Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Development and psychometric validation of a culturally contextualized bullying assessment scale for Indonesian senior high school students

Asep Solikin

ieserjournal@gmail.com

Universitas Muhammadiyah Palangkaraya

Mohammad Fatchurahman

Universitas Muhammadiyah Palangkaraya

Chandra Anugrah Putra

Universitas Muhammadiyah Palangkaraya

Muhammad Andi SETIAWAN

Universitas Muhammadiyah Sampit

Article

Keywords: BullyTRACE, bullying assessment, psychometric validation, cultural contextualization, senior high school students, multidimensional bullying, Indonesia

Posted Date: October 27th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10676861/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Bullying among adolescents remains a significant educational and psychosocial concern, yet existing assessment instruments may not adequately reflect culturally contextualized expressions of bullying in Indonesian school settings. This study aimed to develop and psychometrically validate BullyTRACE, a culturally contextualized bullying assessment scale for Indonesian senior high school students. Item development was informed by three focus group discussions involving school counselors, educational psychologists, and senior teachers in Central Kalimantan, followed by expert content validation and empirical item analysis. The initial 60-item pool was evaluated by seven experts and reduced to 48 items based on content-validity evidence, followed by item analysis that yielded a final 44-item scale. Psychometric validation was conducted with 1,250 senior high school students, with independent subsamples of 625 participants used for exploratory factor analysis (EFA) and 625 for confirmatory factor analysis (CFA). The final scale comprised four correlated dimensions: Physical, Verbal, Relational, and Cyber Bullying. Parallel analysis and EFA supported the four-factor structure, which was strongly confirmed by CFA (CFI = .994, RMSEA = .014, SRMR = .018), substantially outperforming a one-factor model. Standardized factor loadings ranged from .561 to .866. Discriminant validity was supported by HTMT values of .445–.545, while external validity was evidenced by associations with EBIP-Q total scores (r = .819), K10 psychological distress (r = .618), and disciplinary indicators (r = .268–.330). Internal consistency was high (Cronbach's α = .886–.958; McDonald's ω = .907–.962), and test–retest reliability was excellent for the total scale (ICC[2,1] = .981, 95% CI [.978, .982]). Physical and Cyber Bullying showed AVE values below .50, indicating priorities for future refinement. Overall, BullyTRACE provides a culturally contextualized, multidimensional, and psychometrically supported measure of bullying involvement for Indonesian senior high school students and a standardized measurement foundation for future longitudinal and computational research.

1. Introduction

Bullying in educational settings remains a persistent social and educational problem, with substantial implications for students' psychosocial well-being, mental health, and educational experiences (Halliday et al., 2021; Polanin et al., 2021; Samara et al., 2021). In Indonesia, the issue is particularly concerning, with recent evidence indicating that a substantial proportion of students experience different forms of bullying (Borualogo & Casas, 2022; Juli et al., 2025). This presents an important challenge for national educational policies and school-safety initiatives, including the *Sekolah Aman* (Safe School) program, *Profil Pelajar Pancasila* (Pancasila Student Profile), and *Merdeka Belajar* framework, which emphasize safe and supportive educational environments (Rachman et al., 2024).

Despite these commitments, bullying identification and intervention in schools often remain predominantly reactive, relying on manual reporting, teacher observations, and counseling interventions initiated after incidents become visible or are reported (Blakeslee et al., 2023; Kaufman et al., 2020; Lechner et al., 2023; Messman et al., 2024). Such approaches may overlook less visible forms of bullying because students may be reluctant to report their experiences and teachers cannot continuously observe peer interactions across school and digital environments (Iurino et al., 2025; Wójcik & Rzeńca, 2021). Systematic assessment is therefore important for identifying bullying involvement before it develops into more serious educational and psychosocial problems (Dailey et al., 2025; Giuliani et al., 2025).

However, the effectiveness of systematic assessment depends on the quality and contextual relevance of the measurement instruments used. Existing bullying scales provide important foundations for assessing victimization, perpetration, and related behaviors, but several limitations remain. First, many instruments were developed in Western educational and cultural contexts and may not fully capture culturally specific expressions of verbal and

relational bullying in Indonesian peer environments (Zaneva et al., 2023). Bullying behaviors may be embedded in everyday peer interactions, language practices, social hierarchies, and forms of exclusion that are not adequately represented when instruments are transferred across settings without systematic contextualization (Fine et al., 2023; Hu et al., 2025; Nurhalimah et al., 2025). Cross-cultural measurement research therefore emphasizes contextual adaptation to maintain conceptual and measurement relevance across populations (Jukes et al., 2024).

Second, existing instruments vary in the extent of their psychometric validation. Although some have established factor structures and reliability evidence, comprehensive validation addressing content validity, factorial validity, convergent and discriminant validity, criterion-related validity, and temporal stability remains less common, particularly for bullying assessment among Indonesian senior high school students (Álvarez-García et al., 2021; Del Rey et al., 2015; Magaz et al., 2016; Useche et al., 2021). Reliability alone does not demonstrate that an instrument measures the intended construct consistently or is supported by multiple complementary sources of validity evidence.

Third, the increasing use of data-driven approaches in education creates additional requirements for assessment data. Computational applications require standardized item-level data that can be consistently coded, scored, and incorporated into reproducible analytical workflows (Koutsoumpis, 2025; Orrù et al., 2023). However, bullying assessment studies have generally emphasized psychometric measurement, with less attention to explicitly documenting item-level coding, missing-data procedures, variable structures, and machine-readable scoring formats for subsequent computational applications (Chang et al., 2025; Graves et al., 2025; Xie et al., 2023). Technology-oriented bullying research has also frequently focused on post-event detection through text mining and social-media analysis rather than establishing validated individual-level assessment data for subsequent school-based predictive modelling (Batani et al., 2022; Hasan et al., 2023).

These limitations indicate a need for a bullying assessment instrument that combines cultural contextualization, comprehensive psychometric validation, and standardized data structuring. In the Indonesian context, such an instrument can provide a stronger basis for characterizing bullying involvement among senior high school students while generating empirically defensible data for subsequent computational research. Importantly, the present study does not develop or validate a predictive early warning model; it establishes the measurement and data-structuring foundation required for subsequent computational modelling and early-warning applications.

Accordingly, this study aimed to develop and psychometrically validate a culturally contextualized bullying assessment scale for senior high school students in Central Kalimantan, Indonesia. The study makes three contributions. First, it operationalizes bullying involvement across four dimensions—physical, verbal, relational, and cyber bullying—using culturally grounded indicators informed by qualitative inquiry. Second, it establishes comprehensive psychometric evidence through expert content validation, item analysis, exploratory and confirmatory factor analysis, convergent and discriminant validity, criterion-related and external validity, test–retest reliability, and Bland–Altman agreement analysis. Third, it establishes standardized item-level coding and scoring procedures to provide a structured measurement foundation for subsequent computational and data-driven early-warning research.

