

Global Cross-Cultural Validation of a Brief Measure for Identifying Potential Suicide Risk in 42 Countries

Abstract

Objectives: This study aimed to examine the psychometric properties of the P4 suicide screener in a multi-national sample. The primary goal was to evaluate the reliability and validity of the scale and investigate its convergent validity by analyzing its correlation with depression, anxiety, and substance use.

Study Design: The study design is a cross-sectional self-report study conducted across 42 countries.

Methods: A cross-sectional, self-report study was conducted in 42 countries, with a total of 82,243 participants included in the final dataset.

Results: The study provides an overview of suicide ideation rates across 42 countries and confirms the structural validity of the P4 screener. Findings indicated that sexual and gender minority individuals exhibited higher rates of suicidal ideation. The P4 screener showed adequate reliability, convergence and discriminant validity, and a cut-off score of 1 is recommended to identify individuals at risk of suicidal behavior.

Conclusions: The study supports the reliability and validity of the P4 suicide screener across 42 diverse countries, highlighting the importance of using a cross-cultural suicide risk assessment to standardize the identification of high-risk individuals and tailoring culturally sensitive suicide prevention strategies.

Keywords: Suicide; assessment; validation; cross-cultural.

Suicide refers to the act of intentionally causing one's own death, with the risk often indicated by suicidal ideation or intent, especially when accompanied by a well-elaborated suicidal plan (American Psychological Association) and represents a significant societal and healthcare problem. Broader terms encompassing various outcomes related to suicide include "suicide-related outcomes" and "suicidal thoughts and behaviors (STB/SIB)." These terms cover a spectrum of behaviors, including suicidal ideation (thoughts of self-harm or death), suicide attempts (engaging in actions with the intent to die), and completed suicide (resulting in the individual's death). Widely used in research and clinical settings, these terms help capture the diverse range of behaviors and thoughts associated with deliberate self-harm and suicide, playing a crucial role in understanding and addressing these complex issues.

According to the World Health Organization,¹ more than 700,000 people die due to suicide every year, placing suicide as the fourth leading cause of death among 15-29-year-olds. Multiple predisposing and precipitating factors have been identified as contributing to suicides among adults in the general population,^{2,3} with mental health disorders – mood, substance use, and psychotic disorders – being the most important predictors.⁴ Although suicide remains a significant public health concern, previous research has focused predominantly on WEIRD (Western, Educated, Industrialized, Rich and Democratic) populations, leaving a considerable knowledge gap regarding how suicidal ideation manifests in diverse populations.¹ This study addresses this gap through a large-scale, international, multi-lab, multi-language approach using cross-sectional survey methods to examine suicidal ideation and behavior using the P4 suicide screener across 42 countries.

The P4 screener validly and reliably evaluates past suicide attempts, the presence of a suicidal plan, the likelihood of completion, and the presence of protective factors (e.g.,

resilience, access to mental health services).^{5,6} Developed over multiple controlled trials,⁷ the P4 screener possesses adequate clinical validity and was selected for this study due to its culturally sensitive language, which prioritizes ethical considerations in regions with cultural sensitivity towards the topic of suicide.⁸ A recent review highlights the varied use of the P4 screener in diverse settings, such as psychiatry inpatients and veterans, as well as its application for safety monitoring in a clinical trial. The P4 has proven to be an efficient tool for categorizing individuals expressing thoughts of deliberate self-harm into ordinal risk categories.⁶ The language used in the P4 screener is less explicit regarding the topic of suicide, which was deemed appropriate for the various cultural contexts of the population participating in this study.

The Present Study

Validated cross-cultural suicide risk assessment scales are important for accurate and unbiased assessment of suicide risk. These scales standardize and ensure reliability in determining an individual's suicide risk, effectively identify high-risk individuals, and monitor changes in risk levels, leading to better outcomes for those at risk of suicide. The primary aim of this study was to examine the psychometric properties of the P4 suicidality screener in a multinational sample. Our primary goal was to evaluate the reliability and validity of the scale by utilizing appropriate statistical techniques. Secondly, we aimed to investigate the convergent validity of the P4 suicidality screener by analyzing its correlation with established measures of depression and anxiety levels, and substance use. Moreover, we aimed to investigate variations in rates of suicidal ideation across different countries, genders, and sexual orientations.

Method

Procedure

The International Sex Survey (ISS) is a cross-sectional, self-report study conducted in 42 countries. The study design, including the preregistered study protocol, can be found at https://osf.io/uyfra/?view_only=6e4f96b748be42d99363d58e32d511b8. The ISS was initiated in February 2021, following the acquisition of ethical permissions by collaborators. Although the study aimed to utilize extensive and varied samples, achieving a comprehensive representation of all populations within each country proved challenging. Nonetheless, collaborators from each country collected a community sample of adults, striving for a balanced gender ratio of approximately 50–50% for both men and women, and representation of diverse individuals with respect to sexuality and gender within the surveyed population. The English survey battery was translated into 25 other languages following a pre-established procedure for cross-cultural studies.⁹ Data collection took place between October 2021 and May 2022. Participants aged 18 or older (or the legal age of being adult in a given country) were recruited through advertisements and completed the anonymous survey on Qualtrics (a secure online platform), which took approximately 25 to 45 minutes. Rigorous attention checks were implemented to ensure the data's reliability, and participants failing these checks were excluded from the study (see the detailed data cleaning procedure at <https://doi.org/10.17605/OSF.IO/DK78R>). Incentives included a donation of 50 cents (USD) to non-profit, sexuality-related international organizations for each completed survey, with a maximum donation of \$1,000. Participants could choose their preferred organization from a list after completing the survey. The list of collaborating countries, detailed information on the translation and data collection procedures, and eligibility criteria can be found in the study protocol.¹⁰

Participants

After data cleaning (see detailed data cleaning procedure:

<https://doi.org/10.17605/OSF.IO/DK78R>), a total of 82,243 participants ($M_{age} = 32.39$ years, $SD = 12.52$) were included in the final dataset. Concerning participants' gender, 32,549 (39.6%) were men, 46,874 (57.0%) were women, and 2,783 (3.4%) were gender-diverse individuals. Most participants ($n = 56,125$; 68.2%) identified as heterosexual, while 31.5% ($n = 25,777$) of the participants belonged to sexual minorities. Most participants completed tertiary education (e.g., college or university) ($n = 60,896$; 74.0%), worked full-time ($n = 42,981$; 52.3%), and lived in a city or metropolis (i.e., in a city with a population greater than 100,000) ($n = 56,361$; 68.5%). More than half of the participants were in a romantic relationship ($n = 51,778$; 63.0%). Details on participants' sociodemographic characteristics are presented in Table 1, and the same information disaggregated by country can be found at https://osf.io/n3k2c/?view_only=838146f6027c4e6bb68371d9d14220b5.

Measures

Sociodemographic and Sexuality-related Questions

Several sociodemographic questions (e.g., gender) and sexuality-related questions were included in the ISS.¹¹ The translation of the measures in this study can be found at the following link in all available languages:

https://osf.io/jcz96/?view_only=9af0068dde81488db54638a01c8ae118.

Suicide risk was assessed using the P4 screener, which is a brief measure for assessing suicide risk by asking about four key factors: past suicide attempts, a plan for suicide, the probability of completing suicide, and preventive factors that may reduce the likelihood of deliberate self-harm.⁷ The P4 screener begins by asking a screening question: "Have you had thoughts of actually hurting yourself?" (yes/no) If participants answer yes, the screener proceeds

with four additional questions: (1) Have you ever attempted to harm yourself in the past? (yes = 1/no = 0); (2) Have you thought about how you might actually hurt yourself? (yes = 1/no = 0); (3) There's a big difference between having a thought and acting on a thought. How likely do you think it is that you will act on these thoughts about hurting yourself or ending your life sometime over the next month? (not at all likely = 0/somewhat likely = 1/very likely = 2); (4) Is there anything that would prevent or keep you from harming yourself? (yes = 0/no = 1). If respondents chose yes, they were asked to specify what would prevent them from deliberately harming themselves. A sum score was calculated for each participant such that it ranged between 0 and 5 (i.e., item 3 was not collapsed).

Mental Health was assessed using the two subscales (anxiety and depression) of the Brief Symptom Inventory (BSI),¹² a 12-item scale used to assess anxiety and depression. Cronbach's alpha for both subscales was excellent ($\alpha = .90$).

Substance use was assessed using the Alcohol, Smoking, and Substance Involvement Screening Test (ASSIST).^{13,14} Ten ASSIST items were used to assess the frequency and consequences of substance use in the past three months. Cronbach's alpha for the ASSIST scale was acceptable ($\alpha = .63$).

