Assessment tools of non-suicidal self-injury in adolescents 1990-2016: a systematic review

Abstract The purpose of this systematic review was to identify the instruments created or adapted to assess non-suicidal self-injury (NSSI) among adolescents. The Preferred Reporting Items for Systematic Review and Meta-Analyses (PRISMA) methodology was used. Two individual reviewers analyzed the psychometric properties of instruments published in English or Spanish from 1990 to 2016 considering standardized quality criteria. The PsycINFO, PubMed, ISI Web of Knowledge, Scopus, SciELO, ScienceDirect, and EBSCO databases were consulted. Eighteen studies that created or adapted 11 instruments were selected. Most were developed in the United States or Canada, and none were developed in Latin America. Several studies presented no evidence of the psychometric properties of their instruments. Seven of the 18 studies obtained at least one positive score. The Alexian Brothers Urge to Self-Injure Scale (ABUSI) and the Impulse, Self-harm, and Suicide Ideation Questionnaire for Adolescents (ISSIQ-A) obtained the highest positive scores. The limitation of this study is that only seven databases were employed for the literature search in English and Spanish. The reporting of the psychometric properties of NSSI instruments among adolescents should be improved, and adaptations to Latin American countries should be developed for international comparisons.

Key words Self-destructive behavior, Adolescent, Surveys and questionnaires, Cross-cultural comparison, Review
Non-suicidal self-injury (NSSI) was included in the Diagnostic and Statistical Manual of Mental Disorders, fifth edition (DSM-5), as a condition that requires greater research. NSSI is defined as the deliberate self-inflicted destruction of the body or skin tissue without the intention of dying. NSSI behaviors include self-cutting, burning, skin rubbing, hitting, biting, and others.

Internationally, the prevalence of NSSI varies based on the studied population, age group, and criteria used to operationally define this construct. The rates of NSSI among adults during the last 6 months were estimated at 4% for the general population and 21% for the clinical population, whereas those for adolescents over the last 12 months increase to 16.1-18.0% for the general population and 60% for the clinical population. The onset age of NSSI is between 12 and 15 years. The risk of this behavior is higher in adolescents in both the general and clinical populations, which suggests the need for assessment instruments that are validated and appropriate for this age group.

The comprehensive and precise study of NSSI allows, in clinical practice, to support decision making regarding diagnosis and treatment plans. Furthermore, in research field it supports theoretical development and favors agreement among researchers for the denomination and operationalization of the construct. Because of the incidence of NSSI behaviors and health risks associated among adolescents and adults, the number of studies and the development of various evaluation tools (including interviews and self-report tools) have increased since the 1990s. To ensure that the data provided by these reports are useful for the diagnosis and treatment of NSSI, evaluating the psychometric quality of these instruments is paramount to generate evidence about their validity (content, criteria, and construct) and reliability (internal, consistency, and reproducibility) as well as to ensure that evidence is reported according to standardized criteria. Likewise, the use of exploratory or confirmatory factor analyses is required to understand the nature and quantity of the underlying factors of the variables included in an instrument and to provide cross-cultural validity when using instruments from different languages or countries. Cross-cultural validity is understood as how the performance of the items in a translated or culturally adapted instrument reflects the performance of the items in the original version of the instrument.

Systematic reviews attempt to collect all of the empirical evidence that fits previously specified eligibility criteria to answer a specific research question. Although systematic reviews and studies that analyze the psychometric properties of the instruments that evaluate NSSI have been published, they were not specifically performed with regard to adolescents. Data regarding this age group is necessary given that vulnerability to NSSI behaviors is greatest during puberty and adolescence. Therefore, the objective of this research was to execute a systematic review to identify the assessment instruments to assess NSSI among adolescents, created or adapted from 1990 to 2016 as well as to analyze their psychometric properties.

Method

The systematic review protocol applied was approved and registered by the Research and Postgraduate Coordination of the university in charge of this study.

Search strategies

Using the methodology guidelines of Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA; http://www.prisma-statement.org/), an author of the present study conducted (from October 28 to November 28, 2016) a comprehensive search of studies published in English and Spanish between January 1990 and October 2016 and indexed by the PsycINFO, PubMed, ISI Web of Knowledge, Scopus, SciELO, ScienceDirect, and EBSCO databases.