2. Literature Review

2.1. Bullying Involvement as a Multidimensional and Context-Dependent Construct

Bullying involvement is multidimensional, encompassing harmful behaviors expressed physically, verbally, relationally, and through digital environments. These forms are conceptually distinguishable but can co-occur within the same student's experiences or behaviors (Bork-Hüffer et al., 2021; Johansson & Englund, 2021). Assessment should therefore represent their shared domain without assuming that the dimensions are independent.

Cultural context further shapes observable bullying behavior. Peer relationships, communication practices, social norms, and local interpretations of acceptable behavior can influence how verbal and relational bullying are expressed or recognized (Grothaus, 2024). Indonesian contexts may include indirect ridicule, labeling, exclusion, and peer-group practices that are not fully captured when measures are transferred without systematic contextualization (Fine et al., 2023; Hu et al., 2025; Jukes et al., 2024; Nurhalimah et al., 2025). BullyTRACE consequently conceptualizes bullying involvement as four correlated but distinguishable dimensions within the Indonesian senior high school context

2.2. Existing Bullying Assessment Instruments and Psychometric Gaps

Several instruments assess bullying victimization, perpetration, cyberbullying, or bystander responses. The CAI assesses multiple forms of bullying victimization, whereas the CBV focuses on cybervictimization (Álvarez-García et al., 2015; Magaz et al., 2016). The EBIP-Q and ECIP-Q assess bullying and cyberbullying involvement, respectively, while other measures address specific roles such as bystander responses (Álvarez-García et al., 2021; Del Rey et al., 2015; Ortega Ruiz et al., 2016). These instruments provide important foundations but differ in construct definition, dimensionality, population, cultural context, and psychometric evidence.

Comprehensive validation requires evidence of content and factorial validity, convergent and discriminant validity, criterion-related or external validity, and reliability. Cross-cultural use additionally requires evidence that item content is conceptually and contextually appropriate for the target population. In Indonesia, this issue is particularly relevant because established measures developed elsewhere may not capture locally specific expressions of verbal, relational, and digitally mediated bullying. Moreover, the Indonesian NAQ-R was validated for workplace bullying among employees rather than school-age populations (Erwandi et al., 2021). Thus, a contextually developed measure for Indonesian senior high school students remains needed.

BullyTRACE addresses this gap through a culturally contextualized, multidimensional assessment of physical, verbal, relational, and cyber bullying, supported by expert content validation, split-sample EFA and CFA, multiple validity evidence, internal consistency, and test–retest reliability. Table 1 summarizes selected instruments and BullyTRACE to clarify differences in construct, population, dimensionality, and psychometric evidence.

Table 1
Comparison of Selected Bullying Assessment Instruments and the Present Scale

Instrument	Context / Population	Construct / Dimensions	EFA	CFA	Cultural Adaptation
CAI (Magaz et al., 2016)	Spain / Adolescents	Victimization / Physical, verbal, social, psychological, cyber	Yes	Yes	No
CBV (Álvarez-García et al., 2015)	Spain / Students aged 11–19	Cybervictimization / Cyberbullying	Yes	Yes	No
EBIP-Q (Ortega-Ruiz et al., 2016)	Spain / international / Students aged 12–19	Bullying involvement / Victimization, perpetration	Yes	Yes	Yes
ECIP-Q (Del Rey et al., 2015)	Spain / international / Students aged 12–19	Cyber involvement / Cyberaggression, cybervictimization	Yes	Yes	Yes
CIE-A (Useche et al., 2021)	Spain / Secondary students	School bullying / Victimization, symptomatology, intimidation	Yes	Yes	Yes
SBBS (Álvarez-García et al., 2021)	Spain / Primary students	Bystander responses / Defender, passive, pro-bully	Yes	Yes	Yes
NAQ-R (Erwandi et al., 2021)	Indonesia / Employees	Workplace bullying / Person-related, work-related, intimidation	Yes	Yes	Yes
BullyTRACE	Indonesia / Senior high school students (*N* = 1,250)	Bullying involvement / Physical, verbal, relational, cyber	Yes	Yes	Yes

Note. Cultural adaptation refers to explicit contextual adaptation of item content for the target population.

2.3. From Psychometric Measurement to Computationally Structured Assessment

Data-driven educational research requires assessment data that are reliable, valid, consistently coded, and suitable for reproducible analysis (Koutsoumpis, 2025). Although computational bullying research has increasingly focused on automated detection of harmful online content (Batani et al., 2022; Hasan et al., 2023), such approaches differ from individual-level assessment of bullying involvement. BullyTRACE therefore treats computational readiness as a measurement and data-structuring consideration rather than a predictive outcome. Its standardized item identification, response coding, and scoring structure may support future computational analyses, but the scale is not an early-warning or predictive system. Such applications would require independently defined outcomes, model development, validation, and evaluation of predictive performance.

3. Methods

3.1. Research Design

This study employed a Research and Development (R&D) design to develop and psychometrically validate BullyTRACE, a culturally contextualized bullying assessment scale for Indonesian senior high school students. The sequential process comprised construct definition and operationalization, culturally contextualized item generation, expert content validation, empirical item analysis and reduction, exploratory factor analysis (EFA), confirmatory

factor analysis (CFA), construct, external and criterion-related validity assessment, and reliability evaluation. These stages were intended to establish evidence for the content, structural, construct, external, and temporal properties of the final scale. The study was a psychometric instrument-development and validation study; although the item-level dataset was standardized for reproducible statistical and potential future computational applications, no machine-learning or predictive early-warning model was developed or evaluated.

3.2. Construct Definition and Operationalization

The target construct was defined as bullying involvement across four theoretically and behaviorally distinguishable dimensions: Physical Bullying (PB), Verbal Bullying (VB), Relational Bullying (RB), and Cyber Bullying (CB). PB refers to direct physical aggression or harm; VB to insulting, ridiculing, threatening, or derogatory verbal behaviors; RB to behaviors intended to damage peer relationships, social standing, or group inclusion; and CB to bullying-related behaviors in digitally mediated environments. The four dimensions were conceptualized as correlated but distinguishable components of bullying involvement. The final instrument provides both dimension-specific and overall BullyTRACE scores.

3.3. Item Generation and Cultural Contextualization

Item development combined a literature-informed and context-sensitive procedure. Bullying literature was reviewed to identify indicators of physical, verbal, relational, and cyber bullying, followed by three focus group discussions (FGDs) with 18 participants from Central Kalimantan: six school counselors, six educational psychologists, and six senior teachers. The FGDs explored locally relevant bullying expressions, behaviors, interpersonal contexts, and manifestations among senior high school students. Two independent coders conducted thematic analysis, resolving discrepancies through discussion and consensus. Candidate items were retained based on construct relevance, comprehensibility, contextual appropriateness, and minimal conceptual overlap. This process produced 60 items, with 15 items per dimension. All items used a five-point Likert-type response scale (1–5), with higher scores indicating greater endorsement of bullying-related behaviors. The 60-item pool was then subjected to expert content validation.