Statistical Analyses

Descriptive Analyses

As per the preregistered analysis plan (<https://doi.org/10.17605/OSF.IO/DK78R>), we computed descriptive statistics for all items of the P4 screener, which included means with standard deviation, minimum and maximum values, skewness and kurtosis values, and the ratio of "yes" answers. However, missing values were observed in country, language, gender, and sexual orientation variables (ranging from 0 to 0.24%) and were not missing completely at

random. This was supported by the results of Little's Missing Completely at Random Test (MCAR) ($\chi^2 = 1137.0$, $df = 28$, $p < .001$; 15 missing patterns).¹⁵ We excluded participants who had missing data on all items of the P4 screener, allowing partial missingness. To handle the remaining missing data, we used the full-information maximum likelihood (FIML) method based on previous recommendations.¹⁶

Test of Dimensionality

Confirmatory factor analyses (CFA) were conducted to examine the structural validity and dimensionality of the suicidal ideation scale separately for participants' language, country, gender, and sexual orientation. Based on Monte Carlo simulations (see details: <https://doi.org/10.17605/OSF.IO/DK78R>), a minimum of 535 participants were required to be included in analyses. In the first set of CFAs, participants' language was the grouping variable with 20 languages (see all languages in Table 1), as 20 out of the 25 languages had enough participants (i.e., $n > 535$) for these tests. In the second set of analyses, participants' country of residence was the grouping variable with 31 countries (see Table 1 for all countries), as 31 out of the 42 countries had sufficient participants for these analyses. In the third set of analyses, participants' gender identity was the grouping variable with three categories (i.e., men, women, gender-diverse individuals), as the number of participants in different gender minority groups did not allow us to use them as separate groups. In the fourth set of analyses, participants' sexual orientation was the grouping variable with eight sexual orientations (see all sexual orientations in Table 1). Information on the creation of gender-based and sexual-orientation-based groups can be found in the preregistration document (<https://doi.org/10.17605/OSF.IO/DK78R>).

The CFAs were evaluated using common goodness-of-fit indices:¹⁷⁻¹⁹ Comparative Fit Index (CFI; $\geq .90$ adequate; $\geq .95$ good), Tucker-Lewis Index (TLI; $\geq .90$ adequate; $\geq .95$ good),

Root-Mean-Square Error of Approximation with its 90% confidence interval (RMSEA; $.10 \leq$ acceptable, $\leq .08$ adequate, and $\leq .05$ good), and Standardized Root Mean Square Residual (SRMR; $\leq .08$ good).²⁰ The mean and variance adjusted unweighted least squares (ULSMV) was used, which is superior to the maximum likelihood estimation for binary items, particularly when the response categories follow asymmetric thresholds.²¹ CFAs were estimated in Mplus 8.8. In cases where the goodness of fit was inadequate, in-depth exploration was also performed with bootstrap Exploratory Graph Analysis,²² including item and dimensional stability. These analyses were performed in R using the *EGAnet* package.²³

Group differences¹

Differences in suicidal ideation by language, country, gender, and sexual orientation were examined using Kruskal-Wallis tests and eta-squared effect sizes (> 0.01 weak, $>.06$ medium, $>.14$ large) followed by Tukey's Honest Significant Differences post-hoc tests.

Tests of Reliability and Validity

Cronbach's alphas and McDonald's omegas were calculated to assess the reliability of the suicidal ideation scale ($>.70$ good).²⁴ In cases of low reliability, we also calculated coefficients H ($>.70$ good) to appraise whether the low reliability stems from a low factor loading of one of the items.²⁵ We adopted the Multi-trait Multimethod analysis for assessing the convergence and discriminant validity of the P4 screener. The analysis was conducted within the *psy* package with the *mtmm* (Multi-trait Multimethod) function.²⁶ The Multi-trait Multimethod analysis examines within-cluster and between-cluster correlations such that the unit of analysis is a single item. Good psychometric indices are high within-cluster correlations (e.g., high correlations between all items of the P4 screener), moderate between-cluster correlations between items from similar

¹ It was not possible to conduct the pre-registered measurement invariance tests due to the characteristics of the P4 Screener (e.g., binary answer options).

theoretical constructs (i.e., convergent validity), and weak between-cluster correlations between items from unrelated theoretical constructs (i.e., discriminant validity). In the current study, the convergence validity within the Multi-trait Multimethod analysis was estimated between items of the P4 screener and subscales of the BSI scale assessing depressive and anxiety symptoms; the discriminant analysis was appraised by assessing the between-cluster correlations between items of the suicidal ideation scale and items of substance use (ASSIST scale). Correlations around $|.10|$ were considered weak, $|.30|$ moderate, and $|.50|$ strong.²⁷

Establishing a Cut-off Score for the “At-Risk” group

To establish a cut-off score, we examined the effectiveness of the suicidal ideation score in differentiating between participants with suspected depression *and* anxiety (above a T score of 63 in BSI depression and anxiety; $n = 4,908$) and those without suspected depression and anxiety ($n = 61,810$). To do so, we calculated the optimal clinical cut-point by bootstrapping the optimal cut-point while maximizing the sensitivity and specificity (i.e., highest Youden’s index: sensitivity + specificity – 1). We also reported the suggested indexes of the “number needed to diagnose” (NND)²⁸ – the number of participants who need to be examined to correctly detect one person with the disorder of interest in a study population of persons with and without the known disorder – “number needed to misdiagnose” (NNM)²⁹ – number of individuals who need to be tested for one to be misdiagnosed by the test – and the “likelihood to be diagnosed or misdiagnosed” (LDM),³⁰ with LDM higher values (> 1) suggesting that a test is more likely to diagnose than misdiagnose.

Results

Psychometric Properties of the Suicidal Ideation Scale

Descriptive Statistics

Descriptive data for the P4 screener items are presented in Supplementary Table 1.

Among all participants, 20% reported attempting to deliberately harm themselves, and 23% reported having thought about specific ways to deliberately harm themselves. Six percent of participants reported that it was somewhat to very likely that they would act on their thoughts, and 9.4% reported that nothing was preventing them from deliberately harming themselves.

Language-, Country-, Gender-, and Sexual-orientation-based Confirmatory Factor Analyses of the P4 Screener

The results are presented in Supplementary Tables 2 to 5, which correspond to language, country, gender, and sexual orientation, respectively. The CFAs supported the structural validity of the P4 screener in 18 out of 20 languages, excluding Italian and simplified Mandarin languages. To explore the poor fit in these languages, we conducted subsequent bootstrap exploratory graph analyses (EGAs). We found that in Italian, the dimension stability was high but produced a 2-factor solution in 76.02% (factor 1 comprised items 1 and 4; factor 2 comprised items 2 and 3) of the resampling cycles, and a 1-factor solution in the remaining 23.98% of cycles; median dimension of 2, SE = .42, CI 1.16, 2.84. In addition, items 2 “Have you thought about how you might actually hurt yourself?” and 3 “How likely do you think it is that you will act on these thoughts about hurting yourself or ending your life sometime over the next month?” showed poor stability (0.24), indicating that these items Oscillate between 1 and 2 factors. Regarding simplified Mandarin, the bootstrap EGA indicated 100% dimensional and item stabilities, supporting the structural validity of the suicidal ideation scale.

The CFAs supported the structural validity of the suicidal ideation scale in 30 out of 31 countries, excluding Italy, as expected based on the language-based CFAs. The CFAs also corroborated the structural validity of the P4 screener in all genders and sexual orientations.

Standardized factor loadings by language, country, gender, and sexual orientation are presented in Supplementary Tables 6-9.

Group Differences

Kruskal-Wallis tests indicated significant differences in suicidal ideations between languages, $\chi^2_{(25)} = 1598.06$, $p = 1.73^{-322}$, $\eta^2 = .02$ (95% bootstrap CI .02, .02), countries, $\chi^2_{(44)} = 1599.16$, $p = 1.02^{-306}$, $\eta^2 = .02$ (95% bootstrap CI .02, .02), genders, $\chi^2_{(4)} = 2185.85$, $p \sim 0$ (i.e., approximating zero), $\eta^2 = .026$ (95% bootstrap CI .02, .03), and sexual orientations, $\chi^2_{(10)} = 4018.17$, $p \sim 0$, $\eta^2 = .048$ (95% bootstrap CI .045, .05). Post-hoc differences are presented in Supplementary Tables 10-13. Figure 1 presents the scores for suicidal ideation by country.

