

Measures of Suicidal Thoughts and Behaviors in Children and Adolescents: A Systematic Review and Recommendations for Use in Clinical and Research Settings

Assessment
2025, Vol. 32(3) 303–320
© The Author(s) 2024
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/10731911241249438
journals.sagepub.com/home/asm



Richard T. Liu^{1,2,3} , Alexandra H. Bettis⁴, Hannah R. Lawrence⁵, Rachel F. L. Walsh⁶, Ana E. Sheehan⁷, Olivia H. Pollak⁸, Auburn R. Stephenson⁶, Marin M. Kautz⁶, and Rachel M. Marlowe⁴

Abstract

Empirically supported measures of suicidal thoughts and behaviors (STBs) are needed to serve as reference outcomes for suicide risk screening tools and to monitor severity and treatment progress in children and adolescents with STBs. The present paper systematically reviewed existing measures of STBs in youth and studies evaluating their psychometric properties and clinical utility. Measures were then evaluated on reliability, validity, and clinical utility. Sixteen articles (20 independent samples) were found with psychometric data with youth samples for eight measures. Interview-based measures were found to have the strongest psychometric support and clinical utility. Significant limitations exist for all self-report measures due to inherent characteristics of these measures that cannot be remedied through additional psychometric study. There is an urgent need for the development and validation of new self-report measures of STBs, particularly for preadolescent children, sexual and gender minority youth, and racial/ethnic minority youth.

Keywords

adolescents, assessment, children, psychometrics, suicide attempt, suicidal ideation, suicide

Youth suicide is a significant and growing public health concern. Suicide rates in adolescents have tripled in recent years (Curtin et al., 2016; Plemmons et al., 2018). A corresponding trend in related health care use has been observed, with hospital encounters for adolescent suicidal thoughts and behaviors (STBs) increasing significantly in recent years (Arakelyan et al., 2023; Bommersbach et al., 2023; Plemmons et al., 2018). Furthermore, in recent data from the CDC, 19.9% of youth reported experiencing severe suicidal ideation (SI) and 9.0% had attempted suicide within the past year (Jones et al., 2022). Similarly worrying trends have emerged with preadolescent children. In fact, suicide rose from being the 10th leading cause of death among youth of ages 5–12 in 2008 to the fifth leading cause of death in 2019 (Centers for Disease Control and Prevention, 2020), which prompted the NIMH to convene a series of workshops to address this issue (National Institute of Mental Health, 2019, 2021).

Empirically supported measures of STBs are critically needed to serve as reference outcomes for suicide risk

screening tools and to monitor severity and treatment progress in youth with STBs as well as the course of related conditions (e.g., depression). Within this context, the important distinction between suicide risk screening instruments and instruments that serve as measures of STBs—the focus of the current review—should be briefly mentioned. Suicide risk screening instruments and STB outcome measures are often conflated but both

¹Massachusetts General Hospital, Boston, USA

²Harvard Medical School, Boston, MA, USA

³Broad Institute of MIT and Harvard, Cambridge, MA, USA

⁴Vanderbilt University Medical Center, Nashville, TN, USA

⁵Oregon State University, Corvallis, USA

⁶Temple University, Philadelphia, PA, USA

⁷University of Delaware, Newark, USA

⁸The University of North Carolina at Chapel Hill, USA

Corresponding Author:

Richard T. Liu, Department of Psychiatry, Massachusetts General Hospital, Harvard Medical School, 55 Fruit Street, Boston, MA 02114, USA.
Email: rtiupsy@ch@gmail.com

hold distinctly different but complementary functions. Suicide risk screening instruments assess the need for more in-depth evaluation of future risk for suicidal outcomes, whereas the function of STB measures is to assess current and/or prior history of these outcomes. Although several reviews (Baek et al., 2021; Carter et al., 2019; Kreuze & Lamis, 2018; Lowry et al., 2024; Randall et al., 2011) have been conducted on suicide risk screening tools, such as the Patient Health Questionnaire-9 (PHQ-9; Kroenke et al., 2001) no systematic review exists of evidence-based evaluations of STB measures in youth.

The current review aimed to fill this gap by systematically reviewing existing measures of STBs for children and adolescents and studies evaluating their psychometric properties. Youngstrom and colleagues' (Youngstrom et al., 2017) criteria for evaluating the evidence base and pragmatism of clinical measures were used and adapted, supplemented in part by considerations regarding appropriate metrics of (and interpretation of data for) reliability and validity (Cronbach & Shavelson, 2004; Revelle & Condon, 2019; Youngstrom et al., 2019). Finally, recommendations for measures to use in clinical and research settings were made based on findings with these criteria, with particular emphasis given to considerations regarding clinical utility and scalability.

Method

Search Strategy and Selection Criteria

This project was registered in PROSPERO (CRD4 2022381918) and followed Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) reporting guidelines. A two-stage systematic review was conducted. In Stage 1, we systematically identified measures of STBs that conform to currently accepted taxonomies and definitions of self-injurious thoughts and behaviors (Corsby et al., 2011; National Action Alliance for Suicide Prevention, 2014; Posner et al., 2007) and for which psychometric studies may have been published. In Stage 2, we systematically searched for peer-reviewed articles that reported on the psychometric properties in youth for all eligible measures identified in Stage 1.

Throughout both stages, each unique search result was reviewed independently by at least two coders to determine eligibility. In cases where eligibility could not be ruled out based on the title and abstract, the full text was examined. In cases where more information was needed to determine study eligibility or for data extraction, detailed below, every effort was made to obtain additional details in other publications describing the measure or by contacting the study authors and authors

of the relevant measure. Discrepancies in coding article eligibility were resolved by the first author. Details of both stages of our systematic review are presented below.

Stage 1. A systematic search of the literature was conducted in PsycINFO and MEDLINE to identify articles with a focus on measures of potential relevance to the current review. The following search string was applied: (TI ("interview" OR "questionnaire" OR "survey" OR "measure*" OR "assessment" OR "instrument" OR "rating scale")) OR AB ("interview" OR "questionnaire" OR "survey" OR "measure*" OR "assessment" OR "instrument" OR "rating scale") AND ((TI ("psychometric" OR "measure development" OR "measure validation" OR "test construction" OR "test validity") OR AB ("psychometric" OR "measure development" OR "measure validation" OR "test construction" OR "test validity")) AND ((suicid* OR AB suicid*)). These search terms were based on those used in prior systematic reviews evaluating measures of other constructs (Becker-Haimes et al., 2020; Mazefsky et al., 2021). The search results were limited to (a) English-language publications and (b) peer-reviewed journals since inception through March 28, 2024. This was supplemented by a search of peer-reviewed English-language journal articles in the references of prior reviews of overlapping constructs (e.g., self-harm, suicide risk screeners) that were identified in this search (Baek et al., 2021; Batterham et al., 2015; Borschmann et al., 2012; Carter et al., 2019; Erford et al., 2018; Flores-Kanter & Alvarado, in press; Gleeson et al., 2021; Kreuze & Lamis, 2018; Randall et al., 2011).

Identification of eligible measures of STBs were based on the following inclusion criteria: (a) STBs were distinguished from nonsuicidal self-injurious thoughts, urges, and behavior;¹ (b) SI was distinguished from suicidal behaviors in individual measure items and the scoring procedure for the measure (e.g., measures that treated SI and suicide attempts along a spectrum of severity were ineligible); (c) measures that assessed suicide attempts cleanly separated them from other suicidal behaviors (e.g., aborted attempts); (d) individual measure items and the measure scoring procedure distinguished between STBs and other psychopathology, risky behaviors, and other suicide risk factors (e.g., hopelessness); and (e) measures in languages other than English were eligible if presented in an English-language peer-reviewed paper.

Stage 2. A series of systematic searches was then conducted in PsycINFO and MEDLINE of psychometric studies for all the eligible STB measures identified in

Stage 1. Specifically, again modeling prior systematic reviews evaluating the psychometric properties of measures of other constructs (Becker-Haimes et al., 2020; Mazefsky et al., 2021), we identified studies on the psychometric properties of all eligible STB measures from Stage 1 with a series of search strings that was identical to the one used in Stage 1 but with the root of the each measure name and its variants in place of suicid*: ((TI (“psychometric” OR “measure development” OR “measure validation” OR “test construction” OR “test validity”) OR AB (“psychometric” OR “measure development” OR “measure validation” OR “test construction” OR “test validity”)) AND ([root of measure name and its variants] OR AB [root of measure name and its variants])). See Table S1 for the specific search string employed for each STB measure.² Consistent with Stage 1, the search results were limited to (a) English-language publications and (b) peer-reviewed journals since inception through March 28, 2024. This search was then supplemented in two ways. First, the primary studies in Stage 1 that yielded eligible STB measures were systematically reviewed to determine if they provided psychometric evaluations of their STB measure(s). Second, the method section of all psychometric studies identified in the Stage 2 search process were systematically reviewed to identify additional eligible psychometric studies.

To be eligible for Stage 2, studies had to present psychometric data on an STB measure for children and/or adolescents $\leq$ age 19. Studies that also included older participants were eligible, however, if they presented psychometric data for youth separately from adults.

Data Extraction and Calculation

For each STB measure with at least one eligible psychometric study identified in Stage 2, we extracted and computed 10 immutable measure characteristics, most relating to measure scalability: (a) STB types assessed; (b) standard format of the measure (i.e., interview or questionnaire); (c) standard time frame(s) covered by the measure; (d) number of words in the measure; (e) number of items in the measure; (f) whether the measure included reverse-scored items; (g) the measure’s Flesch reading ease formula score; (h) the Flesch-Kincaid grade level formula score for the measure; (i) whether training is required to administer the measure; and (j) whether the measure is freely available or proprietary. For each measure, we also determined whether normative data were available for clinical and nonclinical youth samples in the eligible studies from Stage 2. In cases where the eligible studies did not provide sufficient evidence of normative data based on our measure evaluation criteria (detailed below), every effort was made to obtain the

required information through contacting the study authors, and in the case of proprietary measures, their vendor companies.

For each eligible psychometric study identified in Stage 2, we extracted data for 22 sample, study design, and psychometric characteristics. These included seven sample characteristics: (a) sample size; (b) mean age; (c) age range; (d) percentage of female participants in the sample; (e) racial and ethnic composition of the sample; (f) sample type (i.e., clinical, community, or mixed); and (g) the country in which the study was conducted. The five extracted study design characteristics included (a) the language in which the STB measure was administered; (b) the STBs assessed by the STB measure (to account for the possibility that only part of a measure was used); (c) the format of STB measure administration, to account for the possibility of administration in a format other than originally designed; (d) the reporting source for the measure (i.e., youth or clinician); and (e) the time frame(s) used for assessing STBs, to account for differences across studies using the same STB measure, especially in interview-based assessments of suicidal behavior. Finally, 10 psychometric characteristics were extracted, including (a) internal consistency; (b) test-retest reliability; (c) inter-rater reliability; (d) face validity; (e) content validity; (f) convergent validity; (g) discriminant validity; (h) divergent validity; (i) sensitivity to change over time; and (j) performance of the measure’s recommended clinical cut-off score(s).

