	0.98 (0.01)	1.29 (0.06)	0.05
SP4	0.91 (0.03)	0.94 (0.06)	0.17
Lifetime			
SP 1–3	0.98 (0.01)	0.81 (0.05)	0.04
SP4	0.91 (0.02)	0.50 (0.05)	0.18
Correlations			
	Month	Year	Lifetime
Month	1.00	0.923 (0.02)	0.713 (0.04)
Year	–	1.00	0.790 (0.03)
Lifetime	–	–	1.00

Note: SP1 = Made a plan about how you would attempt suicide; SP2 = made a plan for committing suicide; SP3 = Made a plan to kill yourself; SP4 = Thought about how you might kill yourself (e.g., taking pills, shooting yourself) or work out a plan of how to kill yourself; factor variances set to 1 and all intercepts to 0 for model identification.

Response agreement was greater when considering lifetime versus past month or past year suicidal ideation, planning, and behaviors. This finding may alleviate some concerns that lifetime assessments of these experiences may be particularly prone to recall bias. While contrary to expectations that recent events may be less impacted by recall bias, this pattern of response consistency may be attributed to a few potential explanations. First, it may be the combination of sheer salience of the event (i.e., considering or taking action to end one's life), and the likely repeated recollection of the event over time, that allows individuals to recall its occurrence, regardless of time passed since the event (Wittchen et al. 1989). It is also possible that this consistency is exacerbated by the fact that items utilized in the current study assessed the presence versus absence of suicidal thoughts and behaviors, as opposed to frequency, which may be more susceptible to recall errors and bias (e.g., Graham et al. 2003).

It is also important to note that while response agreement was moderate to good when responses from the entire sample were considered, agreement significantly decreased when examined only among those who endorsed at least one item assessing the construct and time-frame of interest. For example, when considering response agreement for questions assessing lifetime suicidal ideation, if responses from the entire sample were analyzed, the resultant ICC was 0.71, whereas if responses from only those who positively endorsed lifetime suicidal ideation on at least one of the four items, the ICC dropped to 0.29. This highlights the potential that single item assessments of suicidal ideation, planning or attempts may be best suited for distinguishing between those with and without a history of suicidal ideation, planning, or behavior. These single items may be more limited in their use for assessing the presence and parameters of these experiences among those with a history of these thoughts and behaviors.

3.1. Limitations

A strength of the current paper includes the examination of how attentive responding impacts the pattern of results; findings demonstrated that the attentive responding cutoff utilized did not influence responding to suicidal ideation or suicidal planning items. However, findings did indicate that participants who passed 100% of the attention check items, relative to those passing 50–99%, reported higher thresholds (i.e., lower endorsement) across all suicide attempt items. As it is

hard to speculate why this might be the case given the current study design, examination of such attentive responding criterion in relation to consistency of responding on more nuanced self-report measures of suicide risk or across assessment formats (i.e., self-report versus clinician interview) is warranted. Relatedly, the present study employed only self-report single-item measures of suicidal thoughts and behaviors. It is therefore not a study examining the validity of these items, but rather a reliability study conducted to shed light on differences in suicidal thought and behavior endorsement based on nuances in language. Future validation studies that employ gold standard interviews (e.g., C-SSRS; SITBI) in addition to these single item measures would be helpful to provide empirical evidence supporting revised phrasing of specific single items to enhance the classification of specific suicidal thoughts and behaviors. It will be essential for future studies to examine the validity of items developed to help individuals accurately classify their own suicidal thoughts and behaviors. In the present study, we are unable to know which items were helpful in accurately classifying personal experiences. While we evaluated dichotomous single item measures of suicidal ideation, suicidal planning, and suicide attempts employed in well-known large-scale studies (e.g., NCS;/NCS-R; WHO Mental Health Survey Initiative, YRBS) and commonly used assessments (i.e., SITBI; Nock et al., 2007), there are many well-known continuous and dichotomous single-item measures of self-injurious thoughts and behaviors (both standalone items and those that are incorporated within other measures) that were not utilized in the present study (e.g., Patient Health Questionnaire-9 Item 9 [Kroenke et al. 2001]; Suicide Behavior Questionnaire-Revised [Osman et al. 2001]; Beck Depression Inventory-II [Beck et al. 1996]). Future studies should consider evaluating the phrasing of items in these measures to extend the present findings. Finally, given the current sample was recruited via online data collection with an mTurk sample, findings should be replicated both utilizing differing data collection methodologies (e.g., research administered self-reports) and populations (e.g., clinical samples, adolescents).