Reliability and Convergent and Discriminant Validity of the P4 screener

The P4 screener showed adequate reliability in most languages, countries, genders, and sexual orientations (Tables 3-6). The reliability of the scale was less than the recommended level of 0.70 in several languages, and in Canada, Mexico, Spain, and Switzerland with regard to countries. However, Coefficient H revealed high reliability in all of these instances [Czech (.98), Hungarian (.97), Portuguese in Portugal (.97), and Spanish in Spain (.98); Canada (.96), Mexico (.98), Spain (.98), and Switzerland (.96)] indicating that the less-than-adequate alpha and omega scores might stem in several cases from items with lower factor loadings.

The results of the convergence and discriminant validity (i.e., Multi-trait Multimethod analysis) are presented in Figure 2. The analysis supported high convergence and discriminant validity, such that the average correlation between items of the P4 screener and those of the depression or anxiety scales were moderate in size ($r_{mean} = .35$ for depression, $r_{mean} = .27$ for anxiety), and those assessing substance use were weak in size ($r_{mean} = .07$). The average correlation between items of the P4 screener was $r_{mean} = .49$.

Establishing a Cut-off Score for an “At-Risk” Group

Results are summarized in Figure 3. Bootstrapping the optimal cut-point of the P4 score revealed that using a cut-point of 1 produced a maximum Youden’s index of 0.47, with a sensitivity of 70.45% and a specificity of 76.34% when distinguishing participants “at risk” of a co-morbidity of depression and anxiety. Using the cut-point of 1, 2.14 participants were needed to be examined to correctly detect one person with the disorder of interest in a study population of persons with and without the known disorder (i.e., NND value; with a value of 1 as the best possible value). In addition, 4.15 participants needed to be tested for one person to be misdiagnosed by the test (i.e., NNM value). The overall likelihood of being diagnosed compared to being misdiagnosed was 1.94, which indicates adequate effectiveness in the diagnosis process.

Discussion

This study validates the P4 suicide screener as a reliable tool for identifying individuals with suicidal ideation across 42 diverse countries. The cross-cultural validation is particularly important as it allows for the collection of accurate and relevant international data on suicide, which is essential for advancing prevention measures and physician education on suicide prevention.³¹ Quality and accessibility of data can impact healthcare professionals’ ability to recognize patterns and identify suicide risk factors, develop effective treatment plans, and make informed decisions for patient care.

Collecting data on suicide mainly from WEIRD countries may not accurately represent the prevalence and underlying factors of suicide in other parts of the world due to variations in suicide rates, risk factors, cultural and linguistic barriers. Definitions of and discourses on suicide are also largely varying across countries, and not considering these might lead to a biased understanding of the issue. This problem is confounded by the facts that suicide is a sensitive

issue, and suicidal behavior is illegal in some nations. This study contributes to filling this gap by evaluating the reliability and validity of the P4 suicide screener across 42 diverse countries, including non-WEIRD and developing nations. Using a validated cross-cultural suicide risk assessment is crucial in standardizing and ensuring reliability in identifying high-risk individuals, monitoring changes in risk levels, and ultimately leading to better outcomes for those at risk of suicide. Recognizing the contextual differences between countries and cultures can lead to the development of culturally sensitive suicide prevention strategies, tailored to meet the needs of diverse populations, which is crucial for guiding efforts to prevent suicide.^{32,33}

Individuals who identified as a sexual minority had the highest frequencies of suicidal ideation, while heterosexual individuals had the lowest. Findings from the current study also demonstrate that gender-non-conforming individuals exhibited the greatest likelihood of reporting suicidal ideation, followed by individuals who identified as women and men. These results contribute to a growing body of evidence that suggests an increased risk of suicide ideation and behavior among individuals who belong to gender or sexual minorities.³⁴⁻³⁹ Healthcare professionals need to be aware of the higher suicide risk among marginalized groups and provide culturally responsive care that addresses their unique needs and experiences. In the current study, we found that 20% reported attempting deliberate self-harm, while 23% reported thinking about specific ways to do so. These observed rates surpass previous prevalence rates,⁶ potentially suggesting the inclusion of non-suicidal self-harm thoughts or may derive from the present sample's diversity (i.e., more sexually diverse individuals were included in the present study compared to previous ones and a heightened rate of suicidal risk was present among sexual minorities).

Finally, the establishment of a clear-cut cutoff score for identifying individuals as “at-risk” for suicidal ideation is a significant aspect of the study which is essential for early intervention and the provision of appropriate treatment and support in clinical practice. Considering research implications, it is a promising measure to be tested in the clinical field in order to contribute to overcoming the gap of measures that the predictive ability was tested in clinical studies.^{40,41}

Limitations and Future Research

Despite the strengths of the ISS, such as its novelty, large sample size, rigorous methodology, and following open-science practices, some general limitations should be considered in all studies using this dataset (for better overview see https://osf.io/n3k2c/?view_only=838146f6027c4e6bb68371d9d14220b5). Additional limitations include the reliance on online recruitment, which may introduce selection bias and compromise generalizability, especially to clinical populations. Underrepresentation of specific demographics, such as older adults and those with lower socioeconomic status, is acknowledged, likely due to the online survey approach. Geographical and cultural gaps, notably in Africa and certain Islamic countries, raise concerns about the broader applicability of findings. Furthermore, we did not include a specific question related to racial minority status in this study. Study-specific limitations include the inability to directly assess suicide, as no established follow-up protocol for evaluating individuals with a high suicide risk was implemented given the anonymous nature of the study. Additionally, the P4 suicide screener does not distinguish between self-harm with and without the intent to die underscores a limitation, emphasizing the necessity for further research to achieve a more precise differentiation. The utilization of more explicit scales may not be suitable in particular cultures, and their use in online anonymous

surveys would have raised ethical issues. Addressing these limitations in future research is imperative for a nuanced understanding of suicide ideation rates, particularly in clinical contexts.

Conclusions

This study contributes to suicide prevention efforts by demonstrating the validity and reliability of the P4 suicide screener, an efficient tool for identifying individuals at risk for suicidal ideation and behavior across diverse populations. The use of a standardized tool may improve the quality and accessibility of data on suicide, which is important for developing and improving the effectiveness of suicide prevention strategies in clinical practice.

References

1. World Health Organization. Suicide. Published 2021. Accessed January 4, 2023.
<https://www.who.int/news-room/fact-sheets/detail/suicide>
2. Favril L, Yu R, Uyar A, Sharpe M, Fazel S. Risk factors for suicide in adults: Systematic review and meta-analysis of psychological autopsy studies. *Evid Based Ment Health*. 2022;10:1-8. doi:10.1136/ebmental-2022-300549
3. Too LS, Spittal MJ, Bugeja L, Reifels L, Butterworth P, Pirkis J. The association between mental disorders and suicide: A systematic review and meta-analysis of record linkage studies. *J Affect Disord*. 2019;259:302-313. doi:10.1016/J.JAD.2019.08.054
4. Gili M, Castellví P, Vives M, et al. Mental disorders as risk factors for suicidal behavior in young people: A meta-analysis and systematic review of longitudinal studies. Published online 2018. doi:10.1016/j.jad.2018.10.115
5. Holman MS, Williams MN. Suicide Risk and Protective Factors: A Network Approach. <https://doi.org/10.1080/1381111820201774454>. Published online 2020:1-18.
doi:10.1080/13811118.2020.1774454
6. Bullington C, Kroenke K. P4 suicidality screener: Literature synthesis and results from two randomized trials. *Gen Hosp Psychiatry*. 2023;85:177-184.
doi:10.1016/j.genhosppsy.2023.11.004
7. Dube P, Kroenke K, Bair MJ, Theobald D, Williams LS. The P4 Screener: Evaluation of a Brief Measure for Assessing Potential Suicide Risk in 2 Randomized Effectiveness Trials of Primary Care and Oncology Patients. *Prim Care Companion J Clin Psychiatry*. 2010;12(6). doi:10.4088/PCC.10M00978BLU