All data elements were independently extracted by at least two authors, and discrepancies were resolved by the first or second author. To aid interpretability, whenever effect sizes were not reported but could be derived from presented data, standardized mean difference (Cohen’s *d*) was calculated.

Measure Evaluation Criteria

Psychometric evaluations were based on criteria adapted from Youngstrom and colleagues (see Tables 1 and 2 in Youngstrom et al., 2017) and additional considerations regarding reliability and validity (Cronbach & Shavelson, 2004; Revelle & Condon, 2019; Youngstrom et al., 2019). As detailed in Table 1, modifications were made to these criteria more specifically to reflect the nature of STBs. Where notable, these modifications are discussed below.

Given the clinically severe nature of STBs, need for validation of measures is greatest in clinical settings, and the likelihood is modest that any measures would be administered to population-representative samples. The criterion for norms therefore has been modified to allow for maximum range.

Table 1. Criteria for Evidence-Based Evaluations of Suicidal Thoughts and Behaviors Adapted from Youngstrom et al. (2017).

Criterion	Adequate	Good	Excellent	Too excellent
Norms^a	Mean and standard deviation in one large clinical or community sample ($n \geq 300$)	Mean and standard deviation in one large clinical sample ($n \geq 300$)	Mean and standard deviation in multiple large independent samples ($n \geq 300$), at least one of which is a clinical sample	Not applicable
Internal Consistency^a	Most evidence is of $\alpha = .70-.79$	Most evidence is of $\alpha = .80-.89$ in multiple independent samples	Most evidence is of $\alpha \geq .90$ in multiple independent samples, at least one of which is a clinical sample	Consider that α may be inflated by scale length (long) and/or scope (narrow)
Test-Retest Reliability	Most evidence is of $r = .40-.59$, $\kappa = .60-.74$, or $ICC = .70-.79$	Most evidence is of $r = .60-.79$, $\kappa = .75-.84$, or $ICC = .80-.89$ in multiple independent samples	Most evidence is of $r \geq .80$, $\kappa \geq .85$, or $ICC \geq .90$ in multiple independent samples	Not applicable
Inter-Rater Reliability^b	Most evidence is of $\kappa = .60-.74$ or $ICC = .70-.79$	Most evidence is of $\kappa = .75-.84$ or $ICC = .80-.89$ in multiple independent samples	Most evidence is of $\kappa \geq .85$ or $ICC \geq .90$ in multiple independent samples	Consider that estimates of reliability may be inflated by reliability rating conducted with audio recordings or transcripts
Face Validity	Presence of concerns regarding developmental appropriateness AND other face validity issues in measure items	Presence of concerns regarding developmental appropriateness OR other face validity issues in measure items	No face validity issues	Not applicable
Content Validity	Include at least one form of STBs assessed separately from other STBs and other psychopathology	Measures of SI: include both PSI and ASI Measures of suicidal behaviors: include suicidal intent in assessing SA	Measures of SI: same as "good" but also include assessment of thoughts of methods or plans Measures of suicidal behaviors: same as "good" but also include assessments of interrupted SA and aborted SA	Not applicable
Construct Validity (convergent, discriminant, and divergent validity)	Support for multiple aspects of construct validity in one sample	Support for multiple aspects of construct validity in multiple independent samples	Support for multiple aspects of construct validity replicated across multiple independent samples	Consider that estimates of validity may be inflated by shared method and shared source variance
Validity Generalization	Some validity support in multiple demographic groups OR multiple settings	Support for multiple aspects of validity replicated in multiple demographic groups OR multiple settings	Support for multiple aspects of validity replicated in multiple demographic groups AND multiple settings	Not applicable
Sensitivity to Change^c	I sample with evidence of sensitivity to change over time	At least two independent samples with evidence of sensitivity to change over time	Same as "good" but across different settings or sample types (e.g., outpatient and inpatient)	Consider time interval (i.e., long temporal intervals have limited clinical value and may yield inflated estimates of sensitivity to change; ensure no temporal overlap between time covered by measure and the time between consecutive measure administrations)
Clinical Utility	Considerations of practicality/scalability in measure features make it likely to be clinically actionable	Same as "adequate" but with evidence in at least one sample supporting clinical utility (e.g., mean and standard deviation for time to administer; time to score and interpret resulting data; clinician and/or patient rating of the measure)	Same as "good" but with evidence replicated across multiple independent samples	Not applicable

Note. ASI = active suicidal ideation; ICC = intra-class correlation coefficient; PSI = passive suicidal ideation; SA = suicide attempt; SI = suicidal ideation; STBs = suicidal thoughts and behaviors.

^aOnly applicable to measures with (sub)scale scores. ^bOnly applicable to interview-based measures. ^cOnly applicable to measures with suicidal ideation (sub)scale scores.

Table 2. Study Sample Characteristics.

Author(s) (year)	Country	N	Mean age	Age range	% Female	Racial/ethnic composition	Sample type
<i>Beck Scale for Suicidal Ideation (BSS) Clinician-Rated Version</i>							
Allan et al. (1997)	United States	87	—	7–12	25.29	94.25% White, 5.75% Black	clinical
Holi et al. (2005) community subsample	Finland	200	16.5	13–19	81.41	—	community
Holi et al. (2005) clinical subsample	Finland	218	16.4	13–19	81.65	—	clinical
<i>Concise Health Risk Tracking Self-Report (CHRT-SR)</i>							
Mayes et al. (2018)	United States	271	14.93	12–18	82.29	82.66% White, 8.49% Black, 1.48% Asian, 0.37% multiracial, 7.01% unknown, 10.33% Hispanic	clinical
Nandy et al. (2023)	United States	657	14.10	11–17	71.23	68.34% White, 9.89% Black, 21.77% other	mixed
<i>Columbia-Suicide Severity Rating Scale (C-SSRS)</i>							
Kilincaslan et al. (2019)	Turkey	213	14.97	12–18	61.97	—	mixed
Posner et al. (2011) Study 1	United States	124	15.8	12–18	77.42	66.94% White, 12.90% Black, 15.32% Hispanic, 4.84% other	clinical
Posner et al. (2011) Study 2	United States	312	14.60	11–17	59	75.68% White	clinical
<i>Suicidal Ideation Questionnaire (SIQ)</i>							
Jia et al. (2015)	China	505	15.98	12–18	36.60	—	community
Pinto et al. (1997)	United States	226	15.17	13–18	61.06	88.8% White, 8.4% Black, 2.2% Hispanic, 0.6% unknown	clinical
Potard et al. (2014) Study 1	France	871	16.35	14–18	62.47	—	community
Potard et al. (2014) Study 2	France	38	15.66	14–18	60.53	—	clinical
Potard et al. (2014) Study 3	France	56	18.32	17–19	82.14	—	community
Zhang et al. (2014)	China	711	—	15–19	—	—	community
<i>Suicidal Ideation Questionnaire-JR (SIQ-JR)</i>							
Hill et al. (2020)	United States	114	13.64	10–19	67	100% American Indian	clinical
Núñez et al. (2024) Study 1	Chile	619	15.94	14–19	51.21	—	community
Núñez et al. (2024) Study 2	Chile	524	15.41	14–19	43.51	—	community
Reynolds and Mazza (1999)	United States	91	12.5	11–15	58.24	1.1% White, 71.4% Black, 1.1% American Indian, 6.6% other, 19.8% Hispanic	community
Zhang et al. (2014)	China	711	—	15–19	—	—	community
<i>Suicidal Ideation Questionnaire-JR-4 (SIQ-JR-4)</i>							
Zhang et al. (2014)	China	711	—	15–19	—	—	community
<i>Self-Injurious Thoughts and Behaviors Interview (SITBI)</i>							
Fischer et al. (2014)	Germany	111	15.38	12–19	65.8	94.5% German	clinical
Nock et al. (2007)	United States	94	17.1	12–19	77.66	73.4% White, 3.2% Black, 5.3% Asian, 10.6% biracial, 1.1% other, 6.4% Hispanic	mixed
Venta and Sharp (2014)	United States	106	14.63	12–17	64.15	24.8% White, 32.4% Black, 2.9% multiracial, 1.0% other, 39.0% Hispanic	clinical
<i>Self-Injurious Thoughts and Behaviors Interview-Revised (SITBI-R)</i>							
Gratch et al. (2022)	United States	206	16.86	12–19	70.19	33.3% White, 21.7% Black, 25.3% Asian, 18.2% other, 1.5% unknown, 24.4% Hispanic	mixed

Note. Studies differed in whether they observed that race and ethnicity are not mutually exclusive. For this reason, in some studies, race combined with ethnicity will add up to 100%. Wherever possible, values for mean age, % female, and racial/ethnic composition are reported to two decimal places.

Regarding reliability, in cases where internal consistency was found to be very high (e.g., $\alpha > .90$), we closely examined factors that may inflate reliability at the expense of other important considerations (i.e., scale length to the detriment of clinical utility, and narrow measure scope at the cost of content validity; Youngstrom et al., 2019). In cases where concern seemed warranted, the relevant measure was rated “too excellent” (Youngstrom et al., 2017). In addition, the widespread practice of only reporting Cronbach’s α as an index of internal consistency has been increasingly viewed as insufficient, and instead including it in a suite of internal reliability indices is now recommended (Cronbach & Shavelson, 2004; Revelle & Condon, 2019; Youngstrom et al., 2019). We therefore noted where studies included other indices of internal consistency (e.g., ω).

Adjustments were made to the test–retest reliability criterion to accommodate the temporally dynamic nature of SI (Kleiman et al., 2017). That is, unlike with trait constructs, a measure of SI should demonstrate more modest test–retest reliability in clinical samples³ between two administrations that cover nonoverlapping temporal intervals (e.g., a 1-week measure given twice 6 months apart). It is therefore important to ensure reasonable overlap in time covered by consecutive measure administrations, thereby avoiding inaccurately deflated estimates of reliability. Conversely, to ensure accurate assessments of sensitivity to change over time, nonoverlapping temporal intervals between consecutive measure administrations is required. The length of the temporal window used, however, is still an important consideration. Although large temporal intervals (e.g., one year) are more likely to detect significant change, their clinical value is limited. To optimize clinical utility, having the temporal intervals approximate those of regular therapy sessions (e.g., 1 week or 1 month) is needed. Evaluations of sensitivity to change were conducted with these considerations in mind.

