Clinical Psychology Review 112 (2024) 102464

Contents lists available at ScienceDirect

Clinical Psychology Review

journal homepage: www.elsevier.com/locate/clinpsychrev

Review

Assessment of suicidal thoughts and behaviors in adults: A systematic review of measure psychometric properties and implications for clinical and research utility

Olivia H. Pollak [a,*], Ana E. Sheehan [b], Rachel F.L. Walsh [c], Auburn R. Stephenson [c], Holly Zell [d], Jenna Mayes [e], Hannah R. Lawrence [f], Alexandra H. Bettis [e], Richard T. Liu [g,h]

[a] Department of Psychology and Neuroscience, The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
[b] Department of Psychological and Brain Sciences, University of Delaware, Newark, DE, USA
[c] Department of Psychology and Neuroscience, Temple University, Philadelphia, PA, USA
[d] School of Psychological Science, Oregon State University, Corvallis, OR, USA
[e] Department of Psychiatry and Behavioral Sciences, Vanderbilt University Medical Center, Nashville, TN, USA
[f] Google (via Magnit)
[g] Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA
[h] Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA

ARTICLE INFO

Keywords:
Assessment
Psychometrics
Measurement
Suicidal ideation
Suicide attempt
Suicide

ABSTRACT

High-quality clinical care and research on suicidal thoughts and behaviors (STBs) depends on availability and implementation of reliable and valid measures of STBs. In contrast to studies examining STB risk factors, screening instruments, or treatment, little research has rigorously examined the content, characteristics, and psychometric properties of STB measures themselves. This systematic review (1) identified STB measures that conform to empirically supported definitions of STBs, and (2) identified peer-reviewed papers reporting on the psychometric properties of these measures in adults. Data on psychometric properties and other measure characteristics were extracted. A total of 21 eligible measures were identified in the first stage. In the second stage, 70 articles (with 79 independent samples) were included with psychometric data in adult samples for 19 measures. Although there was support for strong internal consistency and content validity across many measures, face validity and clinical utility concerns were prevalent. Few measures comprehensively assessed suicidal behaviors, and interview-based assessments tended to show the strongest psychometric properties and clinical utility. Findings are discussed in the context of recommendations for improving existing measures, including future research to increase utility and translatability across clinical settings, delivery methods, and diverse populations.

1. Introduction

Approximately 4.3% of adults in the United States report having had thoughts of suicide, and 0.6% report having made a suicide attempt, in the prior year (Ivey-Stephenson et al., 2022). The prevalence and clinical seriousness of these outcomes indicate the need for empirically supported and clinically scalable measures of suicidal thoughts and behaviors (STBs) to monitor these outcomes in the general population in epidemiological studies; to assess progress in reducing disparities in historically minoritized populations (e.g., Erlangsen et al., 2023; Haas et al., 2011); to monitor treatment progress with suicidal patients and other patient populations in which STBs are prevalent (e.g., patients with major depression; Liu et al., in press); and to meaningfully improve understanding of STB phenomenology and risk through rigorous research (Coppersmith et al., 2023; Franklin et al., 2017).

As context, it is important to note the distinction between suicide risk screening instruments, and instruments that serve as measures of STBs (i.e., the focus of the current review). Risk screening instruments and outcome measures serve different but complementary functions and should not be conflated. For example, assessment of high blood pressure is commonly conducted as a measure of potential risk for heart disease but is not a valid measure of this condition. Instead, it may indicate the

* Corresponding author at: The University of North Carolina at Chapel Hill, 235 E. Cameron Ave, Chapel Hill, NC 27599, USA.
E-mail address: ohpollak@ad.unc.edu (O.H. Pollak).

https://doi.org/10.1016/j.cpr.2024.102464
Received 2 February 2024; Received in revised form 12 June 2024; Accepted 6 July 2024
Available online 8 July 2024

need for a follow-up with a cardiologist to conduct assessments for the presence of a heart condition or to obtain a clearer picture of future risk of a relevant outcome of concern (e.g., myocardial infarction).

Similarly, suicide risk screening instruments should not be assumed to serve the function of assessing STBs, despite frequent overlap in some of their content. Rather, suicide risk screening tools indicate the need for more thorough assessment of *future risk* for these outcomes. Indeed, this is precisely the guidance provided for commonly used suicide risk screening instruments recommended by the Joint Commission (The Joint Commission, 2020) (e.g., Ask Suicide-Screening Questions [ASQ] and the triage version of the Columbia-Suicide Severity Rating Scale [C-SSRS]) (Ayer et al., 2022; Brahmbhatt et al., 2019). Second, rather than observing distinctions between different forms of STBs (e.g., suicidal ideation versus behaviors, or suicide attempts versus aborted attempts) and non-suicidal self-injury (NSSI), suicide risk screening tools often integrate them together in arriving at a future-oriented assessment of risk. In fact, some widely used suicide risk screening tools recommended by the Joint Commission, such as the Suicide Behaviors Questionnaire-Revised (SBQ-R; Osman et al., 2001), inherently do not permit separation of STB outcomes from each other. Assessing suicide *risk* may also include collecting additional data (e.g., mental health concerns, access to means) or asking about the likelihood of future STBs.

In contrast to future-oriented screeners of suicide risk, measures of STBs are designed to assess current and/or past history of these outcomes—as part of which, the distinction between different forms of STBs is critically important. The ability of measures to distinguish between forms of STBs has important implications for clinical practice and research. In treatment settings, for example, determining whether a patient had attempted suicide versus experienced suicidal thoughts since their last therapy appointment is clinically meaningful insofar as it facilitates accurate determination of clinical severity and guides selection of appropriate treatment approaches (Jobes & Barnett, 2024). In research, distinguishing between types of STBs is critical to facilitate more precise measurement, and therefore understanding, of the phenomenology of STBs, including the temporal course over which suicidal ideation may unfold and how this relates to occurrence of suicidal behavior (Bryan & Rudd, 2016; Coppersmith et al., 2023). Indeed, distinguishing between types of STBs has contributed thus far to identification of important differences in base rates, correlates, and risk factors between types of STBs (Kessler et al., 2005; Nock et al., 2016; Nock et al., 2018) and is critical to understanding STB developmental trajectories, transitions from suicidal thoughts to behaviors, and other future research directions (Glenn et al., 2017; Nock et al., 2016).

Although several systematic reviews of screening tools for risk of self-injurious thoughts and behaviors exist (Baek et al., 2021; Carter et al., 2019; Kreuze & Lamis, 2018; Randall et al., 2011; Thom et al., 2020), much less has been done systematically evaluating measures of STBs. Further, sensitive and accurate assessment of STBs is hindered by several challenges. There is broad misalignment between the clinical nature (i.e., temporal dynamics, phenomenology) of STBs, and the structure (i.e., format, logistical practicalities) of prevailing assessment tools. As noted above, many measures do not cleanly distinguish between distinct forms of STBs, despite data supporting differences in base rates, correlates, risk factors, and treatment approaches between these types of STBs (Kessler et al., 2005; Nock et al., 2016, 2018). Some measures treat occurrence of suicidal behavior as an extreme end of a continuum of suicidal distress, which may be inconsistent with prominent theoretical models of the distinction between suicidal ideation and suicide attempts, such as the Interpersonal Theory of Suicide, according to which these two phenomena have distinctly different underlying risk factors (Van Orden et al., 2010). Second, data show that most individuals who consider suicide do not engage in suicidal behavior (Nock et al., 2008) and that suicidal behavior can occur in the absence of active suicidal ideation (i.e., with passive ideation alone), or even in the absence of any suicidal ideation (Baca-Garcia et al., 2011; Liu et al., 2020). Many measures also assess relatively long time frames that, while perhaps useful in some cases for assessment of suicidal behavior, may less accurately represent characteristics of suicidal ideation (e.g., duration, intensity) that may fluctuate over shorter intervals (Coppersmith et al., 2023; Kleiman et al., 2017) and which may be more susceptible to biased or imprecise recall when reporting on lengthier, more distal time frames (Andrewes et al., 2024; Gratch et al., 2021). Intuitively, it may be more challenging to recall and accurately report on a brief (e.g., one-week) period of passive ideation occurring a month or year ago versus last week, while suicidal behavior, even if distal, may be more memorable.

In light of these challenges, it is important to evaluate the measurement properties of existing STB assessment tools. Several recent studies have tested important STB measurement questions by examining patterns of responses across assessment methods and informants (Deming et al., 2021; Hom et al., 2019; Spears et al., 2023). These studies provide insight into factors that may influence an individual's reporting of STBs or account for multi-informant reporting discrepancies. However, even studies that seek to answer basic measurement questions such as these rely on fundamental assumptions about the characteristics of the assessment tools themselves (i.e., sound psychometric properties).

To date, there has been little effort to rigorously review the format, content, and basic psychometric properties of STB measures. Although a few systematic reviews exist of STB assessment tools (Batterham et al., 2015; Borschmann et al., 2012; Flores-Kanter & Alvarado, 2024; Gleeson et al., 2022), interpretation of their findings are complicated in that they did not observe important distinctions between (1) suicide risk screeners and measures of STBs, (2) suicidal ideation and behavior,[1] and/or (3) suicidal behavior and NSSI. The current study fills this gap by providing a systematic review of measures of STBs in adults, with a focus on measure psychometric properties and utility in clinical and research settings. This effort complements a review of STB measures in youth (Liu et al., 2024) but provides a review specific to adult samples, given age-related differences and developmental considerations in evaluating measure performance (Oppenheimer et al., 2022).

2. Methods

2.1. Search strategy and eligibility criteria

This project was registered in PROSPERO (CRD42022382198), and authors followed Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) reporting guidelines (Page et al., 2021). A systematic review was conducted in two stages. Stage 1 focused on identifying measures of STBs using accepted parameters and definitions (Corsby et al., 2011; National Action Alliance for Suicide Prevention (NAASP), 2014). Stage 2 focused on identifying peer-reviewed papers evaluating the psychometric properties of measures identified in Stage 1. Details about Stage 1 and 2 (e.g., search terms, measure inclusion criteria, data extraction and calculation) are in the Supplement.

2.2. Measure evaluation criteria

Criteria for assessing measures' psychometric properties are presented in Table 1. These were modeled off criteria presented by Youngstrom and colleagues (Youngstrom et al., 2018), with adaptations made to reliability and validity criteria to account for unique considerations for assessment of STBs (Cronbach & Shavelson, 2004; Revelle & Condon, 2019; Youngstrom et al., 2019). Regarding reliability, if data supported very high *internal consistency* (e.g., $\alpha > 0.90$), we considered factors (e.g., scale/measure length, narrow scope, item redundancy) that may inflate reliability at the detriment of other important considerations

[1] For example, the commonly used SBQ-R does not differentiate suicidal thoughts from behaviors and thus did not meet our inclusion criteria (see Supplemental Material).

Table 1
Criteria for evidence-based evaluations of measures of suicidal thoughts and behaviors, adapted from Youngstrom et al. (2017).

Criterion	Adequate	Good	Excellent
Norms[a]	Mean and standard deviation in 1 large clinical or community sample ($n \geq 300$)	Mean and standard deviation in 1 large clinical sample ($n \geq 300$)	Mean and standard deviation in multiple large independent samples ($n \geq 300$), at least 1 of which is a clinical sample
Internal Consistency[a]	Most evidence is of $\alpha = 0.70–0.79$	Most evidence is of $\alpha = 0.80–0.89$	Most evidence is of $\alpha \geq .90$[c]
Test-Retest Reliability	Most evidence is of $r = 0.40–0.59$, $\kappa = 0.60–0.74$, or ICC = 0.70–0.79	Most evidence is of $r = 0.60–0.79$, $\kappa = 0.75–0.84$, or ICC = 0.80–89	Most evidence is of $r \geq 0.80$, $\kappa \geq 0.85$, or ICC ≥ 0.90
Inter-Rater Reliability[b]	Most evidence is of $\kappa = 0.60–0.74$ or ICC = 0.70–0.79	Most evidence is of $\kappa = 0.75–0.84$ or ICC = 0.80–89	Most evidence is of $\kappa \geq 0.85$ or ICC ≥ 0.90
Face Validity	Presence of multiple (more than one) face validity concerns and/or validity issues in multiple measure items	Minimal (i.e., one) face validity concerns and/or face validity issues in one measure items	No face validity issues
Content Validity	Include at least 1 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
Construct Validity (convergent, discriminant, and divergent validity)	Support for multiple aspects of construct validity in 1 sample	Support for multiple aspects of construct validity in multiple independent samples	Support for multiple aspects of construct validity replicated across multiple independent samples
Sensitivity to Change[d]	1 sample with evidence of sensitivity to change over time	At least 2 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)
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 1 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

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.

[a] Only applicable to measures with (sub)scale scores.

[b] Only applicable to interview-based measures.

[c] For measures rated as "excellent" for internal consistency, presence of factors that may inflate internal consistency at the expense of other psychometric considerations was evaluated. If relevant, a designation of "too excellent" was noted to indicate that internal consistency may be inflated (e.g., $\alpha > 0.90$) by scale length (long) and/or scope (narrow). For more details regarding this consideration, see Youngstrom et al. (2017).