8. Chu JP, Goldblum P, Floyd R, Bongar B. The cultural theory and model of suicide. *Applied and Preventive Psychology*. 2010;14(1-4):25-40. doi:10.1016/j.appsy.2011.11.001
9. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine (Phila Pa 1976)*. 2000;25(24):3186-3191..
10. Bóthe B, Koós M, Nagy L, Kraus SW, Potenza MN, Demetrovics Z. International Sex Survey: Study protocol of a large, cross-cultural collaborative study in 45 countries. *J Behav Addict*. 2022;10:632-645. doi:10.1556/2006.2021.00063
11. Bóthe B, Koós M, Nagy L, Kraus SW, Potenza MN, Demetrovics Z. International Sex Survey: Study protocol of a large, cross-cultural collaborative study in 45 countries. *J Behav Addict*. 2021;10(3):632-645. doi:10.1556/2006.2021.00063
12. Asner-Self KK, Schreiber JB, Marotta SA. A cross-cultural analysis of the Brief Symptom Inventory-18. *Cultur Divers Ethnic Minor Psychol*. 2006;12(2):367-375. doi:10.1037/1099-9809.12.2.367
13. Ali R, Awwad E, Babor TF, et al. The Alcohol, Smoking and Substance Involvement Screening Test (ASSIST): development, reliability and feasibility. *Addiction*. 2002;97(9):1183-1194. doi:10.1046/J.1360-0443.2002.00185.X
14. Lee CT, Lin CY, Koós M, et al. The eleven-item Alcohol, Smoking and Substance Involvement Screening Test (ASSIST-11): Cross-cultural psychometric evaluation across 42 countries. *J Psychiatr Res*. 2023;165:16-27. doi:10.1016/j.jpsychires.2023.06.033

15. Little RJ. A test of missing completely at random for multivariate data with missing values. *J Am Stat Assoc.* 1988;83(404):1198-1202.
16. Newman DA. Missing Data: Five Practical Guidelines. *Organ Res Methods.* 2014;17(4):372-411. doi:10.1177/1094428114548590
17. Browne MW, Cudeck R. Alternative ways of assessing model fit. *Testing structural equation models.* 1993;21(2):136-162. doi:10.1167/iovs.04-1279
18. Marsh HW, Hau KT, Grayson D. Goodness of fit evaluation in structural equation modeling. In: Maydeu-Olivares A, McArdle J, eds. *Contemporary Psychometrics.* Erlbaum; 2005:275-340.
19. Schermelleh-Engel K, Moosbrugger H, Müller H. Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *MPR-online.* 2003;8(2):23-74.
20. Sahoo M. Structural equation modeling: Threshold criteria for assessing model fit. In: *Methodological Issues in Management Research: Advances, Challenges, and the Way Ahead.* . Emerald Publishing Limited; 2019.
21. KILIÇ A, UYSAL İ, ATAR B. Comparison of confirmatory factor analysis estimation methods on binary data. *International Journal of Assessment Tools in Education.* Published online August 24, 2020:451-487. doi:10.21449/ijate.660353
22. Golino H, Shi D, Christensen AP, et al. Investigating the performance of Exploratory Graph Analysis and traditional techniques to identify the number of latent factors: A simulation and tutorial. *Psychol Methods.* 2019;25(3):292-320. doi:10.1037/met0000255

23. Golino HCAP, MR. EGAnet: Exploratory graph analysis: A framework for estimating the number of dimensions in multivariate data using network psychometrics. . *R package version 09, 5* . Published online 2020.
24. Revelle W, CDM. Supplemental Material for Reliability From α to ω : A Tutorial. *Psychol Assess*. Published online 2019. doi:10.1037/pas0000754.supp
25. Kalkbrenner MT. Alpha, Omega, and H Internal Consistency Reliability Estimates: Reviewing These Options and When to Use Them. *Counseling Outcome Research and Evaluation*. Published online 2021. doi:10.1080/21501378.2021.1940118
26. Falissard B. psy: Various Procedures Used in Psychometrics. In: *Version R Package Version 1.2.* ; 2022.
27. Cohen J. *A Power Primer*. Vol 112.; 1992.
28. Linn S, Grunau PD. New patient-oriented summary measure of net total gain in certainty for dichotomous diagnostic tests. *Epidemiologic Perspectives and Innovations*. 2006;3. doi:10.1186/1742-5573-3-11
29. Habibzadeh F, Yadollahie M. Number needed to misdiagnose: A measure of diagnostic test effectiveness. *Epidemiology*. 2013;24(1):170. doi:10.1097/EDE.0b013e31827825f2
30. Citrome L, Ketter TA. When does a difference make a difference? Interpretation of number needed to treat, number needed to harm, and likelihood to be helped or harmed. *Int J Clin Pract*. 2013;67(5):407-411. doi:10.1111/ijcp.12142
31. Mann JJ, Apter A, Bertolote J, et al. Suicide Prevention Strategies A Systematic Review. *JAMA*. 2005;294(16):2064-2074. <https://jamanetwork.com/>

32. Hjelmeland H. Cultural context is crucial in suicide research and prevention. *Crisis*. 2011;32(2):61-64. doi:10.1027/0227-5910/a000097
33. Dückers MLA, Reifels L, De Beurs DP, Brewin CR. The vulnerability paradox in global mental health and its applicability to suicide. *British Journal of Psychiatry*. 2019;215(4):588-593. doi:10.1192/bjp.2019.41
34. Ramchand R, Schuler MS, Schoenbaum M, Colpe L, Ayer L. Suicidality Among Sexual Minority Adults: Gender, Age, and Race/Ethnicity Differences. *Am J Prev Med*. 2022;62(2):193-202. doi:10.1016/j.amepre.2021.07.012
35. Mcdaniel JS, Purcell D, D'augelli AR. The Relationship Between Sexual Orientation and Risk for Suicide: Research Findings and Future Directions for Research and Prevention. *Suicide Life Threat Behav*. Published online 2001:84-105.
36. Plöderl M, Wagenmakers EJ, Tremblay P, et al. Suicide risk and sexual orientation: A critical review. In: *Archives of Sexual Behavior*. Vol 42. ; 2013:715-727. doi:10.1007/s10508-012-0056-y
37. Salway T, Plöderl M, Liu J, Gustafson P. Effects of Multiple Forms of Information Bias on Estimated Prevalence of Suicide Attempts According to Sexual Orientation: An Application of a Bayesian Misclassification Correction Method to Data from a Systematic Review. *Am J Epidemiol*. 2019;188(1):239-249. doi:10.1093/aje/kwy200
38. McNeil J, Ellis SJ, Eccles FJR. Suicide in trans populations: A systematic review of prevalence and correlates. *Psychol Sex Orientat Gend Divers*. 2017;4(3):341-353. doi:10.1037/sgd0000235

39. Turban JL, Beckwith N, Reisner SL, Keuroghlian AS. Association between Recalled Exposure to Gender Identity Conversion Efforts and Psychological Distress and Suicide Attempts among Transgender Adults. *JAMA Psychiatry*. 2020;77(1):68-76.
doi:10.1001/jamapsychiatry.2019.2285
40. Bolton JM, Gunnell D, Turecki G. Suicide risk assessment and intervention in people with mental illness. doi:10.2307/26524477
41. Randall JR, Colman I, Rowe BH. A systematic review of psychometric assessment of self-harm risk in the emergency department. *J Affect Disord*. 2011;134(1-3):348-355.
doi:10.1016/j.jad.2011.05.032

Table 1

Participants' Sociodemographic Characteristics

Characteristic	N = 73,531 ¹
Country of residence	
Algeria	18 (<0.1%)
Australia	577 (0.8%)
Austria	689 (0.9%)
Bangladesh	328 (0.4%)
Belgium	584 (0.8%)
Bolivia	313 (0.4%)
Brazil	3,140 (4.3%)
Canada	2,285 (3.1%)
China	2,395 (3.3%)
Colombia	1,771 (2.4%)
Croatia	2,133 (2.9%)
Czech Republic	1,618 (2.2%)
Ecuador	233 (0.3%)
Egypt	50 (<0.1%)
France	1,527 (2.1%)
Germany	3,026 (4.1%)
Gibraltar	47 (<0.1%)
Hungary	9,974 (14%)
India	151 (0.2%)
Iraq	84 (0.1%)
Ireland	1,435 (2.0%)
Israel	1,167 (1.6%)
Italy	2,047 (2.8%)
Japan	495 (0.7%)
Lithuania	1,840 (2.5%)
Malaysia	1,085 (1.5%)
Mexico	1,821 (2.5%)
New Zealand	2,488 (3.4%)
North Macedonia	1,147 (1.6%)
Pakistan	0 (0%)
Panama	279 (0.4%)
Peru	2,291 (3.1%)
Poland	8,653 (12%)
Portugal	1,981 (2.7%)
Slovakia	994 (1.4%)
South Africa	1,620 (2.2%)
South Korea	1,328 (1.8%)