Adjustments were also made in expectations for “too excellent” ratings for inter-rater reliability to account for the fact that categorizing and coding for the presence of STBs is generally easier than for psychiatric disorders in diagnostic interviews. Criteria for SI and suicide attempts, respectively, are more straightforward and require less clinical judgment than many psychiatric disorders with their multiple symptoms with severity thresholds. It is therefore reasonable to expect higher inter-rater reliability with STBs. For this reason, when reliability was very high, we based “too excellent” ratings on study reliance on recordings and transcripts to generate reliability estimates.

We have introduced a face validity criterion, for which we conducted an item-level inspection to identify

any issues with item wording, either in accuracy (e.g., does the item accurately reflect SI or suicide attempts) or developmental appropriateness of content (e.g., concerns about reading level; items that appear more appropriate for adults). Measures received a rating of adequate for face validity if there were concerns regarding developmental appropriateness and other face validity concerns. To achieve a rating of good, a measure could have concerns regarding developmental appropriateness or other face validity concerns, but not both. Measures rated as excellent for face validity needed to have no face validity concerns.

Psychometrics evaluation of STB measures also included assessments of content validity. For measures of SI, this included determining whether both passive SI (i.e., a desire to be dead) and active SI (i.e., a desire to kill oneself) were assessed, as well as thoughts of methods or plans. For measures of suicidal behaviors, this meant evaluating whether suicidal intent was included in evaluations of suicide attempts, and if interrupted and aborted suicide attempts were separately assessed.

Construct validity consisted of convergent validity, discriminant validity, and divergent validity. Another criterion, validity generalization, refers to whether an STB measure has received validity support in multiple demographic groups and settings.

Clinical utility, including developmental appropriateness and scalability in clinical settings, was informed by eight metrics (in addition to the prerequisites regarding norms, reliability, and validity): (a) whether the measure is free (versus proprietary); (b) whether no training (or a modest amount in the case of interviews) is required for administration, scoring, or interpretation of results; (c) use of appropriate and nonstigmatizing language to describe STBs (Silverman, 2006); (d) empirical support for recommended clinical cut-off score(s) (if applicable); (e) evidence of sensitivity to change (if applicable); (e) two indices of time required for measure administration (i.e., number of items and words); (f) two indices of measure readability (i.e., Flesch Reading Ease Formula [Flesch, 1948],⁴ a widely used metric, which uses average word and sentence length to provide an estimate of reading level with scores ranging from 0 [*practically unreadable*] to 100 [*easy for any literate person*]; and Flesch-Kincaid Grade Level Formula [Flesch, 1949]⁵); and (g) temporal window covered by the measure, particularly in the case of SI. The last two clinical utility metrics warrant elaboration. Readability is an especially important consideration for youth measures that does often not to receive sufficient attention (Krause et al., 2022). In evaluating measures in the current review, we observed the recommendation that youth self-report measures should aim for a readability level of two years below the youngest age for which the measure is used (Patalay et al.,

2018). In addition, given recognition of the short temporal scale in which SI can fluctuate significantly (Kleiman et al., 2017), the finding that variability in SI may uniquely predict risk for suicide attempts (Oquendo et al., 2021), and the clinical relevance of these changes in SI immediately prior to a suicide attempt (Bryan et al., 2014; Miller & Prinstein, 2019; Rudd, 2006), the time frame covered by measures of SI was reviewed as part of an evaluation of their clinical utility. Measures assessing SI on a temporal scale of ≤ 1 week are rated high in clinical utility and measures only covering ≥ 1 month rated poor in clinical utility.

If any significant issues affecting clinical utility were detected (e.g., a measure is at the reading level of adults), its utility was rated limited. Otherwise, the lowest rating was adequate. For limited ratings, particular weight was given to aspects of individual measures that were immutable (i.e., adequate ratings for face and/or content validity; readability; administration time; and time-frame for SI severity assessment), as opposed to criteria that simply required more data. This meant if only adequate ratings were present for any immutable feature of an individual measure, the clinical utility for the measure was limited, regardless of strengths in any other aspects of the measure.

The overall measure quality classification system used in the current review was adapted from that developed by Mazefsky and colleagues (Mazefsky et al., 2021). Overall quality ratings of each measure were based on performance on four overarching categories: norms, reliability, validity, and clinical utility. Given our interest in forming recommendations for use in clinical practice, this system places greatest weight on clinical utility. To be rated excellent overall, a measure needed to meet for at least good in clinical utility and excellent in at least two other categories. To be good overall, a measure was required to be at least adequate in clinical utility and good in at least two other categories. For a measure to be adequate overall, it was required to be at least adequate across all four categories. Measures that do not meet the standard for adequate received an overall rating of limited.

Results

Stage 1 yielded a total of 1,407 articles, of which 895 were unique reports. Of these unique records, 609 records were excluded based on their titles and abstracts. An additional 196 articles were excluded based on a detailed full-text review, leaving 88 publications with 21 STB measures satisfying the Stage 1 eligibility criteria: Adult Suicidal Ideation Questionnaire (ASIQ); Adult Suicidal Ideation Questionnaire-4 (ASIQ-4); Beck Scale for Suicidal Ideation (BSS);

Concise Health Risk Tracking Self-Report (CHRT-SR); Columbia-Suicide Severity Rating Scale (C-SSRS); Depressive Symptom Index (DSI-SS) Suicidality Subscale; Geriatric Suicide Ideation Scale (GSIS); Multidimensional Adolescent Assessment Scale (MAAS); Multi-Problem Screening Inventory (MPSI) suicidal thoughts subscale; Multidimensional Suicide Inventory-28 (MSI-28) suicidal ideation subscale; Personality Assessment Inventory (PAI) suicide ideation subscale; Suicidal Ideation Questionnaire (SIQ); Suicidal Ideation Questionnaire-JR (SIQ-JR); Suicidal Ideation Questionnaire-JR-4 (SIQ-JR-4); Self-Injurious Thoughts and Behaviors Interview (SITBI); Self-Injurious Thoughts and Behaviors Interview-Revised (SITBI-R); Self-Injurious Thoughts and Behaviors Interview-Short Form (SITBI-SF); Ultra-Short-Form Suicidal Thinking Measure (USSIS); Wish to be Dead Scale (WDS); Youth Suicide Ideation Screen-3 (YSIS-3); and Forkmann et al.'s (2018) ecological momentary assessment (EMA) of suicidal ideation. See Figure 1 for the Step 1 PRISMA flowchart.

Stage 2 yielded a total of 442 articles, of which 212 were unique reports. Of these unique records, 67 records were excluded based on their titles and abstracts. An additional 128 were excluded based on full-text review, leaving 17 articles (with 21 independent samples) for eight STB measures (see Figure 2 for the Stage 2 PRISMA flowchart). Four of these measures were interview-based or clinician-rated measures of SI and/or suicidal behaviors (i.e., BSS, C-SSRS, SITBI, and SITBI-R) and another four self-report questionnaires of SI (CHRT-SR, SIQ, SIQ-JR, and SIQ-JR-4). Table 2 presents these eight measures and 17 associated articles (Allan et al., 1997; Fischer et al., 2014; Gratch et al., 2022; Hill et al., 2020; Holli et al., 2005; Jia et al., 2015; Kilincaslan et al., 2019; Mayes et al., 2018; Nandy et al., 2023; Nock et al., 2007; Núñez et al., 2024; Pinto et al., 1997; Posner et al., 2011; Potard et al., 2014; Reynolds & Mazza, 1999; Venta & Sharp, 2014; Zhang et al., 2014), including their sample characteristics. Relevant psychometric studies were conducted in seven countries: Chile, China, Finland, France, Germany, Turkey, and the United States.

Study-level measure reliability and validity characteristics and clinical cut-point performance data are presented in Tables S2 through S7. Below, we briefly describe each measure and present main findings on scalability characteristics (see Table 3 for a summary), as well as their psychometric and overall ratings (see Table 4 for a summary), highlighting major considerations that informed our overall ratings, as well as other important considerations.

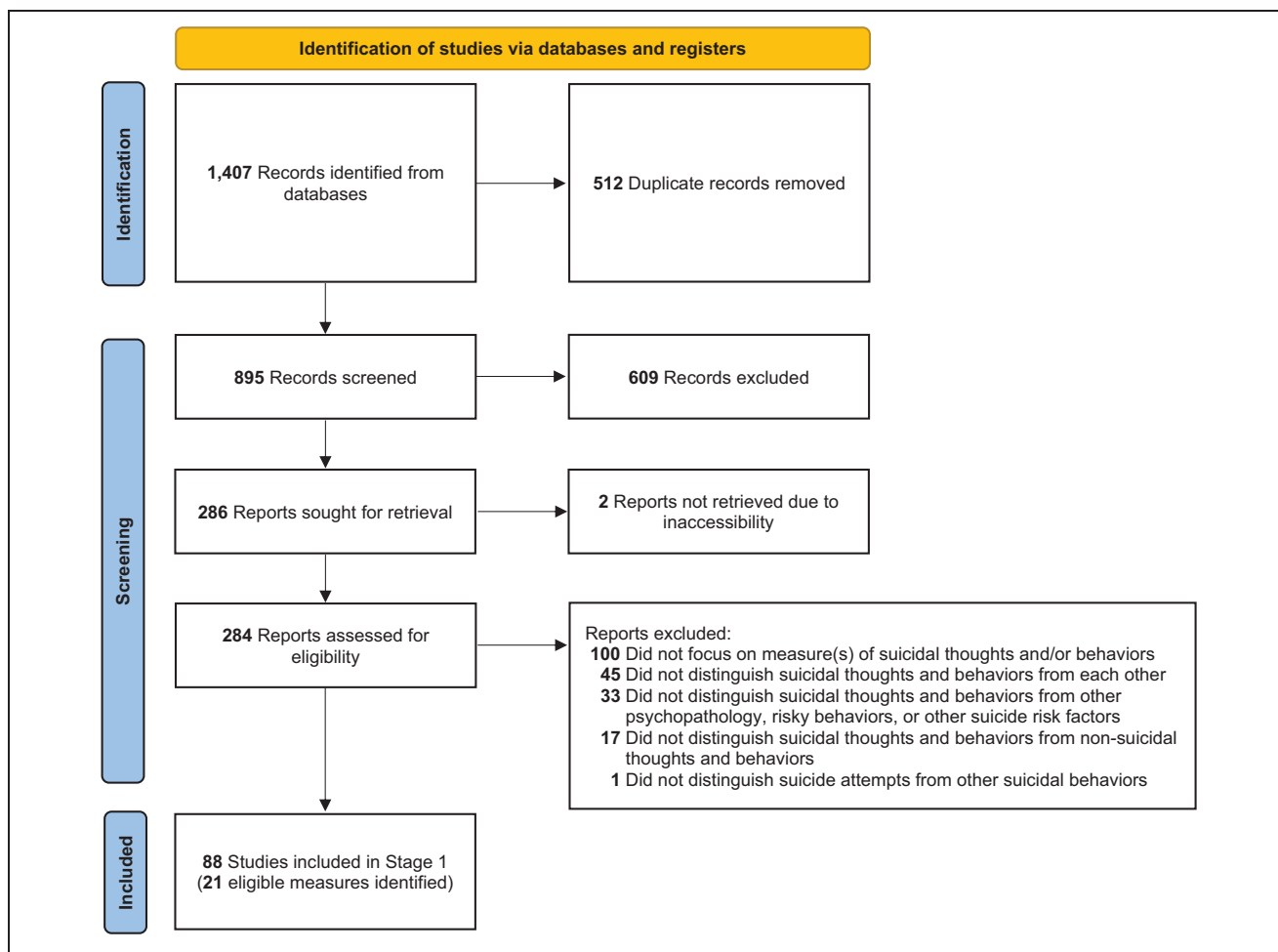


Figure 1. PRISMA Flowchart of Literature Search for STAGE I (Identifying Suicide Measures).