[d] Only applicable to measures with suicidal ideation (sub)scale scores.

(e.g., clinical utility, content validity, local independence; Boyle, 1991; Youngstrom et al., 2018). If concerns were present, internal consistency was flagged as potentially "too excellent," again modeling criteria presented by Youngstrom and colleagues (Youngstrom et al., 2018). For *test-retest reliability*, adjustments were made to criteria given the temporally dynamic nature of suicidal ideation, in contrast to trait-like constructs. Because it is expected that presence and severity of suicidal ideation may fluctuate over short time intervals (Kleiman et al., 2017), it is important that consecutive administrations of a measure involve temporal overlap in assessment timeframe to provide a valid index of test-retest reliability. Thus, psychometric data were considered indices of test-retest reliability only if they captured reasonable temporal overlap in consecutive measure administrations (e.g., consecutive administrations of a one-month measure, given two weeks apart). This also ensures evaluations of test-retest reliability are not confounded by memory or recall bias.[2]

Second, criteria for *sensitivity to change* required that data represent a comparison of scores on a measure administered consecutively over non-overlapping timeframes. As long as assessment timeframes were non-overlapping, there were no formal criteria for the length of time between measure administrations. However, it should be noted that although longer intervals (e.g., over a span of a year) may be more likely to produce evidence of sensitivity to change, shorter temporal intervals may be of greater clinical utility (i.e., detecting changes in suicidal ideation over consecutive weekly therapy sessions). Finally, a face validity criterion was added, capturing item-level evaluations of the extent to which a measure appears to assess STBs.

Clinical utility was informed by an evaluations of a measure's norms, reliability, and validity, as well as the following metrics: (1) whether the measure is free (versus proprietary); (2) whether no training is required in administration, scoring, or interpretation of results; (3) empirical support for recommended clinical cut-off score(s) (if applicable); (4) evidence of sensitivity to change (if applicable); (5) two indices of time required for measure administration (i.e., number of items and words); (6) two indices of measure readability: the Flesch Reading Ease Formula (Flesch, 1948) and the Flesch-Kincaid Grade Level Formula (Flesch, 1949); and (7) temporal window covered by the measure, particularly in the case of suicidal ideation. Regarding this last metric, the timeframe of assessment captured by a measure is paramount when assessing suicidal ideation, given evidence that suicidal thoughts can fluctuate across short timeframes (Kleiman et al., 2017), and that this variation may predict suicidal behavior (Wang et al., 2021). Thus, measures of suicidal ideation with shorter assessment timeframes (e.g., past week or less) were considered to be more useful clinically.

Finally, using a system similar to Mazefsky and colleagues (Mazefsky et al., 2021), an *overall* measure evaluation was made by synthesizing ratings across four psychometric categories: norms, reliability, validity, and clinical utility. Clinical utility was weighed most heavily in determining the overall evaluation, for several reasons. First, reliable assessment of STBs facilitates comprehensive case conceptualization and is critical to effectively monitor patient safety and manage risk (Chu et al., 2015; Ryan & Oquendo, 2020). Second, assessment of STBs in clinical settings is often challenging, given fears or misconceptions

[2] For the importance of this consideration in reporting of psychopathology and related phenomena, see Liu, 2017 and Moffitt et al., 2010.

about asking about STBs; lack of training or familiarity with the nuances of STBs, perhaps particularly among providers outside psychology or psychiatry; or qualities of particular clinical settings, such as the time-limited nature of emergency and acute care settings (Fowler, 2012; Jobes, 1995). Several of these considerations (e.g., limited time available to conduct assessments with patients with complex STB presentations in acute care settings) make assessment of STBs in clinical settings more challenging than in research contexts. To receive an *overall* evaluation of "adequate," a measure needed to receive evaluations of at least "adequate" across all four psychometric categories (i.e., norms, reliability, validity, and clinical utility). To receive an *overall* evaluation of "good," a measure was required to be at least "adequate" in clinical utility and "good" in at least two other categories. Finally, an *overall* evaluation of "excellent" required an evaluation of at least "good" in clinical utility, and "excellent" in at least two other categories.

3. Results

Among the 1439 publications obtained in Stage 1, 909 unique records were identified. Of these, 622 were excluded based on titles and abstracts, and 197 articles were excluded based on a detailed full-text review. This resulted in a final total of 88 publications with 21 STB measures that met Stage 1 eligibility criteria and were therefore included (see Table S1 for a list of included measures and Sage 1 PRISMA flow chart [Fig. S1]). Stage 2 yielded 460 search results (216 unique records). Of these, 68 were excluded based on titles and abstracts, and 78 articles were excluded based on full text reviews. This resulted in a final total of 70 articles (see Appendix for references) with 19 STB measures included in Stage 2 (see Fig. S2 for the Stage 2 PRISMA flow chart). Table S2 presents these measures, their associated articles, and study sample characteristics. Measures ultimately included in this review are presented below, alongside a summary of key takeaways of their respective psychometric evaluation. Additionally, Table S3 presents these measures scalability characteristics.

3.1. Adult Suicidal Ideation Questionnaire (ASIQ) and ASIQ-4

The ASIQ is a self-report measure of passive and active suicidal ideation (i.e., a desire for death and a desire to kill oneself, respectively) and suicide plans in the prior month. The ASIQ-4 is an abbreviated, 4-item version assessing active suicidal ideation. This review identified two papers reporting on the psychometric properties of the English or Chinese-language versions of the ASIQ and/or ASIQ-4 across two samples (Fu et al., 2007; Reynolds, 1991).

The English and Chinese-language versions of the ASIQ showed excellent internal consistency, although this may be inflated due to high alpha values, length, and some item redundancy (e.g., "I wondered if I had the nerve to kill myself"; "I wished I had the nerve to kill myself"). The English version showed excellent test-retest reliability over a two-week period ($r = 0.86$; Reynolds, 1991). Data were unavailable for sensitivity to change. Content validity was excellent and face validity was adequate. As with other measures, like the suicide subscale of the Personality Assessment Inventory (PAI; see below), face validity was impacted by concerns about multiple items that assess motivations or sequalae of STBs, as well as items asking about thoughts that could plausibly be associated with psychopathology or clinical symptoms other than STBs (e.g., thoughts about having an accident may reflect generalized anxiety disorder). Construct validity was adequate for the English and Chinese-language versions of the ASIQ: measure scores were significantly correlated with constructs such as depression, anxiety, and hopelessness (Fu et al., 2007). When considering both language versions, construct validity was stronger (i.e., 'excellent') given support for associations with multiple of the same constructs across multiple independent samples (i.e., the two samples captured by the two included studies).

The ASIQ-4 in Chinese showed overall similar psychometric properties, including excellent internal consistency and adequate face, content, and construct validity. Content validity was considered adequate because ASIQ-4 assesses active suicidal ideation but does not cleanly assess passive suicidal ideation. That is, one item (i.e., asking if the individual thought others would be happier if the individual were dead) more accurately reflects feelings of connectedness with others rather than passive ideation. Indeed, concerns about this and another item (i.e., thoughts of killing oneself as a solution to one's problems) also had an impact on face validity. Clinical utility for both the ASIQ and ASIQ-4 was limited, primarily due to concerns with face validity as well as the one-month timeframe, which is a longer temporal window for assessment of suicidal ideation than may be useful in many clinical settings (e.g., to evaluate symptom change between therapy sessions). The ASIQ and ASIQ-4 are also proprietary, further limiting accessibility and clinical scalability, especially in under-resourced clinical settings, and thus potentially presenting a barrier heightening disparities in communities served in these settings.

3.2. Beck Scale for Suicide Ideation (BSS)

The BSS assesses passive and active suicidal ideation, suicide plans, and suicidal behaviors (i.e., preparatory acts) in the past week and includes both self-report and clinician-rated versions. In this review, one paper was included reporting on the psychometric properties of the BSS Clinician-Rated Version in an adult sample (Beck et al., 1973), and eight papers were identified reporting on the psychometric properties of the BSS Self-Report Version across 10 adult samples (Beck et al., 1988; Chioqueta & Stiles, 2006; Cochrane-Brink et al., 2000; Esfahani et al., 2015; Keliat et al., 2023; Kleiman et al., 2017; McCall et al., 2021; Zhang & Brown, 2007).

3.2.1. Clinician-Rated

The BSS Clinician-Rated Version is completed by a clinician or other administrator to reflect respondents' answers during a semi-structured interview. The BSS Clinician-Rated Version showed good internal consistency and interrater reliability. As psychometric data were available for only one sample, construct validity was considered adequate, reflecting support for positive correlations of measure scores with constructs such as hopelessness and depression. No evaluation was conducted for sensitivity to change. Content validity was excellent, as the measure assesses both passive and active suicidal ideation, as well as thoughts of plans and methods. However, face validity was only adequate due to concerns with several items (for details, see the *Self-Report* section below). Data were not available for test-retest reliability. Clinical utility was considered limited, primarily due to concerns with face validity and other immutable characteristics (i.e., proprietary; some readability concerns).

3.2.2. Self-Report

The BSS Self-Report Version showed excellent internal consistency in English and German, and good internal consistency in other language translations (i.e., Chinese, Indonesian, Norwegian, and Persian). The English language version exhibited excellent construct validity. In other languages, the BSS Self-Report Version showed adequate construct validity, with support for multiple aspects of construct validity in singular samples. More studies examining the psychometric properties of translations of the BSS Self-Report Version are needed to establish construct validity across multiple independent samples for these non-English language translations. As with the BSS Clinician-Version, content validity for the Self-Report Version was excellent, but face validity was only adequate and affected by a number of concerns. Several items assess motivations or contextual factors for STBs that may extend beyond explicit assessment of presence or frequency of STBs themselves. For example, one item assesses reasons for wanting to die by suicide (e.g., to influence others) rather than thoughts of or desire for suicide. Additionally, some of the language used in this scale is no longer

considered appropriate (e.g., "commit suicide"). Data were not available for test-retest reliability. Although one study reported a significant correlation between two consecutive administrations three months apart for the Norwegian version (Chioqueta & Stiles, 2006), suggesting support for sensitivity to change, change in suicidal ideation over this lengthy temporal interval is of limited utility in treatment contexts, where clinical concern is often about change over much smaller time windows. Finally, the clinical utility of the BSS Self-Report Version was considered limited, due to similar concerns as those impacting the Clinician-Rated Version (in particular, face validity concerns). Clinical utility of the BSS would also be improved through establishment of scoring guidelines that isolate items directly assessing STBs from those assessing motivations, contextual factors, or other secondary factors or sequalae.

3.3. Columbia-Suicide Severity Rating Scale (C-SSRS)

The C-SSRS is a traditionally interview-based assessment of a range of suicidal outcomes, including both passive and active suicidal ideation as well as multiple types of suicidal behavior (e.g., preparatory behaviors, interrupted and aborted suicide attempts). In this review, a total of 11 papers were identified reporting on the psychometric properties of the C-SSRS across 11 adult samples (Al-Halabí et al., 2016; Azcurra, 2017; Brown et al., 2020; Franks et al., 2021; Ji et al., 2023; Madan et al., 2016; Matarazzo et al., 2019; McCall et al., 2021; Posner et al., 2011; Tabares et al., 2021; Youngstrom et al., 2015). Of these, two papers across two adult samples (Franks et al., 2021; Tabares et al., 2021) reported psychometric properties of the C-SSRS when administered as a self-report measure (vs. interview). Because the C-SSRS is designed to be administered via interview, typically by a clinician, psychometric evaluations from the two studies that administered the measure as a self-report are described separately. While there are multiple versions of the C-SSRS assessing STBs over different time frames (e.g., lifetime, past month, since last assessment),[3] psychometric evaluations reported here are based on all available study data within a given psychometric category, regardless of time frame(s) of assessment, to maximize available data and provide higher-level takeaways across studies.

For the interview-based C-SSRS, studies reported a range of values for internal consistency, reflecting evidence ranging from 'limited' and/or 'adequate,' to 'good' and/or 'excellent' for assessment of suicidal ideation. This was the case for both the English and Spanish-language versions; the Chinese-language version, however, showed good internal consistency. In English, the interview-based C-SSRS showed overall good interrater reliability for suicidal ideation but mixed interrater reliability for suicidal behavior (Youngstrom et al., 2015). There was support for good construct validity for suicidal ideation for the English and Spanish-language versions in multiple independent samples (English: Madan et al., 2016; Matarazzo et al., 2019; McCall et al., 2021; Posner et al., 2011, Study 3. Spanish: Al-Halabí et al., 2016; Azcurra, 2017). Construct validity was adequate for suicidal ideation for the Chinese language version and for suicidal behavior for the English language version, primarily as relevant data were available for only one sample (Ji et al., 2023 and Posner et al., 2011, Study 3, respectively). No papers reported analyses of test-retest reliability or sensitivity to change for the interview-based C-SSRS. When administered as a self-report, the C-SSRS showed at least adequate internal consistency for suicidal ideation (Franks et al., 2021). Data were not available for test-retest and interrater reliability, construct validity, and sensitivity to change for the C-SSRS when given as a self-report.