3.4. Expert Content Validation

The 60-item pool was evaluated by seven experts with expertise in educational psychology, psychometrics, and data architecture. Experts assessed item relevance, representativeness, cultural appropriateness, and clarity. Content validity was evaluated using Aiken's V and the item-level content validity index (I-CVI). An Aiken's V ≥ .80 was the primary retention criterion, while I-CVI was calculated as the proportion of experts assigning a relevance rating of 3 or 4 on a four-point scale. Quantitative results were considered alongside qualitative expert comments. Items failing the predefined criterion or requiring substantive modification were excluded or revised before empirical administration, yielding the preliminary pool for subsequent item analysis.

3.5. Participants, Sampling, and Data Collection

Participants were senior high school students enrolled in schools under the jurisdiction of the Department of Education of Central Kalimantan, Indonesia. Data were collected from February to July 2025 using stratified sampling to represent school location, grade level, and gender. Of 1,420 questionnaires distributed, 1,310 were returned. After data-quality screening, 60 responses were excluded because of systematic missingness, unengaged response patterns, or insufficient enrollment duration, yielding a final sample of 1,250 students. To reduce sample-specific capitalization during factor development and validation, the sample was divided into two independent subsamples: 625 for EFA and 625 for CFA. Participants were from Grades 10–12 and included both male and female students.

3.6. Ethical Considerations

The study received ethical approval from the Institutional Review Board of the Faculty of Education and Social Sciences (Approval No. xxx). Written informed consent was obtained from participants' legal guardians, together with student assent. Participation was voluntary, and anonymity and confidentiality were maintained throughout data collection, data management, and analysis. Participants were informed that individual responses would not be disclosed to school authorities.

3.7. Statistical Analysis

Item performance was examined using descriptive statistics, response distributions, floor and ceiling effects, and corrected item–total correlations (CITCs). A CITC ≥ .30 was used as the minimum retention criterion, with CITCs recalculated after item deletion. Exploratory factor analysis (EFA) was conducted in the first independent subsample (n = 625). Factorability was assessed using the Kaiser–Meyer–Olkin (KMO) measure and Bartlett's test of sphericity. Principal Axis Factoring with Promax rotation was used because the dimensions were expected to correlate. Factor retention was based primarily on parallel analysis, supported by scree structure and substantive interpretability. Factor loadings, cross-loadings, and communalities were examined for item retention. Confirmatory factor analysis (CFA) was then conducted in the independent subsample (n = 625) to cross-validate the EFA structure. A one-factor model was compared with a correlated four-factor model representing Physical, Verbal, Relational, and Cyber Bullying. Model fit was evaluated using χ^2, CFI, RMSEA, and SRMR, with standardized factor loadings examined for item–factor relationships.

Convergent validity was assessed using standardized factor loadings, Average Variance Extracted (AVE), and Composite Reliability (CR), with AVE ≥ .50 and CR ≥ .70 used as conventional benchmarks. Discriminant validity was assessed using the heterotrait–monotrait ratio (HTMT) and latent factor correlations; HTMT < .85 was interpreted as evidence of empirical distinction. These indices were interpreted jointly rather than in isolation. Internal consistency was evaluated using Cronbach's alpha (α), McDonald's omega (ω), and CR for the four dimensions and, where appropriate, the total scale. External and criterion-related validity was examined using the European Bullying Intervention Project Questionnaire (EBIP-Q), Kessler Psychological Distress Scale (K10), and school disciplinary indicators. Pearson correlations were used primarily for EBIP-Q and K10, with Spearman correlations as robustness checks. Point-biserial correlation was used for the binary disciplinary-record indicator and Spearman correlations for incident number and severity. Effect magnitude, direction, statistical significance, and confidence intervals were considered in interpretation.

Known-groups validity was examined across gender and grade groups using Welch's independent-samples t-test and one-way ANOVA, respectively, with effect sizes reported alongside significance tests. Temporal stability was assessed across two measurement occasions. Pearson correlation assessed relative stability, while ICC(2,1), a two-way random-effects, absolute-agreement, single-measure coefficient, was the primary reliability estimate. Paired-samples t-tests and Cohen's dz examined systematic mean change. Bland–Altman analysis assessed absolute agreement using the mean difference and 95% limits of agreement, together with the standard error of measurement (SEM) and minimum detectable change at the 95% confidence level (MDC95). These estimates were interpreted jointly to distinguish temporal reliability, systematic change, and measurement error.

3.8. Data Coding and Computational Readiness

To facilitate reproducible downstream analysis, each retained item was assigned a standardized machine-readable identifier corresponding to its dimension and item sequence. The final dataset contained 44 item-level variables

representing Physical Bullying (12 items), Verbal Bullying (13 items), Relational Bullying (9 items), and Cyber Bullying (10 items). Responses were coded consistently across the five response categories, and the dataset was organized in a rectangular item-level structure in which each row represented one participant and each column represented one item. Dimension-specific scores and the overall BullyTRACE score were calculated from the retained items. Data-quality procedures, item identification, scoring structure, and variable organization were documented to support reproducible statistical and computational workflows. This data-structuring procedure was intended to establish a standardized measurement layer for subsequent computational research. It did not constitute the development, training, or validation of a predictive machine-learning or early-warning model.

4. Result

4.1. Data Quality and Sample Characteristics

The final validation sample comprised 1,250 Indonesian senior high school students who completed the 44-item BullyTRACE scale. The sample was equally divided into EFA (*n* = 625) and CFA (*n* = 625) subsamples. Females comprised 52.9% (*n* = 661) and males 47.1% (*n* = 589); 33.0% were in Grade 10 (*n* = 412), 35.1% in Grade 11 (*n* = 439), and 31.9% in Grade 12 (*n* = 399) (Table 2). The final scale comprised 44 items across four dimensions: PB (12), VB (13), RB (9), and CB (10), with a theoretical total-score range of 44–220. No missing responses occurred on the final items, all participant identifiers were unique, and complete data were available for EBIP-Q, K10, disciplinary indicators, and all 1,250 T1–T2 test–retest pairs (Table 2). Detailed item composition is provided in the Supplementary Material.

Table 2
Sample Characteristics and Data Quality

Characteristic	n	%
Total sample	1,250	100.0
Analysis sample		
EFA subsample	625	50.0
CFA subsample	625	50.0
Gender		
Female	661	52.9
Male	589	47.1
Grade		
Grade 10	412	33.0
Grade 11	439	35.1
Grade 12	399	31.9
Data quality		
Complete responses on 44 items	1,250	100.0
Missing responses on 44 items	0	0.0
Unique participant IDs	1,250	100.0
External-validity data		
EBIP-Q, K10, and disciplinary data	1,250	100.0
Test–retest data		
Complete T1–T2 pairs	1,250	100.0

4.2. Item Development and Content Validity

The initial pool comprised 60 items, with 15 items per dimension. Seven experts evaluated item relevance and representativeness using a four-point relevance scale. As shown in Table 3, mean Aiken's V was .805 (range = .095–1.000), resulting in 48 retained and 12 removed items. Retained items showed substantially higher mean Aiken's V (.920 vs. .345) and mean I-CVI (1.000 vs. .262) than removed items. Across the full pool, mean I-CVI was .852 and S-CVI/UA was .800; all 48 retained items achieved I-CVI = 1.000. Thus, expert evaluation reduced the pool from 60 to 48 items before empirical item analysis. Item-level results are provided in Supplementary Table S1.