Beck Scale for Suicidal Ideation (BSS) Clinician-Rated Version

The BSS is a widely used interview-based measure of SI severity, originally created as a 19-item clinician-rated version (Beck et al., 1979), from which a self-report version was later developed (Beck et al., 1988). Although developed for adults, the clinician-rated version has also been employed with youth. Two studies with three independent samples of youth (Allan et al., 1997; Holli et al., 2005) were found reporting on the psychometric properties of the clinician-rated version in English and Finnish (Table S2). Only internal consistency was available as an index of reliability, and although ranging from adequate to good, was limited to α . Content validity was excellent, as the BSS included items reflecting passive SI, active SI, and suicide plans. Construct validity is supported for the English version but needs evaluation for the Finnish version. Its 1-week assessment period has potential for clinical utility.

This measure received an overall rating of limited for issues regarding face validity and thus clinical utility. Specifically, several items do not cleanly assess SI, but instead reflect related concepts (e.g., practical factors that instrumentally facilitate suicide attempts within the broader concept of capability for suicide [Klonsky et al., 2016], such as access to suicide means). There are also concerns regarding the wording of several of its items that, although perhaps more normative when the scale was created over 40 years ago, are no longer viewed as appropriate (e.g., having “courage” to “commit” suicide). In addition, although it may initially seem to be a strength of this measure that it was the only one found in the current review to have been administered to a pre-teen sample (Allan et al., 1997), the vendor company’s (Pearson Assessments) current position is that the BSS is more appropriate for adult populations (personal communication on January 3, 2023). Its sensitivity to change has not yet been evaluated in youth, an implication of

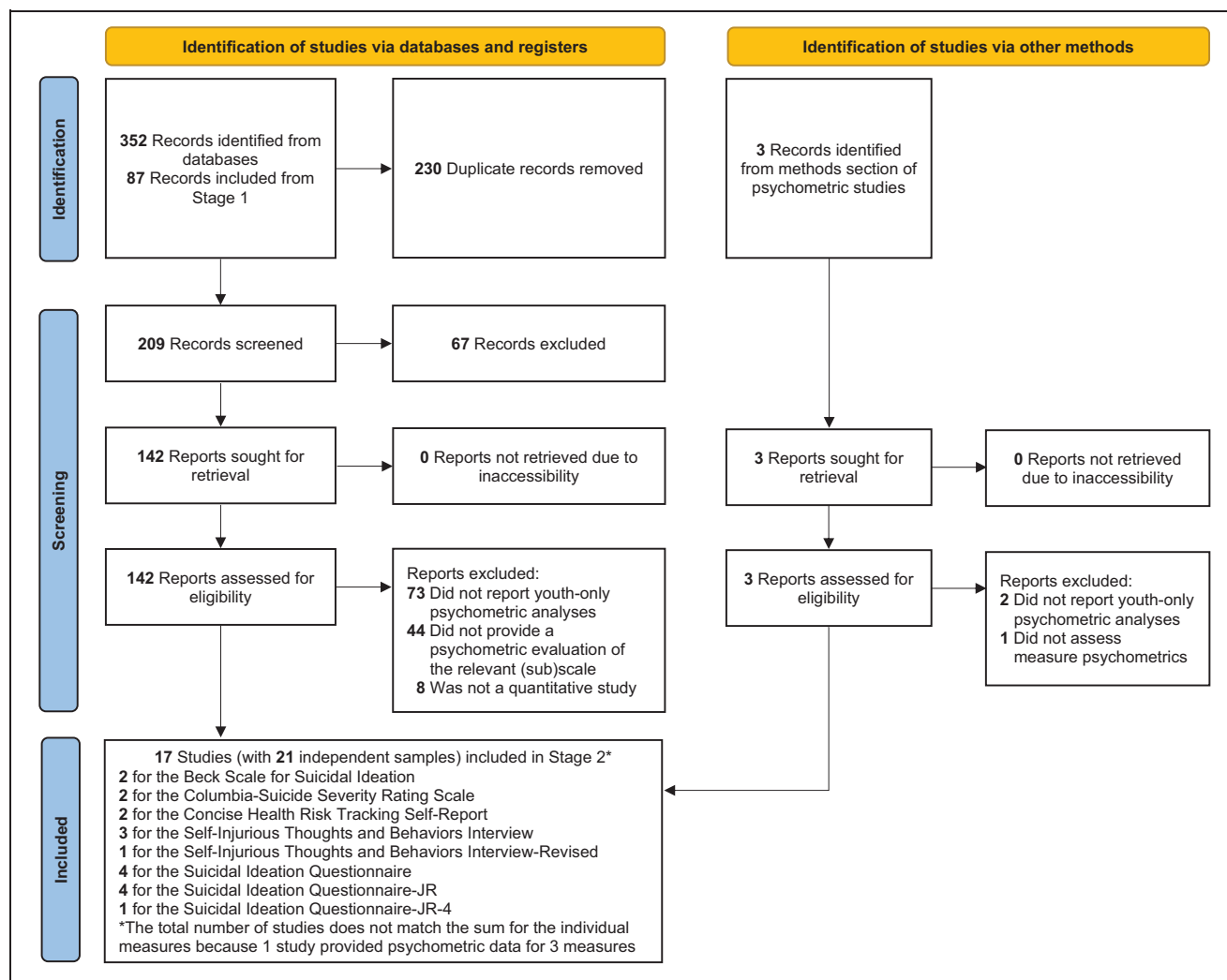


Figure 2. PRISMA Flowchart of Literature Search for Stage 2 (Identifying Psychometric Studies).

which is that its use for assessing change across treatment does not currently have empirical support. Also important to note is that psychometric evaluation was performed almost entirely with White samples (e.g., the sample in the single English-language study was 94.25% White), limiting generalization of validity across race and ethnicity. There are currently no normative data for this measure with youth. Finally, the BSS is a propriety measure, which is a limiting factor for its scalability.

Concise Health Risk Tracking Self-Report (CHRT-SR)

The CHRT-SR is a self-report measure that includes a three-item SI severity subscale. Although the CHRT-SR was originally evaluated in adults, two studies (Mayes et al., 2018; Nandy et al., 2023) were identified that provided psychometric data for the SI subscale with youth (Table S3). As was the case with the BSS, reliability was

good but limited to α . Face validity was excellent, as no issues of concern were discerned. Generally, good support for construct validity was found. In addition, general replication of sensitivity to change—receiving a good rating—was a unique strength among self-report measures. Its 1-week assessment period increases clinical utility, and its brevity facilitates scalability.

Nonetheless, the CHRT-SR received an overall limited rating because its clinical utility is significantly affected by issues of content validity, specifically the absence of items assessing passive SI. The importance of this lies in the fact that passive SI (a) is not uncommon, with estimates ranging from 5.81% for 1-year prevalence to 10.57% for lifetime prevalence, and (b) is strongly associated with risk for suicide, comparable in this regard to active SI in head-to-head comparisons (Liu, Bettis, et al., 2020). Also absent are data on norms and validity generalization.

Table 3. Measure Scalability Characteristics.

Measure	BSS Clinician-Rated Version		CHRT-SR	C-SSRS		SIQ	SIQ-JR	SIQ-JR-4	SITBI	SITBI-R
STBs	PSI, ASI, plans, preparatory acts	ASI, plans	PSI, ASI, SA, interrupted SA, aborted SA, plans, preparatory acts	PSI, ASI, SA, interrupted SA, aborted SA, plans, preparatory acts	PSI, ASI	PSI, ASI	PSI, ASI	PSI, ASI	ASI, SA, interrupted SA, aborted SA, plans, preparatory acts	PSI, ASI, SA, interrupted SA, aborted SA, plans, preparatory acts
Format	Interview	Questionnaire	Questionnaire	Interview	Questionnaire	Questionnaire	Questionnaire	Questionnaire	Interview	Interview
Standard Time	1 Week	1 Week	1 Week	Current, Lifetime	1 Month	1 Month	1 Month	1 Month	1 Week, 1 Month, 1 Year, Lifetime	1 Week, 1 Month, 1 Year, Lifetime
Frame(s)										
# of Words ^b	833	136		1,718	438	230		–	1,742; 5,003 ^a	5,449
# of Items	19	3		16	30	15		4	72; 198 ^a	111
Reverse-Scored Items	No	No		No	No	No		No	No	No
Flesch Reading Ease	68.5	77.7		64.4	85.4	82.9		–	82.6; 78.8 ^a	74.1
Formula ^{c,d}										
Flesch-Kincaid Grade	5.6	5.2		8.2	2.6	4.0		–	3.5, 6.6 ^a	5.7
Level Formula ^c										
Training Required	No ^e	No		Yes	No	No		No	Yes	Yes
Proprietary	Yes	No		No	Yes	Yes		Yes	No	No
Normative Data:	No ^{2,3}	No ²		No ^{2,6}	No ^{1,2,4}	No ^{1,2,5}		No ¹	Not applicable	Not applicable
Clinical	No ^{2,3}	No ²		No ^{2,6}	Yes ^{1,2,4}	Yes ^{2,5}		No ¹	Not applicable	Not applicable
Nonclinical										

Note. ASI = active suicidal ideation; BSS = Beck Scale for Suicidal Ideation; CHRT-SR = Concise Health Risk Tracking Self-Report; C-SSRS = Columbia-Suicide Severity Rating Scale; PSI = passive suicidal ideation; SA = suicide attempts; SIQ = Suicidal Ideation Questionnaire, SIQ-JR = Suicidal Ideation Questionnaire-JR, SIQ-JR-4 = Suicidal Ideation Questionnaire-JR-4; SITBI = Self-Injurious Thoughts and Behaviors Interview; SITBI-R = Self-Injurious Thoughts and Behaviors Interview-Revised; STBs = suicidal thoughts and behaviors.