The search terms were divided into three groups using the Boolean operators “AND” and “OR”. The terms in English were (“self-injury OR self-injurious behavior OR self-mutilation OR non-suicidal self-injury * OR NSSI OR self-poisoning * OR self-harm * OR deliberate self-harm * OR parasuicide OR deliberate self-injury * OR deliberate self-determination OR self-aggression * OR autoaggression * OR autoaggressive “*”) AND (“ instrument * OR measure * OR questionnaire * OR scale * OR assessments * OR inventory * OR psychometric * OR validity OR reliability OR psychometric properties OR factor analysis) AND (“Adolescence OR youth OR teenager OR teens OR adolescents”). The terms in Spanish were (“autolesi* OR comportamientos autolesivos OR autoagresión OR lesiones autoprovocadas OR automutilación OR autodaño...
Selection of studies

Based on the outcome of the main, directed and manual searches (which were performed independently), the research team agreed select studies based on the following inclusion and exclusion criteria. Studies that developed, adapted, or evaluated the psychometric properties of instruments or domains of NSSI with regard to adolescents (10 to 19 years of age) were included. Studies that used the term “self-injury” to evaluate suicidal behavior or those in which the evaluation of psychometric properties could not be performed because of insufficient data or the authors’ lack of response were excluded. To reduce bias with regard to the search and study selection, the definition of self-injury was reviewed in each article, and its theoretical foundation was identified to ensure that NSSI was evaluated.

Data extraction

Two of the authors of this study separately used an electronic form previously designed and piloted by the research team to collect the data. Data from each study concerning the characteristics of the instrument (i.e., name, author, country, year of publication, dimensions, and number of items), the study characteristics (i.e., population type, sample size, average age, and sex distribution), and reported evidence of the psychometric properties of each instrument (considering the criteria of Terwee et al.12,13) were extracted.

Results

The main search identified 282 studies, and the manual and directed searches identified 29, for a total of 311. A total of 32 duplicate studies were eliminated. Based on the inclusion and exclusion criteria, 207 studies were eliminated, and 72 were selected for a full-text review. Based on the full review, 54 studies were excluded. The main grounds for rejection were the lack of psychometric characteristics for any instrument (71%), the evaluated instruments were created based on a population of adults or young adults (23%), or that the evaluation method included fewer than three items (5.7%). Finally, 18 studies were included in the systematic review to evaluate the psychometric properties (Figure 1).

Table 1 shows the characteristics of the studies selected for evaluation. Of these 18 studies, 11 were instruments for evaluate NSSI in adolescents: the Self-Injurious Thoughts and Behaviors Interview (SITBI)22 with one adaptation23; the Self-Harm Behavior Questionnaire (SHBQ)24; the Functional Assessment of Self-Mutila-
tion (FASM)\(^6\) with three adaptations\(^{25-27}\); the Self-Injury Motivation Scale, Adolescent version (SIMS-A)\(^{28}\); the Alexian Brothers Assessment of Self-Injury (ABASI)\(^{29}\); the Alexian Brothers Urge to Self-Injure Scale (ABUSI)\(^{30}\), four adaptations of the instrument Ottawa Self-Injury Inventory (OSI)\(^{31-34}\); the Risk-Taking and Self-Harm Inventory for Adolescents (RTSHIA)\(^{35}\); the Repetitive Non-Suicidal Self-Injury Questionnaire (R-NSI-Q)\(^{36}\); the 6-item measure of Prinstein\(^{37}\); and the Impulse, Self-harm, and Suicide Ideation Questionnaire for Adolescents (ISSIQ-A)\(^{38}\). With its publication in 2007, the SITBI\(^{22}\) became the first instrument intended to measure NSSI among adolescents. The SITBI is an interview-format instrument with 169 items that was developed in the United States using a general population of adolescents and young adults. Of the 18 included studies, eleven (61.1%) were performed in the United States, and four (22.2%) were performed in Canada. Eight (44.4%) were adapted cross-culturally. No instruments created in or adapted for Latin America were found. Of the total studies, six (33.3%) were performed with clinical populations, ten (55.5%) were performed with students, one (5.6%) was performed with a general population, and one (5.6%) was performed with both patients and students. Regarding the use of a factor analysis, three studies (16.6%) conducted an exploratory factor analysis, eight (44.4%) conducted a confirmatory factor analysis, and seven (38.8%) did not report such an analysis.