Table 3
Summary of expert judgment and content validity

Indicator	Overall (60 items)	Retained (48 items)	Removed (12 items)
Number of items	60	48	12
Mean Aiken's V	0.805	0.920	0.345
Minimum Aiken's V	0.095	0.762	0.095
Maximum Aiken's V	1.000	1.000	0.571
Mean I-CVI	0.852	1.000	0.262
Minimum I-CVI	0.000	1.000	0.000
Maximum I-CVI	1.000	1.000	0.714
S-CVI/UA	0.800	1.000	—

Note. Aiken's V was calculated from ratings provided by seven experts on a four-point relevance scale. I-CVI was calculated as the proportion of experts assigning a rating of 3 or 4 to each item. S-CVI/UA represents the proportion of items achieving universal expert agreement.

4.3. Item Analysis and Item Reduction

The 48 items retained after expert evaluation were subjected to empirical item analysis. The final 44-item BullyTRACE had a total mean of 94.984 (SD = 25.334; range = 45–202), with item means ranging from 1.686 to 2.673 and item SDs from 0.824 to 1.043 (Table 4). At the dimension level, Physical Bullying had the lowest mean (M = 22.487, SD = 7.777), whereas Verbal Bullying had the highest (M = 31.416, SD = 10.680). Response distributions were generally positively skewed, particularly for Physical Bullying. A floor effect of at least 15% was observed for 42 of the 44 items, with the largest effects for PB01 (50.32%), PB08 (48.56%), PB10 (45.04%), PB03 (44.96%), and PB05 (43.76%). No item showed a ceiling effect of ≥ 15%. The observed floor effects were retained because low endorsement of specific bullying behaviors was considered substantively plausible in the school-based sample. Complete item-level statistics and response distributions are provided in Supplementary Table S2. At the 48-item screening stage, four items had CITCs below .30 and were removed: RB04 (.100), RB09 (.140), CB05 (.114), and CB11 (.130). The remaining 44 items initially showed CITCs of .358–.633. After deletion and recalculation, final CITCs ranged from .442 to .728, with all items exceeding the .30 criterion. The final scale therefore comprised 12 PB, 13 VB, 9 RB, and 10 CB items and was used for subsequent factor and reliability analyses.

Table 4
Descriptive Statistics of the Final BullyTRACE Scale

Dimension	Items	Mean	SD	Minimum	Maximum	Skewness	Kurtosis
Physical Bullying	12	22.487	7.777	12	57	1.021	1.196
Verbal Bullying	13	31.416	10.680	13	64	0.538	−0.056
Relational Bullying	9	19.226	6.641	9	44	0.669	0.183
Cyber Bullying	10	21.855	6.666	10	48	0.629	0.524
Total	44	94.984	25.334	45	202	0.919	1.595

Note. SD = standard deviation. Higher scores indicate greater endorsement of bullying-related behaviors. Complete item-level statistics, response distributions, and CITCs are provided in Supplementary Table S2.

4.4. Internal Consistency Reliability

The final 44-item BullyTRACE demonstrated high internal consistency across all four dimensions and the overall scale (Table 5). Cronbach's α ranged from .886 to .958 and McDonald's ω from .907 to .962 across dimensions. For PB, VB, RB, and CB, α/ω coefficients were .919/.931, .958/.962, .913/.929, and .886/.907, respectively. The total scale showed α = .959 and ω = .962. These estimates support strong internal consistency and were interpreted alongside the factor-analytic and validity evidence rather than as evidence that the four dimensions are interchangeable.

Table 5
Internal consistency reliability of the BullyTRACE scale

Scale	No. of items	Cronbach's α	McDonald's ω
Physical Bullying	12	.919	.931
Verbal Bullying	13	.958	.962
Relational Bullying	9	.913	.929
Cyber Bullying	10	.886	.907
Total BullyTRACE	44	.959	.962

Note. α = Cronbach's alpha; ω = McDonald's omega.

4.5. Structural Validity

Structural validity was examined using EFA and CFA in independent subsamples of 625 participants each. Factorability was first assessed using the KMO measure and Bartlett's test of sphericity. The KMO was .977, and Bartlett's test was significant, $\chi^2(946) = 16{,}661.12$, $p < .001$, supporting the suitability of the correlation matrix for factor analysis (Table 6).

Table 6
Assessment of Factorability of the BullyTRACE Item Correlation Matrix

Test	Statistic	df	p
Kaiser–Meyer–Olkin (KMO)	**.977**	—	—
Bartlett's test of sphericity	**16,661.12**	**946**	**< .001**

Note. EFA sample = 625. KMO = Kaiser–Meyer–Olkin measure of sampling adequacy.

4.5.1. Factor Retention

Parallel analysis supported a four-factor solution. The first four observed eigenvalues (17.159, 3.250, 2.965, and 2.260) exceeded their corresponding 95th-percentile random-data eigenvalues (1.599, 1.528, 1.482, and 1.442), whereas the fifth (.752 vs. 1.404) and sixth (.727 vs. 1.372) did not. The retained factors corresponded to Physical, Verbal, Relational, and Cyber Bullying (Table 7).

Table 7
Parallel analysis for factor retention

Factor	Observed eigenvalue	Random-data 95th percentile	Decision
1	**17.159**	1.599	Retain
2	**3.250**	1.528	Retain
3	**2.965**	1.482	Retain
4	**2.260**	1.442	Retain
5	0.752	1.404	Reject
6	0.727	1.372	Reject

Note. Factors were retained when the observed eigenvalue exceeded the corresponding 95th-percentile eigenvalue generated from random data.

4.5.2. Exploratory Factor Analysis

The four-factor EFA produced an interpretable structure in which items primarily loaded on their hypothesized Physical, Verbal, Relational, and Cyber Bullying factors. Primary loadings ranged from .439 to .838, and communalities from .284 to .769. Most items showed clear primary loadings with substantially smaller secondary loadings. PB09, PB11, RB08, and RB10 had comparatively higher secondary loadings but retained dominant primary loadings and were therefore retained. Complete loadings, cross-loadings, and communalities are provided in Supplementary Table S3. The EFA findings supported the hypothesized four-dimensional structure, which was subsequently cross-validated through CFA in the independent subsample.