Characteristic	N = 73,531 ¹
Spain	2,101 (2.9%)
Switzerland	1,070 (1.5%)
Taiwan	2,603 (3.5%)
Turkey	689 (0.9%)
United Kingdom	1,249 (1.7%)
United States of America	2,147 (2.9%)
Other	975 (1.3%)
Chile	1,083 (1.5%)
Language	
Arabic	121 (0.2%)
Bangla	300 (0.4%)
Croatian	2,248 (3.1%)
Czech	1,568 (2.1%)
Dutch	469 (0.6%)
English	12,310 (17%)
French	3,575 (4.9%)
German	3,262 (4.4%)
Hebrew	1,147 (1.6%)
Hindi	14 (<0.1%)
Hungarian	9,724 (13%)
Italian	2,080 (2.8%)
Japanese	409 (0.6%)
Korean	1,304 (1.8%)
Lithuanian	1,913 (2.6%)
Macedonian	1,193 (1.6%)
Mandarin – simplified	2,437 (3.3%)
Mandarin – traditional	2,618 (3.6%)
Polish	9,044 (12%)
Portuguese – Brazil	3,207 (4.4%)
Portuguese – Portugal	1,984 (2.7%)
Slovak	1,901 (2.6%)
Spanish – Latin American	7,829 (11%)
Spanish – Spain	2,091 (2.8%)
Turkish	718 (1.0%)
Sex assigned at birth	
Male	29,733 (40%)
Female	43,788 (60%)
Missing values	10
Gender (original answer options in the survey)	
Masculine/Man	29,106 (40%)
Feminine/Woman	41,881 (57%)
Indigenous or other cultural gender minority identity (e.g., two-spirit)	151 (0.2%)

Characteristic	N = 73,531¹
Non-binary, gender fluid, or something else (e.g., genderqueer)	2,089 (2.8%)
Other (If you wish, tell us how you personally describe your gender):	272 (0.4%)
Missing values	32
Gender (categories used in the analyses)	
Man	29,106 (40%)
Woman	41,881 (57%)
Gender diverse individual	2,512 (3.4%)
Missing values	32
Trans status	
No, I am not a trans person	70,833 (96%)
Yes, I am a trans man	329 (0.4%)
Yes, I am a trans woman	271 (0.4%)
Yes, I am a non-binary trans person	798 (1.1%)
I am questioning my gender identity	1,032 (1.4%)
I don't know what it means	248 (0.3%)
Missing values	20
Sexual orientation (original answer options in the survey)	
Heterosexual/Straight	50,205 (68%)
Gay or lesbian or homosexual	4,156 (5.7%)
Heteroflexible	5,538 (7.5%)
Homoflexible	483 (0.7%)
Bisexual	6,804 (9.3%)
Queer	851 (1.2%)
Pansexual	1,747 (2.4%)
Asexual	978 (1.3%)
I do not know yet or I am currently questioning my sexual orientation	1,731 (2.4%)
None of the above, specify:	722 (1.0%)
I don't want to answer	288 (0.4%)
Missing values	28
Sexual orientation (categories used in the analyses)	
Heterosexual	50,205 (69%)
Homosexual	4,156 (5.7%)
Bisexual	6,804 (9.3%)
Queer and pansexual	2,598 (3.5%)
Homo- and heteroflexible identities	6,021 (8.2%)
Asexual	978 (1.3%)
Questioning	1,731 (2.4%)
Other	722 (1.0%)
Missing values	316
Highest level of education	
Primary (e.g., elementary school)	879 (1.2%)
Secondary (e.g., high school)	17,943 (24%)
Tertiary (e.g., college or university)	54,693 (74%)

Characteristic	N = 73,531 ¹
Missing values	16
Currently being in education	
No	44,779 (61%)
Yes, in primary education (e.g., elementary school)	55 (<0.1%)
Yes, in secondary education (e.g., high school)	1,315 (1.8%)
Yes, in tertiary education (e.g., college or university)	27,342 (37%)
Missing values	40
Work status	
No	18,764 (26%)
Yes, full time	38,379 (52%)
Yes, part-time	10,121 (14%)
Yes, I do odd jobs	6,247 (8.5%)
Missing values	20
Socioeconomic status	
My life circumstances are among the worst	201 (0.3%)
My life circumstances are much worse than average	693 (0.9%)
My life circumstances are worse than average	3,833 (5.2%)
My life circumstances are average	23,908 (33%)
My life circumstances are better than average	28,182 (38%)
My life circumstances are much better than average	13,172 (18%)
My life circumstances are among the best	3,534 (4.8%)
Missing values	8
Residence	
Metropolis (population is over 1 million people)	23,767 (32%)
City (population is between 100,000-999,999 people)	26,721 (36%)
Town (population is between 1,000-99,999 people)	18,808 (26%)
Village (population is below 1,000 people)	4,222 (5.7%)
Missing values	13
Relationship status	
Single	24,709 (34%)
In a relationship	24,271 (33%)
Married or common-law partners	21,871 (30%)
Widow or widower	394 (0.5%)
Divorced	2,264 (3.1%)
Missing values	22
Having children	
No	51,655 (70%)
yes, 1	7,528 (10%)
yes, 2	9,328 (13%)
yes, 3	3,476 (4.7%)
yes, 4	921 (1.3%)
yes, 5	262 (0.4%)
yes, 6-9	116 (0.2%)

Characteristic	N = 73,531 ¹
yes, 10 or more	20 (<0.1%)
Missing values	225
Present religion	
Christian	21,348 (29%)
Buddhist	1,445 (2.0%)
Hindu	198 (0.3%)
Muslim	890 (1.2%)
Spiritual but not committed to one religion	10,086 (14%)
I am not religious	35,369 (48%)
Other, please specify:	1,958 (2.7%)
Jewish	1,133 (1.5%)
Taoist	605 (0.8%)
Confucianist	18 (<0.1%)
Sikh	29 (<0.1%)
Spiritist	381 (0.5%)
Jain	14 (<0.1%)
Missing values	57
Age	33 (13), 18 - 99
Missing values	13

¹n (%); Mean (SD), Range

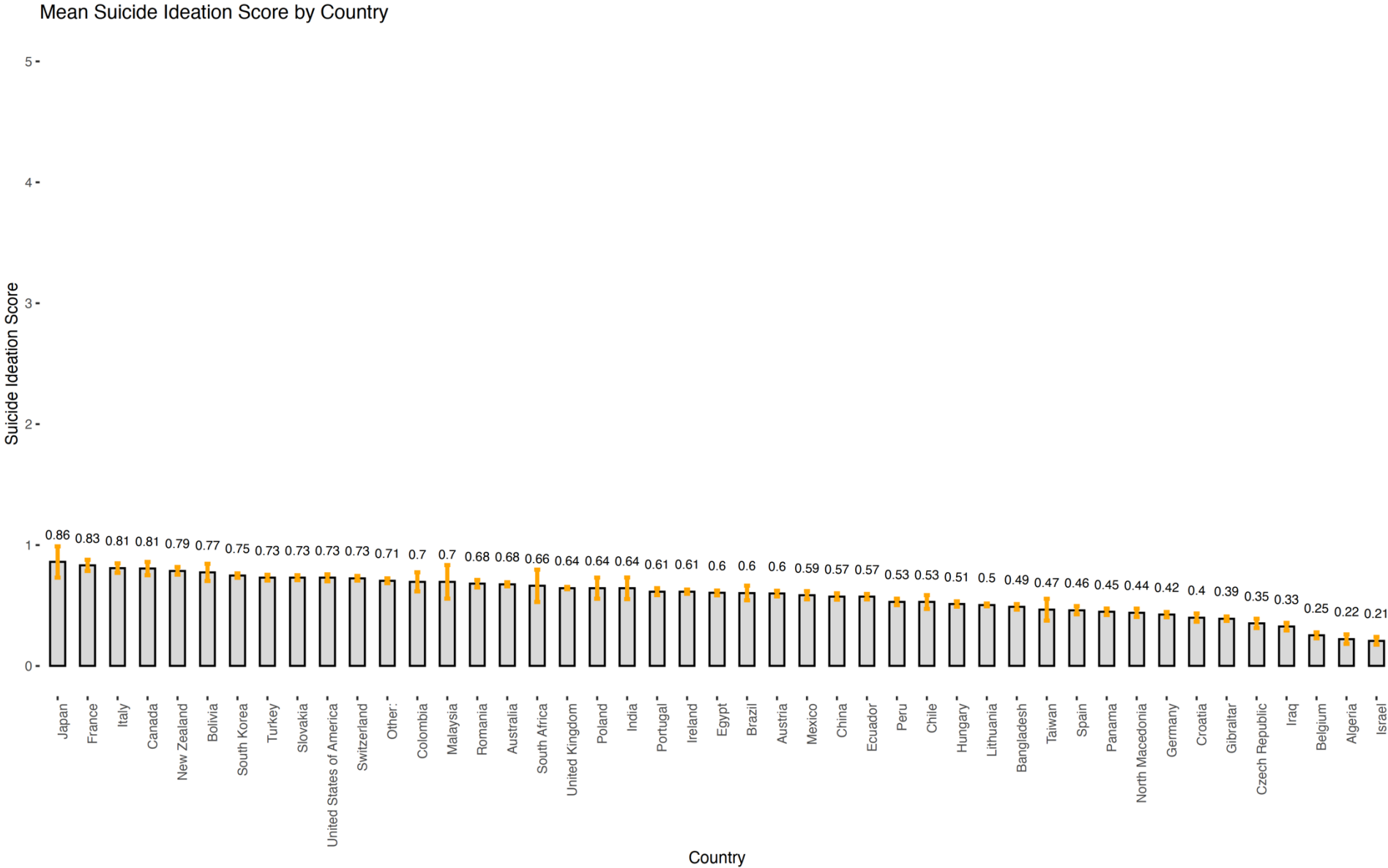


Figure 1. Suicidal ideation scores by country.