Regarding the quality and risk of bias of the selected studies, all listed clear objectives. Concerning their methods, all showed appropriate designs and clearly defined samples with a sampling frame representing the researched population as well as risk factors and outcome variables appropriate for their objectives. Only 33% justified their sample sizes, 28% ensured a random selection, 39% took action to address non-respondents, 78% evaluated risk factors and outcome variables using previously tested or published instruments, 78% specified what was applied to determine the significance or precision estimates, and 94% described the methodology to replicate them. Regarding the results, all studies described the basic data and delivered the results of the analyses described in the methods. Although 61% provided information about non-respondents, 44% provided insufficient information about a non-response bias. Approximately 83% of the studies had outcomes that were internally consistent. Limitations were indicated in all of the studies’ discussions, and 94% of the results and authors’ discussions were justified. No study reported conflicts of interest, and 94% secured ethical approval or the consent of the participants.

**Results of the evaluation of the psychometric properties**

Table 2 shows the studies that created instruments that evaluate NSSI among adolescents and their psychometric properties. Of the seven instruments, only the ISSIQ-A\(^{18}\) obtained positive content validity scores, whereas five instruments (71.4%) obtained poor ratings. When evaluating internal consistency, two of the seven instruments (28.5%), the ABUSI\(^{10}\) and ISSIQ-A\(^{18}\), obtained positive scores, and two instruments presented no data (28.5%). Five instruments (71.4%) provided data on criteria validity, but none obtained positive ratings. Regarding construct validity, five instruments (71.4%) obtained a positive score, and two instruments (28.5%) had no available data. Concerning reliability, two instruments (28.5%), the ABUSI\(^{10}\) and SITBI\(^{22}\), obtained positive scores, whereas the ABASI\(^{29}\) (14.2%) obtained poor scores, and the remaining...
Instruments presented no data. Only two instruments (28.5%), the R-NSSI-Q and RTSHIA, presented evidence of responsiveness and obtained a positive score. None of the instruments provided data concerning agreement or floor or ceiling effects. Five instruments (71.4%) provided data on interpretability, but none obtained a positive score.

Of the 11 studies that described an instrument adaptation (Table 3), only the SHBQ obtained a positive score for internal consistency. Six instrument adaptations (54.5%) obtained poor scores, and four adaptations (36.3%) presented no data. With regard to criteria validity, only the C-FASM provided evidence and obtained an intermediate score. Regarding construct validity, only two instruments (18.1%), the C-FASM and SHBQ, obtained positive scores, and the SITBI-G (9.0%) scored poorly. Concerning reliability, only the SITBI-G presented evidence and obtained a poor score. None of the instruments adaptations provided evidence of

Table 1. Description of the NSSI instruments and sample characteristics.