4.6. Confirmatory Factor Analysis

CFA was conducted in the independent subsample (n = 625) to compare the hypothesized correlated four-factor model with a competing one-factor model. The one-factor model showed poor fit, $\chi^2(902) = 6485.27$, CFI = .636, RMSEA = .100, and SRMR = .115, whereas the four-factor model showed substantially better fit, $\chi^2(764) = 853.17$, CFI = .994, RMSEA = .014, and SRMR = .018 (Table 8). The findings therefore supported the correlated four-factor structure over a unidimensional model.

Table 8
Comparison of Competing Factor Structures of the BullyTRACE Scale

Model	χ^2	df	CFI	RMSEA	SRMR
1-factor	6485.27	902	.636	.100	.115
4-factor	**853.17**	**764**	**.994**	**.014**	**.018**

Note. CFA sample = 625. CFI = comparative fit index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual.

Standardized loadings in the four-factor model ranged from .600 to .735 for PB, .708 to .866 for VB, .648 to .809 for RB, and .561 to .726 for CB (Table 9). Thus, all 44 items loaded substantially on their hypothesized dimensions. Complete item-level estimates are provided in Supplementary Table S4.

Table 9
Standardized Factor Loadings of the Four-Factor CFA Model

Dimension	No. of items	Standardized loading
Physical Bullying	12	.600–.735
Verbal Bullying	13	.708–.866
Relational Bullying	9	.648–.809
Cyber Bullying	10	.561–.726
Overall	**44**	**.561–.866**

Note. Complete item-level standardized factor loadings are reported in Supplementary Table S4.

The CFA results replicated the four-factor structure identified by EFA in the independent subsample, providing internal cross-validation of the proposed measurement structure.

4.7. Convergent Validity

Convergent validity was assessed using standardized loadings, Average Variance Extracted (AVE), and Composite Reliability (CR) (Table 10). Loadings ranged from .600 to .866, AVE from .424 to .626, and CR from .879 to .956. VB (AVE = .626; CR = .956) and RB (AVE = .517; CR = .906) met both conventional criteria, whereas PB (AVE = .477; CR = .916) and CB (AVE = .424; CR = .879) met the CR criterion but not the AVE ≥ .50 criterion. Thus, convergent-validity evidence was stronger for VB and RB and qualified for PB and CB.

Table 10
Convergent Validity of the Four-Factor BullyTRACE Measurement Model

Factor	No. of items	Standardized loading range	AVE	CR	AVE ≥ .50	CR ≥ .70
Physical Bullying	12	.600–.735	.477	.916	No	Yes
Verbal Bullying	13	.708–.866	.626	.956	Yes	Yes
Relational Bullying	9	.648–.809	.517	.906	Yes	Yes
Cyber Bullying	10	.561–.726	.424	.879	No	Yes

Note. AVE = average variance extracted; CR = composite reliability.

4.8. Discriminant Validity

Discriminant validity was assessed using HTMT and correlations among the four latent factors. HTMT values ranged from .445 to .545, with the highest value between Physical and Verbal Bullying (.545); all six values were below the .85 criterion, supporting empirical distinctiveness among the four dimensions (Table 11).

Table 11
Heterotrait–Monotrait Ratios for Discriminant Validity

Factor pair	**HTMT**
Physical–Verbal	.545
Physical–Relational	.532
Physical–Cyber	.445
Verbal–Relational	.534
Verbal–Cyber	.478
Relational–Cyber	.480
Maximum HTMT	**.545**

Note. HTMT = heterotrait–monotrait ratio of correlations. All coefficients were below .85.

Latent factor correlations ranged from .419 to .531, with the strongest association between Physical and Verbal Bullying (r = .531) and the weakest between Physical and Cyber Bullying (r = .419). Together, the HTMT and latent correlations indicate that the four dimensions share meaningful variance while remaining empirically distinguishable. The complete correlation matrix is provided in Supplementary Table S6.

4.9. External and Criterion-Related Validity

External and criterion-related validity was examined in the full sample (N = 1,250) using EBIP-Q, K10 psychological distress, and school disciplinary indicators (Table 12). Associations were strongest with the closely related EBIP-Q, followed by K10 and the more distal disciplinary indicators.

Table 12
External and Criterion-Related Validity of BullyTRACE

External criterion	Method	Correlation	95% CI	*p*
EBIP-Q Victimization	Pearson *r*	.731	[.704, .755]	< .001
EBIP-Q Aggression	Pearson *r*	.687	[.657, .715]	< .001
EBIP-Q Total	Pearson *r*	.819	[.800, .836]	< .001
K10 Total	Pearson *r*	.618	[.583, .651]	< .001
Disciplinary—Any record	Point-biserial *r*	.330	[.280, .379]	< .001
Disciplinary—Number of incidents	Spearman ρ	.268	—	< .001
Disciplinary—Severity	Spearman ρ	.270	—	< .001

Note. N = 1,250. EBIP-Q = European Bullying Intervention Project Questionnaire; K10 = Kessler Psychological Distress Scale. Pearson correlations were used as the primary estimates where appropriate. Spearman correlations were used for ordinal/count-based disciplinary indicators and were also available as robustness checks for the EBIP-Q and K10 associations.

The strongest association was between BullyTRACE and EBIP-Q total scores (r = .819, 95% CI [.800, .836], p < .001), followed by EBIP-Q victimization (r = .731) and aggression (r = .687). Corresponding Spearman correlations (ρ = .778, .688, and .642) supported the robustness of these associations. Because EBIP-Q assesses closely related bullying constructs, the results provide strong related-construct evidence while remaining below unity, indicating that the measures are not interchangeable.

BullyTRACE was also positively associated with K10 psychological distress (r = .618, 95% CI [.583, .651], p < .001; Spearman ρ = .563, p < .001), providing nomological evidence with a related but distinct psychosocial construct. This association does not indicate that BullyTRACE measures psychological distress or establish causality. Associations with disciplinary indicators were smaller but consistently positive: any disciplinary record, r = .330 (95% CI [.280, .379], p < .001); number of incidents, ρ = .268 (p < .001); and severity, ρ = .270 (p < .001). Their smaller magnitude is consistent with disciplinary records being indirect and more distal indicators that may reflect behaviors beyond bullying. Overall, the graded association pattern—strongest for EBIP-Q, moderate for K10, and smaller for disciplinary indicators—supports the external validity of BullyTRACE while indicating that these criteria represent related but non-identical constructs.

4.10. Known-Groups Validity

Known-groups validity was examined across gender and grade level. Male students had a mean BullyTRACE score of 95.96 (SD = 25.65), compared with 94.11 (SD = 25.03) for females. The gender difference was not significant, Welch's t = − 1.283, p = .200, Cohen's d = − .073. Grade-level means were 95.17 (Grade 10, n = 412), 95.45 (Grade 11, n = 439), and 94.27 (Grade 12, n = 399). The one-way ANOVA was also nonsignificant, $F(2, 1247) = 0.245$, p = .783, η^2 = .0004. Thus, BullyTRACE total scores did not differ meaningfully by gender or grade in this sample. These findings represent one component of validity evidence and should be interpreted alongside the other psychometric evidence.