Note. Error bars represent standard errors.

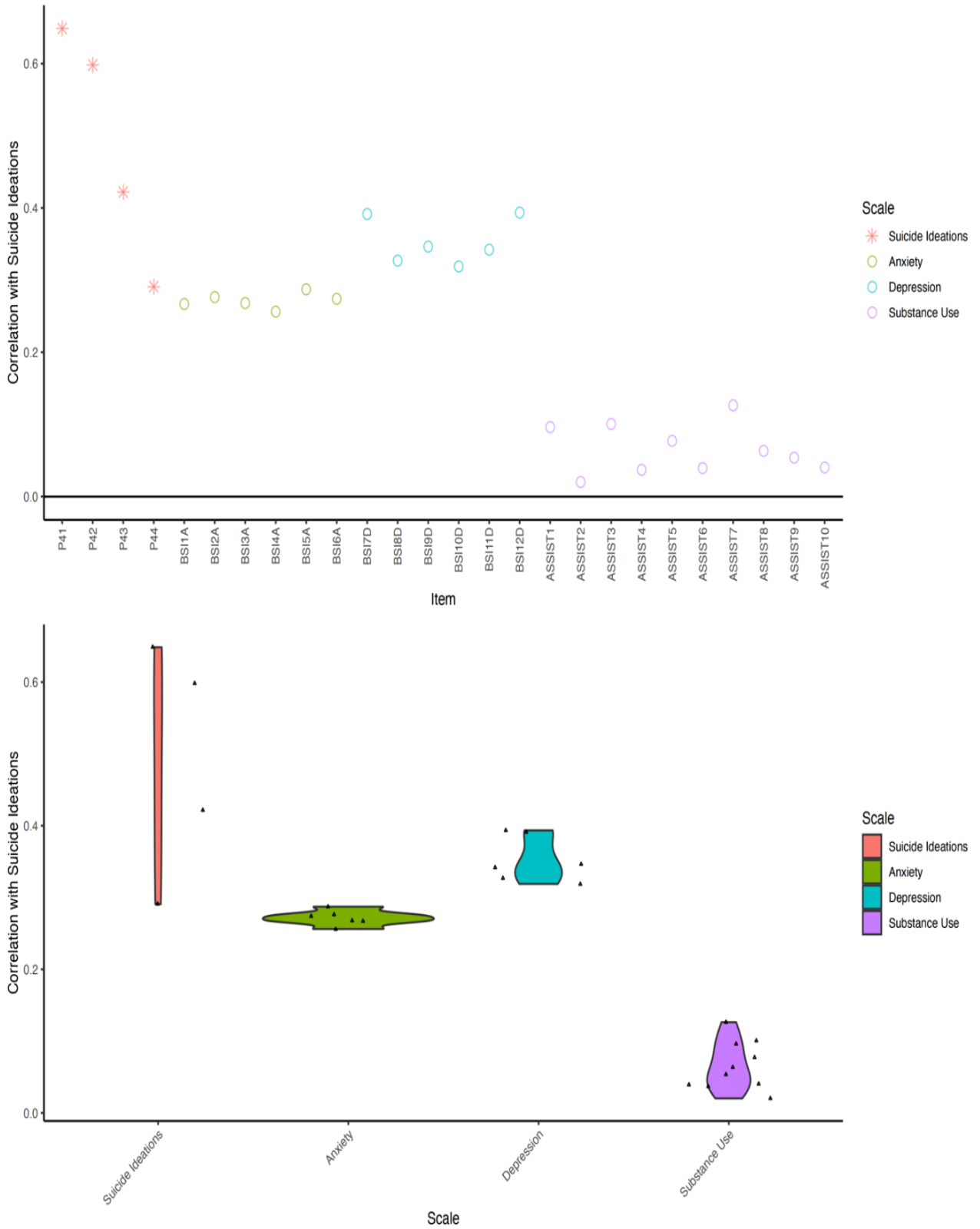


Figure 2. Multi-trait Multimethod analysis for assessing convergence and discriminant validity.

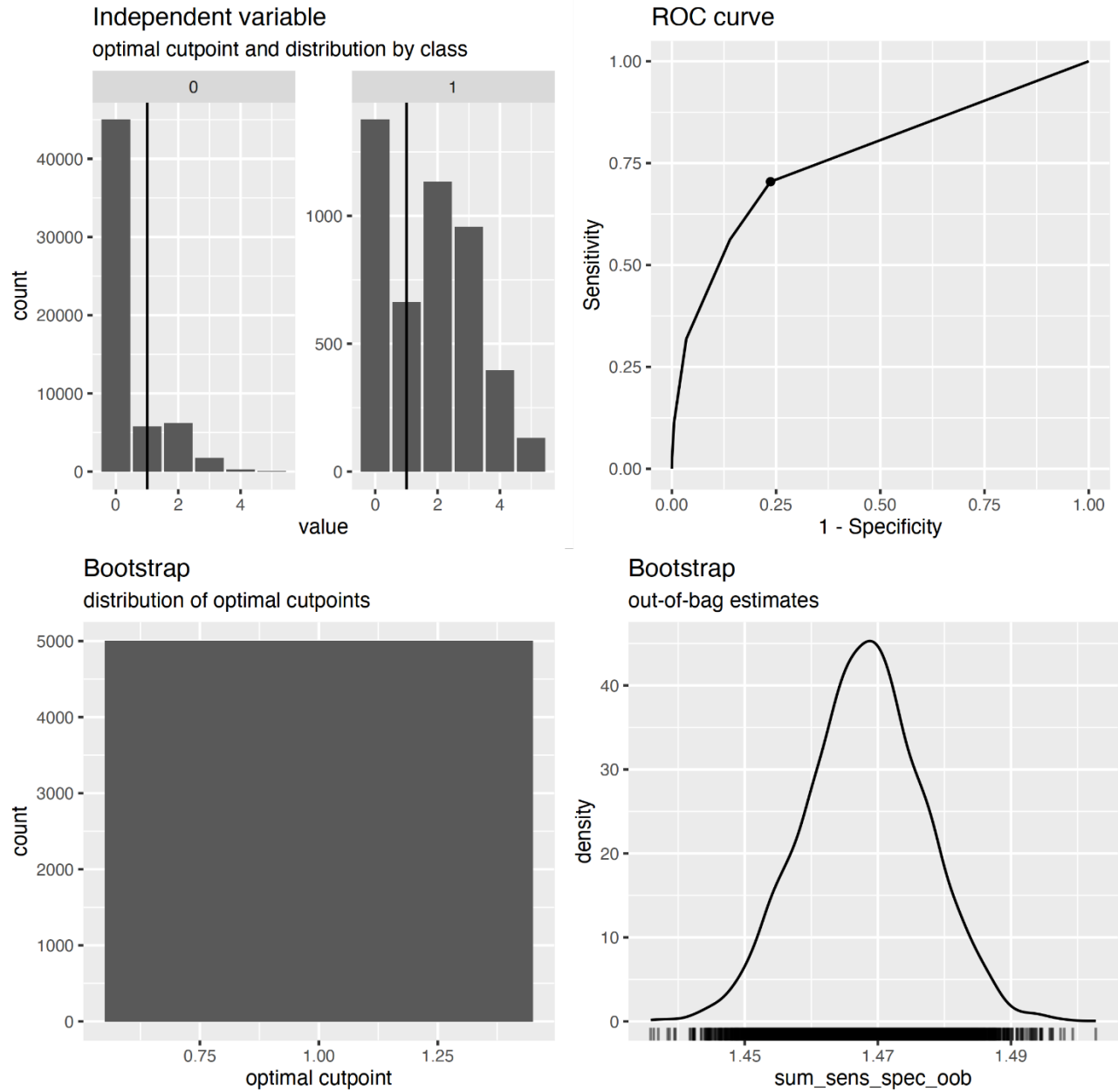


Figure 3. Results of the cut-point optimization process using bootstrap analysis. Top-left: density of participants below and above the suggested cut-point of 1 in the scale among participants with “at risk” depression and anxiety (1) and those at low risk (0). Top-right: the ROC curve for the estimation process of the optimal cut-point is presented with the black dot indicating the highest Youden’s index. Bottom-left: Density of the optimal cut-point in the estimation process. The optimal cut-point was identified as 1 with a standard error of 0, attributable to the discrete nature of the scale. The x-axis spread solely

represents a bar over the discrete score of 1. Bottom-right: density of the highest summed sensitivity and specificity scores of the suicidal ideation scale.