^aValues are presented for the SITBI short-form and SITBI long-form, respectively. ^bValues are for English versions of each measure. ^cValues are presented for English versions of each measure and calculated to one decimal point. These values were not calculated for the SIQ-JR-4 because this measure only exists in Chinese. ^dValues range from 1 to 100, with higher scores indicating greater readability. ^eBeyond the training required of a clinician, no measure-specific training is required. ¹Chinese version. ²English version. ³Finnish version. ⁴French version. ⁵Spanish version. ⁶Turkish version.

Table 4. Summary of Norms, Reliability, Validity, and Clinical Utility Ratings by Measure.

Criterion	BSS Clinician-Rated Version	CHRT-SR	C-SSRS	SIQ	SIQ-JR	SIQ-JR-4	SITBI	SITBI-R
Overall	Limited ^{2,3}	Limited ²	Adequate ^{2,7}	Limited ^{1,2,4}	Limited ^{1,2,6}	Limited ¹	Adequate ^{2,5}	Adequate ²
Norms ^a	No data ^{2,3}	No data ²	No data ^{2,7}	Adequate ^{1,2,4}	No data ¹ Adequate ^{2,6}	No data ¹	Not applicable	Not applicable
Internal Consistency ^a	Adequate ² Good ³	Good ²	Adequate ^{2,7}	Adequate ² Good ¹	Adequate ^{1,2} Excellent ⁶	No data ¹	Not applicable	Not applicable
Test-Retest Reliability	No data ^{2,3}	No data ²	No data ^{2,7}	Too Excellent ⁴ No data ^{1,2}	No data ^{1,6} Adequate ²	No data ¹	Adequate ^{2b,5}	No data ²
Inter-Rater Reliability ^c	No data ^{2,3}	Not applicable	No data ² Adequate ⁷	Adequate ⁴ Not applicable	Not applicable	Not applicable	Adequate ^{2,5}	Adequate ²
Face Validity Content Validity	Adequate Excellent	Excellent Adequate	Excellent Excellent	Adequate Excellent	Adequate Excellent	Excellent Excellent	Excellent Adequate (SI)/ Excellent	Excellent Excellent
Construct Validity (convergent, discriminant, and divergent validity)	Insufficient data ³ Adequate ²	Good ²	Adequate ⁷ Excellent ²	Adequate ² Good ^{1,4}	Adequate ^{1,2,6}	Adequate ¹	(suicidal behaviors) Adequate ⁵ Excellent ²	Adequate ²
Validity Generalization	Insufficient data ^{2,3}	Insufficient data ²	Insufficient data ^{2,7}	Insufficient data ^{1,2} Adequate ⁴	Insufficient data ^{1,2,6}	Insufficient data ¹	Insufficient data ^{2,5}	Insufficient data ²
Sensitivity to Change ^d	No data ^{2,3}	Good ²	No data ⁷ Excellent ²	No data ^{1,2,4}	No data ^{1,2,6}	No data ¹	Not applicable	Not applicable
Clinical Utility	Limited ^{2,3}	Limited ²	Adequate ^{2,7}	Limited ^{1,2,4}	Limited ^{1,2,6}	Limited ¹	Adequate ^e	Adequate ^e

Note. BSS = Beck Scale for Suicidal Ideation; CHRT-SR = Concise Health Risk Tracking Self-Report; C-SSRS = Columbia-Suicide Severity Rating Scale; SI = suicidal ideation; SIQ = Suicidal Ideation Questionnaire; SIQ-JR = Suicidal Ideation Questionnaire-JR; SIQ-JR-4 = Suicidal Ideation Questionnaire-JR-4; SITBI = Self-Injurious Thoughts and Behaviors Interview; SITBI-R = Self-Injurious Thoughts and Behaviors Interview-Revised.

^aOnly applicable to measures with (sub)scale scores. ^bAdequate reliability for lifetime presence (but below-adequate reliability for lifetime frequency) of suicidal thoughts and behaviors. ^cOnly applicable to interview-based measures. ^dOnly applicable to measures with suicidal ideation (sub)scale scores. ^eSelective reduction of the measure to core questions would be required for use in clinical settings.

¹Chinese version. ²English version. ³Finnish version. ⁴French version. ⁵German version. ⁶Spanish version. ⁷Turkish version.

Columbia-Suicide Severity Rating Scale (C-SSRS)

The C-SSRS is an interview-based measure designed to assess SI and suicidal behavior, including suicide attempts, interrupted attempts, aborted attempts, and preparatory behavior. Two psychometric papers (with three independent samples [Kilincaslan et al., 2019; Posner et al., 2011]) were identified for two versions of the C-SSRS interview (English and Turkish; Table S4). For both the English and Turkish C-SSRS, reliability was good but limited to internal consistency (specifically α). Inter-rater reliability for the Turkish C-SSRS was adequate. Face validity was excellent, as was content validity, as the C-SSRS is comprehensive in assessing passive SI, active SI, SI methods, plans, and multiple forms of suicidal behaviors. Construct validity was excellent for the English version of the measure and adequate for the Turkish version. This measure was also unique among all instruments in this review for having excellent evidence of sensitivity to change (for the English version; no sensitivity-to-change data are currently available for the Turkish version). In support of the clinical utility, the English C-SSRS is unique among the measures in this review in possessing strong evidence of scalability; it is nonproprietary and already recommended by the Joint Commission (albeit for suicide risk screening with the triage version of the C-SSRS, rather than with the full version for assessment of STBs, the focus of the current review; The Joint Commission, n.d.) and so many health care settings already employ this measure in their standard protocols.

As for overall measure ratings, the English and Turkish versions were both adequate, primarily because of the absence of normative data for the suicidal ideation scales. Finally, both versions are lacking evidence of validity generalization.

Suicidal Ideation Questionnaire (SIQ), SIQ-JR, SIQ-JR-4

Only two self-report measures in the current review were originally designed for adolescents, the SIQ (30 items) and SIQ-JR (15 items), both measures of SI severity. Four studies (seven independent samples; see Table S5) provided data on the SIQ (in Chinese, English, and French; Jia et al., 2015; Pinto et al., 1997; Potard et al., 2014; Zhang et al., 2014) and four studies (with five independent samples; see Table S6) provided evaluations of the SIQ-JR (in Chinese, English, and Spanish; Hill et al., 2020; Núñez et al., 2024; Reynolds & Mazza, 1999; Zhang et al., 2014). A unique strength of the SIQ (in all languages) and SIQ-JR (in English and Spanish, but not Chinese) relative to other measures in this review is that they had normative data. This is a particularly

notable in the case of the English SIQ-JR because of the large minority representation in its study samples (100% American Indian in one sample [Hill et al., 2020] and 98.9% minority identity in the other [Reynolds & Mazza, 1999]), affording greater generalizability to demographics traditionally underreported in suicide research (Cha et al., 2018). Internal consistency was adequate for the English SIQ and SIQ Jr, as well as the Chinese SIQ Jr; good for the Chinese SIQ, and too excellent for the French SIQ on account of its very high α combined with its length. Again, internal consistency for these measures was almost entirely based on α . The French SIQ and English SIQ-JR also had adequate test-retest reliability. Content validity was excellent and construct validity was rated adequate to good for both measures across all languages.

Despite the mostly unique strengths mentioned above, both measures were rated limited overall, chiefly because of multiple face validity concerns. As a few examples, both measures included an item that may not be developmentally appropriate (thoughts of writing a will). They also include an item on thoughts of people dying, which is unlikely to reflect SI and thus likely to generate false positives. Yet another item on general thoughts of death could be similarly endorsed for reasons unrelated to SI (e.g., concern about the health of a loved one) or even for reasons antithetical to SI (e.g., concerns about one's own health). This is particularly true in certain populations (e.g., general community or patients with serious physical health conditions).

Two other concerns with this measure warrant mention that factored in the limited overall rating for these measures. The one-month period covered by these measures poses challenges to clinical utility, in that it does not afford sufficient temporal resolution for observing relatively short-term fluctuations in SI, and it is much larger than the typical window between psychotherapy sessions (e.g., 1–2 weeks), limiting its utility for assessing SI change across treatment. In addition, the frequency-response anchors for these measures also introduce challenges of their own. In practice, the lowest anchor (no lifetime history of the item), especially when considered with the second lowest anchor (lifetime history of the item, but not in the prior month) can be nonintuitive, introducing variance in responding; some youth with even severe past SI may endorse the lowest anchor when they no longer experience SI. Alternatively, if they do accurately adhere to the anchors, their lowest possible score will be higher than that of peers with no SI history, thus confounding current SI with past SI.

Except for the French version of the SIQ, these measures do not currently have support for validity generalization, and no versions of either measure have evidence of sensitivity to change. Finally, although multiple

studies evaluated sensitivity and specificity of recommended clinical cut-points for both measures, the performance of these cut-points was highly variable, possibly in part a function of differences in study sample demographics.

A third form of the SIQ, the SIQ-JR-4 (four items), was derived from the SIQ-JR and exists only in one study with a Chinese sample (Zhang et al., 2014). It was found to have excellent face validity and content validity, but its psychometrics were otherwise largely undefined, with data limited to evaluations of convergent validity and a clinical cut-point in this single study. The paucity of psychometric data resulted in this measure receiving a limited rating for clinical utility and for the measure overall.

Self-Injurious Thoughts and Behaviors Interview (SITBI), SITBI-Revised (SITBI-R)

Like the C-SSRS, the SITBI and SITBI-R are interviews designed to capture active SI (and passive SI in the SITBI-R) and suicidal behavior, including suicide attempts, interrupted attempts, aborted attempts, and preparatory acts. Although they are the longest measures in this review, they are also the most comprehensive in coverage of topography, functions, and other aspects of STBs. Three studies (see Table S7) evaluated the psychometric properties of the SITBI (in English and German; Fischer et al., 2014; Nock et al., 2007; Venta & Sharp, 2014) and one the SITBI-R (in English; Gratch et al., 2022). One study for each of the English SITBI (Venta & Sharp, 2014) and SITBI-R (Gratch et al., 2022) included large racial/ethnic minority representation (66.6%–75.2% minority identity), enhancing generalizability. Test–retest reliability but was generally adequate for the SITBI, but no test–retest reliability data are currently available for the SITBI-R. Inter-rater reliability was also adequate for both measures. No issues with face validity were detected. Content validity was mixed for the SITBI, in that it did not assess passive SI, but no issues were found in its assessment of suicidal behaviors. The absence of passive SI was addressed to a degree in the SITBI-R, although here it was assessed only if active SI is first endorsed. Construct validity ranged from adequate (German SITBI and SITBI-R) to excellent (English SITBI). The clinical utility of both SITBI and SITBI-R was adequate, given feasibility in clinical settings is facilitated by the fact that these measures can be reduced to only the items of principal clinical interest.