4.11. Test–Retest Reliability

Test–retest reliability was assessed among 1,250 participants who completed BullyTRACE at two measurement occasions. Pearson correlations assessed relative stability, while ICC(2,1) was the primary estimate of absolute agreement. Paired-samples t-tests and Cohen's dz assessed systematic mean changes (Table 13).

Table 13
Test–Retest Reliability of the BullyTRACE Scale

Scale	Items	T1 Mean	T2 Mean	Mean Difference	Pearson *r*	ICC(2,1)	95% CI for ICC	Paired *t*	Cohen's *dz*
Physical Bullying	12	22.487	22.953	0.466	.975	**.967**	[.962, .971]	−8.944***	.253
Verbal Bullying	13	31.416	31.706	0.290	.979	**.973**	[.970, .976]	−4.388***	.124
Relational Bullying	9	19.226	19.490	0.264	.971	**.964**	[.960, .968]	−5.589***	.158
Cyber Bullying	10	21.855	22.139	0.284	.965	**.958**	[.954, .962]	−5.536***	.157
Total BullyTRACE	**44**	**94.984**	**96.287**	**1.303**	**.989**	**.981**	**[.978, .982]**	**−10.189***	**.288**

Note. N = 1,250. ICC(2,1) = two-way random-effects, absolute-agreement, single-measure intraclass correlation coefficient. Mean difference = T2 − T1. **p < .001. Cohen's dz = standardized paired mean difference.

BullyTRACE demonstrated high temporal reliability across dimensions, with ICC(2,1) values of .958–.973. For the total scale, Pearson *r* = .989 and ICC(2,1) = .981 (95% CI [.978, .982]). Despite this high stability, scores increased modestly at retest: the total mean increased by 1.303 points, $t(1249) = -10.189$, $p < .001$, $dz = .288$. Subscale increases were similarly small ($dz^* = .124–.253$). Thus, the results indicate high temporal reliability alongside a modest systematic increase in mean scores across occasions.

4.12. Bland–Altman Agreement Analysis

Bland–Altman analysis complemented the ICC by examining individual-level agreement between T1 and T2 total scores. The mean difference was 1.303 points (SD = 4.522), with 95% limits of agreement from − 7.560 to 10.167. The positive difference indicated a small systematic increase at retest. SEM was 3.327 points and MDC95 was 9.221 points, indicating that individual changes of approximately 9.22 points or less should be interpreted cautiously because they may fall within expected measurement variability. Overall, the findings indicate high temporal reliability with a small systematic shift in retest scores.

5. Discussion

This study developed and psychometrically validated a 44-item bullying assessment scale for Indonesian senior high school students. Evidence from content validation, item analysis, factor analysis, construct validity, external validity, and reliability supports bullying involvement as a multidimensional construct comprising Physical, Verbal, Relational, and Cyber Bullying. Parallel analysis and EFA supported a four-factor solution, while CFA in an independent subsample showed substantially better fit for the correlated four-factor model than for a one-factor model. These findings are consistent with research conceptualizing bullying and cyberbullying as related but distinct forms of behavior (Bork-Hüffer et al., 2021; Del Rey et al., 2015; Johansson & Englund, 2021; Magaz et al., 2016). The four dimensions showed adequate empirical distinctiveness, with latent factor correlations of .419–.531 and HTMT values of .445–.545. The item-development process reduced the initial pool from 60 to 48 items through expert evaluation and to 44 items after empirical analysis, with final CITCs ranging from .442 to .728. Standardized loadings ranged from .561 to .866. Although VB and RB met the AVE criterion (.626 and .517), PB and CB showed lower AVE values (.477 and .424); nevertheless, composite reliability remained high (.879–.956), as did internal consistency (α

= .886–.958; ω = .907–.962). Thus, the lower AVE values for PB and CB indicate priorities for further refinement rather than sufficient grounds for rejecting these dimensions.

External associations further supported construct coherence. BullyTRACE total scores correlated most strongly with EBIP-Q total scores (r = .819, p < .001), followed by K10 psychological distress (r = .618, p < .001) and disciplinary indicators (r = .268–.330, all p < .001). This graded pattern is theoretically plausible because EBIP-Q assesses closely related bullying involvement, whereas psychological distress and disciplinary records represent more distal constructs (Ariani et al., 2025; Hysing et al., 2021). Overall, these findings support BullyTRACE as a multidimensional, culturally contextualized measure of bullying involvement, while highlighting the need for further validation across regions and in prospective applications.

A central contribution of BullyTRACE is its cultural contextualization rather than direct translation or adaptation of an existing instrument. Three focus group discussions with school counselors, educational psychologists, and senior teachers in Central Kalimantan identified locally relevant behavioral expressions, interpersonal contexts, and manifestations of bullying among senior high school students. These findings informed the initial items, which were subsequently evaluated by multidisciplinary experts and empirically tested. This approach reflects the principle that valid measurement requires conceptual fidelity alongside linguistic and sociocultural appropriateness (Fine et al., 2023; Jukes et al., 2024). Cultural contextualization is particularly relevant because bullying may be expressed differently across language, peer relationships, social norms, and communication practices. Verbal and relational bullying may involve ridicule, labeling, exclusion, or indirect behaviors, while cyberbullying may reflect locally familiar digital practices. Contextualization therefore does not redefine bullying as culture-specific but operationalizes established dimensions through behaviors that are understandable and meaningful to the target population. Compared with previous Indonesian research on adapted bullying measures (Bowes et al., 2019), BullyTRACE incorporated contextual inquiry before large-scale validation. The development sequence—60 initial items, 48 retained after seven-expert evaluation, and 44 after empirical item analysis—demonstrates the convergence of contextual information, expert judgment, and empirical evidence.

Importantly, BullyTRACE did not propose a uniquely Indonesian taxonomy; it retained four related but distinguishable dimensions—Physical, Verbal, Relational, and Cyber Bullying—which were supported by the factor-analytic findings. Its contribution is therefore the contextualized operationalization and psychometric evaluation of established bullying dimensions for Indonesian senior high school students. However, because both qualitative contributors and participants were drawn from Central Kalimantan, the findings should not be generalized automatically to other Indonesian regions. Cross-regional replication is needed to determine whether additional contextual expressions require item refinement. Thus, cultural contextualization and psychometric validation are complementary: the former supports meaningful item development, while the latter establishes measurement quality.

Construct-validity evidence supported the quality and distinctiveness of the four BullyTRACE dimensions. Convergent validity was assessed using standardized factor loadings, AVE, and CR, whereas discriminant validity was examined using HTMT and latent factor correlations. Convergent validity was supported to different degrees across dimensions. Standardized loadings ranged from .561 to .866. Verbal and Relational Bullying met the AVE criterion (.626 and .517), whereas Physical and Cyber Bullying yielded lower values (.477 and .424), despite high CR across all dimensions (.879–.956). The lower AVE values may reflect the behavioral breadth of physical and cyber bullying and identify these dimensions for future refinement without requiring automatic item removal. Discriminant validity was consistently supported, with HTMT values of .445–.545 and latent factor correlations of .419–.531, all indicating related but empirically distinguishable dimensions. Thus, the findings support retaining the four-factor structure rather than collapsing the dimensions into a single construct. For interpretation, dimension-specific scores can

characterize the behavioral form of bullying involvement, while the total score may represent overall involvement when theoretically appropriate. Overall, construct-validity evidence supports the multidimensional structure while identifying Physical and Cyber Bullying as priorities for further psychometric refinement.