3.2. Implications & recommendations

Despite limitations, this study has several implications. Prior to establishing gold standard single item measures, it will be important for researchers to first better define constructs, and to do so based on empirical evidence. For example, while the present evidence demonstrates that using the word “serious” impacts item endorsement, we do not know whether it is useful to know if a thought is considered “serious” (or not) or what the use of this term means to respondents. Nevertheless, the findings of this study suggest that researchers and clinicians alike should pay close attention to the wording of self-report single items used to assess suicide risk, both in designing and interpreting findings of theirs and others' studies. This is of particular importance as we see that even minor language changes in suicidal ideation and planning items may inflate or restrict prevalence estimates in a meaningful way. For researchers, this can pose serious challenges for comparing rates of suicidal ideation and planning across studies, as well as for assessing risk factors for these outcomes in a consistent way and with valid statistical testing across studies (Millner et al. 2015). For clinicians, imprecise language can result in an underestimation or overestimation of risk, the former of which arguably represents a significant threat to patient safety. Further, our findings indicate that imprecise language may potentially differentially impact risk estimation based on individuals' histories of suicidal thoughts and behaviors. Our findings align with prior research (Hom et al. 2019; Millner et al. 2015) suggesting that single-item self-report assessments often, and may inherently, lack the clarity needed to ensure that individuals comprehend the definition of the STB being assessed. Evidence suggests that clinician-rated

Pattern					Past Month			Past Year			Lifetime		
	SA1	SA2	SA3	SA4	<i>n</i>	Total % (<i>n</i> = 613)	Within % (<i>n</i> = 54)	<i>n</i>	Total % (<i>n</i> = 613)	Within % (<i>n</i> = 63)	<i>n</i>	Total % (<i>n</i> = 613)	Within % (<i>n</i> = 106)
1					12	2.0%	22.2%	11	1.8%	17.5%	61	10.0%	57.5%
2					2	0.3%	3.7%	5	0.8%	7.9%	1	0.2%	0.9%
3					2	0.3%	3.7%	3	0.5%	4.8%	3	0.5%	2.8%
4					2	0.3%	3.7%	5	0.8%	7.9%	3	0.5%	2.8%
5					0	0.0%	0.0%	3	0.5%	4.8%	4	0.7%	3.8%
6					4	0.2%	7.4%	3	0.5%	4.8%	3	0.5%	2.8%
7					4	0.2%	7.4%	1	0.2%	1.6%	4	0.7%	3.8%
8					3	0.5%	5.6%	0	0.0%	0.0%	2	0.3%	1.9%
9					3	0.5%	5.6%	1	0.2%	1.6%	3	0.5%	2.8%
10					1	0.2%	1.9%	2	0.3%	3.2%	2	0.3%	1.9%
11					1	0.2%	1.9%	5	0.8%	7.9%	3	0.5%	2.8%
12					7	1.1%	13.0%	10	1.6%	15.9%	2	10.3%	1.9%
13					6	1.0%	11.1%	5	0.8%	7.9%	4	0.7%	3.8%
14					3	0.5%	5.6%	5	0.8%	7.9%	7	1.1%	6.6%
15					4	0.7%	7.4%	4	0.7%	6.3%	4	0.7%	3.8%
16					559	91.2%	--	550	89.7%	--	507	82.7%	--

Fig. 4. Suicide attempt endorsement patterns by item.

Note: SA1 = Attempted to kill yourself; SA2 = Attempted suicide; SA3 = Made a suicide attempt (i.e., purposefully hurt yourself with at least some intention to die); SA4 = Actually attempted suicide; Grey boxes represent presence (i.e., “yes” to the specific item) and white boxed represent the absence (i.e., “no” to the specific item) of each suicide attempt experience; Within sample descriptive statistics are based on the sample reporting the presence of the outcome of interest on at least one item within the specified timeframe.

Table 5

Factor loadings, correlations, and threshold estimates for strong invariance suicide attempt model.