The overall ratings were adequate for all versions of the SITBI and SITBI-R, because of adequate reliability for all measures, and adequate construct validity in the cases of the German SITBI and the SITBI-R. Finally, as

with almost all measures in this review, both the SITBI and SITBI-R do not currently have data supporting validity generalization.

Discussion

In systematically evaluating existing measures of STBs in youth, a primary aim of the current review was to offer recommendations on measures for use in clinical practice and research settings. Self-report measures and interviews are complementary to one another, both offering unique advantages. Self-report measures are often more efficient, which is valuable in many health care settings because of the inherent advantages in clinical utility and scalability this provides. They also provide a medium more conducive to responding openly to questions about personally sensitive content, such as STBs, especially in youth (Tourangeau & Yan, 2007; Turner et al., 1998). Indeed, self-report measures have been found to yield higher reporting of STBs than interview-based measures (Deming et al., 2021; Kaplan et al., 1994). However, interviews allow for more accurate categorization of STBs, particularly in differentiating between different forms of suicidal behavior and between suicide attempts and nonsuicidal self-injury, because interviewers can clarify and validate affirmative responses relative to operational definitions of these behaviors recognized within the field (Posner et al., 2011, 2014). Therefore, it may not be a coincidence that all self-report measures identified in the current review only assess SI, whereas several interview-based measures are more comprehensive, assessing STBs. For these reasons, when considering protocols for assessing STBs, clinicians and researchers may benefit most from incorporating into their repertoire both a self-report dimensional measure of SI severity and an interview-based measure that includes suicidal behaviors.

Among the eight measures identified in the current review, only three achieved overall ratings of adequate. All others were rated limited. Notably, the measures that performed the best (i.e., adequate) were three of the four interview-based ones—the C-SSRS, SITBI, and SITBI-R. All self-report STB measures were designated as limited. Of note, the aforementioned interviews were also among the most recently developed measures, whereas the SIQ, and SIQ-JR (and BSS) are at least 35 years old. The difference in overall ratings between interview-based and self-report measures may therefore be partly a function of advances in measure construction and consensus definitions of STBs that have emerged over this span of time.

For interviews, both the C-SSRS and SITBI-R offer unique strengths. Although both may have promise for use in clinical and research contexts, the C-SSRS may

have particular strengths in clinical settings, given its clinical potential and scalability (i.e., brief administration, training is freely available and relatively brief, and there is strong evidence of scalability in clinical settings), in addition to generally excellent validity in clinical samples for the English version. Nonetheless, greater evaluations supporting clinical utility are required. The SITBI-R may have potential utility in clinical settings too, but given its length, would first require trimming to central questions of concern. In addition, as the one psychometric study conducted to date with youth featured a mixed sample recruited from the community (Gratch et al., 2022), further psychometric evaluation with clinical youth samples is required before it can be recommended for use in treatment settings. The SITBI-R is also more comprehensive, however, and so may be most suitable for research contexts where greater detail regarding STB phenomenology is desired and time is not a major constraint. Although both the C-SSRS and SITBI-R assess suicide plans, for example, the SITBI-R is unique among the measures covered in this review in specifically assessing three critical components of suicide plans: how, where, and when.

Regarding self-report measures, unfortunately no measure met the threshold to recommend their use as a complement to interview-based approaches. Concerningly, each has major limitations affecting clinical utility that are inherent to the measure and cannot be remedied through additional psychometric study. Also worth mentioning is that several commonly used measures of STBs are absent from the current review because they are inconsistent with currently accepted taxonomies and definitions of self-injurious thoughts and behaviors (Corsby et al., 2011; National Action Alliance for Suicide Prevention, 2014) and thus did not meet our inclusion criteria.⁶ Although this consideration may be less an issue for suicide risk screening, as is often done with such measures as the SBQ-R, it becomes a significant concern if the intention is to assess SI severity or the occurrence of suicidal behaviors, particularly for accurately monitoring symptom course and treatment progress. Collectively, these findings are part of a broader pattern in the field of a lack of adherence to definitional and measurement precision in the assessment and study of STBs. Indeed, even in a review of randomized controlled trials for suicidal behavior (in adults), only approximately 20% used validated measures (Brown & Green, 2014). The lack of adequate self-report SI severity measures in youth has clear and strong implications. If accurate assessment is a basic foundation of clinical care and research, the absence of suitable measures jeopardizes these operations, and research findings on which existing measures are based become questionable.

The one identified study that reported on the psychometrics of an EMA measure of STBs (Forkmann et al., 2018) was with an adult sample and thus ineligible for inclusion in the current review. Development of EMA youth measures of STBs to complement traditional questionnaire and interview-based measures is particularly needed, given recent theoretical and empirical support for the view that SI may fluctuate significantly over brief temporal scales and that these fluctuations may be clinically important (Bryan et al., 2014; Kleiman et al., 2017; Miller & Prinstein, 2019; Oquendo et al., 2021; Rudd, 2006). Psychometrically supported EMA STB measures may hold unique value for close monitoring of individuals during periods of symptom acuity and elevated risk (e.g., post-discharge from an emergency department or inpatient hospitalization).

Also of note is the critical need for development and/or validation of STB measures with three populations understudied or altogether unaccounted for in the studies in the current review: preadolescent children, sexual and gender minority (SGM) youth, and racial/ethnic minority youth. Only one study included in the current review featured a preadolescent sample (Allan et al., 1997). As noted above, even in this study, the measure used (i.e., the clinician-rated BSS) was one that its developers have not recommended for use with youth. Yet, preadolescent children have received growing attention as a population of concern (National Institute of Mental Health, 2021), as suicide rates in this demographic have increased over the last decade (Centers for Disease Control and Prevention, 2020), and prevalence of SI in the general community is high (Liu et al., 2022). Not only are SGM and racial/ethnic minority youth health disparity (National Institutes of Health, 2021) and underserved (National Institutes of Health, 2019) populations that have historically been understudied in suicide research (Cha et al., 2018), but SGM youth have experienced persistent disparities in STBs (Liu, Walsh, et al., 2020; Thoma et al., 2019), and Black youth have experienced an alarming increase in suicide rates in recent years (Bridge et al., 2015; Sheftall et al., 2022).

It is always important to take into account contextual factors that may affect reporting of STBs, but this is especially true for minoritized populations (Molock et al., 2023). Providing some support for the need to evaluations and STB measures, how they are administered, and the potential need for adaptation of the measures themselves and/or their administration, one study found that minority youth in general, and Black youth in particular, at elevated risk for STBs may be less inclined to answer questions about these outcomes (Anderson et al., 2015). Also supporting the need to consider cultural appropriateness of specific items in existing measures or the development of new ones, one of the studies

included in this review noted that the SIQ item of “writing a will” was incongruent with normative practices among Native American and Indigenous communities (Hill et al., 2020). In addition, incorporating consideration of contextual factors (e.g., stigma, cultural expectations regarding mental health in general and suicide specifically) may enhance reporting of STBs among racial/ethnic minority individuals (Molock et al., 2023).

In summary, our findings highlight important limitations of measures commonly used to assess STBs in youth. Although the C-SSRS and SITBI have evidence to support their use in clinical and research settings, there are currently no self-report measures that adequately assess SI severity in youth. Particularly concerning too is the want of support for the clinical utility of existing measures. There is a need rigorously to assess their clinical utility, and particularly to develop and evaluate self-report measures of STBs for use in clinical settings, including behavioral health and nonbehavioral health contexts. As health care systems are increasingly recognizing the need to identify and treat STBs in youth, they are underequipped to assess STBs and their course over treatment.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Preparation of this manuscript was supported, in part, by the National Institute of Mental Health of the National Institutes of Health under Award Numbers RF1MH120830, R01MH115905, R01MH124899, and R21MH130767 to R.T.L.; K23MH122737 to A.H.B.; an American Foundation for Suicide Prevention grant (PDF-0-095-19) to H.R.L.; and National Science Foundation Graduate Research Fellowships to R.F.L.W. and A.E.S. R.T.L. serves as a consultant to Relmada Therapeutics. The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies or Relmada Therapeutics.

ORCID iD

Richard T. Liu  <https://orcid.org/0000-0002-1367-9716>

Supplemental Material

Supplemental material for this article is available online.

Notes

1. Suicide attempts involves deliberate self-injury with non-zero suicidal intent, whereas nonsuicidal self-injury occurs

in the absence of suicidal intent. (Corsby et al., 2011; Posner et al., 2007)

2. One study (Forkmann et al., 2018) presented an ecological momentary assessment measure of STBs. As no name was provided for this measure, it was not possible to conduct a systematic search for potential studies providing additional data on its psychometric properties.
3. Specification of a clinically severe sample is made because in general community samples, the relatively absence of SI may be expected to be fairly stable over time.
4. Flesch Reading Ease Formula = $206.835 - 1.015 \times (\text{words/sentences}) - 84.6 \times (\text{syllables/words})$.
5. Flesch-Kincaid Grade Level Formula = $0.39 \times (\text{words/sentences}) + 11.8 \times (\text{syllables/words}) - 15.59$.
6. As examples, the Suicide Behaviors Questionnaire-Revised (SBQ-R; Osman et al., 2001) does not differentiate suicidal thoughts from suicidal behaviors; and the Kiddie Schedule for Affective Disorders and Schizophrenia (K-SADS; Kaufman et al., 2016), without modification, does not differentiate suicide attempts from other suicidal behaviors.