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Author(s), year</th>
<th>Country of creation and/or adaptation</th>
<th>n</th>
<th>Type of participant</th>
<th>Age, M (SD)</th>
<th>No. Items</th>
<th>Dimensions</th>
<th>Factor Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABASI</td>
<td>Washburn et al., 2015</td>
<td>USA</td>
<td>511</td>
<td>P</td>
<td>17.3 (6.2)</td>
<td>28</td>
<td>Criteria A, B, C and D of the DSM-5</td>
<td>EFA</td>
</tr>
<tr>
<td>ABUSI</td>
<td>Washburn et al., 2010</td>
<td>USA</td>
<td>36</td>
<td>P</td>
<td>18.7 (7.5)</td>
<td>5</td>
<td>Unidimensional</td>
<td>EFA</td>
</tr>
<tr>
<td>FASM</td>
<td>Nock &amp; Prinstein, 2004</td>
<td>USA</td>
<td>108</td>
<td>P</td>
<td>14.8 (1.4)</td>
<td>22</td>
<td>Methods, frequency and self-injury function</td>
<td>CFA</td>
</tr>
<tr>
<td></td>
<td>Leong et al., 2014</td>
<td>USA, China</td>
<td>825</td>
<td>St</td>
<td>11.4 (1.3)</td>
<td>CFA</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Calvete et al., 2015</td>
<td>USA, China</td>
<td>1864</td>
<td>St</td>
<td>15.3 (2.0)</td>
<td>CFA</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dahlstrom et al., 2015</td>
<td>USA, Sweden</td>
<td>3097</td>
<td>St</td>
<td>15-17*</td>
<td>CFA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ISSIQ-A</td>
<td>Carvalho et al., 2015</td>
<td>Portugal</td>
<td>1722</td>
<td>St</td>
<td>16.7 (1.3)</td>
<td>56</td>
<td>Impulsivity, self-injury, risk behaviors, suicidal ideation</td>
<td>CFA</td>
</tr>
<tr>
<td>R-NSSI-Q</td>
<td>Manca et al., 2014</td>
<td>Italy</td>
<td>383</td>
<td>St</td>
<td>23.3 (4.0)</td>
<td>38</td>
<td>Frequency and non-suicidal self-injury methods</td>
<td>NR</td>
</tr>
<tr>
<td>RTSHIA</td>
<td>Vrouva et al., 2010</td>
<td>Greece</td>
<td>651</td>
<td>P</td>
<td>15.3 (2.1)</td>
<td>27</td>
<td>Self-injury and risk taking</td>
<td>CFA</td>
</tr>
<tr>
<td>SHBQ</td>
<td>Muehlenkamp et al., 2010</td>
<td>USA, USA</td>
<td>1386</td>
<td>St</td>
<td>15.4 (1.4)</td>
<td>22</td>
<td>NSSI behaviors, suicidal ideation, suicide threat and past suicide attempts</td>
<td>CFA</td>
</tr>
<tr>
<td>SIMS-A</td>
<td>Swannell et al., 2008</td>
<td>USA-Australia</td>
<td>38</td>
<td>P</td>
<td>M 15.7 (1.0)</td>
<td>22</td>
<td>Emotions Regulation, communicating/influencing others, punishment/excitement and psychosis/emptiness of insight</td>
<td>EFA</td>
</tr>
</tbody>
</table>

it continues
agreement, responsiveness or floor or ceiling effects. Regarding interpretability, five instrument adaptations (45.5%) presented evidence and obtained intermediate scores.

Only the FASM adaptation in China obtained a positive evaluation regarding cross-cultural validity, and the remaining adaptations (87.5%) were evaluated as poor. These poor evaluations were for not describing or poorly describing the translator’s expertise regarding the tested construct; for not clarifying whether there was more than one translator, whether they worked independently, or how they resolved their differences; or for not clarifying whether the translation was reviewed by a committee involving the researchers or the creators of the original instrument. Some studies did not pilot the translated instrument or perform a confirmatory factor analysis.

Discussion

This systematic review identified the NSSI assessment instruments created or adapted for adolescents between 1990 and 2016 as well as analyzed the outcomes of their psychometric properties. A total of 18 studies were identified; seven created instruments, and 11 adaptations. Seven of the 18 studies obtained at least one positive score. The ABUSI and the ISSIQ-A obtained the greatest number of positive scores, fulfilling three of the nine evaluation criteria. However, no evidence regarding the psychometric properties of the various instruments was found using Terwee et al.’s evaluation criteria, and no data were presented regarding agreement or floor or ceiling effects.

Because NSSI is a widespread phenomenon reported across different countries, it is associa-
Table 2. Analysis of the psychometric properties of studies reporting the creation of instruments to assess NSSI among adolescents considering the criteria of Terwee et al.12.