BullyTRACE showed a theoretically coherent pattern of associations with bullying, psychosocial, and behavioral indicators. The strongest association was with EBIP-Q total scores (r = .819, 95% CI [.800, .836], p < .001), followed by victimization (r = .731) and aggression (r = .687). Corresponding Spearman correlations (ρ = .778, .688, and .642) supported the robustness of these associations. The strong but sub-unity correlations indicate substantial correspondence while preserving distinction between the two measures, which organize bullying differently: EBIP-Q emphasizes victimization and aggression/perpetration, whereas BullyTRACE assesses Physical, Verbal, Relational, and Cyber Bullying (Ortega Ruiz et al., 2016). BullyTRACE was also positively associated with K10 psychological distress (r = .618, 95% CI [.583, .651], p < .001; ρ = .563), providing nomological evidence consistent with the relationship between bullying involvement and psychosocial functioning. Because K10 assesses a distinct construct, this association should not be interpreted as evidence that BullyTRACE measures psychological distress or as evidence of causality. Associations with disciplinary indicators were smaller but consistently positive: any disciplinary record (r = .330), number of incidents (ρ = .268), and severity (ρ = .270; all p < .001). Their lower magnitude is plausible because disciplinary records are more distal indicators and may include behaviors unrelated to bullying, while bullying incidents may remain unobserved or unreported. Overall, the graded pattern—from strongest associations with the closely related EBIP-Q to smaller associations with K10 and disciplinary indicators—is consistent with the expected conceptual proximity of the criteria. These findings support the external and criterion-related validity of BullyTRACE but do not establish causality or predictive validity; longitudinal research using independently collected criteria is needed to examine temporal and predictive relationships.

The temporal reliability findings provide strong evidence of score stability across measurement occasions. For the total scale, the test–retest correlation was r = .989 and ICC(2,1) = .981 (95% CI [.978, .982]), while ICC values across the four dimensions ranged from .958 to .973. These findings indicate consistently high temporal reliability across the BullyTRACE dimensions. Importantly, high reliability did not imply identical scores across occasions. The total mean increased from 94.984 at T1 to 96.287 at T2, a difference of 1.303 points, which was statistically significant, t(1249) = − 10.189, p < .001, but small in magnitude (dz = .288). Similar small increases occurred across all dimensions. Thus, BullyTRACE showed high temporal stability alongside a modest systematic increase at retest. Bland–Altman analysis provided complementary evidence of absolute agreement. The mean T2–T1 difference was 1.303 points (SD = 4.522), with 95% limits of agreement from − 7.560 to 10.167. The SEM was 3.327 points and MDC95 was 9.221 points, indicating that individual changes of approximately 9.22 points or less should be interpreted cautiously because they may fall within expected measurement error. These estimates add practical information for repeated assessment by distinguishing measurement variability from potentially meaningful individual change. Overall, the findings support BullyTRACE as a temporally stable measure while showing that reliability does not imply complete score invariance across occasions. The modest retest increase may reflect repeated exposure to the items or contextual differences between occasions, although the present design cannot establish its cause. Future longitudinal research should examine stability across longer intervals and diverse educational contexts.

BullyTRACE provides a structured, multidimensional profile of Physical, Verbal, Relational, and Cyber Bullying that may help counselors, teachers, and student-support personnel identify less visible forms of bullying. It should complement, rather than replace, observation, reporting, interviews, and professional judgment; scores should not be used alone for diagnostic or predictive decisions or to label students as perpetrators or victims without contextual assessment. The standardized item-level structure also supports reproducible longitudinal and computational

research through consistent coding and scoring. However, the present study establishes measurement readiness rather than predictive capability: no machine-learning or prospective prediction model was developed or evaluated. Unlike computational systems that detect bullying-related digital content (Batani et al., 2022; Hasan et al., 2023), BullyTRACE assesses self-reported experiences across behavioral dimensions. Future computational applications should therefore use prospective outcomes, independent validation and external testing, with attention to calibration, error rates, subgroup performance, interpretability, privacy, and data governance. BullyTRACE is best positioned as a standardized assessment component within broader school-based prevention and support, with human oversight remaining central to educational decision making.

Several limitations should be considered. First, although the sample was substantial (*N* = 1,250), participants were drawn from Central Kalimantan, limiting generalization to other Indonesian regions. Replication across provinces is therefore needed to examine the stability of item content and psychometric properties. The reliance on student self-report also introduces potential recall, social desirability, interpretation, and disclosure biases; future studies should incorporate multimethod evidence where ethically appropriate. Second, convergent validity was weaker for PB and CB (AVE = .477 and .424) than for VB and RB (.626 and .517), despite high CR (.916 and .879). These dimensions warrant further refinement and replication while preserving content coverage. Because BullyTRACE uses five ordered response categories, future studies should also apply ordinal CFA with an estimator such as WLSMV. Formal measurement-invariance and DIF analyses across gender, grade, and regional groups are likewise needed, particularly for subgroup comparisons. Third, associations with EBIP-Q, K10, and disciplinary indicators provide external and criterion-related evidence but do not establish causality or prospective predictive validity. Longitudinal studies using independently defined behavioral, psychosocial, and educational outcomes are needed to examine temporal and predictive relationships. Similarly, test–retest analysis showed excellent reliability (ICC[2,1] = .981) but a modest systematic increase at retest, with Bland–Altman limits of agreement from − 7.56 to + 10.17. Longer retest intervals and different developmental and school contexts should be examined to determine the stability of this pattern. Finally, the standardized computational structure should not be interpreted as evidence of a predictive or early-warning system, as no machine-learning or prospective prediction model was evaluated. Future computational research should use prospective datasets, independent outcome criteria, and external validation, with safeguards for privacy, fairness, interpretability, and data governance. Overall, future validation should prioritize geographic replication, multimethod assessment, refinement of the Physical and Cyber dimensions, ordinal CFA, measurement equivalence and DIF, longitudinal validation, and prospective computational evaluation.