References

- Allan, W. D., Kashani, J. H., Dahlmeier, J., Taghizadeh, P., & Reid, J. C. (1997). Psychometric properties and clinical utility of the scale for suicide ideation with inpatient children. *Journal of Abnormal Child Psychology*, 25, 465–473. <https://doi.org/10.1023/A:1022633714888>
- Anderson, L. M., Lowry, L. S., & Wuensch, K. L. (2015). Racial differences in adolescents' answering questions about suicide. *Death Studies*, 39, 600–604. <https://doi.org/10.1080/07481187.2015.1047058>
- Arakelyan, M., Freyleue, S., Avula, D., McLaren, J. L., O'Malley, A. J., & Leyenaar, J. K. (2023). Pediatric mental health hospitalizations at acute care hospitals in the US, 2009-2019. *JAMA*, 329, 1000–1011. <https://doi.org/10.1001/jama.2023.1992>
- Baek, I. C., Jo, S., Kim, E. J., Lee, G. R., Lee, D. H., & Jeon, H. J. (2021). A review of suicide risk assessment tools and their measured psychometric properties in Korea. *Frontiers in Psychiatry*, 12, Article 679779. <https://doi.org/10.3389/fpsyt.2021.679779>
- Batterham, P. J., Ftanou, M., Pirkis, J., Brewer, J. L., Mackinnon, A. J., Beautrais, A., Kate Fairweather-Schmidt, A., & Christensen, H. (2015). A systematic review and evaluation of measures for suicidal ideation and behaviors in population-based research. *Psychological Assessment*, 27, 501–512. <https://doi.org/10.1037/pas0000053>
- Beck, A. T., Kovacs, M., & Weissman, A. (1979). Assessment of suicidal intention: The Scale for Suicide Ideation. *Journal of Consulting and Clinical Psychology*, 47, 343–352. <https://doi.org/10.1037/0022-006X.47.2.343>
- 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](https://doi.org/10.1002/1097-4679(198807)44:4<499::AID-JCLP2270440404>3.0.CO;2-6)

- Becker-Haimes, E. M., Tabachnick, A. R., Last, B. S., Stewart, R. E., Hasan-Granier, A., & Beidas, R. S. (2020). Evidence base update for brief, free, and accessible youth mental health measures. *Journal of Clinical Child and Adolescent Psychology, 49*, 1–17. <https://doi.org/10.1080/15374416.2019.1689824>
- Bommersbach, T. J., McKean, A. J., Olfson, M., & Rhee, T. G. (2023). National trends in mental health-related emergency department visits among youth, 2011–2020. *JAMA, 329*, 1469–1477. <https://doi.org/10.1001/jama.2023.4809>
- Borschmann, R., Hogg, J., Phillips, R., & Moran, P. (2012). Measuring self-harm in adults: A systematic review. *European Psychiatry, 27*, 176–180. <https://doi.org/10.1016/j.eurpsy.2011.04.005>
- Bridge, J. A., Asti, L., Horowitz, L. M., Greenhouse, J. B., Fontanella, C. A., Sheftall, A. H., Kelleher, K. J., & Campo, J. V. (2015). Suicide trends among elementary school-aged children in the United States from 1993 to 2012. *JAMA Pediatrics, 169*, 673–677. <https://doi.org/10.1001/jamapediatrics.2015.0465>
- Brown, G. K., & Green, K. L. (2014). A review of evidence-based follow-up care for suicide prevention: Where do we go from here? *American Journal of Preventive Medicine, 47*, S209–S215. <https://doi.org/10.1016/j.amepre.2014.06.006>
- Bryan, C. J., Bryan, A. O., Ray-Sannerud, B. N., Etienne, N., & Morrow, C. E. (2014). Suicide attempts before joining the military increase risk for suicide attempts and severity of suicidal ideation among military personnel and veterans. *Comprehensive Psychiatry, 55*, 534–541. <https://doi.org/10.1016/j.comppsy.2013.10.006>
- Carter, T., Walker, G. M., Aubeeluck, A., & Manning, J. C. (2019). Assessment tools of immediate risk of self-harm and suicide in children and young people: A scoping review. *Journal of Child Health Care, 23*, 178–199. <https://doi.org/10.1177/1367493518787925>
- Centers for Disease Control and Prevention. (2020). *Web-Based Injury Statistics Query and Reporting System (WISQARS)*. <http://www.cdc.gov/injury/wisqars/Leading-Causes.html>
- Cha, C. B., Tezanos, K. M., Peros, O. M., Ng, M. Y., Ribeiro, J. D., Nock, M. K., & Franklin, J. C. (2018). Accounting for diversity in suicide research: Sampling and sample reporting practices in the United States. *Suicide and Life-Threatening Behavior, 48*, 131–139. <https://doi.org/10.1111/sltb.12344>
- Corsby, A. E., Ortega, L., & Melanson, C. (2011). *Self-directed violence surveillance: Uniform definitions and recommended data elements* (Version 1.0). Centers for Disease Control and Prevention, National Center for Injury Prevention and Control.
- Cronbach, L. J., & Shavelson, R. J. (2004). My current thoughts on coefficient alpha and successor procedures. *Educational and Psychological Measurement, 64*, 391–418. <https://doi.org/10.1177/0013164404266386>
- Curtin, S. C., Hedegaard, H., Minino, A. M., & Warner, M. (2016). QuickStats: Death rates for motor vehicle traffic injury, suicide, and homicide among children and adolescents aged 10–14 years—United States, 1999–2014. *MMWR. Morbidity and Mortality Weekly Report, 65*, 1203. <https://doi.org/10.15585/mmwr.mm6543a8>
- Deming, C. A., Harris, J. A., Castro-Ramirez, F., Glenn, J. J., Cha, C. B., Millner, A. J., & Nock, M. K. (2021). Inconsistencies in self-reports of suicidal ideation and attempts across assessment methods. *Psychological Assessment, 33*, 218–229. <https://doi.org/10.1037/pas0000976>
- Erford, B. T., Jackson, J., Bardhoshi, G., Duncan, K., & Atalay, Z. (2018). Selecting suicide ideation assessment instruments: A meta-analytic review. *Measurement and Evaluation in Counseling and Development, 51*, 42–59. <https://doi.org/10.1080/07481756.2017.1358062>
- Fischer, G., Ameis, N., Parzer, P., Plener, P. L., Groschwitz, R., Vonderlin, E., Kölch, M., Brunner, R., & Kaess, M. (2014). The German version of the Self-Injurious Thoughts and Behaviors Interview (SITBI-G): A tool to assess non-suicidal self-injury and suicidal behavior disorder. *BMC Psychiatry, 14*, 1–8. <https://doi.org/10.1186/s12888-014-0265-0>
- Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology, 32*, 221–233. <https://doi.org/10.1037/h0057532>
- Flesch, R. (1949). *The art of readable writing*. Harper.
- Flores-Kanter, P. E., & Alvarado, J. M. (in press). The state of open science practices in psychometric studies of suicide: A systematic review. *Assessment*. Advance online publication. <https://doi.org/10.1177/10731911241236315>
- Forkmann, T., Spangenberg, L., Rath, D., Hallensleben, N., Hegerl, U., Kersting, A., & Glaesmer, H. (2018). Assessing suicidality in real time: A psychometric evaluation of self-report items for the assessment of suicidal ideation and its proximal risk factors using ecological momentary assessments. *Journal of Abnormal Psychology, 127*, 758–769. <https://doi.org/10.1037/abn0000381>
- Gleeson, H., Roesch, C., Hafford-Letchfield, T., & Ellmers, T. (2021). Assessing suicide ideation among older adults: A systematic review of screening and measurement tools. *International Psychogeriatrics, 34*, 439–452. <https://doi.org/10.1017/S1041610221002659>
- Gratch, I., Fernandes, S. N., Bell, K. A., Pollak, O. H., Fox, K. R., Tezanos, K., Ebo, T., & Cha, C. B. (2022). Self-Injurious Thoughts and Behaviors Interview–Revised (SITBI-R): Reliability, validity, and inter-informant agreement in an adolescent sample. *Journal of Clinical Child and Adolescent Psychology, 51*, 484–494. <https://doi.org/10.1080/15374416.2021.1901229>
- Hill, K., Van Eck, K., Goklish, N., Larzelere-Hinton, F., & Cwik, M. (2020). Factor structure and validity of the SIQ-JR in a southwest American Indian tribe. *Psychological Services, 17*, 207–216. <https://doi.org/10.1037/ser0000298>
- Holi, M. M., Pelkonen, M., Karlsson, L., Kiviruusu, O., Ruutu, T., Heilä, H., Tuisku, V., & Marttunen, M. (2005). Psychometric properties and clinical utility of the Scale of Suicidal Ideation (SSI) in adolescents. *BMC Psychiatry, 5*, 1–8. <https://doi.org/10.1186/1471-244X-5-8>
- Jia, C. X., Wang, X. T., & Zhao, Z. T. (2015). Psychometric properties of the Suicidal Ideation Questionnaire among Chinese high school students. *Omega, 70*, 195–207. <https://doi.org/10.2190/OM.70.2.c>