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Content Validity</th>
<th>Internal Consistency</th>
<th>Criteria Validity</th>
<th>Construct Validity</th>
<th>Agreement</th>
<th>Reliability</th>
<th>Responsiveness</th>
<th>Floor-Ceiling Effect</th>
<th>Interpretability</th>
<th>Positive Scores Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABASI</td>
<td>-</td>
<td>-</td>
<td>?</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>?</td>
<td>0</td>
</tr>
<tr>
<td>ABUSI</td>
<td>-</td>
<td>+</td>
<td>?</td>
<td>+</td>
<td>0</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>?</td>
<td>3</td>
</tr>
<tr>
<td>ISSIQ-A</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>?</td>
<td>3</td>
</tr>
<tr>
<td>R-NSSI-Q</td>
<td>?</td>
<td>0</td>
<td>?</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>0</td>
<td>?</td>
<td>2</td>
</tr>
<tr>
<td>RTSHIA</td>
<td>-</td>
<td>-</td>
<td>?</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>0</td>
<td>?</td>
<td>2</td>
</tr>
<tr>
<td>SITBI</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>0</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6-item measure</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Note: ABASI, Alexian Brothers Assessment of Self-Injury; ABUSI, Alexian Brothers Urge to Self-Injure Scale; ISSIQ-A, Impulse, Self-harm and Suicide Ideation Questionnaire for Adolescents; R-NSSI-Q, Repetitive Non-Suicidal Self-Injury Questionnaire; RTSHIA, Risk-taking and self-harm inventory for adolescents; SITBI, Injurious Thoughts and Behavior Interview. Criteria: + positive; ? intermediate; - poor; 0 unavailable.

Table 3. Analysis of the psychometric properties of the adaptations of the instruments evaluating NSSI among adolescents considering the criteria of Terwee et al.12.

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Content Validity</th>
<th>Internal Consistency</th>
<th>Criteria Validity</th>
<th>Construct Validity</th>
<th>Agreement</th>
<th>Reliability</th>
<th>Responsiveness</th>
<th>Floor-Ceiling Effect</th>
<th>Interpretability</th>
<th>Positive Scores Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>FASM</td>
<td>EU</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C-FASM</td>
<td>Ch</td>
<td>-</td>
<td>?</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>?</td>
<td>1</td>
</tr>
<tr>
<td>FASM-E</td>
<td>E</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>?</td>
<td>0</td>
</tr>
<tr>
<td>FASM</td>
<td>S</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SHBQ</td>
<td>EU</td>
<td>+</td>
<td>0</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>?</td>
<td>2</td>
</tr>
<tr>
<td>SIMS-A</td>
<td>Au</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SITBI-G</td>
<td>A</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>OSI</td>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>?</td>
<td>0</td>
</tr>
<tr>
<td>OSI</td>
<td>Hu</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>OSI</td>
<td>Is</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>?</td>
<td>0</td>
</tr>
<tr>
<td>OSI</td>
<td>C</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Note: FASM, Functional Assessment of Self-Mutilation; SHBQ, Self-Harm Behavior Questionnaire; SIMS-A, Self-Injury Motivation Scale Adolescent; SITBI, Injurious Thoughts and Behaviors Interview; OSI, Ottawa Self-Injury Inventory. G, Germany; Au, Australia; C, Canada; Ch, China; S, Spain; USA, United States of America; Hu, Hungary; Is, Israel; Sw, Sweden. Criteria: + positive; ? intermediate; - poor; 0 unavailable.
tected with negative well-being\textsuperscript{39,40}, higher probability of suicide\textsuperscript{41,42}, and a higher prevalence of NSSI has been reported among adolescents compared with adult populations, and that the present review identified 11 instruments, more than half of which were created over the last 7 years (as well as seven of eight adaptations), suggest that this field of study is still incipient. No instruments created for or adapted to Latin America were identified, despite evidence that NSSI is a common phenomenon in these populations\textsuperscript{43-47}. Eleven instruments were created or adapted in the United States; only the FASM\textsuperscript{4-25-27}, the SITBI\textsuperscript{22,23}, and the OSI\textsuperscript{31-34} were adapted for application among the adolescents of several countries to allow for international comparisons\textsuperscript{48}. Therefore, it is necessary to make progress toward culturally adapting the instruments with the best psychometric properties for several regions of the world.

Regarding the target population of the analyzed instruments, seven were originally designed for adolescents, and four were designed for adults. It is necessary that the instruments applied to adolescents be designed specifically for this age group; if they were created for adults, then a semantic adaptation should be made because adolescents might interpret and answer questions from an instrument in a different way because of developmental differences compared with adults\textsuperscript{49}.