Supplementary Table 1

Descriptive Statistics of the Items of the Suicidal Ideation Scale

	<i>N</i>	<i>M</i>	<i>SD</i>	<i>Mdn</i>	<i>Min</i>	<i>Max</i>	<i>Skew</i>	<i>Kurt</i>	<i>SE</i>	<i>n (%) Yes</i>
1. Have you ever attempted to harm yourself in the past?	73,351	1.20	0.40	1	1	2	1.48	0.19	<0.01	14,870 (20.0%)
2. Have you thought about how you might actually hurt yourself?	73,362	1.23	0.42	1	1	2	1.25	-0.43	<0.01	17,212 (23.0%)
3. There's a big difference between having a thought and acting on a thought. How likely do you think it is that you will act on these thoughts about hurting yourself or ending your life some time over the next month?	73,462	1.07	0.28	1	1	3	4.4	20.43	<0.01	4,445 (6.0%)
4. Is there anything that would prevent or keep you from harming yourself?	72,815	1.09	0.29	1	1	2	2.78	5.72	<0.01	6,860 (9.4%)

Note. *M* = mean, *SD* = standard deviation, *Mdn* = median, *Skew.* = skewness, *SE* = standard error, *Kurt.* = kurtosis.

Supplementary Table 2

Language-based Confirmatory Factor Analyses

Language	<i>ULSMV</i> χ^2	<i>p</i>	<i>Robust CFI</i>	<i>Robust TLI</i>	<i>Robust RMSEA</i>	90% CI				
						<i>LB</i>	<i>UB</i>	<i>SRMR</i>	<i>Alpha</i>	<i>Omega</i>
1 Croatian	0.39	0.530	1.00	1.00	0.00	0.00	0.05	0.00	0.80	0.84
2 <i>Czech</i>	3.26	0.071	1.00	1.00	0.02	0.00	0.04	0.02	0.63	0.66
3 English	10.54	0.001	1.00	1.00	0.03	0.02	0.04	0.01	0.71	0.77
4 French	6.63	0.010	1.00	0.98	0.04	0.02	0.07	0.01	0.69	0.74
5 German	0.11	0.740	1.00	1.00	0.00	0.00	0.03	0.00	0.64	0.72
6 Hebrew	0.57	0.450	1.00	1.00	0.00	0.00	0.07	0.01	0.77	0.83
7 <i>Hungarian</i>	4.84	0.028	1.00	0.99	0.02	0.01	0.04	0.01	0.56	0.61
8 Italian	32.25	0.000	0.90	0.41	0.12	0.09	0.16	0.08	0.74	0.80
9 Japanese	1.30	0.250	1.00	1.00	0.02	0.00	0.08	0.01	0.70	0.78
10 Korean	5.46	0.019	1.00	1.00	0.03	0.01	0.05	0.01	0.67	0.73
11 Lithuanian	3.36	0.067	1.00	1.00	0.02	0.00	0.04	0.01	0.64	0.71
12 Macedonian	6.77	0.009	0.99	0.96	0.04	0.02	0.08	0.03	0.67	0.73
13 Mandarin – simplified	58.42	0.000	0.98	0.85	0.09	0.07	0.11	0.05	0.64	0.70
14 Mandarin – traditional	20.44	0.000	0.99	0.94	0.05	0.03	0.07	0.03	0.78	0.83
15 Polish	20.22	0.000	1.00	0.99	0.03	0.02	0.04	0.01	0.63	0.70
16 Portuguese – Brazil	0.30	0.583	1.00	1.00	0.00	0.00	0.02	0.00	0.71	0.79
17 <i>Portuguese – Portugal</i>	8.52	0.004	1.00	0.98	0.04	0.02	0.06	0.02	0.55	0.59
18 Spanish – Latin American	21.76	0.000	1.00	0.99	0.03	0.02	0.04	0.01	0.85	0.90
19 <i>Spanish – Spain</i>	0.30	0.582	1.00	1.00	0.00	0.00	0.02	0.00	0.33	0.40
20 Turkish	5.08	0.024	0.99	0.93	0.06	0.02	0.11	0.02	0.76	0.82

Note. *ULSMV* = mean and variance adjusted unweighted least squares; *CFI* = Comparative Fit Index; *TLI* = Tucker-Lewis Index; *RMSEA* = Root-Mean-Square Error of Approximation; *LB* = 90% lower bound; *UB* = 90% upper bound; *SRMR* = Standardized Root Mean Square Residual; *Alpha* = Cronbach's alpha; *Omega* = McDonald's omega; countries in bold have poor-fitted CFA; countries in an italicized font have a reliability below .70

Supplementary Table 3

Country-based Confirmatory Factor Analyses

		90% CI									
Country	<i>ULSMV</i>	χ^2	<i>p</i>	<i>Robust CFI</i>	<i>Robust TLI</i>	<i>Robust RMSEA</i>	<i>LB</i>	<i>UB</i>	<i>SRMR</i>	<i>Alpha</i>	<i>Omega</i>
1	Australia	0.23	0.629	1.00	1.00	0.00	0.00	0.04	0.00	0.67	0.78
2	Austria	0.20	0.652	1.00	1.00	0.00	0.00	0.03	0.00	0.72	0.77
3	Brazil	0.40	0.525	1.00	1.00	0.00	0.00	0.02	0.00	0.77	0.82
4	<i>Canada</i>	0.85	0.356	1.00	1.00	0.00	0.00	0.03	0.01	0.55	0.63
5	China	54.87	0.000	0.98	0.86	0.09	0.07	0.11	0.05	0.78	0.83
6	Chile	2.83	0.093	1.00	1.00	0.02	0.00	0.50	0.01	0.76	0.82
7	Colombia	4.36	0.037	1.00	1.00	0.02	0.01	0.05	0.01	0.78	0.78
8	Croatia	0.74	0.391	1.00	1.00	0.00	0.00	0.03	0.01	0.70	0.78
9	Czech Republic	0.68	0.411	1.00	1.00	0.00	0.00	0.03	0.01	0.76	0.80
10	Germany	0.27	0.601	1.00	1.00	0.00	0.00	0.02	0.00	0.65	0.73
11	Hungary	6.05	0.014	1.00	1.00	0.01	0.00	0.02	0.01	0.71	0.76
12	Israel	0.67	0.415	1.00	1.00	0.00	0.00	0.04	0.01	0.66	0.71
13	Italy	33.97	0.000	0.90	0.38	0.13	0.09	0.17	0.08	0.67	0.74
14	Japan	1.65	0.199	1.00	1.00	0.02	0.00	0.07	0.01	0.66	0.71
15	Lithuania	4.19	0.041	1.00	0.99	0.02	0.00	0.04	0.01	0.73	0.80
16	Malaysia	4.27	0.039	1.00	1.00	0.03	0.01	0.06	0.01	0.67	0.74
17	<i>Mexico</i>	0.16	0.694	1.00	1.00	0.00	0.00	0.02	0.00	0.64	0.69
18	New Zealand	5.29	0.021	1.00	1.00	0.02	0.01	0.04	0.01	0.77	0.77
19	North Macedonia	6.33	0.012	0.99	0.95	0.04	0.02	0.08	0.02	0.62	0.69
20	Peru	11.45	0.010	1.00	0.99	0.03	0.02	0.05	0.02	0.76	0.81
21	Poland	19.41	0.000	1.00	0.99	0.03	0.02	0.04	0.01	0.69	0.77
22	Portugal	8.64	0.003	1.00	0.98	0.04	0.02	0.06	0.02	0.69	0.76
23	Slovakia	0.81	0.368	1.00	1.00	0.00	0.00	0.06	0.01	0.64	0.71