Face validity and content validity were excellent for both the interview-based and self-report C-SSRS. Of note, several items in the C-SSRS ideation intensity subscale mirror items from other measures reviewed herein in their assessment of particular characteristics of suicidal thoughts, such as controllability or deterrents to STBs. While such items were considered to negatively impact face validity for other measures (e.g., BSS, PAI), these items present less concern to face validity for the C-SSRS given they are included in a distinct, separate 'ideation intensity' subscale that more specifically assesses particular qualities of ideation (i.e., those contributing to 'intensity'). Finally, overall clinical utility of the C-SSRS was considered adequate. While the interview-based version may require some training to administer, the measure's comprehensive assessment of multiple types of STBs across multiple timeframes (e.g., current, lifetime, etc. depending on version) are significant strengths and greatly enhance its clinical practicality. Currently, the C-SSRS is widely used across healthcare settings, which provides evidence supporting clinical scalability.

3.4. Concise Health Risk Tracking (CHRT) suicidal ideation subscale

Although the CHRT is a self-report measure of "suicide risk," it contains a suicidal thoughts subscale that explicitly assesses active suicidal ideation. Items in the suicidal ideation subscale ask about thoughts and plans related to killing oneself. In this review, nine papers were found reporting on the psychometric properties of the CHRT suicidal ideation subscale across 10 adult samples (De La Garza et al., 2019; Nandy et al., 2023; Ostacher et al., 2015; Sanchez et al., 2018; Trivedi et al., 2011; Trombello et al., 2019, 2023; Villegas et al., 2018; Weissman et al., 2022). One of these papers (Trivedi et al., 2011) also reported on the psychometric properties of the suicidal ideation subscale when administered by clinicians, in addition to the traditional administration as a self-report measure.

The CHRT suicidal ideation subscale generally showed overall good internal consistency, although one study found evidence of poor internal consistency (Trombello et al., 2019). As no studies compared the suicidal ideation subscale across consecutive administrations with temporal overlap in items, test-retest reliability could not be evaluated. While face validity was excellent (i.e., the three suicidal ideation subscale items directly assess thoughts of suicide and of suicide methods and plans), content validity was only adequate because the subscale does not assess passive ideation.[4] Construct validity was excellent: the CHRT suicidal ideation subscale was correlated with constructs ranging from depressive symptoms and hopelessness to life satisfaction and resilience, in expected directions (e.g., De La Garza et al., 2019; Ostacher et al., 2015; Sanchez et al., 2018; Trivedi et al., 2011; Trombello et al., 2019, 2023; Villegas et al., 2018; Weissman et al., 2022). The one study that also examined an interviewer-based version of the CHRT found support for adequate construct validity of the suicidal ideation subscale in this format (Trivedi et al., 2011).

In contrast to other measures included in this review, for which data on sensitivity to change were largely unavailable, studies of the CHRT reported a range of analyses reflective of aspects of sensitivity to change that warrant mention here. Two studies reported only modest correlations between CHRT suicidal ideation subscale scores obtained at consecutive assessments given one week apart (Sanchez et al., 2018; Trombello et al., 2023), providing tentative evidence of sensitivity to change. Another study showed a statistically significant difference in

[3] A majority of studies reported on the C-SSRS asking about lifetime time frame (e.g., Al-Halabi et al., 2016; Brown et al., 2020; Franks et al., 2021; Ji et al., 2023; Matarazzo et al., 2019; Posner et al., 2011, Study 3; Tabares et al., 2021; Youngstrom et al., 2015). Some studies assessed other time frames, such as past month, past week, or since last assessment (Brown et al., 2020, Ji et al., 2023, Posner et al., 2011; Study 3; Youngstrom et al., 2015), in addition to lifetime. Several did not specify time frame(s) (e.g., Azcurra et al., 2017; Madan et al., 2016; McCall et al., 2021).

[4] Several items of the CHRT appear relevant to passive suicidal ideation (e.g., *"I feel that there is no reason to live"*). However, these items are grouped as part of the "despair" subscale of the CHRT and are not included in the "suicidal thoughts" subscale of interest (see De La Garza et al., 2019).

CHRT suicidal ideation subscale scores (i.e., reflecting declines) between assessments spaced, on average, 40 days apart (Nandy et al., 2023). While it is possible this captured at least some consecutive assessments with temporal overlap (i.e., given less than one week apart), the mean and median number of days between assessments (40 and 30 days, respectively) suggests that this analysis largely captures comparison of assessments covering non-overlapping temporal intervals, consistent with this review's criteria for evaluating sensitivity to change. Additionally, another study found a significant main effect of time when comparing CHRT suicidal ideation subscale scores over multiple weeks, across groups distinguished by varying levels of baseline suicidal ideation (De La Garza et al., 2019). Another study reported no significant change in CHRT suicidal ideation subscale scores in two independent samples over either an 8- or 12-week interval (Villegas et al., 2018; Study 1 and Study 2). Collectively, these findings appear to provide mixed support for sensitivity to change for the CHRT suicidal ideation subscale, which was considered adequate given evidence of sensitivity to change in at least one sample (e.g., Nandy et al., 2023).

Clinical utility of the CHRT was limited, primarily as the suicidal ideation subscale does not assess passive suicidal ideation. That is, while the subscale's short length may aid scalability, the absence of subscale items assessing passive suicidal ideation is a significant limitation to clinical utility given that passive thoughts of suicide are relatively common and show similar associations with suicide risk to active ideation (Liu et al., 2020).

3.5. Depressive Symptom Index-Suicidality Subscale (DSI-SS)

The DSI-SS is brief self-report measure of active suicidal ideation (i.e., with and without a plan) and suicidal urges (i.e., "impulses"). In this review, two papers were identified reporting on the psychometric properties of the DSI-SS across two[5] samples (Jeon et al., 2024; Stanley et al., 2021). Internal consistency was excellent. DSI-SS scores were correlated with suicidal ideation assessed via other measures and with constructs such as depression and impulsivity, supporting good construct validity. Data also supported changes in DSI-SS scores over time (i.e., evidence of a statistically significant decrease in DSI-SS scores from study enrollment to discharge; Stanley 2021). However, as the interval between assessments was not cleanly distinct from the temporal period assessed by the DSI-SS (i.e., past two weeks), this finding does not sufficiently indicate sensitivity to change in line with this review's criteria. No issues with face validity were identified, although content validity was considered only adequate due to absence of items assessing passive suicidal ideation. Clinical utility was limited, primarily for this reason, even though the measure's brevity and relatively short temporal window of assessment (i.e., prior two weeks) offer some advantages with regards to clinical utility.

3.6. Geriatric Suicide Ideation Scale (GSIS) and Brief GSIS suicidal ideation subscales

The GSIS is a self-report measure designed to assess multiple suicide-relevant constructs, with one subscale specifically assessing current suicidal ideation (i.e., passive and active) and suicide plans. It is designed for assessment of STBs in older adults. An abbreviated version, the Brief GSIS, consists of a subset of 10 items from the full GSIS measure and, like the full GSIS, contains a suicidal ideation subscale (two items). In this review, four papers were found reporting on the psychometric properties of the suicidal ideation subscale of the GSIS across five adult samples (Chou et al., 2005; Heisel & Flett, 2006, 2016; O'Rourke et al., 2018). An additional paper was included reporting on the psychometric properties of the Brief GSIS in one adult sample (Heisel & Flett, 2022).

3.6.1. GSIS

Internal consistency was good for the Chinese and English-language versions of the GSIS suicidal ideation subscale. No studies assessed test-retest reliability in a way that captured overlapping time intervals, in line with this review's criteria. One study stated that adequate-to-good reliability was found when comparing baseline GSIS suicidal ideation subscale scores to scores across multiple follow-up assessments spanning approximately one month to over one year later (ICC range: 0.53–0.68, $ps < 0.001$; Heisel & Flett, 2016), but since the GSIS evaluates "current" suicidal ideation, these analyses more accurately reflect stability of suicidal ideation than test-retest reliability (i.e., an aspect of the phenomenology of suicidal ideation rather than of psychometric properties of the GSIS). Another study also reported strong reliability when comparing GSIS suicidal ideation subscale scores assessed approximately one to two months apart ($r = 0.78$, $p < .01$; Heisel & Flett, 2006). For the Chinese-language version of the GSIS, the correlation between suicidal ideation subscale scores at consecutive administrations over a shorter time interval (i.e., two weeks) was robust ($r = 0.72$, $p < .01$; Chou et al., 2005). Construct validity was excellent for the English version of the GSIS. Studies provided evidence of convergent validity with constructs such as depression and hopelessness, and divergent validity with constructs such as life satisfaction and subjective well-being across multiple independent samples (Heisel & Flett, 2006, 2016). Data were available from one sample for the Chinese version, for which construct validity was adequate.

Content validity for the GSIS suicidal ideation subscale was excellent, as it assesses both passive and active suicidal ideation and includes an item about thoughts of specific suicide methods or plans. Several issues with face validity were identified, including multiple items that appear to capture constructs closer to hopelessness (e.g., *"Nothing left for me in this world"*) or negative feelings whose clear relevance to STBs is more remote, or which may more accurately reflect physical health concerns that may be more common in older age groups that are the focus of this measure (e.g., *"Feels like I am wasting away"*). Clinical utility was considered limited, primarily due to concerns about face validity and readability (e.g., approximately 8th to 9th grade reading level). Additionally, clinical utility was impacted by mixed findings for internal consistency, at least for the English version, including some evidence of limited and/or adequate internal consistency. Also negatively affecting clinical utility was that the timeframe for "current" suicidal ideation is undefined, introducing idiosyncratic variability between respondents in terms of how they operationalize "current." However, the measure's design for use in older adults is a unique strength and pertinent to those working with geriatric populations.

3.6.2. Brief GSIS

The one study assessing the Brief GSIS found support for excellent internal consistency. Construct and face validity were adequate. Content validity was good, as the measure included both passive and active suicidal ideation. While not within this review's criteria for test-retest reliability (i.e., requiring overlapping intervals), the Brief GSIS suicidal ideation subscale exhibited adequate-to-good reliability when comparing assessments given approximately one-month to one-year apart (ICCs = 0.68–0.89; Heisel & Flett, 2022). There were no data on sensitivity to change. Clinical utility was limited for the Brief GSIS, primarily because of the reading level required for this measure and the lack of clarity of the timeframe for this measure.

3.7. Multidimensional Suicide Inventory-28 (MSI-28) suicidal ideation subscale

The MSI-28 is self-report measure of suicidal thoughts and associated risk and protective factors, with one subscale (i.e., suicidal ideation subscale) assessing active suicidal ideation and suicide plans. One study

[5] Jeon and colleagues (2024) report on the psychometric properties of the DSI-SS in a large data set that was obtained by combining three datasets (i.e., three samples).

was included examining the psychometric properties of the MSI-28, including the suicidal ideation subscale, in one sample (Bezdjian et al., 2015). The MSI-28 suicidal ideation subscale exhibited excellent internal consistency. Data supported adequate construct validity for the MSI-28 suicidal ideation subscale, including convergent validity with negative affect and other mental health symptoms, and divergent validity with positive self-perception. Content validity was only adequate, as this measure does not assess passive suicidal ideation. Face validity was only adequate, as concerns were identified with several overly general items (e.g., an item that asks about death, and another asking about intrusion of unspecified thoughts). Data were not available for test-retest reliability or sensitivity to change. Clinical utility was considered limited. The unspecified timeframe of assessment, including variation and lack of temporal specificity across items, adequate face validity, and the absence of items assessing passive ideation hinder clinical utility.

3.8. Personality Assessment Inventory (PAI) and PAI-Short Form (PAI-SF) suicide subscale

The PAI is a self-report measure of psychopathology and personality dynamics, with one subscale assessing suicidal ideation (i.e., passive and active) and suicide planning. In this review, 22 papers were identified reporting on the psychometric properties of the suicide subscale of the PAI, including its translations into multiple languages (e.g., Spanish, French, Italian), across 24 samples (Abilleira & Rodicio-García, 2019; Alterman et al., 1995; Bradley et al., 2007; Burneo-Garcés et al., 2020; Busse et al., 2014; Corsica et al., 2010; Frazier et al., 2006; Hopwood et al., 2008; Jeffay et al., 2020; Karlin et al., 2005; Lyrakos, 2011; Mabilia et al., 2019; Morey et al., 2011; Patry & Magaletta, 2015; Pignolo et al., 2018; Ruchensky et al., 2021; Sims et al., 2013; Stover et al., 2015; Tasca et al., 2002; Udala et al., 2022; Wang et al., 2021; Ward et al., 2018). The PAI-SF is an abbreviated version of the PAI containing approximately fewer than half as many items and generating prorated scores for full measure scales. This review identified 3 papers reporting on the psychometric properties of the suicide subscale of the PAI-SF across 4 adult samples (Frazier et al., 2006; Udala et al., 2022; Ward et al., 2018).

