

APPENDIX

Adaptation of the ACSS (Rangel-Villafañe et al., 2023) for Mexican adolescents.

In the first instance, an adaptation procedure was carried out on the items that make up the scale with the aim of making them easily understandable for adolescents.

Evidence of content validity was obtained from the judgment of experts, who evaluated two aspects: a) the relevance of each of the items, in terms of its congruence with its theoretical meaning, and b) the coverage of the dimension evaluated, ensuring that it was made up of necessary and sufficient items.

Method

Participants

To gather expert opinions, 21 professionals with experience in adolescent psychology or education and the construction of psychometric instruments were recruited as judges, in three different and iterative rounds. Additionally, 18 adolescents aged 13 to 18 were recruited to participate in the study: Ten of them from public schools were involved in cognitive laboratories, while eight adolescents from the Hospital Psiquiátrico Infantil “Dr. Juan N. Navarro” were recruited to pilot the scale.

Measure

Acquired Capability for Suicide Scale for adult Mexican population (Rangel-Villafañe et al., 2023) comprises 14 items with five response options, grouped into two factors: Fearlessness about death and Pain tolerance. The scale showed excellent adjustment indices for the Mexican university population, with $RMSEA = .011$, $CFI = .99$, $TLI = .99$, $IFI = .99$, $NFI = .91$, $\chi^2 = 86.75$, $DF = 84$, and $p = .39$, and explained 58% of the variance.

Procedure

The phrasing of the items on the ACSS was evaluated based on criteria outlined by DeVellis (2017) and Furr (2018), considering language, length, grammatical simplicity, precision of terms, double negations, and directionality. Modifications were made to ensure comprehension by the adolescent population. The response options were adjusted from five to six points to eliminate the neutral response option, aiming to increase the accuracy, reliability, and discriminative capacity of the scale (Calleja et al., 2019; Grande & Abascal, 2017; Simms et al., 2019).

A first version of the scale was obtained by submitting it to three expert judges in three iterative rounds to evaluate the inter-judge agreement and provide their qualitative opinions on content validity.

An online cognitive laboratory was conducted with a group of adolescents to evaluate the understanding of the items. The strategy used the concurrent, group, and scrutiny method (Nolin & Chandler, 1996). The participants and their legal guardians provided verbal and written consent and assent, respectively, prior to participating. Based on the participants' feedback, the items were reformulated to produce a second version of the scale. This version included a commitment question to detect cases of random answers (Fonseca-Pedrero et al., 2009). Eight additional adolescents evaluated the design, instructions, and response options of the scale (Streiner et al., 2015), leading to the final print and electronic versions.

Data analytic strategy

A scale ranging from 0 to 14 points was utilized to assess each item (Furr, 2018; DeVellis, 2017). Those items that scored less than 14 points were identified for modification. Inter-judge reliability was calculated for each expert round, with a criterion of 80% satisfaction. The first round evaluated (a) the clarity of the wording, (b) the theoretical consistency with the defined dimension, and (c) the logical relationship between the items and the dimension. The second and third rounds focused solely on the understanding of the items by adolescents (Argimon-Pallás & Jiménez-Villa, 2013).

To establish content validity, the relevance of each item and its coverage of the underlying dimension were evaluated using a 5-point Likert scale. Aiken's V coefficient was calculated, with a desired value of at least .70 and Type I Error $p < .05$ (Charter, 2003). The confidence interval for Aiken's V was also computed, with a lower limit of at least .70 (Merino & Livia, 2009). Whenever an item scored less than 5 points, expert judges were asked to provide suggestions for modifying

the wording. These proposals were then discussed with the research team.

Results

The qualitative evaluation of the items revealed that eight out of the original 14 items required adjustments (57.14%). The results of inter-judge agreements for each of the expert rounds are presented in **Table S1**.

Table S1
Percentage of items that reached an 80% inter-judge agreement

		<i>Pain tolerance (Items in agreement/total)</i>	<i>Fearlessness about death (Items in agreement/total)</i>
Round 1	Clarity	57.14% (4/7)	42.85% (3/7)
	Theoretical maintenance	71.4% (5/7)	100% (7/7)
	Logic	100% (7/7)	100% (7/7)
Round 2	Understanding	61.53% (8/13)	100% (7/7)
Round 3	Understanding	100% (8/8)	67% (4/6)

Note: Percentages that did not reach an 80% agreement are highlighted in bold. The Fearlessness about death factor did not reach an 80% inter-judge agreement in the third round. This outcome occurred because two of the items evaluated still posed challenges in understanding for adolescents, as noted by the experts. A consecutive increase from round to round was not anticipated, as the evaluations focused solely on items that proved difficult for adolescents to comprehend.

In the first round, the clarity of the wording for both factors was insufficient (inter-judge agreement < 80%), and there were issues with maintaining the theoretical concept of Pain Tolerance. Consequently, they suggested the inclusion of new items to cover the domain of this dimension (DeVellis, 2017), resulting in 13 items for the second group of experts. However, according to the second group of experts, adolescents still had difficulties in comprehending the factor, and therefore, they requested the inclusion of new items to improve the clarity of the scale. For the third round, only newly created and problematic items from the previous rounds (eight for Pain Tolerance and six for Fearlessness about death) were evaluated, so a consecutive improvement in the inter-judge agreement compared to the previous round was not expected. The understanding was better for the first factor, but for the second factor, two items were still not understandable, so the criterion of 80% agreement was not reached. These items were further discussed, and a consensus reached by the research team, and then asked to adolescents for their opinions in the cognitive laboratories. The relevance of the items and coverage of the dimensions for the three rounds of judges to ensure content validity are shown in **Table S2**. The table indicates that the