	Loading (SE)	Threshold (SE)	Variance
Month	0.94 (0.02)	1.66 (0.07)	0.11
Year	0.94 (0.01)	1.58 (0.07)	0.12
Lifetime	0.99 (0.00)	1.11 (0.06)	0.03
Correlations			
	Month	Year	Lifetime
Month	1.00	0.993 (0.02)	0.721 (0.04)
Year	–	1.00	0.722 (0.05)
Lifetime	–	–	1.00

Note: Factor variances set to 1 and all intercepts to 0 for model identification; all items constrained equal within timeframes.

interviews, such as the Columbia Suicide Self-Rating Scale (C-SSRS; Posner et al., 2011), may result in reduced rates of false negatives and false positives as compared to some single item self-report measures given their use of follow-up questions (Hom et al. 2019). While this may be true, the benefits of single-item assessments continue to result in widespread use for both research and clinical purposes. Given this, it is incumbent upon researchers to develop and test brief self-report measures that are both valid and reliable and minimize false positives and negatives.

Declaration of Competing Interest

The authors have no conflicts of interests or funding sources to disclose.

Table 6
Factor loadings, correlations, and threshold estimates for suicide attempt attention check metric invariance.

	Loading (SE)	Threshold cutoff (SE)	Threshold100% (SE)	Variance
Month				
SA1	0.95 (0.02)	1.38 (0.12)	1.78 (0.12)	0.10
SA2	0.96 (0.02)	1.41 (0.12)	1.89 (0.13)	0.07
SA3	0.97 (0.02)	1.47 (0.12)	1.85 (0.13)	0.06
SA4	0.93 (0.03)	1.57 (0.13)	1.89 (0.13)	0.14
Year				
SA1	0.94 (0.02)	1.33 (0.11)	1.72 (0.12)	0.12
SA2	0.92 (0.03)	1.35 (0.12)	1.85 (0.11)	0.15
SA3	0.97 (0.02)	1.44 (0.08)	1.67 (0.12)	0.06
SA4	0.94 (0.03)	1.44 (0.08)	1.75 (0.08)	0.13
Lifetime				
SA1	0.99 (0.01)	1.05 (0.10)	1.18 (0.08)	0.02
SA2	0.99 (0.01)	1.01 (0.10)	1.19 (0.09)	0.02
SA3	0.98 (0.01)	0.98 (0.10)	1.16 (0.08)	0.05
SA4	0.99 (0.01)	1.00 (0.10)	1.19 (0.09)	0.03
Correlations				
	Group = cut	Month	Year	Lifetime
Month		1.00	0.96 (0.04)	0.72 (0.06)
Year		–	1.00	0.78 (0.06)
Lifetime		–	–	1.00
	Group = 100%	Month	Year	Lifetime
Month		1.00	1.01 (0.03)	0.72 (0.07)
Year		–	1.00	0.66 (0.08)
Lifetime		–	–	1.00

Note: SA1 = Attempted to kill yourself; SA2 = Attempted suicide; SA3 = Made a suicide attempt (i.e., purposefully hurt yourself with at least some intention to die); SA4 = Actually attempted suicide; Cutoff = participants above 50% attention checks but below 100%; 100% = participants passing all attention checks; factor variances set to 1 and all intercepts to 0 for model identification; loadings constrained equal across groups.

Table 7
Attentive responding inclusion criteria group measurement invariance.

Model	Group metric invariance	Group strong invariance
SI Configural	Yes ($\chi^2_{12} = 9.01; p = 0.70$)	Yes ($\chi^2_{12} = 15.27; p = 0.23$)
SI metric invariance	Yes ($\chi^2_3 = 4.88; p = 0.18$)	Yes ($\chi^2_3 = 1.54; p = 0.67$)
SI modified constraint	Yes ($\chi^2_6 = 5.36; p = 0.50$)	Yes ($\chi^2_6 = 8.50; p = 0.20$)
SP Configural	Did not converge	Yes ($\chi^2_{12} = 12.59; p = 0.40$)*
SP metrical invariance	No ($\chi^2_3 = 9.25; p = 0.003$)	Yes ($\chi^2_3 = 2.39; p = 0.50$)*
SP modified constraint	Yes ($\chi^2_6 = 8.62; p = 0.20$)	Yes ($\chi^2_6 = 5.84; p = 0.44$)
SA Configural	Yes ($\chi^2_{12} = 8.23; p = 0.77$)	No ($\chi^2_{12} = 29.75; p = 0.003$)

Note: SI = suicidal ideation; SP = suicidal planning; SA = suicide attempt; * = interpretation of these condition should be met with caution given lack of group metric invariance; Yes indicates that significant invariance was demonstrated in the model.

References

Beck, A.T., Steer, R.A., Brown, G., 1996. Beck depression inventory-II [Database record]. APA PsycTests.