- The Joint Commission. (n.d.). *Suicide prevention*. <https://www.jointcommission.org/resources/patient-safety-topics/suicide-prevention/>
- Jones, S. E., Ethier, K. A., Hertz, M., DeGue, S., Le, V. D., Thornton, J., Lim, C., Dittus, P. J., & Geda, S. (2022). Mental health, suicidality, and connectedness among high school students during the COVID-19 Pandemic—Adolescent behaviors and experiences survey, United States, January–June 2021. *Morbidity and Mortality Weekly Report*, 71, 16–21.
- Kaplan, M. L., Asnis, G. M., Sanderson, W. C., Keswani, L., de Lecuona, J. M., & Joseph, S. (1994). Suicide assessment: Clinical interview vs. self-report. *Journal of Clinical Psychology*, 50, 294–298. [https://doi.org/10.1002/1097-4679\(199403\)50:2<294::AID-JCLP2270500224>3.0.CO;2-R](https://doi.org/10.1002/1097-4679(199403)50:2<294::AID-JCLP2270500224>3.0.CO;2-R)
- Kaufman, J., Birmaher, B., Axelson, D., Perepletchikova, F., Brent, D., & Ryan, N. (2016). Schedule for Affective Disorders and Schizophrenia for School-Aged Children: Present and Lifetime Version (K-SADS-PL) DSM-5 November 2016 working draft. New Haven: Yale University, Child and Adolescent Research and Education.
- Kilincaslan, A., Gunes, A., Eskin, M., & Madan, A. (2019). Linguistic adaptation and psychometric properties of the Columbia-Suicide Severity Rating Scale among a heterogeneous sample of adolescents in Turkey. *International Journal of Psychiatry in Medicine*, 54, 115–132. <https://doi.org/10.1177/0091217418791454>
- Kleiman, E. M., Turner, B. J., Fedor, S., Beale, E. E., Huffman, J. C., & Nock, M. K. (2017). Examination of real-time fluctuations in suicidal ideation and its risk factors: Results from two ecological momentary assessment studies. *Journal of Abnormal Psychology*, 126, 726–738. <https://doi.org/10.1037/abn0000273>
- Klonsky, E. D., May, A. M., & Saffer, B. Y. (2016). Suicide, suicide attempts, and suicidal ideation. *Annual Review of Clinical Psychology*, 12, 307–330. <https://doi.org/10.1146/annurev-clinpsy-021815-093204>
- Krause, K. R., Jacob, J., Szatmari, P., & Hayes, D. (2022). Readability of commonly used quality of life outcome measures for youth self-report. *International Journal of Environmental Research and Public Health*, 19, 9555. <https://doi.org/10.3390/ijerph19159555>
- Kreuze, E., & Lamis, D. A. (2018). A review of psychometrically tested instruments assessing suicide risk in adults. *Omega*, 77, 36–90. <https://doi.org/10.1177/0030222816688151>
- Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, 16, 606–613. <https://doi.org/10.1046/j.1525-1497.2001.016009606.x>
- Liu, R. T., Bettis, A. H., & Burke, T. A. (2020). Characterizing the phenomenology of passive suicidal ideation: A systematic review and meta-analysis of its prevalence, psychiatric comorbidity, correlates, and comparisons with active suicidal ideation. *Psychological Medicine*, 50, 367–383. <https://doi.org/10.1017/S003329171900391X>
- Liu, R. T., Walsh, R. F. L., Sheehan, A. E., Cheek, S. M., & Carter, S. M. (2020). Suicidal ideation and behavior among sexual minority and heterosexual youth: 1995–2017. *Pediatrics*, 145, Article e20192221. <https://doi.org/10.1542/peds.2019-2221>
- Liu, R. T., Walsh, R. F. L., Sheehan, A. E., Cheek, S. M., & Sanzari, C. M. (2022). Prevalence and correlates of suicide and non-suicidal self-injury in children: A systematic review and meta-analysis. *JAMA Psychiatry*, 79, 718–726. <https://doi.org/10.1001/jamapsychiatry.2022.1256>
- Lowry, N. J., Goger, P., Hands Ruz, M., Ye, F., & Cha, C. B. (2024). Suicide risk screening tools for pediatric patients: A systematic review of test accuracy. *Pediatrics*, 153(3), e2023064172. <https://doi.org/10.1542/peds.2023-064172>
- Mayes, T. L., Kennard, B. D., Killian, M., Carmody, T., Grannemann, B. D., Rush, A. J., Jha, M. K., Hughes, J., Emslie, G. J., & Trivedi, M. H. (2018). Psychometric properties of the Concise Health Risk Tracking (CHRT) in adolescents with suicidality. *Journal of Affective Disorders*, 235, 45–51. <https://doi.org/10.1016/j.jad.2018.03.007>
- Mazefsky, C. A., Conner, C. M., Breitenfeldt, K., Leezenbaum, N., Chen, Q., Bylsma, L. M., & Pilkonis, P. (2021). Evidence base update for questionnaires of emotion regulation and reactivity for children and adolescents. *Journal of Clinical Child and Adolescent Psychology*, 50, 683–707. <https://doi.org/10.1080/15374416.2021.1955372>
- Miller, A. B., & Prinstein, M. J. (2019). Adolescent suicide as a failure of acute stress-response systems. *Annual Review of Clinical Psychology*, 15, 425–450. <https://doi.org/10.1146/annurev-clinpsy-050718-095625>
- Molock, S. D., Boyd, R. C., Alvarez, K., Cha, C., Denton, E.-G., Glenn, C. R., Katz, C. C., Mueller, A. S., Meca, A., Meza, J. I., Miranda, R., Ortin-Peralta, A., Polanco-Roman, L., Singer, J. B., Zullo, L., & Miller, A. B. (2023). Culturally responsive assessment of suicidal thoughts and behaviors in youth of color. *American Psychologist*, 78, 842–855. <https://doi.org/10.1037/amp0001140>
- Nandy, K., Rush, A. J., Carmody, T. J., Kulikova, A., Mayes, T. L., Emslie, G., & Trivedi, M. H. (2023). The Concise Health Risk Tracking—Self-Report (CHRT-SR)—A measure of suicidal risk: Performance in adolescent outpatients. *International Journal of Methods in Psychiatric Research*, 32, Article e1944. <https://doi.org/10.1002/mpr.1944>
- National Action Alliance for Suicide Prevention. (2014). *A prioritized research agenda for suicide prevention: An action plan to save lives*. National Institute of Mental Health and the Research Prioritization Task Force.
- National Institute of Mental Health. (2019). *Identifying research priorities in child suicide risk*. <https://www.nimh.nih.gov/news/events/2019/identifying-research-priorities-in-child-suicide-risk-0>
- National Institute of Mental Health. (2021). *Understanding suicide risk among children and pre-teens: A synthesis workshop*. <https://www.nimh.nih.gov/news/events/2021/understanding-suicide-risk-among-children-and-pre-teens-a-synthesis-workshop>
- National Institutes of Health. (2019). *Effectiveness of implementing sustainable evidence-based mental health practices in low-resource settings to achieve mental health equity for traditionally underserved populations*. <https://grants.nih.gov/grants/guide/rfa-files/RFA-MH-20-400.html>

- National Institutes of Health. (2021). *Transformative research to address health disparities and advance health equity*. <https://grants.nih.gov/grants/guide/rfa-files/RFA-RM-21-021.html>
- Nock, M. K., Holmberg, E. B., Photos, V. I., & Michel, B. D. (2007). Self-injurious thoughts and behaviors interview: Development, reliability, and validity in an adolescent sample. *Psychological Assessment*, 19, 309–317. <https://doi.org/10.1037/1040-3590.19.3.309>
- Núñez, D., Villacura-Herrera, C., Gaete, J., Meza, D., Andaur, J., & Robinson, J. (2024). Psychometric assessment of the Suicidal Ideation Questionnaire Junior: A two-study validation in Spanish-speaking adolescents. *Current Psychology*, 43, 14411–14424. <https://doi.org/10.1007/s12144-023-05422-2>
- Oquendo, M. A., Galfalvy, H. C., Choo, T. H., Kandlur, R., Burke, A. K., Sublette, M. E., Miller, J. M., Mann, J. J., & Stanley, B. H. (2021). Highly variable suicidal ideation: A phenotypic marker for stress induced suicide risk. *Molecular Psychiatry*, 26, 5079–5086. <https://doi.org/10.1038/s41380-020-0819-0>
- 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>
- Patalay, P., Hayes, D., & Wolpert, M. (2018). Assessing the readability of the self-reported Strengths and Difficulties Questionnaire. *BJPsych Open*, 4, 55–57. <https://doi.org/10.1192/bjo.2017.13>
- Pinto, A., Whisman, M. A., & McCoy, K. J. M. (1997). Suicidal ideation in adolescents: Psychometric properties of the Suicidal Ideation Questionnaire in a clinical sample. *Psychological Assessment*, 9, 63–66. <https://doi.org/10.1037/1040-3590.9.1.63>
- Plemmons, G., Hall, M., Doupnik, S., Gay, J., Brown, C., Browning, W., Casey, R., Freundlich, K., Johnson, D. P., Lind, C., Rehm, K., Thomas, S., & Williams, D. (2018). Hospitalization for suicide ideation or attempt: 2008–2015. *Pediatrics*, 141, Article e20172426. <https://doi.org/10.1542/peds.2017-2426>
- Posner, K., Brodsky, B., Yershova, K., & Mann, J. J. (2014). Classification of self-injurious behaviors. In M. K. Nock (Ed.), *The Oxford handbook of suicide and self-injury* (pp. 7–22). Oxford University Press.
- Posner, K., Brown, G. K., Stanley, B., Brent, D. A., Yershova, K. V., Oquendo, M. A., Currier, G. W., Melvin, G. A., Greenhill, L., Shen, S., & 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, 1266–1277. <https://doi.org/10.1176/appi.ajp.2011.10111704>
- Posner, K., Oquendo, M. A., Gould, M., Stanley, B., & Davies, M. (2007). Columbia Classification Algorithm of Suicide Assessment (C-CASA): Classification of suicidal events in the FDA's pediatric suicidal risk analysis of antidepressants. *American Journal of Psychiatry*, 164, 1035–1043. <https://doi.org/10.1176/appi.ajp.164.7.1035>
- Potard, C., Kubiszewski, V., Gimenes, G., & Courtois, R. (2014). Validation of the French version of the Suicidal Ideation Questionnaire among adolescents. *Psychiatry Research*, 215, 471–476. <https://doi.org/10.1016/j.psychres.2013.11.025>
- Randall, J. R., Colman, I., & Rowe, B. H. (2011). A systematic review of psychometric assessment of self-harm risk in the emergency department. *Journal of Affective Disorders*, 134, 348–355. <https://doi.org/10.1016/j.jad.2011.05.032>
- Revelle, W., & Condon, D. (2019). Reliability from alpha to omega: A tutorial. *Psychological Assessment*, 31, 1395. <https://doi.org/10.1037/pas0000754>
- Reynolds, W. M., & Mazza, J. J. (1999). Assessment of suicidal ideation in inner-city children and young adolescents: Reliability and validity of the Suicidal Ideation Questionnaire-JR. *School Psychology Review*, 28, 17–30.
- Rudd, M. D. (2006). Fluid vulnerability theory: A cognitive approach to understanding the process of acute and chronic suicide risk. In T. E. Ellis (Ed.), *Suicide and cognition: Theory, research, and therapy* (pp. 355–368). American Psychological Association.
- Sheftall, A. H., Vakil, F., Ruch, D. A., Boyd, R. C., Lindsey, M. A., & Bridge, J. A. (2022). Black youth suicide: Investigation of current trends and precipitating circumstances. *Journal of the American Academy of Child & Adolescent Psychiatry*, 61, 662–675. <https://doi.org/10.1016/j.jaac.2021.08.021>
- Silverman, M. M. (2006). The language of suicidology. *Suicide and Life-Threatening Behavior*, 36, 519–532. <https://doi.org/10.1521/suli.2006.36.5.519>
- Thoma, B. C., Salk, R. H., Choukas-Bradley, S., Goldstein, T. R., Levine, M. D., & Marshal, M. P. (2019). Suicidality disparities between transgender and cisgender adolescents. *Pediatrics*, 144, Article e20191183. <https://doi.org/10.1542/peds.2019-1183>
- Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, 133, 859–883. <https://doi.org/10.1037/0033-2909.133.5.859>
- Turner, C. F., Ku, L., Rogers, S. M., Lindberg, L. D., Pleck, J. H., & Sonenstein, F. L. (1998). Adolescent sexual behavior, drug use, and violence: Increased reporting with computer survey technology. *Science*, 280, 867–873. <https://doi.org/10.1126/science.280.5365.867>
- Venta, A., & Sharp, C. (2014). Extending the concurrent validity of the Self-Injurious Thoughts and Behaviors Interview to inpatient adolescents. *Journal of Psychopathology and Behavioral Assessment*, 36, 675–682. <https://doi.org/10.1007/s10862-013-9402-1>
- Youngstrom, E. A., Salcedo, S., Frazier, T. W., & Perez Algorta, G. (2019). Is the finding too good to be true? Moving from “More is better” to thinking in terms of simple predictions and credibility. *Journal of Clinical Child and Adolescent Psychology*, 48, 811–824. <https://doi.org/10.1080/15374416.2019.1669158>
- Youngstrom, E. A., Van Meter, A., Frazier, T. W., Hunsley, J., Prinstein, M. J., Ong, M. L., & Youngstrom, J. K. (2017). Evidence-based assessment as an integrative model for applying psychological science to guide the voyage of treatment. *Clinical Psychology: Science and Practice*, 24, 331–363. <https://doi.org/10.1111/cpsp.12207>
- Zhang, Y., Yip, P. S. F., & Fu, K.-W. (2014). Validation of the Chinese version of the Reynolds' Suicidal Ideation Questionnaire: Psychometric properties and its short version. *Health and Quality of Life Outcomes*, 12, 1–9. <https://doi.org/10.1186/1477-7525-12-33>