Of the evaluated instruments, those with the greatest evidence of validity and reliability were the ABUSI\textsuperscript{10} and ISSIQ-A\textsuperscript{38}. These instruments evaluated the different dimensions of the NSSI. The ABUSI\textsuperscript{10} evaluates criteria A, B, C, and D of the DSM-5, whereas the ISSIQ-A\textsuperscript{38} is useful for determining impulsivity, NSSI, risky behaviors, and the intrapersonal or interpersonal functions of the NSSI. More dimensions are addressed in other instruments that might be useful in clinical practice given the relevant information they provide; however, it is necessary to provide greater evidence of validity and reliability. Seven of the 18 analyzed studies did not report the use of factor analysis methods. These methods provide evidence of the exploration or confirmation of the dimensions evaluated by an instrument.

Eleven of the 18 analyzed instruments were created for or adapted to general or school populations; therefore, it is important to evaluate the psychometric properties of the instruments across different types of vulnerable populations (e.g., clinical populations), which would enable researchers to establish cut-off points and broaden the analysis of psychometric characteristics such as responsiveness and specificity; however, one possible difficulty would be accessing this population, partially because of the low prevalence of NSSI or the lack of seeking help among those who show these behaviors.

Assessing the psychometric properties of the instruments is important because of its relevance regarding the evaluation and treatment of NSSI. Regarding the internal consistency of the criteria, the ABUSI\textsuperscript{10}, the ISSIQ-A\textsuperscript{38}, and the adaptation of the SHBQ\textsuperscript{24} obtained positive scores. This criterion was met by only three instruments. From our perspective, this finding suggests that difficulties (e.g., those regarding content validity) were present during previous phases of the construction of the instruments or their adaptations. Thus, instruments with internal consistency issues also scored low on content validity.

The application of Cronbach’s alpha to evaluate the internal consistency of the instruments was criticized because this statistic improves as the number of items in the instrument increases\textsuperscript{49,50}. The evidence for content validity was weak because only one of the seven created instruments, the ISSIQ-A\textsuperscript{38}, obtained positive scores. The participation of experts and the target population in focus groups is critical for the items’ final outcomes (and the domains to which they contribute) to be included in the instrument. Although most of the analyzed studies did not report satisfactory evidence with regard to the two evaluated criteria, they contribute to an understanding of how the scales behave.

No instrument obtained positive scores regarding criteria validity because there are few gold standards in psychology. Therefore, the need to create standardized evaluation criteria and updates of the psychometric properties of the instruments is evident because the criteria of Terwee et al.\textsuperscript{12} were elaborated from a medical context. Regarding construct validity, obtained positive scores for the instruments created ISSIQ-A\textsuperscript{38}, ABUSI\textsuperscript{10}, R-NSSI-Q\textsuperscript{36}, RTSHIA\textsuperscript{35}, and SITBI\textsuperscript{22}, and those adapted the C-FASM\textsuperscript{37} and SHBQ\textsuperscript{24}. Because this type of validity helps prove that a specific construct is being evaluated, it is an essential feature among instruments designed to evaluate the set of indicators that comprise or relate to a specific diagnostic criterion. Based on all of the analyzed studies, it is noteworthy that this value was not reported by more than half. This fact might be related to the shortage of validated instruments that evaluate NSSI among adolescents, through which a convergent validity analysis could be performed.
Only the created instruments the ABUSI\textsuperscript{30} and SITBI\textsuperscript{22} obtained positive reliability scores. Terwee et al.\textsuperscript{12} distinguished between the relative and absolute measurement of reliability, which he called reliability and agreement, respectively. No study reported having calculated this latter condition, which shows that an analysis like the minimum important change (MIC), which describes the certainty that the magnitude of a change is not the result of chance, is not widely used among researchers in this field even though its use has increased, most likely because it requires more dissemination and training for this community of researchers to become familiar with it, compared with traditional analysis Kappa. Although reliability is much less complex to calculate, presents difficulties because of the need for multiple measurements of the same participants, so it was only evaluated by four instruments: the ABASIS\textsuperscript{29}, ABUSI\textsuperscript{30}, and SITBI\textsuperscript{22} as well as the adaptation SITBI-G\textsuperscript{23}. Therefore, we also think that the journals where these studies are published should play a primary role in requiring that certain important analyses are executed by the researchers to guarantee the quality of the creation or adaptation of an instrument.