24	South Africa	0.54	0.463	1.00	1.00	0.00	0.00	0.03	0.01	0.67	0.71
25	South Korea	0.59	0.016	1.00	1.00	0.03	0.01	0.05	0.01	0.68	0.74
26	<i>Spain</i>	1.18	0.277	1.00	1.00	0.00	0.00	0.03	0.01	0.63	0.63
27	<i>Switzerland</i>	4.03	0.045	1.00	0.99	0.03	0.00	0.07	0.02	0.34	0.55
28	Taiwan	20.61	0.000	0.99	0.94	0.05	0.03	0.07	0.03	0.77	0.82
29	Turkey	6.85	0.009	0.99	0.91	0.07	0.03	0.12	0.03	0.71	0.77
30	United Kingdom	0.39	0.531	1.00	1.00	0.00	0.00	0.03	0.00	0.73	0.77
31	United States of America	0.72	0.396	1.00	1.00	0.00	0.00	0.03	0.00	0.65	0.70

Note. ULMSV = mean and variance adjusted unweighted least squares; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; RMSEA = Root-Mean-Square Error of Approximation; LB = 90% lower bound of Robust RMSEA; UB = 90% upper bound of Robust RMSEA; SRMR = Standardized Root Mean Square Residual; Alpha = Cronbach's alpha; Omega = McDonald's omega; countries in bold have poor-fitted CFA; countries in an italicized font have reliability below .70

Supplementary Table 4

Gender-based Confirmatory Factor Analyses

Gender		<i>ULSMV</i> χ^2	<i>p</i>	<i>Robust CFI</i>	<i>Robust TLI</i>	<i>Robust RMSEA</i>	90% CI		<i>SRMR</i>	<i>Alpha</i>	<i>Omega</i>
							<i>LB</i>	<i>UB</i>			
1	Men	29.04	0.000	1.00	0.99	0.02	0.01	0.03	0.01	0.67	0.71
2	Women	180.35	0.000	1.00	0.98	0.04	0.03	0.05	0.02	0.68	0.75
3	Gender-diverse individuals	2.62	0.105	1.00	1.00	0.02	0.00	0.04	0.01	0.71	0.78

Note. *ULMSV* = mean and variance adjusted unweighted least squares; *CFI* = Comparative Fit Index; *TLI* = Tucker-Lewis Index; *RMSEA* = Root-Mean-Square Error of Approximation; *LB* = 90% lower bound of Robust *RMSEA*; *UB* = 90% upper bound of Robust *RMSEA*; *SRMR* = Standardized Root Mean Square Residual; *Alpha* = Cronbach's alpha; *Omega* = McDonald's omega; countries in bold have poor-fitted CFA; countries in an italicized font have reliability below .70

Supplementary Table 5

Sexual-orientation-based Confirmatory Factor Analyses

Sexual orientation		<i>ULSMV</i> χ^2	<i>p</i>	<i>Robust</i> <i>CFI</i>	<i>Robust</i> <i>TLI</i>	<i>Robust</i> <i>RMSEA</i>	90% CI				
							<i>LB</i>	<i>UB</i>	<i>SRMR</i>	<i>Alpha</i>	<i>Omega</i>
1	Heterosexual	150.84	0.000	1.00	0.97	0.03	0.03	0.04	0.02	0.70	0.76
2	Gay or lesbian	5.41	0.020	1.00	1.00	0.02	0.01	0.03	0.01	0.72	0.77
3	Bisexual	8.21	0.004	1.00	1.00	0.02	0.01	0.03	0.01	0.65	0.71
4	Queer and pansexual	3.95	0.047	1.00	1.00	0.02	0.00	0.04	0.01	0.70	0.77
5	Homo- and heteroflexible identities	7.89	0.005	1.00	1.00	0.02	0.01	0.04	0.01	0.69	0.75
6	Asexual	0.34	0.560	1.00	1.00	0.00	0.00	0.05	0.01	0.69	0.77
7	Questioning	4.44	0.035	1.00	0.99	0.03	0.01	0.06	0.01	0.71	0.76
8	Other	7.47	0.006	0.99	0.98	0.06	0.03	0.10	0.03	0.67	0.73

Note. ULSMV = mean and variance adjusted unweighted least squares; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; RMSEA = Root-Mean-Square Error of Approximation; LB = 90% lower bound of Robust RMSEA; UB = 90% upper bound of Robust RMSEA; SRMR = Standardized Root Mean Square Residual; Alpha = Cronbach's alpha; Omega = McDonald's omega; countries in bold have poor-fitted CFA; countries in an italicized font have reliability below .70

Supplementary Table 6

Standardized factor loadings of the P4 screener by language

Language	P4-1	P4-2	P4-3	P4-4
Arabic	0.68	0.61	0.62	0.40
Croatian	0.83	0.68	0.30	0.40
Czech	0.78	0.73	0.87	0.58
Dutch	0.65	0.60	0.60	0.38
French	0.87	0.81	0.42	0.37
German	0.55	0.48	0.61	0.55
Hebrew	0.79	0.78	0.30	0.22
Hungarian	0.66	0.67	0.43	0.40
Italian	0.71	-0.29	-0.01	0.18
Japanese	0.87	0.82	0.72	0.46
Korean	0.76	0.74	0.67	0.51
Lithuanian	0.83	0.84	0.42	0.50
Macedonian	0.63	0.42	0.57	0.41
Mandarin – simplified	0.80	0.47	0.47	0.41
Mandarin – traditional	0.72	0.46	0.60	0.43
Polish	0.60	0.58	0.67	0.31
Portuguese – Brazil	0.81	0.76	0.42	0.36
Portuguese – Portugal	0.81	0.66	0.39	0.35
Slovak	0.53	0.48	0.67	0.31
Spanish – Latin American	0.74	0.70	0.75	0.39
Spanish – Spain	0.77	0.75	0.66	0.39
Turkish	0.61	0.50	0.57	0.19

Supplementary Table 7

Standardized factor loadings of the P4 by country

Country	P4-1	P4-2	P4-3	P4-4
Australia	0.65	0.53	0.63	0.42
Austria	0.61	0.45	0.54	0.50
Belgium	0.88	0.81	0.62	0.36
Brazil	0.80	0.75	0.42	0.36
Canada	0.81	0.76	0.39	0.33
China	0.80	0.46	0.48	0.41
Colombia	0.71	0.65	0.80	0.38
Croatia	0.68	0.61	0.62	0.40
Czech Republic	0.94	0.79	0.30	0.34
France	0.80	0.71	0.49	0.45
Germany	0.57	0.52	0.61	0.49
Hungary	0.63	0.64	0.46	0.40
Ireland	0.66	0.58	0.66	0.46
Israel	0.84	0.87	0.30	0.19
Italy	0.71	-0.29	-0.01	0.18
Japan	0.88	0.82	0.70	0.46
Lithuania	0.86	0.86	0.41	0.49
Malaysia	0.57	0.56	0.71	0.36
Mexico	0.73	0.72	0.74	0.40
New Zealand	0.76	0.68	0.44	0.33
North Macedonia	0.60	0.42	0.63	0.42
Peru	0.77	0.72	0.71	0.33
Poland	0.59	0.58	0.68	0.32
Portugal	0.87	0.71	0.38	0.33
Slovakia	0.60	0.58	0.58	0.27
South Africa	0.72	0.69	0.58	0.29
South Korea	0.74	0.72	0.68	0.50
Spain	0.75	0.72	0.65	0.42
Switzerland	0.90	0.89	0.38	0.31
Taiwan	0.72	0.46	0.61	0.43
Turkey	0.62	0.53	0.56	0.19
United Kingdom	0.68	0.60	0.63	0.41
United States of America	0.65	0.62	0.62	0.40
Other	0.57	0.57	0.65	0.40
Chile	0.75	0.69	0.69	0.42

Supplementary Table 8

standardized factor loadings of the P4 by gender

Gender	P4-1	P4-2	P4-3	P4-4
Man	0.75	0.64	0.50	0.38
Woman	0.70	0.63	0.56	0.36
Gender-diverse individuals	0.60	0.58	0.66	0.42

Supplementary Table 9

Standardized factor loadings of the P4 screener by sexual orientation

Sexual orientation	P4-1	P4-2	P4-3	P4-4
Heterosexual	0.74	0.61	0.48	0.35
Gay or lesbian	0.71	0.65	0.60	0.43
Bisexual	0.66	0.62	0.62	0.38
Queer and pansexual	0.65	0.60	0.61	0.41
Homo- and heteroflexible identities	0.68	0.65	0.53	0.37
Asexual	0.62	0.58	0.69	0.41
Questioning	0.59	0.55	0.70	0.38
Other	0.66	0.72	0.61	0.28