Berman, A.L., Silverman, M.M., 2014. Rethinking suicide risk assessment and risk formulation. In: Koslow, S.H., Ruiz, P., Nemeroff, C.B. (Eds.), *A Concise Guide to Understanding Suicide*. Cambridge University Press, Cambridge, MA, USA.

Brown, G.K., Currier, G., Stanley, B., 2008. Suicide attempt registry pilot project. In: National Institute of Mental Health Annual Meeting of the Developing Centers for Intervention and Prevention of Suicide, September 2008, Canandaigua, NY.

Center for Disease Control and Prevention, 2020. Preventing Suicide. Retrieved from. <https://www.cdc.gov/violenceprevention/suicide/fastfact.html>. Accessed on October 15, 2020.

Centers for Disease Control and Prevention. (n.d.) Youth Risk Behavior Survey Data. Available at: www.cdc.gov/yrbps. Accessed on October 15, 2020.

Desseilles, M., Perroud, N., Guillaume, S., Jaussent, I., Genty, C., Malafosse, A., Courtet, P., 2012. Is it valid to measure suicidal ideation by depression rating scales? *J. Affect. Disord.* 136, 398–404. <https://doi.org/10.1016/j.jad.2011.22.013>.

Franklin, J.C., Ribeiro, J.D., Fox, K.R., Bentley, K.H., Kleiman, E.M., Huang, X., Nock, M.K., 2017. Risk factors for suicidal thoughts and behaviors: a meta-analysis of 50 years of research. *Psychological Bulletin* 143, 187–232. <https://doi.org/10.1027/bul0000084>.

Graham, C.A., Catania, J.A., Brand, R., Duong, T., Canchola, J.A., 2003. Recalling sexual behavior: a methodological analysis of memory recall bias via interview using the diary as the gold standard. *J. Sex Res.* 40, 325–332. <https://doi.org/10.1080/00224490209552198>.

Harris, Kathleen Mullan, Udry, J. Richard, 2018. National Longitudinal Study of Adolescent to Adult Health (Add Health), 1994–2008 [Public Use]. Carolina Population Center, University of North Carolina-Chapel Hill [distributor], Inter-university Consortium for Political and Social Research [distributor], 2018-08-06. <https://doi.org/10.3886/ICPSR21600.v21>.

Hom, M.A., Stanley, I.H., Duffy, M.E., Rogers, M.L., Hanson, J.E., Gutierrez, P.M., Joiner, T.E., 2019. Investigating the reliability of suicide attempt history reporting across five measures: a study of US military service members at risk of suicide. *J. Clin. Psychol.* 75, 1332–1349. <https://doi.org/10.1002/jclp.22776>.

Kessler, R.C., Haro, J.M., Heeringa, S.G., Pennell, B.E., Üstün, T.B., 2006. The World Health Organization world mental health survey initiative. *Epidemiol. Psychiatr. Sci.* 15, 161–166. <https://doi.org/10.1017/S1121189X00004395>.

Kessler, R.C., Merikangas, K.R., 2004. The national comorbidity survey replication (NCS-R): background and aims. *International journal of methods in psychiatric research* 13 (2), 60–68.

Kessler, Ronald, 2015. National Comorbidity Survey: Reinterview (NCS-2), 2001–2002. Inter-university Consortium for Political and Social Research [distributor], 2015-03-31, Ann Arbor, MI. <https://doi.org/10.3886/ICPSR35067.v2>.

Kessler, Ronald C., 2018. National Comorbidity Survey: Baseline (NCS-1), 1990–1992. Inter-university Consortium for Political and Social Research [distributor], 2008-09-12, Ann Arbor, MI. <https://doi.org/10.3886/ICPSR06693.v6>.

Kessler, R.C., Aguilar-Gaxiola, S., Alonso, J., Chatterji, S., Lee, S., Üstün, T.B., 2009. The WHO world mental health (WMH) surveys. *Psychiatrie (Stuttgart, Germany)* 6 (1), 5–9.

Koo, T.K., Li, M.Y., 2016. A guideline of selecting and reporting Intraclass correlation coefficients for reliability research. *J. Chiropractic Med.* 15, 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>.

Kroenke, K., Spitzer, R.L., Williams, J.B., 2001. The PHQ-9: validity of a brief depression severity measure. *J. Gen. Intern. Med.* 16, 606–613. <https://doi.org/10.1046/j.1525-1497>.