3.8.1. PAI

For the English-language version of the PAI, internal consistency for the suicide subscale was mixed and ranged from 'limited' and/or 'adequate,' to 'good' and/or 'excellent.' Papers examining translations of the PAI into other languages (e.g., Spanish, Greek, Italian) tended to report good internal consistency, with the exception of the French version, for which there was evidence of weak internal consistency. Data were unavailable for test-retest reliability or sensitivity to change. Construct validity was good for the English-language version of the PAI, and adequate for the Italian and Spanish-language translations, although data on construct validity data across multiple samples was lacking for these translations. Content validity was excellent. Face validity was only adequate, as determined by several concerns. First, a few items assess thoughts (e.g., how others would respond to one's suicide) and constructs (e.g., reasons for living) related to, but not directly of, suicidal ideation. Clinical utility was limited given face validity concerns, as well as variable or unspecific timeframes provided for measure items.

3.8.2. PAI-SF

The PAI-SF suicide subscale showed good internal consistency and adequate construct validity, including convergent validity with STBs assessed via other instruments. Content validity and face validity were excellent. The items impacting face validity for the suicide subscale of the full PAI are not present in the corresponding subscale of the PAI-SF. Instead, four items cleanly and straightforwardly assess passive and active suicidal ideation, and thoughts of suicide methods or plans. Data were unavailable for test-retest reliability and sensitivity to change. Clinical utility was adequate given strong face and content validity, as well as at least adequate evaluations for all other psychometric properties for which data were available in the studies included here. The PAI-SF also showed evidence of good readability (i.e., high Flesch Reading Ease score) and its 4-item suicide subscale is brief, facilitating quick administration and scoring.[6] However, it is worth noting that the temporal window of assessment is somewhat unclear; thus, clinicians may need to ask clarification questions if item(s) are endorsed.

3.9. Self-Injurious Thoughts and Behaviors Interview (SITBI) and SITBI adaptations

The SITBI is an interview-based measure assessing a comprehensive spectrum of STBs, including suicidal ideation and suicide plans, gestures, and attempts. In this review, two papers reported on the psychometric properties of the SITBI in other languages (i.e., Korean, Spanish) across two adult samples (García-Nieto et al., 2013; Lee et al., 2021). An additional two papers evaluating adaptations of the SITBI were included (Fox et al., 2020; Stanley et al., 2023). One paper (Stanley et al., 2023) examined the SITBI-Short Form (SITBI-SF), which contains fewer items but assesses presence and frequency of the same STB outcomes. Another paper (Fox et al., 2020) examined the SITBI-Revised (SITBI-R) across two independent adult samples. Details about the SITBI-R are briefly described below and further elaborated in Fox et al. (2020).

3.9.1. SITBI

Because the SITBI includes multiple, discrete sections of varying formats to assess the presence, frequency, and other characteristics of distinct STBs, evaluations of internal consistency are not relevant to this measure. Analyses of test-retest reliability in both the Korean and Spanish translations of the SITBI yielded mixed findings across types of STBs and for assessment of presence versus frequency. For the Korean version, for example, test-retest reliability across two months was excellent for lifetime frequency of suicide planning, but adequate for suicide attempts and limited for suicidal ideation (Lee et al., 2021). For the Spanish version, test-retest reliability was good for lifetime presence of suicidal ideation, plans, and attempts, but limited for suicide gestures (Garcia-Nieto et al., 2013). Of note, both studies conducted test-retest reliability analyses in a small subset of the full sample. Additionally, interrater reliability for the Korean version was perfect for lifetime presence and frequency over multiple timeframes of suicidal ideation, plans, gestures, and attempts, although this finding was based on a very small subset of the sample ($n = 5$; Lee et al., 2021). For the Spanish version, interrater reliability was excellent for presence of STBs (i.e., suicidal ideation, plans, gestures, attempts) across multiple timeframes (i.e., lifetime, past year, past month).

Construct validity for the Korean and Spanish versions was adequate; both showed convergent validity with corresponding STBs assessed via other measures. Content validity and face validity were excellent in the case of suicidal behaviors, but content validity was only adequate in the case of suicidal ideation, given the absence of passive ideation. Clinical utility was adequate (e.g., excellent face validity, adequate to excellent content validity) and provided that the measure is first reduced to primary items of clinical interest. That the SITBI clearly specifies multiple timeframes of assessment was a further asset with regards to clinical utility. Of note, while the SITBI (and its adaptations, discussed below) received overall ratings of "adequate," there were relatively few studies examining psychometric properties of these interview-based measures

[6] For the full PAI-SF measure (approximately 160 items), one study included in this review reported an average completion time of 12.35 min (range = 6–24 min; Ward et al., 2018). This suggests that the average completion time for the 4-item suicide subscale would be extremely brief.

(versus other interview-based measures like the C-SSRS), and the SITBI showed relatively strong properties compared to many self-report measures reviewed here. Although the full interview-based format poses a potential barrier to scalability and practicality in clinical settings, the SITBI provides unique value in research settings in its comprehensive assessment of a wide spectrum of specific types of STBs.

3.9.2. SITBI-SF

The SITBI-SF assesses the same spectrum of STBs as the original SITBI with fewer items. Mirroring findings for the SITBI, the SITBI-SF showed overall excellent interrater reliability for both presence and frequency of suicidal ideation, plans, gestures, and attempts (Stanley et al., 2023). The SITBI-SF showed adequate construct validity in the one sample represented in this review. SITBI-SF items assessing suicidal ideation showed strong correlations with other measures of suicidal ideation, and items across multiple SITBI-SF sections correlated with related constructs such as depression. Data were not available for test-retest reliability, and other psychometric evaluations, including clinical utility, mirrored those of the SITBI.

3.9.3. SITBI-R

The SITBI-R differs from the SITBI in several ways, including length and assessment of additional types of suicidal behaviors (for more details, see Fox et al., 2020). Perhaps most importantly, it includes an assessment of passive ideation (albeit only if active ideation is first endorsed). In this review, one paper was found reporting on the psychometric properties of the SITBI-R across two independent adult samples.

In one study (Fox et al., 2020 Study 1), the test-reliability of the SITBI-R ranged from adequate to excellent for presence of STBs and was strongest for presence of suicidal ideation and suicide attempts, as well as select other types of STBs (e.g., aborted suicide attempts). For STB frequency, reliability indices tended to be stronger for past-year than lifetime frequency. Of note, these test-retest reliability data are based on comparisons between assessments that used different formats (i.e., SITBI-R given via in-person interview, compared to an online, self-report version given two weeks later). Other analyses of test-retest reliability using a consistent format across two administrations (i.e., SITBI-R as online self-report) were mixed (i.e., ranging from limited to excellent) for lifetime presence of STBs (Fox et al., 2020; Study 2). Stronger test-retest reliability was found for lifetime and past-year frequency of STBs. Interrater reliability for the interview-based SITBI-R was excellent for presence and frequency of all types of STBs across lifetime, past year, and past month timeframes, except for lifetime presence of suicide gesture, for which interrater reliability was lower but still good.

Content validity and face validity were excellent. The interview-based SITBI-R showed adequate construct validity, with support for convergent validity with corresponding STBs assessed using the C-SSRS. While the interview-based SITBI-R showed excellent agreement with corresponding C-SSRS items for assessment of multiple types of STBs, replication in additional samples is required to further establish construct validity. Finally, like the SITBI and SITBI-SF, clinical utility for the SITBI-R was adequate, provided that the measure is reduced to items of core clinical interest. Data also support correspondence between in-person interview and online self-report versions of the SITBI-R (Fox et al., 2020), suggesting both formats may operate similarly (e.g., elicit similar responses). Because online measures may be more time and cost efficient, certain formats of the SITBI-R may have even stronger clinical utility, provided that evidence of construct validity is observed for online administration.

3.10. Suicidal Ideation Questionnaire (SIQ)

The SIQ is a self-report measure of passive and active suicidal ideation in the prior month. This review identified one paper assessing the psychometric properties of the SIQ in two independent samples (Abdel-Khalek & Lester, 2007), with the English-language version examined in one sample and the Arabic translation in the other sample.

The SIQ showed excellent internal consistency in both languages, although similar to the ASIQ and BSS, very high alpha values, longer measure length, and some item redundancy (e.g., "I thought that people would be happier if I were not around"; "I thought about how people would feel if I killed myself") may indicate inflated (i.e., "too excellent") internal consistency. Content validity was excellent. Face validity was considered adequate and affected by a number of concerns, primarily about items assessing thoughts or experiences that are unrelated or tangential to suicidal ideation. For example, one item assesses thoughts about others' deaths. Another item may arguably better reflect thoughts of NSSI than of suicide (i.e., thoughts of hurting oneself but not killing oneself). Item response anchors raise further concerns: the second lowest anchor (i.e., had the thought but not in the one-month period covered by the measure), when considered with the lowest anchor (i.e., never had the thought), means that current suicidal ideation is confounded with its past occurrence among individuals with a history of ideation in that the former should always have higher scores on this measure than the latter. Construct validity was adequate for the English and Arabic versions of the SIQ. For example, scores on the Arabic translation were positively associated with anxiety, death obsession, and pessimism, and negatively associated with optimism (Abdel-Khalek & Lester, 2007). When considering the English and Arabic versions together, construct validity was excellent. Data were unavailable for test-retest reliability. Clinical utility was limited due to significant face validity concerns and because the longer temporal window of assessment (i.e., past month) may pose limitations when greater temporal granularity is clinically warranted.

3.11. Ultra-Short Suicidal Ideation Scale (USSIS)

The USSIS is a self-report measure of current active suicidal ideation and plans. In this review, one paper was identified reporting on the psychometric properties of the USSIS in a sample recruited across two clinical sites (Nugent & Cummings, 2014). Construct validity was adequate, as USSIS scores were shown to be positively correlated with depressive symptoms and with clinicians' ratings of suicidal thinking (Nugent & Cummings, 2014). Data were not available for internal consistency, test-retest reliability, inter-rater reliability, and sensitivity to change. Content validity was adequate, as the measure consists of just 4 items that directly assess active suicidal ideation and plans (with excellent face validity), but not passive suicidal ideation. That the USSIS assesses current suicidal ideation and plans adds to its clinical utility.

3.12. Wish to Be Dead Scale (WDS)

The WDS is a self-report measure of thoughts or desires about death or passive suicidal ideation over an unspecified timeframe. Three papers were found reporting on the psychometric properties of the WDS across three adult samples (Dadfar et al., 2017, 2018; Lester, 2013).

Internal consistency was good for the English version and mixed for the Farsi version of the WDS (i.e., internal reliability ranged from adequate to good; Dadfar et al., 2017, 2018). Because the WDS does not specify a timeframe, it is difficult to evaluate the measure's test-retest reliability and sensitivity to change. However, for both the English and Farsi versions, there was evidence of strong correlations between scores across consecutive administrations spaced one to two weeks apart. Across both languages, the WDS showed adequate to good construct validity, including positive correlations with constructs such as psychological distress and depression, and negative correlations with hopefulness and life satisfaction. Content validity was only adequate, as the WDS assesses passive, but not active, suicidal ideation. Face validity was only also adequate, as several items assess thoughts that may more closely capture related constructs, such as meaning and purpose in life (i.e., *"I sometimes think that there is no purpose to life"*). In light of these content and face validity concerns, clinical utility was considered

limited. That WDS items lack specificity in assessing the time period over which an individual has experienced these passive thoughts further hinders the measure's clinical utility.

3.13. Forkmann et al. (2018) EMA measure

One study included in this review evaluated the psychometric properties of a set of EMA items assessing suicidal ideation in one adult sample (Forkmann et al., 2018). The four-item set assesses passive and active suicidal ideation. Data supported overall good reliability of the suicidal ideation items at both the person and prompt levels, as reflected by good reliability across the full item set assessing suicidal ideation (i.e., all reliability indices across both passive and active item sets >0.80). Content validity was good. Of note, however, passive and active suicidal ideation items (two items each) were analyzed as separate scales, and the two-item active suicidal ideation set included a passive ideation item (i.e., *"I want to die"*). Although the active ideation set has face validity insofar as its items both reflect suicidal ideation, this issue regarding misclassification of a passive ideation item as active ideation presents a face validity concern. There was evidence of adequate construct validity, including positive correlations for suicidal ideation with thwarted belongingness, perceived burdensomeness, depression, and hopelessness (i.e., convergent validity), and negative correlations with positive affect (i.e., divergent validity). Clinical utility was limited given face validity concerns, and more data are needed further evaluating this item set developed specifically for this study. As EMA-based assessments are used most often in research, their translation to clinical care requires evaluation.

3.14. Overall measure evaluations

The majority of measures received an overall rating of "limited." While the primary limitations driving this designation varied, measures frequently did not receive ratings of at least "adequate" for clinical utility. Clinical utility was frequently impacted by face validity concerns (e.g., ASIQ, BSS, SIQ); longer, variable, or unspecified timeframes for individual items or the full measure (e.g., ASIQ, ASIQ-4, PAI, PAI-SF, WDS); and other immutable characteristics (i.e., poorer readability, proprietary) impacting accessibility (e.g., ASIQ, ASIQ-4, BSS, GSIS, Brief GSIS). The SITBI, SITBI-SF, and SITBI-R were the only three measures to receive an overall rating of "adequate," and the C-SSRS the only one to receive an overall rating of "good." No measures were rated "excellent" overall.