Table S2
Aiken's V and Lower Limit of CI for item Relevance and Coverage of ACSS dimensions

		<i>Pain tolerance</i>	<i>Fearlessness about death</i>
Round 1	R- Aiken's V (Items/total)	85.7% (6/7)	100% (7/7)
	C- Aiken's V (LL CI)	.75 (.40)	1 (.67)
Round 2	R- Aiken's V (Items/total)	92.3% (12/13)	85.7% (6/7)
	C- Aiken's V (LL CI)	.94 (.83)	.90 (.77)
Round 3	R- Aiken's V (Item/total)	100% (8/8)	100% (6/6)
	C- Aiken's V (LL CI)	.8 (.83)	.8 (.83)

Note: R refers to the relevance of each item that makes up the dimension. The number of items that reached an inter-judge agreement of 80% is expressed in percentage. In parentheses, the number of items that obtained it (numerator) is expressed in comparison with the total of items evaluated (denominator). C corresponds to the coverage of the dimension by the items and in parentheses, the Lower Limit of the Confidence Interval (LL CI) is expressed for Aiken's V.

relevance of the items achieved the minimum agreement of 80% (Aiken's $V \geq .70$ for each item). For the coverage of each dimension, the value of Aiken's V also met the satisfactory criterion ($\geq .70$). Because the items evaluated in each round were only those that the judges considered to be poorly understood, no consecutive round-to-round increase was expected for item relevance and dimension coverage.

Although the Lower Limit of the Confidence Interval (LL CI) was below .70 for the first round, it was because only two judges evaluated it, and the calculation is sensitive to the number of participating judges. In the end, both Aiken's V and LL CI met the expected cut-off point, resulting in content validity with a 23-item version.

Regarding the cognitive laboratories, the adolescents provided feedback suggesting simplifying the wording of some items. However, not all suggestions were incorporated as they were related to other theoretical dimensions.

Table S3 presents the minimum and maximum values of the descriptive statistics for the items, along with the percentage of items that fall outside the acceptable range (considering 23 items as a total).

Table S3
Descriptive statistics of items

	$\bar{x}$	SD	Answer options	High freq. opt.	Floor/Ceiling	Skewness / Kurtosis	Inter-item corr.	Item-Total corr.	Extr. group discrim.
Rank	1.45 - 4.19	1.24 -1.85	0 - 5				-.15 - .75		
Proportion of items out of acceptable range	39.1%	21.7%	0%	8%	8% / 21.7%	8% / 60.8%		0%	4%

Note: $\bar{x}$ = Mean; SD = Standard Deviation; High freq. opt. = High frequent options; Inter-item corr. = Inter – item correlation; Item-total corr. = Item – total correlation; Extr. group discrim. = Extreme group discrimination.

REFERENCES

- Argimon-Pallás, J. M. y Jiménez-Villa, J. (2013). *Métodos de investigación clínica y epidemiológica* (4th ed.). Elsevier.
- Calleja, N., Reskala Sánchez, F. S., Rivera-Fong, L. & Buenrostro Mercado, D. (2019). Effect of the number of answer choices in the psychometric properties of four psychosocial scales. *Journal of Psychology and Behavioral Sciences of the UACJS (2007-1833)*, 10(2), 100-113.
- Charter, R. A. (2003). A Breakdown of Reliability Coefficients by Test Type and Reliability Method, and the Clinical Implications of Low Reliability. *The Journal of General Psychology*, 130(3), 290-304. <https://doi.org/10.1080/00221300309601160>
- DeVellis, R. F. (2017). *Scale Development, Theory and Applications*. Sage.
- Fonseca-Pedrero, E., Paíno-Piñero, M., Lemos-Giráldez, S., Villazón-García, Ú., & Muñiz, J. (2009). Validation of the Schizotypal Personality Questionnaire—Brief Form in adolescents. *Schizophrenia Research*, 111(1-3), 53-60. <https://doi.org/10.1016/j.schres.2009.03.006>
- Furr, R. M. (2018). *Psychometrics. An Introduction* (3rd ed.). Sage publications.
- Grande, I., & Abascal, E. (2017). *Fundamentos y técnicas de investigación comercial* (13th ed.). Esic.
- Merino, C., & Livia, J. (2009). Intervalos de confianza asimétricos para el índice la validez de contenido: Un programa Visual Basic para la V de Aiken. *Anales de Psicología*, 25(1), 169–171. <https://revistas.um.es/analesps/article/view/71631>
- Nolin, M. J., & Chandler, K. (1996). *Use of cognitive laboratories and recorded interviews in the National Household Education Survey*. National Center for Education Statistics.
- Rangel Villafañá, J. N., Silva, C., & Jurado Cárdenas, S. (2023). Spanish Validation of the Interpersonal Needs Questionnaire and the Acquired Capability for Suicide Scale Among Mexican University Students. *Revista Internacional de Investigación en Adicciones*, 9(1), 22-27. <https://doi.org/10.28931/riiad.2023.1.03>
- Simms, L. J., Zelazny, K., Williams, T. F., & Bernstein, L. (2019). Does the number of response options matter? Psychometric perspectives using personality questionnaire data. *Psychological Assessment*, 31(4), 557-566. <https://doi.org/10.1037/pas0000648>
- Streiner, D. L., Norman, G. R., & Cairney, J. (2015). *Health Measurement Scales: A practical guide to their development and use* (5th ed.). Oxford University Press. <https://doi.org/10.1093/med/9780199685219.001.0001>