Liu, Y., Millsap, R.E., West, S.G., Tein, J.-Y., Tanaka, R., Grimm, K.J., 2017. Testing measurement invariance in longitudinal data with ordered-categorical measures. *Psychol. Methods* 22, 486–506. <https://doi.org/10.1037/met0000075>.

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), e0141606. <https://doi.org/10.1371/journal.pone.0141606>.

Millner, A.J., Lee, M.D., Nock, M.K., 2017. Describing and measuring the pathway to suicide attempts: a preliminary study. *Suicide Life Threat. Behav.* 47, 353–369. <https://doi.org/10.1111/sltb.12284>.

National Institute of Mental Health, 2017. Dysregulation and Proximal Risk for Suicide. Retrieved from: <https://www.nimh.nih.gov/funding/grant-writing-and-application-process/concept-clearances/2017/dysregulation-and-proximal-risk-for-suicide.shtml>. Accessed on October 15, 2020.

Nock, M.K., Kessler, R.C., 2006. Prevalence of and risk factors for suicide attempts versus suicide gestures: analysis of the National Comorbidity Survey. *J. Abnorm. Psychol.* 115, 616–623. <https://doi.org/10.1037/0021-843X.115.3.616>.

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>.

Osman, A., Bagge, C.L., Gutierrez, P.M., Konick, L.C., Kopper, B.A., Barrios, F.X., 2001. The suicidal behaviors questionnaire-revised (SBQ-R): validation with clinical and nonclinical samples. *Assessment* 8, 443–454. <https://doi.org/10.1177/107319110100800409>.

Plöderl, M., Kravlovec, K., Yazdi, K., Fartacek, R., 2011. A closer look at self-reported suicide attempts: false positives and false negatives. *Suic. Life-Threat. Behav.* 41, 1–5. <https://doi.org/10.1111/j.1943-278X.2010.000005.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. *American journal of psychiatry* 168 (12), 1266–1277.

Posner, K., Buchanan, J., Amira, L., Yershova, K., Lesser, A., Goldstein, E., 2014. Identification and screening of suicide risk. In: Koslow, S.H., Ruiz, P., Nemeroff, C.B. (Eds.), *A Concise Guide to Understanding Suicide*. Cambridge University Press, Cambridge, MA, USA.

R Core Team, 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL: <https://www.R-project.org/>.

Revelle, W., 2019. psych: Procedures for Personality and Psychological Research. Northwestern University, Evanston, Illinois, USA. <https://CRAN.R-project.org/package=psych>. Version = 1.9.12.

Rosseel, Y., 2012. lavaan: An R Package for Structural Equation Modeling. *J. Stat. Softw.* 48, 1–36. URL: <http://www.jstatsoft.org/v48/i02/>.

- Schmier, J.K., Halpern, M.T., 2004. Patient recall and recall bias of health state and health status. *Exp. Rev. Pharmacoecon. Outcomes Resear.* 4, 159–163. <https://doi.org/10.1586/14737167.4.2.159>.
- Silverman, M.M., Berman, A.L., Sanddal, N.D., O'Carroll, P.W., Joiner, T.E., 2007. Rebuilding the tower of babel: a revised nomenclature for the study of suicide and suicidal behaviors, part 2: suicide-related ideations, communications, and behaviors. *Suicide Life Threat. Behav.* 37, 264–277.
- Svetina, D., Rutkowski, L., Rutkowski, D., 2020. Multiple-group invariance with categorical outcomes using updated guidelines: an illustration using *M plus* and the lavaan/semTools packages. *Struct. Equ. Model. Multidiscip. J.* 27, 111–130. <https://doi.org/10.1080/10705511.2019.1602776>.
- Ursano, Robert J., Stein, Murray B., Kessler, Ronald C., Heeringa, Steven G., Wagner, James, 2020. Army Study to Assess Risk and Resilience in Servicemembers (STARRS). Inter-university Consortium for Political and Social Research [distributor], 2020-08-27. <https://doi.org/10.3886/ICPSR35197.v7>.
- Wittchen, H.U., Essau, C.A., Hecht, H., Teder, W., Pfister, H., 1989. Reliability of life event assessments: test—retest reliability and fall-off effects of the Munich interview for the assessment of life events and conditions. *J. Affect. Disord.* 16, 77–91.
- World Health Organization, 2020. Suicide in the world: Global Health Estimates. Retrieved from: https://www.who.int/mental_health/prevention/suicide/suicideprevent/en/. Accessed on October 15, 2020.