4. Discussion

Suicidal thoughts and behaviors (STBs) are prevalent, often undetected, and potentially lethal, underscoring the importance of psychometrically sound measures to assess and monitor their presence and frequency. The extent to which currently available measures reliably and accurately capture STBs has direct implications for research quality and clinical care of suicidal individuals. Despite this, surprisingly little work has rigorously examined the reliability, validity, and other measurement-related characteristics of prevailing STB assessment tools (Spears et al., 2023). As a step toward addressing this gap, the present systematic review identified measures of STBs and evaluated their psychometric properties by extracting and reviewing psychometrics-relevant data available in published studies. In addition to reliability, validity, and other psychometric properties and measure characteristics described in the results section and presented in Tables 2 and S2, this paper also evaluated the impact of these properties on measures' clinical utility. Results of this review also yield critical implications for current research on STBs, which are further discussed here.

This review made apparent that a majority of STB measures assess suicidal ideation, with fewer also assessing suicidal behaviors (e.g., suicide attempts) and very few assessing other more granular or specific suicide-relate outcomes (e.g., suicide gestures, aborted or interrupted suicide attempts). Among the measures reviewed here, a majority assessed suicidal ideation only (i.e., passive and/or active suicidal ideation, with or without plans). While some of these measures included one item assessing suicide attempts, few measures (e.g., C-SSRS, SITBI and its variants) assessed suicidal behaviors with multiple items or with a distinct subscale or section. This imbalance mirrors patterns in the broader suicide research literature, with studies examining suicidal ideation (with or without suicidal behavior) far outweighing those that focus on suicidal behavior (Franklin et al., 2017). The relative lack of measures assessing suicidal behavior (i.e., with more than one item) is concerning given evidence that single-item assessments of suicide attempts may result in misclassification of prior suicidal behaviors (Hom et al., 2016).

Given fewer measures of suicidal behaviors are available to clinicians and researchers, psychometric conclusions about these measures may have particularly impactful implications. Two multi-item measures of suicidal behavior in this review—the C-SSRS and the SITBI (and its variants: the SITBI-SR and SITBI-R)—are interview-based and were also among the few measures to show generally strong validity (i.e., excellent content and face validity; at least adequate construct validity) and adequate (versus limited) clinical utility. On the one hand, interviews permit use of interviewer (e.g., clinician) judgment during assessment, which may be advantageous when the suicidal nature or intent of thoughts or behaviors is ambiguous. On the other hand, the time and expertise required for this format is a potential barrier to their implementation in clinical settings. An avenue for future research will be continued translation of interview-based measures into abbreviated and/or alternative (e.g., self-report, computer, or AI-delivered) formats. Of note, studies in this review that administered these interview-based measures as self-reports, including online self-reports, tended to show roughly equivalent psychometric properties with the traditional, interview-based format (Fox et al., 2020; Franks et al., 2021; Tabares et al., 2021). While adaptability into alternative formats (e.g., self-report) may increase the practicality of administering these measures in clinical settings, assessment of certain STBs may necessitate more nuanced judgment (e.g., determining whether dangerous behaviors constitute suicide attempts). Balancing reliability and validity with clinical practicality and scalability will be an ongoing priority for the development, translation, and evaluation of tools that assess a full range of STBs, including behaviors. Further phenomenological study of STBs for which current definitions are broader or more ambiguous, like suicide gestures, is also needed. Indeed, Fox et al. (2020) found comparatively weaker psychometric properties of the SITBI-R for evaluation of suicide gestures, compared to other STBs.

Regarding types of STBs assessed, this review also demonstrated marked variability in whether and how measures assessed suicide plans. While measures more frequently asked about suicide plans than about behavior, they did so in varied or idiosyncratic ways: together with items assessing suicidal ideation (e.g., C-SSRS) or in separate questions (e.g., PAI, PAI-SF) or sections/modules (e.g., SITBI-SF, SITBI-R). Language describing suicide plans or planning varied: some measure items directly asked about "suicide plans" (i.e., without defining this term) and others referenced aspects of suicide planning, such as consideration of ways/means and/or methods. As small differences in wording can impact endorsement when assessing STBs (Ammerman et al., 2021), it will be important for researchers to investigate how best to assess suicide plans (or planning-related thoughts or behaviors), which may involve considering whether plans are best conceptualized and measured similarly to suicidal thoughts or behaviors.

This review also uncovered a number of face validity concerns across measures, including those that are more commonly used or well-known. While concerning, this is not surprising given marked variability in how researchers and clinicians have conceptualized STBs (De Leo et al., 2006; Goodfellow et al., 2018). Issues with face validity were most apparent in self-report measures and for items assessing suicidal

Table 2
Summary of norms, reliability, validity, and clinical utility ratings by measure.

Criterion	ASIQ	ASIQ-4	BSS (Clinician-Rated Version)[a]	BSS (Self-Report Version)	C-SSRS[d]	CHRT[a]
Overall	Limited	Limited	Limited	Limited	Good	Limited
Norms[b]	Adequate[2,3] No data[2]	No data[2]	No data	Adequate[10,11] No data[2,3,5,7]	Excellent[3,12] Good[2] No data[vi]	Excellent No data[v]
Internal Consistency[b]	Excellent[2,3; viii]	Excellent[2]	Good	Excellent[3,5, viii] Good[2,7,10,11]	Mixed[i,vi,3,12] Good[2]	Good[ii] No data[v]
Test-Retest Reliability	Excellent[3] No data[2]	No data[2]	No data	No data[2,3,5,7,10,11]	No data[2,3,12;vi]	No data
Inter-Rater Reliability[c]	Not applicable	Not applicable	Good	Not applicable	Good (SI)[3] Mixed (suicidal behavior)[3] No data[2,12,vi]	Not applicable
Face Validity	Adequate	Adequate	Adequate	Adequate	Excellent	Excellent
Content Validity	Excellent	Adequate	Excellent	Excellent	Excellent	Adequate
Construct Validity (e.g., convergent, discriminant, divergent validity)	Adequate[2,3]	Adequate[2]	Adequate	Excellent[3] Adequate[2,5,10,11] No data[7]	Good (SI)[3,12] Adequate (SI)[2] Adequate (suicidal behavior)[3] No data[vi]	Excellent Adequate[v]
Sensitivity to Change[d]	No data[2,3]	No data[2]	No data	No data[2,3,5,7,11] Limited[10]	No data[2,3,12, vi]	Adequate
Clinical Utility	Limited	Limited	Limited	Limited	Adequate	Limited

Criterion	DSI-SS[a]	GSIS	Brief GSIS[a]	MSI-28[a]	PAI	PAI-SF[a]
Overall	Limited	Limited	Limited	Limited	Limited	Limited
Norms[b]	Excellent	No data[2,3]	Adequate	Adequate	Excellent[3] Adequate[6,8,12] No data[4]	No data
Internal Consistency[b]	Excellent	Good[2,3]	Excellent	Excellent	Mixed[i;3] Good[6,8,12] Limited[4]	Good
Test-Retest Reliability	No data	No data[2,3]	No data	No data	No data[3,4,6,8,12]	No data
Inter-Rater Reliability[c]	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable
Face Validity	Excellent	Adequate	Adequate	Adequate	Adequate	Excellent
Content Validity	Adequate	Excellent	Good	Adequate	Excellent	Excellent
Construct Validity (e.g., convergent, discriminant, divergent validity)	Good	Excellent[3] Adequate[2]	Adequate	Adequate	Good[3] Adequate[12] Limited[8] No data[4,6]	Adequate
Sensitivity to Change[d]	No data	No data[2,3]	No data	No data	No data[3,4,6,8,12]	No data
Clinical Utility	Limited	Limited	Limited	Limited	Limited	Adequate

Criterion	SIQ	SITBI	SITBI-SF[a]	SITBI-R[a]	USSIS[a]	WDS	Forkmann et al. (2018) EMA items[a]
Overall	Limited	Adequate	Adequate	Adequate	Limited	Limited	Limited
Norms[b]	No data[1,3]	Not applicable	Not applicable	Not applicable	No data	No data[3,11]	No data

(*continued on next page*)

Table 2 (*continued*)

Criterion	SIQ	SITBI	SITBI-SF[a]	SITBI-R[a]	USSIS[a]	WDS	Forkmann et al. (2018) EMA items[a]
Internal Consistency[b]	Excellent[1,3; viii]	Not applicable	Not applicable	Not applicable	No data	Good[3] Mixed[11, i]	Good[vii]
Test-Retest Reliability	No data[1,3]	Mixed[iii] (Excellent: suicide plan frequency; Adequate: SA frequency; Limited: SI frequency)[9]; (Good: SI, suicide plan, SA presence; Limited: suicidal gesture presence)[12]	No data	Mixed[iii]	No data	No data/unclear[3,11, iv]	No data
Inter-Rater Reliability[c]	Not applicable	Excellent[9,12]	Excellent	Excellent	No data	Not applicable	Not applicable
Face Validity	Adequate	Excellent	Excellent	Excellent	Excellent	Adequate	Adequate
Content Validity	Excellent	Adequate (SI)/ Excellent (suicidal behaviors)	Adequate (SI)/ Excellent (suicidal behaviors)	Excellent	Adequate	Adequate	Good
Construct Validity (e.g., convergent, discriminant, divergent validity)	Adequate[1,3]	Adequate[9,12]	Adequate	Adequate	Adequate	Adequate[3] Good[11]	Adequate
Sensitivity to Change[d]	No data[1,3]	Not applicable[9,12]	Not applicable	Not applicable	No data	No data[3,11]	No data
Clinical Utility	Limited	Adequate	Adequate	Adequate	Adequate	Limited	Limited

Note. In cases where multiple language versions of a measure were assessed, superscripts are used to denote different language version. When only the English version of a measure was reviewed, a superscript is not used. ASIQ = Adult Suicidal Ideation Questionnaire; BSS = Beck Scale for Suicide Ideation; CHRT = Concise Health Risk Tracking; C-SSRS = Columbia-Suicide Severity Rating Scale; DSS-SS = Depressive Symptom Index-Suicidality Subscale; EMA = ecological momentary assessment; GSIS = Geriatric Suicide Ideation Scale; MSI-28 = Multidimensional Suicide Inventory-28; PAI = Personality Assessment Inventory; PAI-SF = Personality Assessment Inventory-Short Form; SA = suicide attempt, SI = suicidal ideation; SIQ = Suicidal Ideation Questionnaire; SITBI = Self-Injurious Thoughts and Behaviors Interview; SITBI-R = Self-Injurious Thoughts and Behaviors Interview-Revised; SITBI-SF = Self-Injurious Thoughts and Behaviors Interview-Short Form; USSIS = Ultra-Short Suicidal Ideation Scale; WDS = Wish to Be Dead Scale.

[a] Psychometrics reported in this column reflect only the measure in English.

[b] Only applicable to measures with (sub)scale scores.

[c] Only applicable to interview-based measures.

[d] Psychometrics reported for the C-SSRS capture assessment time frames ranging from lifetime to past-week, as well as time since last assessment. Ratings for norms and internal consistency pertain to suicidal ideation subscales. Psychometric ratings reported for suicidal behavior excluded studies that explicitly included an item assessing self-injurious behavior without intent to die in analyses of suicidal behavior.

[1] Arabic version.

[2] Chinese version.

[3] English version.

[4] French version.

[5] German version.

[6] Greek version.

[7] Indonesian version.

[8] Italian version.

[9] Korean version.

[10] Norwegian version.

[11] Persian/Farsi version.

[12] Spanish version.

[i] "Mixed" here indicates that studies reported a range of values for internal consistency, reflecting evidence ranging from "limited" and/or "adequate," to "good" and/or "excellent" across either measure total and/or subscale scores.

[ii] Internal consistency for CHRT-SR suicide subscale was overall good, with one exception. Trombello et al. (2019) found evidence of limited internal consistency, as reflected by low Cronbach's alpha.

[iii] 'Mixed' here indicates that statistics assessing test-retest reliability ranged from "limited" and/or "adequate," to "good" and/or "excellent" across types of STBs.

[iv] As the WDS does not specify a timeframe of assessment, test-retest reliability cannot be rigorously evaluated according to this review's criteria. However, some papers reported correlations indicative of excellent-to-good reliability across consecutive assessments for the English and Farsi-language versions, respectively.

[v] Reflects the CHRT when administered by a clinician/interviewer (Trivedi et al. 2011), in cases where the psychometric evaluations for a given category differed from the self-report format. For the interviewer-based CHRT, other psychometric evaluations (i.e., those not indicated by this superscript) are the same as those for the self-report CHRT.

[vi] Reflects the C-SSRS when administered as a self-report questionnaire (Franks et al., 2021; Tabares et al., 2021).

[vii] This captures an overall evaluation of multiple reliability indices, across both 'person' and 'prompt' levels, for the item set assessing suicidal ideation, administered via ecological momentary assessment (Forkmann et al. 2018).