Positive scores in responsiveness were only reported for the created instruments R-NSS-Q\textsuperscript{36} and RTSHIA\textsuperscript{34}. Previous systematic reviews such as that concerning self-injury instruments for adults\textsuperscript{19} have argued that the importance of reporting responsiveness has received little attention. It is necessary to insist on the importance of evaluating these properties for the creation or adaptation of instruments because, rather than being an exclusive failure of the NSSI instruments aimed at adolescents, it is a general weakness of the studies and reviews of the creation or adaptation of instruments.

Our results showed that of the seven studies that culturally adapted an instrument, only one adaptation, the C-FASM\textsuperscript{25}, obtained a good evaluation, and six adaptations obtained poor evaluations. This result occurred because the process of translating the instrument and verifying its adaptation through a review by a committee, pilot study, or confirmatory factor analysis was not described or only briefly described. Thus, the instruments evaluated in this review were not culturally adapted based on established standards. In this regard, Leong et al.\textsuperscript{25} mentioned that the translation of psychological instruments without empirical validation is common because the psychometric properties are not exhaustively verified or the procedure is not described in detail. It cannot be assumed that a translated version of an instrument has the same psychometric properties as the original version; therefore, the validity of a translated instrument should be evaluated before applying it to another country or culture.

The lack of high scores regarding the evaluation of an instrument does not necessarily denote poor psychometric properties; rather, it represents an absence of evidence because no available data were found for any of the nine evaluated criteria for any of the instruments. One relevant aspect is that several instruments showed evidence of their psychometric characteristics, although they differed from Terwee et al.’s\textsuperscript{12} criteria. Therefore, the evidence was considered as insufficient. In the case of the RTSHIA\textsuperscript{34}, reliability was evaluated using Pearson’s $r$ coefficient, which is not an incorrect procedure\textsuperscript{51}. In the case of the SITBI-G\textsuperscript{23}, inter-rater reliability was evaluated by agreement; although this method delivers excellent Kappa coefficients, it is not possible to evaluate it using standardized criteria. In addition, the considered evaluating criteria were designed for self-report instruments; therefore, because the SITBI\textsuperscript{22,23} is an interview, it was not possible to evaluate its internal consistency or perform a factor analysis.

Regarding the current review’s limitations, this search was conducted using only seven databases and two languages; therefore, subsequent studies might seek to broaden this search. It was not possible to collect data from one study, even after the corresponding author was contacted. Future studies should address important practical aspects when selecting an evaluation instrument, such as the time to respond according to the number of items, the required resources for qualifications, or the training needed for its application\textsuperscript{52}.

One challenge for Ibero-American researchers is to quickly become familiar with the process toward measuring the properties of instruments and their evaluation techniques to generate knowledge. Likewise, it is important that the knowledge and techniques required to evaluate the properties of instruments in developing countries (e.g., most of Latin America), especially those that are least likely to be reported in scientific articles, be promoted in research and training centers. If the researchers know and value the important role of instrument adaptations, they will achieve intercultural comparability and a greater understanding of the studied phenomenon, in this case the NSSI, even if they
are developed and designed in other countries out of necessity.

The use of different nomenclatures to refer to the NSSI, the different ways of managing it, and the lack of adequately standardized and validated assessment instruments might trigger potential biases in the estimates of NSSI, which limits our knowledge of its incidence and prevalence. This review found that little or insufficient evidence regarding the psychometric properties of the instruments evaluating NSSI among adolescents is available, which justifies greater research because it is during this developmental stage that higher prevalence of NSSI is observed, in both general and clinical populations. This outcome is consistent with that regarding the NSSI instruments for adults, where a lack of developed instruments and insufficient evidence of psychometric properties is also observed. This lack of data hinders the evaluation of NSSI using standardized methods.
Collaborations

YV Chávez-Flores worked on the design, analysis, interpretation, drafting, review and approval of the final version of the article. CA Hidalgo-Rasmussen worked on the conception, design, analysis, interpretation, drafting, review and approval of the final version of the article. LY Yáñez-Penúñuri worked on the analysis of the studies, interpretation, review and approval of the final version of the article.

Acknowledgments

The author expresses gratitude to Dr. Karina Franco Paredes for her orientation and support with regard to the methodological revision of this study.

The main author received grant from the National Council of Science and Technology. The funding source played no role in conducting this study. No conflicts of interest are present.

References