[viii] Consider that internal consistency may be "too excellent" and inflated by, e.g., scale length (long) and/or scope (narrow, item redundancy). For more details regarding this consideration, see Youngstrom et al. (2017).

thoughts. While many measures directly ask about desire for death, thoughts of killing oneself, or thoughts of suicide, a number of items use more ambiguous or tangential language, or better reflect related but different constructs (e.g., thwarted belongingness). Some other items appear unrelated to suicide, and may in relatively rare circumstances reflect homicidal ideation, or obsessive-compulsive thoughts insofar as they are egodystonic (e.g., thoughts of other people dying). Additionally, some measures include items that ask about thoughts regarding others' reactions to suicide or thoughts of sharing suicidal thoughts with others. While such items may elicit clinically relevant information, the extent to which endorsing these items is indicative of STB severity is unclear. That individuals often cite confidentiality concerns and fear of repercussions as primary reasons for reluctance to report STBs (Deming et al., 2021; Fulginiti & Frey, 2019; McGillivray et al., 2022) suggests that disclosure willingness may be relatively independent of actual suicidal ideation severity. For example, a positive response to an item asking about thoughts of disclosing suicidal ideation with others may in some cases indicate higher suicidal ideation (e.g., an individual recognizing their thoughts are severe and indicating a need to seek help). In some other cases, a negative response to the same item may reflect greater suicidal ideation (e.g., an individual recognizes their thoughts are severe but does not wish to disclose out of fear of hospitalization and/or stigma; Podlogar et al., 2016; Podlogar & Joiner, 2020). Inclusion of these items in a total score may obscure the extent to which measure scores reflect meaningful variability in STB severity.

Measures that adhere more closely to parsimonious and direct assessment of STBs, using language that explicitly names these outcomes, tended to show stronger face validity. For suicidal ideation, measures such as the CHRT (suicidal thoughts subscale), C-SSRS, DSI-SS, SITBI, SITBI-SF, and SITBI-R had on average a greater proportion of items that directly asked about thoughts of killing oneself or suicide—or closely related thoughts capturing more passive ideation that still imply a desire to die (e.g., "I wish I could go to sleep and not wake up"). This is in keeping with recommendations for nomenclature in suicidology that emphasize use of simple, commonly understood terms that describe or define the basic, clinical phenomenological aspects of STBs (e.g., Silverman, 2006). Indeed, an important function of some assessment measures—in addition to evaluation of presence or severity of clinical phenomena—is to classify thoughts or behaviors into (sub)types (e.g., classification of a suicide attempt as an aborted or interrupted attempt). For these reasons, although items included in measures such as the BSS and SIQ that tap related but different phenomena may be important if they facilitate more specific classification of STBs into categories, in the absence of subscales, discrete sections, or an underlying classification framework, however, these tangential or more descriptive items undermine overall face validity and measurement precision.

Unsurprisingly, measures with stronger face validity were rated as having better clinical utility. Particularly for measures of suicidal ideation, those with stronger clinical utility also assessed current STBs and/or STBs occurring over shorter, recent intervals (e.g., past week). These timeframes are important given research suggesting that retrospective recall impacts reports of an array of psychiatrically relevant phenomena (Liu, 2017; Moffitt et al., 2010), including STBs, and this may be even more true for thoughts (e.g., suicidal ideation) than for behaviors (Andrewes et al., 2024). Individuals may also be more likely to endorse STBs, particularly suicidal ideation, when asked about shorter, proximal timeframes (versus longer, more distal intervals; Gratch et al., 2021), and endorsement of STBs has been shown to be higher across repeated assessments covering shorter intervals (versus fewer or one assessment covering a longer interval; Esposito et al., 2022; Czyz et al., 2018). Shorter timeframes may therefore enhance a measure's utility by facilitating more precise risk management and monitoring of treatment course in clinical settings, and greater measurement precision in research studies.

It will also be important for future work to further examine sensitivity to change, as this review revealed a stark lack of data rigorously examining the extent to which measures can detect changes in STBs over temporally distinct (i.e., non-overlapping) timeframes. Several studies reported analyses that were presented as tests of sensitivity to change, but which compared measures across overlapping timeframes. Studies should also test sensitivity to change across multiple measure modalities and assessment schedules, to evaluate the extent to which measures reflect changes in STBs across a range of timescales—particularly given popularity of ambulatory assessment methods that rely on many repeated assessments over micro-timescales.

Many of these considerations have direct implications for research efforts to advance understanding of STBs. Critically, empirical study of the phenomenology of STBs requires valid and reliable measures with specificity to identify distinct forms (e.g., suicidal ideation versus behavior) and characteristics (e.g., severity, duration) of STBs. Recent studies have begun to shed light on increasingly fine-grained aspects of STB phenomenology, such as their short-term temporal dynamics (Coppersmith et al., 2023). While use of single items to assess STBs in research is not uncommon, other recent studies demonstrate that single-item assessment may impact endorsement (Ammerman et al., 2021), further highlighting the importance of psychometrically sound, multi-item measures, even if brief, for use in research. As current studies aim to characterize patterns of STBs across time, settings, and transitions between STBs (Oakey-Frost et al., 2023), continued scrutiny will be critical to ensure that measures (e.g., format, psychometrics) align as closely as possible with the phenomenological structure of STBs.

The clinical and research utility of STB measures could be strengthened through additional changes. For instance, utility might be improved through collection and analysis of more data to establish norms and clinical cut-off scores, or other standardized metrics for ascertaining clinical severity. To start, studies that administer STB measures should prioritize reporting of basic descriptive statistics (e.g., means and standard deviations of full measure and subscale scores). Researchers might also capitalize on availability of large, representative datasets that include data from a measure (or subscale) of STBs to obtain population norms. It may be possible to leverage these data to inform derivation or validation of candidate cut-off scores using integrative data analysis or other data harmonization methods (Hussong et al., 2013). Studies might also derive or test candidate cut-off scores empirically. One study in this review proposed cut-off scores for the BSS Self-Report Version based on values that optimized measure sensitivity, specificity, and predictive values, although these values were derived post hoc in a small, clinically acute sample (Cochrane-Brink et al., 2000). The STB assessment literature would benefit from studies that develop alternative frameworks for ascertaining STB severity based on measure data across clinical and nonclinical populations.

While additional research may facilitate improvements, the present review highlights an urgent need for the development of *new* STB measures, particularly self-report measures. There were significant concerns inherent to the content of many existing measures, which cannot be remedied with more data. Content or face validity issues, such as lack of items assessing passive suicidal ideation and questionable item language, were prevalent among the self-report measures reviewed here. A key priority will be to develop new—particularly self-report—measures not only for clinical assessment, but also to ensure that future research findings are grounded in measurement tools that reflect current knowledge of STB phenomenology. A second priority is to establish psychometric characteristics of existing and new (i.e., future) STB measures, and ensure measure accuracy in particular, with at-risk minoritized populations (e.g., sexual and gender minority individuals; Erlangsen et al., 2023; Haas et al., 2011) to reduce disparities in STBs and treatment of minoritized individuals. Only one study included in this review (Jeon et al., 2024) found support for measurement invariance (i.e., of the DSI-SS) across minoritized racial, ethnic, and sexual orientation identities compared to majority-aligning individuals. There is a problematic lack of research examining how STB measures, and their psychometric properties, relate to cultural, linguistic, racial/ethnic,

sexual, and gender identity and other aspects of diversity. Substantially more research is needed. Among studies in this review that reported racial/ethnic demographic data, most were conducted in majority (i.e., >50%) White samples, consistent with evidence of poor sample representativeness in STB research (Cha et al., 2018; Guzmán et al., 2024) and emphasizing the need for sample diversity. Measure translation into alternative delivery modes (e.g., technology-assisted) and cultural formats (e.g., non-English languages, wording adaptations to reflect culturally specific conceptions of suicide; Molock et al., 2023) is also needed to facilitate scalability and accessibility across diverse backgrounds.

Role of funding sources

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 RTL; F31MH134599-01A1 to OHP; K23MH122737 to AHB; by an American Foundation for Suicide Prevention grant (PDF-0-095-19) to HRL; and by National Science Foundation Graduate Research Fellowships to RFLW and AES. RTL 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.

Declaration of competing interest

The authors report no conflict of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cpr.2024.102464.

References

Abilleira, M. P., & Rodicio-García, M. L. (2019). Psychometric properties of the abbreviated version of Personality Assessment Inventory (PAI-R) in a sample of inmates in Spanish prisons. *Journal of Forensic Psychology Research and Practice, 19*(1), 9–23.

Al-Halabí, S., Sáiz, P. A., Burón, P., Garrido, M., Benabarre, A., Jiménez, E., … Bobes, J. (2016). Validation of a Spanish version of the Columbia-Suicide Severity Rating Scale (C-SSRS). Validación de la versión en español de la Columbia-Suicide Severity Rating Scale (Escala Columbia para Evaluar el Riesgo de Suicidio). *Revista de Psiquiatria y Salud Mental, 9*(3), 134–142.

Alterman, A. I., Zaballero, A. R., Lin, M. M., Siddiqui, N., Brown, L. S., Rutherford, M. J., & McDermott, P. A. (1995). Personality Assessment Inventory (PAI) scores of lower-socioeconomic African American and Latino methadone maintenance patients. *Assessment, 2*(1), 91–100.

Ammerman, B. A., Burke, T. A., Jacobucci, R., & McClure, K. (2021). How we ask matters: The impact of question wording in single-item measurement of suicidal thoughts and behaviors. *Preventive Medicine, 152*(Pt 1), Article 106472.

Andrewes, H. E., Cavelti, M., Hulbert, C., Cotton, S. M., Betts, J. K., Jackson, H. J., … Chanen, A. M. (2024). An analysis of real-time suicidal ideation and its relationship with retrospective reports among young people with borderline personality disorder. *Suicide and Life-Threatening Behavior*.

Ayer, L., Horowitz, L. M., Colpe, L., Lowry, N. J., Ryan, P. C., Boudreaux, E., … Schoenbaum, M. (2022). Clinical pathway for suicide risk screening in adult primary care settings: Special recommendations. *Journal of the Academy of Consultation-Liaison Psychiatry, 63*(5), 497–510.

Azcurra, D. (2017). Psychometric validation of the Columbia-Suicide Severity rating scale in Spanish-speaking adolescents. *Colombia Medica, 48*(4), 174–182.

Baca-Garcia, E., Perez-Rodriguez, M. M., Oquendo, M. A., Keyes, K. M., Hasin, D. S., Grant, B. F., & Blanco, C. (2011). Estimating risk for suicide attempt: Are we asking the right questions?: Passive suicidal ideation as a marker for suicidal behavior. *Journal of Affective Disorders, 134*(1–3), 327–332.

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.

Batterham, P. J., Ftanou, M., Pirkis, J., Brewer, J. L., Mackinnon, A. J., Beautrais, A., … Christensen, H. (2015). A systematic review and evaluation of measures for suicidal ideation and behaviors in population-based research. *Psychological Assessment, 27*(2), 501–512.

Beck, A. T., Davis, J. H., Frederick, C. J., Perlin, S., Pokorny, A. D., Schulman, R. E., et al. (1973). Classification and nomenclature. In H. L. P. Resnick, & B. C. Hathore (Eds.), *Suicide prevention in the seventies* (pp. 7–12). Washington, DC: U.S. Government Printing Office.

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*(4), 499–505.

Bezdjian, S., Burchett, D., Schneider, K. G., Baker, M. T., & Garb, H. N. (2015). Multidimensional Suicide Inventory-28 (MSI-28) within a sample of military basic trainees: An examination of psychometric properties. *Military Psychology, 27*(6), 325–334.

Borschmann, R., Hogg, J., Phillips, R., & Moran, P. (2012). Measuring self-harm in adults: A systematic review. *European Psychiatry, 27*(3), 176–180.

Boyle, G. J. (1991). Does item homogeneity indicate internal consistency or item redundancy in psychometric scales? *Personality and Individual Differences, 12*(3), 291–294.

Bradley, R., Hilsenroth, M., Guarnaccia, C., & Westen, D. (2007). Relationship between clinician assessment and self-assessment of personality disorders using the SWAP-200 and PAI. *Psychological Assessment, 19*(2), 225–229.

Brahmbhatt, K., Kurtz, B. P., Afzal, K. I., Giles, L. L., Kowal, E. D., Johnson, K. P., … Workgroup, P. C. C. (2019). Suicide risk screening in pediatric hospitals: Clinical pathways to address a Global Health crisis. *Psychosomatics, 60*(1), 1–9.

Brown, L. A., Boudreaux, E. D., Arias, S. A., Miller, I. W., May, A. M., Camargo, C. A., Jr., … Armey, M. F. (2020). C-SSRS performance in emergency department patients at high risk for suicide. *Suicide and Life-Threatening Behavior, 50*(6), 1097–1104.

Bryan, C. J., & Rudd, M. D. (2016). The importance of temporal dynamics in the transition from suicidal thought to behavior. *Clinical Psychology: Science and Practice, 23*(1), 21–25.

Burneo-Garcés, C., Fernández-Alcántara, M., Aguayo-Estremera, R., & Pérez-García, M. (2020). Psychometric Properties of the Spanish Adaptation of the Personality Assessment Inventory in Correctional Settings: An ESEM Study. *Journal of Personality Assessment, 102*(1), 75–87.

Busse, M., Whiteside, D., Waters, D., Hellings, J., & Ji, P. (2014). Exploring the reliability and component structure of the personality assessment inventory in a neuropsychological sample. *The Clinical Neuropsychologist, 28*(2), 237–251.

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*(2), 178–199.

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*(2), 131–139.

Chioqueta, A. P., & Stiles, T. C. (2006). Psychometric properties of the Beck Scale for Suicide Ideation: a Norwegian study with university students. *Nordic Journal of Psychiatry, 60*(5), 400–404.

Chou, K. L., Jun, L. W., & Chi, I. (2005). Assessing Chinese older adults' suicidal ideation: Chinese version of the Geriatric Suicide Ideation Scale. *Aging and Mental Health, 9*(2), 167–171.

Chu, C., Klein, K. M., Buchman-Schmitt, J. M., Hom, M. A., Hagan, C. R., & Joiner, T. E. (2015). Routinized assessment of suicide risk in clinical practice: An empirically informed update. *Journal of Clinical Psychology, 71*(12), 1186–1200.

Cochrane-Brink, K. A., Lofchy, J. S., & Sakinofsky, I. (2000). Clinical rating scales in suicide risk assessment. *General Hospital Psychiatry, 22*(6), 445–451.

Coppersmith, D. D., Ryan, O., Fortgang, R. G., Millner, A. J., Kleiman, E. M., & Nock, M. K. (2023). Mapping the timescale of suicidal thinking. *Proceedings of the National Academy of Sciences, 120*(17), Article e2215434120.

Corsby, A. E., Ortega, L., & Melanson, C. (2011). *Self-directed violence surveillance: Uniform definitions and recommended data elements, version 1.0*. Atlanta, GA: Centers for Disease Control and Prevention, National Center for Injury Prevention and Control. https://www.cdc.gov/violenceprevention/pdf/self-directed-violence-a.pdf.

Corsica, J. A., Azarbad, L., McGill, K., Wool, L., & Hood, M. (2010). The Personality Assessment Inventory: clinical utility, psychometric properties, and normative data for bariatric surgery candidates. *Obesity Surgery, 20*(6), 722–731.

Cronbach, L. J., & Shavelson, R. J. (Eds.). (2004). *Educational and Psychological Measurement, 64*(3), 391–418.

Czyz, E. K., King, C. A., & Nahum-Shani, I. (2018). Ecological assessment of daily suicidal thoughts and attempts among suicidal teens after psychiatric hospitalization: Lessons about feasibility and acceptability. *Psychiatry Research, 267*, 566–574.

De La Garza, N., Rush, A. J., Killian, M. O., Grannemann, B. D., Carmody, T. J., & Trivedi, M. H. (2019). The Concise Health Risk Tracking Self-Report (CHRT-SR) assessment of suicidality in depressed outpatients: A psychometric evaluation. *Depression and Anxiety, 36*(4), 313–320.

De Leo, D., Burgis, S., Bertolote, J. M., Kerkhof, A. J., & Bille-Brahe, U. (2006). Definitions of suicidal behavior: Lessons learned from the WHO/EURO multicentre study. *Crisis, 27*(1), 4–15.

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*(3), 218–229.

Erlangsen, A., Jacobsen, A. L., Ranning, A., Delamare, A. L., Nordentoft, M., & Frisch, M. (2023). Transgender identity and suicide attempts and mortality in Denmark. *JAMA, 329*, 2145.

Esfahani, M., Hashemi, Y., & Alavi, K. (2015). Psychometric assessment of beck scale for suicidal ideation (BSSI) in general population in Tehran. *Medical Journal of the Islamic Republic of Iran, 29*, 268.

Esposito, E. C., Duan, A. M., Kearns, J. C., Kleiman, E. M., Conwell, Y., & Glenn, C. R. (2022). Measuring Adolescents' self-injurious thoughts and behaviors: Comparing ecological momentary assessment to a traditional interview. *Research on Child and Adolescent Psychopathology, 50*(8), 1095–1105.

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology, 32*(3), 221–233.
Flesch, R. (1949). *The art of readable writing*. New York: Harper.
Flores-Kanter, P. E., & Alvarado, J. M. (2024). State of Open Science practices in psychometric studies of suicide: A systematic review. *Assessment*, 10731911241236315.
Fowler, J. C. (2012). Suicide risk assessment in clinical practice: Pragmatic guidelines for imperfect assessments. *Psychotherapy (Chicago, Ill.), 49*(1), 81–90.
Fox, K. R., Harris, J. A., Wang, S. B., Millner, A. J., Deming, C. A., & Nock, M. K. (2020). Self-injurious thoughts and behaviors interview-revised: Development, reliability, and validity. *Psychological Assessment, 32*(7), 677–689.
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*(2), 187.
Franks, M., Cramer, R. J., Cunningham, C. A., Kaniuka, A. R., & Bryan, C. J. (2021). Psychometric assessment of two suicide screeners when used under routine conditions in military outpatient treatment programs. *Psychological Services, 18*(3), 433–439.
Frazier, T. W., Naugle, R. I., & Haggerty, K. A. (2006). Psychometric adequacy and comparability of the short and full forms of the Personality Assessment Inventory. *Psychological Assessment, 18*(3), 324–333.
Fu, K. W., Liu, K. Y., & Yip, P. S. (2007). Predictive validity of the Chinese version of the Adult Suicidal Ideation Questionnaire: psychometric properties and its short version. *Psychological Assessment, 19*(4), 422–429.
Fulginiti, A., & Frey, L. M. (2019). Exploring suicide-related disclosure motivation and the impact on mechanisms linked to suicide. *Death Studies, 43*(9), 562–569.
Gleeson, H., Roesch, C., Hafford-Letchfield, T., & Ellmers, T. (2022). Assessing suicide ideation among older adults: A systematic review of screening and measurement tools. *International Psychogeriatrics, 34*(5), 439–452.
Glenn, C. R., Lanzillo, E. C., Esposito, E. C., Santee, A. C., Nock, M. K., & Auerbach, R. P. (2017). Examining the course of suicidal and nonsuicidal self-injurious thoughts and behaviors in outpatient and inpatient adolescents. *Journal of Abnormal Child Psychology, 45*, 971–983.
Goodfellow, B., Kölves, K., & de Leo, D. (2018). Contemporary nomenclatures of suicidal behaviors: A systematic literature review. *Suicide & Life-Threatening Behavior, 48*(3), 353–366.
Gratch, I., Choo, T. H., Galfalvy, H., Keilp, J. G., Itzhaky, L., Mann, J. J., … Stanley, B. (2021). Detecting suicidal thoughts: The power of ecological momentary assessment. *Depression and Anxiety, 38*(1), 8–16.
Guzmán, E. M., LeDuc, M. K., Cha, C. B., Goger, P., Ng, M. Y., Huang, X., … Fox, K. R. (2024). Accounting for diversity in the treatment of suicide and self-injury: A systematic review of the past 50 years of randomized controlled trials. *Suicide & Life-Threatening Behavior*.
Haas, A. P., Eliason, M., Mays, V. M., Mathy, R. M., Cochran, S. D., D'Augelli, A. R., … Clayton, P. J. (2011). Suicide and suicide risk in lesbian, gay, bisexual, and transgender populations: Review and recommendations. *Journal of Homosexuality, 58*, 10–51.
Heisel, M. J., & Flett, G. L. (2006). The development and initial validation of the geriatric suicide ideation scale. *The American Journal of Geriatric Psychiatry, 14*(9), 742–751.
Heisel, M. J., & Flett, G. L. (2016). Investigating the psychometric properties of the Geriatric Suicide Ideation Scale (GSIS) among community-residing older adults. *Aging and Mental Health, 20*(2), 208–221.
Heisel, M. J., & Flett, G. L. (2022). Screening for suicide risk among older adults: assessing preliminary psychometric properties of the Brief Geriatric Suicide Ideation Scale (BGSIS) and the GSIS-Screen. *Aging and Mental Health, 26*(2), 392–406.
Hom, M. A., Joiner, T. E., Jr., & Bernert, R. A. (2016). Limitations of a single-item assessment of suicide attempt history: Implications for standardized suicide risk assessment. *Psychological Assessment, 28*(8), 1026–1030.
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. *Journal of Clinical Psychology, 75*(7), 1332–1349.
Hopwood, C. J., Baker, K. L., & Morey, L. C. (2008). Extratest validity of selected personality assessment inventory scales and indicators in an inpatient substance abuse setting. *Journal of Personality Assessment, 90*(6), 574–577.
Hussong, A. M., Curran, P. J., & Bauer, D. J. (2013). Integrative data analysis in clinical psychology research. *Annual Review of Clinical Psychology, 9*, 61–89.
Ivey-Stephenson, A. Z., Crosby, A. E., Hoenig, J. M., Gyawali, S., Park-Lee, E., & Hedden, S. L. (2022). Suicidal thoughts and behaviors among adults aged ≥18 years - United States, 2015-2019. In *, 71(1). Morbidity and mortality weekly report. Surveillance summaries (Washington, D.C.: 2002)* (pp. 1–19).
Jeffay, E., Sekely, A., Lacerte, M., & Zakzanis, K. K. (2020). Reliability of the French-Canadian adaptation of the Personality Assessment Inventory: Medical-legal implications. *Psychiatry, Psychology, and Law, 28*(1), 135–148.
Jeon, M. E., Udupa, N. S., Potter, M. R., Robison, M., Robertson, L., Rogers, M. L., & Joiner, T. E. (2024). Measurement invariance of the Depressive Symptom Inventory–Suicidality Scale across race, ethnicity, sexual orientation, and plurality of minoritized identities. *Psychological Assessment, 36*(4), 303–310.
Ji, Y., Liu, X., Zheng, S., Zhong, Q., Zheng, R., Huang, J., & Yin, H. (2023). Validation and application of the Chinese version of the Columbia-Suicide Severity Rating Scale: Suicidality and cognitive deficits in patients with major depressive disorder. *Journal of Affective Disorders, 342*, 139–147.
Jobes, D. A. (1995). The challenge and the promise of clinical suicidology. *Suicide and Life-Threatening Behavior, 25*(4), 437–449.
Jobes, D. A., & Barnett, J. E. (2024). Evidence-based care for suicidality as an ethical and professional imperative: How to decrease suicidal suffering and save lives. *American Psychologist*. Advance online publication.
Karlin, B. E., Creech, S. K., Grimes, J. S., Clark, T. S., Meagher, M. W., & Morey, L. C. (2005). The Personality Assessment Inventory with chronic pain patients: Psychometric properties and clinical utility. *Journal of Clinical Psychology, 61*(12), 1571–1585.
Keliat, B. A., Nasution, R. A., Handini, I. T., & Falani, I. (2023). Psychometric properties evaluation of the Indonesian version of the Beck Scale for Suicide Ideation (BSSI) questionnaire using a Rasch model. *European Review for Medical and Pharmacological Sciences, 27*(21), 10669–10677.
Kessler, R. C., Berglund, P., Borges, G., Nock, M., & Wang, P. S. (2005). Trends in suicide ideation, plans, gestures, and attempts in the United States, 1990-1992 to 2001-2003. *JAMA, 293*(20), 2487–2495.
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*(6), 726–738.
Kreuze, E., & Lamis, D. A. (2018). A review of psychometrically tested instruments assessing suicide risk in adults. *Omega, 77*(1), 36–90.
Liu, R. T. (2017). Childhood adversities and depression in adulthood: Current findings and future directions. *Clinical Psychology: Science and Practice, 24*, 140–153.
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*(3), 367–383.
Liu, R.T., Walsh, R.F.L., Kautz, M., Stephenson, A., Pollak, O.H., & Clayton, M.G. (in press). Suicide and depression: Epidemiology, theory, assessment, and treatment. In Pettit, J., & Olino, T. (Eds.), *APA Handbook of Depression*. American Psychological Association.
Liu, R.T., Bettis, A.H., Lawrence, H.R., Walsh, R.F.L., Sheehan, A.E., Pollak, O.H., … Marlowe, R.M. (2024). Measures of suicidal thoughts and behaviors in children and adolescents: A systematic review and recommendations for use in clinical and research settings. *Assessment*.
Lyrakos, D. G. (2011). The development of the Greek Personality Assessment Inventory. *Psychology, 2*(8), 797–803.
Mabilia, D., Di Riso, D., Lis, A., & Bobbio, A. (2019). A prediction model for separation anxiety: The role of attachment styles and internalizing symptoms in young adults. *Journal of Adult Development, 26*, 286–294.
Madan, A., Frueh, B. C., Allen, J. G., Ellis, T. E., Rufino, K. A., Oldham, J. M., & Fowler, J. C. (2016). Psychometric Reevaluation of the Columbia-Suicide Severity Rating Scale: Findings From a Prospective, Inpatient Cohort of Severely Mentally Ill Adults. *The Journal of Clinical Psychiatry, 77*(7), e867–e873.
Matarazzo, B. B., Brown, G. K., Stanley, B., Forster, J. E., Billera, M., Currier, G. W., … Brenner, L. A. (2019). Predictive Validity of the Columbia-Suicide Severity Rating Scale among a Cohort of At-risk Veterans. *Suicide and Life-Threatening Behavior, 49*(5), 1255–1265.
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*(6), 683–707.
McCall, W. V., Porter, B., Pate, A. R., Bolstad, C. J., Drapeau, C. W., Krystal, A. D., … Nadorff, M. R. (2021). Examining suicide assessment measures for research use: Using item response theory to optimize psychometric assessment for research on suicidal ideation in major depressive disorder. *Suicide and Life-Threatening Behavior, 51*(6), 1086–1094.
McGillivray, L., Rheinberger, D., Wang, J., Burnett, A., & Torok, M. (2022). Non-disclosing youth: A cross sectional study to understand why young people do not disclose suicidal thoughts to their mental health professional. *BMC Psychiatry, 22*(1), 3.
Moffitt, T. E., Caspi, A., Taylor, A., Kokaua, J., Milne, B. J., Polanczyk, G., & Poulton, R. (2010). How common are common mental disorders? Evidence that lifetime prevalence rates are doubled by prospective versus retrospective ascertainment. *Psychological Medicine, 40*(6), 899–909.
Molock, S. D., Boyd, R. C., Alvarez, K., Cha, C., Denton, E.-G., Glenn, C. R., … Miller, A. B. (2023). Culturally responsive assessment of suicidal thoughts and behaviors in youth of color. *American Psychologist, 78*(7), 842–855.
Morey, L. C., Lowmaster, S. E., Coldren, R. L., Kelly, M. P., Parish, R. V., & Russell, M. L. (2011). Personality Assessment Inventory profiles of deployed combat troops: an empirical investigation of normative performance. *Psychological Assessment, 23*(2), 456–462.
Nandy, K., Rush, A. J., Carmody, T. J., Mayes, T. L., & Trivedi, M. H. (2023). The 9-item Concise Health Risk Tracking - Self-Report (CHRT-SR$_9$) measure of suicidal risk: Performance in adult primary care patients. *Frontiers in Psychiatry, 14*, 1014766.
National Action Alliance for Suicide Prevention (NAASP). (2014). *A prioritized research agenda for suicide prevention: An action plan to save lives*. Rockville, MD: National Institute of Mental Health and the Research Prioritization Task Force.
Nock, M. K., Borges, G., Bromet, E. J., Alonso, J., Angermeyer, M., Beautrais, A., & Williams, D. (2008). Cross-national prevalence and risk factors for suicidal ideation, plans and attempts. *The British Journal of Psychiatry, 192*(2), 98–105.
Nock, M. K., Kessler, R. C., & Franklin, J. C. (2016). Risk factors for suicide ideation differ from those for the transition to suicide attempt: The importance of creativity, rigor, and urgency in suicide research. *Clinical Psychology: Science and Practice, 23*(1), 31–34.
Nock, M. K., Millner, A. J., Joiner, T. E., Gutierrez, P. M., Han, G., Hwang, I., … Kessler, R. C. (2018). Risk factors for the transition from suicide ideation to suicide

attempt: Results from the Army study to assess risk and resilience in Servicemembers (Army STARRS). *Journal of Abnormal Psychology, 127*(2), 139.
Nugent, W. R., & Cummings, S. (2014). A validity and measurement equivalence study of the ultra-short suicidal ideation scale with older adults. *Journal of the Society for Social Work and Research, 5*(4), 439–459.
Oakey-Frost, N., Moscardini, E. H., Cowan, T., Cohen, A., & Tucker, R. P. (2023). The temporal dynamics of wish to live, wish to die, and their short-term prospective relationships with suicidal desire. *Behavior Therapy, 54*(3), 584–594.
Oppenheimer, C. W., Glenn, C. R., & Miller, A. B. (2022). Future directions in suicide and self-injury revisited: Integrating a developmental psychopathology perspective. *Journal of Clinical Child and Adolescent Psychology, 51*(2), 242–260.
O'Rourke, N., Heisel, M. J., Canham, S. L., Sixsmith, A., Yaghoubi-Shahir, H., King, D. B., & BADAS Study team. (2018). Psychometric validation of the Geriatric Suicide Ideation Scale (GSIS) among older adults with bipolar disorder. *Aging and Mental Health, 22*(6), 794–801.
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*(4), 443–454.
Ostacher, M. J., Nierenberg, A. A., Rabideau, D., Reilly-Harrington, N. A., Sylvia, L. G., Gold, A. K., … Trivedi, M. H. (2015). A clinical measure of suicidal ideation, suicidal behavior, and associated symptoms in bipolar disorder: Psychometric properties of the Concise Health Risk Tracking Self-Report (CHRT-SR). *Journal of Psychiatric Research, 71*, 126–133.
Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, Article n71.
Patry, M. W., & Magaletta, P. R. (2015). Measuring suicidality using the personality assessment inventory: a convergent validity study with federal inmates. *Assessment, 22*(1), 36–45.
Pignolo, C., Di Nuovo, S., Fulcheri, M., Lis, A., Mazzeschi, C., & Zennaro, A. (2018). Psychometric properties of the Italian version of the Personality Assessment Inventory (PAI). *Psychological Assessment, 30*(9), 1226–1236.
Podlogar, M. C., & Joiner, T. E. (2020). Allowing for nondisclosure in high suicide risk groups. *Assessment, 27*(3), 547–559.
Podlogar, M. C., Rogers, M. L., Chiurliza, B., Hom, M. A., Tzoneva, M., & Joiner, T. (2016). Who are we missing? Nondisclosure in online suicide risk screening questionnaires. *Psychological Assessment, 28*(8), 963–974.
Posner, K., Brown, G. K., Stanley, B., Brent, D. A., Yershova, K. V., Oquendo, M. A., … Mann, J. J. (2011). The Columbia-Suicide Severity Rating Scale: initial validity and internal consistency findings from three multisite studies with adolescents and adults. *The American Journal of Psychiatry, 168*(12), 1266–1277.
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*(1–3), 348–355.
Revelle, W., & Condon, D. M. (2019). Reliability from α to ω: A tutorial. *Psychological Assessment, 31*(12), 1395–1411.
Reynolds, W. M. (1991). Psychometric characteristics of the Adult Suicidal Ideation Questionnaire in college students. *Journal of Personality Assessment, 56*(2), 289–307.
Ruchensky, J. R., Balsis, S., Edens, J. F., & Douglas, K. S. (2021). Suicidal ideation across race in a justice-involved sample: An item response theory approach. *Suicide and Life-Threatening Behavior, 51*(3), 385–393.
Ryan, E. P., & Oquendo, M. A. (2020). Suicide risk assessment and prevention: Challenges and opportunities. *Focus (American Psychiatric Publishing), 18*(2), 88–99.
Sanchez, K., Killian, M. O., Mayes, T. L., Greer, T. L., Trombello, J. M., Lindblad, R., … Trivedi, M. H. (2018). A psychometric evaluation of the Concise Health Risk Tracking Self-Report (CHRT-SR)- a measure of suicidality-in patients with stimulant use disorder. *Journal of Psychiatric Research, 102*, 65–71.
Silverman, M. M. (2006). The language of suicidology. *Suicide and Life-Threatening Behavior, 36*(5), 519–532.
Sims, J. A., Thomas, K. M., Hopwood, C. J., Chen, S. H., & Pascale, C. (2013). Psychometric properties and norms for the Personality Assessment Inventory in egg donors and gestational carriers. *Journal of Personality Assessment, 95*(5), 495–499.
Spears, A. P., Gratch, I., Nam, R. J., Goger, P., & Cha, C. B. (2023). Future directions in understanding and interpreting discrepant reports of suicidal thoughts and behaviors among youth. *Journal of Clinical Child and Adolescent Psychology, 52*(1), 134–146.
Stanley, I. H., Hom, M. A., Christensen, K., Keane, T. M., Marx, B. P., & Björgvinsson, T. (2021). Psychometric properties of the Depressive Symptom Index-Suicidality Subscale (DSI-SS) in an adult psychiatric sample. *Psychological Assessment, 33*(10), 987–997.
Stover, J. B., Solano, A. C., & Liporace, M. F. (2015). Personality assessment inventory: Psychometric analyses of its Argentinean version. *Psychological Reports, 117*(3), 799–823.
Tabares, J. V., Butner, J. E., Bryan, C. J., Harris, A., & J.. (2021). Mokken scale analysis of lifetime responses on the Columbia suicide severity rating Scale's severity of ideation subscale. *Assessment, 28*(6), 1624–1634.
Tasca, G. A., Wood, J., Demidenko, N., & Bissada, H. (2002). Using the PAI with an eating disordered population: scale characteristics, factor structure, and differences among diagnostic groups. *Journal of Personality Assessment, 79*(2), 337–356.
The Joint Commission. (2020). Suicide prevention resources to support Joint Commission Accredited organizations implementation of NPSG 15.01.01. www.jointcommission.org/-/media/tjc/documents/resources/patient-safety-topics/suicide-prevention/pages-from-suicide_prevention_compendium_5_11_20_updated-july2020_ep2.pdf.
Thom, R., Hogan, C., & Hazen, E. (2020). Suicide risk screening in the hospital setting: A review of brief validated tools. *Psychosomatics, 61*(1), 1–7.
Trivedi, M. H., Wisniewski, S. R., Morris, D. W., Fava, M., Gollan, J. K., Warden, D., … Rush, A. J. (2011). Concise Health Risk Tracking scale: a brief self-report and clinician rating of suicidal risk. *The Journal of Clinical Psychiatry, 72*(6), 757–764.
Trombello, J. M., Killian, M. O., Grannemann, B. D., Rush, A. J., Mayes, T. L., Parsey, R. V., … Trivedi, M. H. (2019). The Concise Health Risk Tracking-Self Report: Psychometrics within a placebo-controlled antidepressant trial among depressed outpatients. *Journal of Psychopharmacology, 33*(2), 185–193.
Trombello, J. M., Kulikova, A., Mayes, T. L., Nandy, K., Carmody, T., Bart, G., … Trivedi, M. H. (2023). Psychometrics of the Concise Health Risk Tracking Self-Report ($CHRT\text{-}SR_{16}$) Assessment of Suicidality in a Sample of Adults with Moderate to Severe Methamphetamine Use Disorder: Findings from the ADAPT-2 Randomized Trial. *Neuropsychiatric Disease and Treatment, 19*, 1443–1454.
Udala, M., Ohlhauser, L., Campbell, M., Langlois, A., Leitner, D., Libben, M., & Miller, H. (2022). A psychometric examination of the PAI-SF in persons with recent stroke. *The Clinical Neuropsychologist, 36*(6), 1471–1492.
Van Orden, Witte, T. K., Cukrowicz, K. C., Braithwaite, S. R., Selby, E. A., & Joiner, T. E., Jr. (2010). The interpersonal theory of suicide. *Psychological Review, 117*(2), 575–600.
Villegas, A. C., DuBois, C. M., Celano, C. M., Beale, E. E., Mastromauro, C. A., Stewart, J. G., … Hoeppner, B. B. (2018). A longitudinal investigation of the Concise Health Risk Tracking Self-Report (CHRT-SR) in suicidal patients during and after hospitalization. *Psychiatry Research, 262*, 558–565.
Wang, S. B., Coppersmith, D. D. L., Kleiman, E. M., Bentley, K. H., Millner, A. J., Fortgang, R., … Nock, M. K. (2021). A pilot study using frequent inpatient assessments of suicidal thinking to predict short-term Postdischarge suicidal behavior. *JAMA Network Open, 4*(3), Article e210591.
Ward, T., Arnold, K., Cunningham, M. C., & Liljequist, L. (2018). Three validation studies of the personality assessment inventory short form. *Journal of Clinical Psychology, 74*(12), 2264–2275.
Weissman, R. S., Scott, B. G., Schure, M. B., McCrory, B., Greist, J., Kuntz, M., & Wilcox, H. C. (2022). Psychometric properties of the 7-item version of the Concise Health Risk Tracking Self-Report Scale for rural adults enrolled in an Internet-delivered self-help program for depression. *The Journal of Rural Health, 38*(3), 574–582.
Youngstrom, E. A., Halverson, T. F., Youngstrom, J. K., Lindhiem, O., & Findling, R. L. (2018). Evidence-based assessment from simple clinical judgments to statistical learning: Evaluating a range of options using pediatric bipolar disorder as a diagnostic challenge. *Clinical Psychological Science, 6*(2), 243–265.
Youngstrom, E. A., Hameed, A., Mitchell, M. A., Van Meter, Freeman, A. J., … Meyer, R. E. (2015). Direct comparison of the psychometric properties of multiple interview and patient-rated assessments of suicidal ideation and behavior in an adult psychiatric inpatient sample. *The Journal of Clinical Psychiatry, 76*(12), 1676–1682.
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*(6), 811–824.
Zhang, J., & Brown, G. K. (2007). Psychometric properties of the scale for suicide ideation in China. *Archives of Suicide Research, 11*(2), 203–210.