Conclusion

This study developed and psychometrically evaluated BullyTRACE, a culturally contextualized 44-item bullying assessment scale for Indonesian senior high school students. The final instrument comprises four correlated but distinguishable dimensions: Physical, Verbal, Relational, and Cyber Bullying. Developed through contextual inquiry, expert judgment, and empirical evaluation, the instrument was supported by independent EFA and CFA, evidence of convergent and discriminant validity, high internal consistency, strong temporal reliability, and theoretically coherent associations with external indicators. The evidence was not uniformly optimal: Physical and Cyber Bullying showed AVE values below .50, and test–retest analysis identified a small systematic increase despite high temporal reliability. These findings highlight priorities for further refinement rather than undermining the overall four-dimensional structure. BullyTRACE should therefore be regarded as a psychometrically supported assessment instrument and measurement foundation, not a predictive or early-warning system. Future research should extend validation across Indonesian regions, examine ordinal CFA and measurement invariance/DIF, refine the Physical and Cyber dimensions, and evaluate prospective applications using independently collected outcomes. Overall,

BullyTRACE provides a culturally contextualized multidimensional measure that can support structured school-based assessment and future research in educational psychology and psychometrics.

Declarations

Data availability

The data that support the findings of this study are not publicly available because they contain potentially sensitive information concerning school students and are subject to ethical and privacy restrictions. De-identified data may be available from the corresponding author upon reasonable request and subject to approval by the relevant institutional authority.

Competing interests

The authors declare that they have no competing interests.

Author contributions

Author contributions are reported according to the CRediT taxonomy: xxx: Conceptualization, Methodology, Investigation, Supervision, and Writing – Review & Editing; xxx: Conceptualization, Methodology, Validation, Supervision, and Writing – Review & Editing; xxx: Methodology, Validation, Formal Analysis, and Writing – Review & Editing; xxx: Conceptualization, Data Curation, Formal Analysis, Investigation, Writing – Original Draft, Writing – Review & Editing, and Visualization. All authors have read and approved the final manuscript and agree to be accountable for all aspects of the work.

Acknowledgements

The authors gratefully acknowledge the participating schools, school personnel, and students for their valuable cooperation and contribution to the data collection process.

Declaration of AI Use

During the preparation of this manuscript, Grammarly was used solely for grammar checking, spelling correction, and improving language readability. No generative artificial intelligence (AI) tools were used to generate the scientific content, analyze the data, interpret the findings, or formulate the conclusions. The authors remain fully responsible for the scientific content, analyses, interpretations, and conclusions presented in this manuscript.

Ethics declaration

This study was conducted in accordance with established ethical principles for research involving human participants. Ethical approval was granted by the Institutional Review Board of the Faculty of Teacher Training and Education, Universitas Muhammadiyah Palangkaraya, Indonesia (Ethical Approval No. 122/UMPR/EC/VIII/2025). Written informed consent was obtained from participants' legal guardians, accompanied by documented student assent, before data collection. Participation was voluntary, and participants were informed about the study purpose and the confidentiality of their responses.

Funding

This publication was supported by the Badan Riset dan Inovasi Nasional (BRIN) through the Program Riset dan Inovasi untuk Indonesia Maju (RIIM) Kompetisi Gelombang 11 under Decision No. 106/II.7/HK/2026.

References

1. Álvarez-García, D., Dobarro, A., & Núñez, J. C. (2015). Validez y fiabilidad del Cuestionario de cibervictimización en estudiantes de Secundaria. *Aula Abierta*, *43*(1), 32–38. https://doi.org/10.1016/j.aula.2014.11.001
2. Álvarez-García, D., Thornberg, R., & Suárez-García, Z. (2021). Validation of a Scale for Assessing Bystander Responses in Bullying. *Psicothema*, *4*(33), 623–630. https://doi.org/10.7334/psicothema2021.140
3. Ariani, T. A., Putri, A. R., Firdausi, F. A., & Aini, N. (2025). Global prevalence and psychological impact of bullying among children and adolescents: A meta-analysis. *Journal of Affective Disorders*, *385*, 119446. https://doi.org/10.1016/j.jad.2025.119446
4. Batani, J., Mbunge, E., Muchemwa, B., Gaobotse, G., Gurajena, C., Fashoto, S., Kavu, T., & Dandajena, K. (2022). A Review of Deep Learning Models for Detecting Cyberbullying on Social Media Networks. In R. Silhavy (Ed.), *Cybernetics Perspectives in Systems* (pp. 528–550). Springer International Publishing.
5. Blakeslee, T., Snethen, J., Schiffman, R. F., Gwon, S. H., Sapp, M., & Kelber, S. (2023). Adolescent Characteristics, Suicide, and Bullying in High School. *The Journal of School Nursing*, *39*(6), 463–474. https://doi.org/10.1177/10598405211038235
6. Bork-Hüffer, T., Mahlknecht, B., & Kaufmann, K. (2021). (Cyber)Bullying in schools – when bullying stretches across cON/FFlating spaces. *Children's Geographies*, *19*(2), 241–253. https://doi.org/10.1080/14733285.2020.1784850
7. Borualogo, I. S., & Casas, F. (2022). Understanding Bullying Cases in Indonesia. In H. Tiliouine, D. Benatuil, & M. K. W. Lau (Eds.), *Handbook of Children's Risk, Vulnerability and Quality of Life: Global Perspectives* (pp. 187–199). Springer International Publishing. https://doi.org/10.1007/978-3-031-01783-4_12
8. Bowes, L., Aryani, F., Ohan, F., Haryanti, R. H., Winarna, S., Arsianto, Y., Budiyawati, H., Widowati, E., Saraswati, R., Kristianto, Y., Suryani, Y. E., Ulum, D. F., & Minnick, E. (2019). The development and pilot testing of an adolescent bullying intervention in Indonesia – the ROOTS Indonesia program. *Global Health Action*, *12*(1), 1656905. https://doi.org/10.1080/16549716.2019.1656905
9. Chang, Y.-H., Chen, F.-C., & Lee, C.-I. (2025). Developing an Early Warning System with Personalized Interventions to Enhance Academic Outcomes for At-Risk Students in Taiwanese Higher Education. *Education Sciences*, *15*(10), 1321. https://doi.org/10.3390/educsci15101321
10. Dailey, S. F., Roche, R. R., & Sharkey, M. C. (2025). Addressing Bullying and Cyberbullying in Public Health: A Systematic Review of Interventions for Healthcare and Public Health Professionals. *International Journal of Environmental Research and Public Health*, *22*(11), 1682. https://doi.org/10.3390/ijerph22111682
11. Del Rey, R., Casas, J. A., Ortega-Ruiz, R., Schultze-Krumbholz, A., Scheithauer, H., Smith, P., Thompson, F., Barkoukis, V., Tsorbatzoudis, H., Brighi, A., Guarini, A., Pyżalski, J., & Plichta, P. (2015). Structural validation and cross-cultural robustness of the European Cyberbullying Intervention Project Questionnaire. *Computers in Human Behavior*, *50*, 141–147. https://doi.org/10.1016/j.chb.2015.03.065
12. Erwandi, D., Kadir, A., & Lestari, F. (2021). Identification of Workplace Bullying: Reliability and Validity of Indonesian Version of the Negative Acts Questionnaire-Revised (NAQ-R). *International Journal of Environmental Research and Public Health*, *18*(8), 3985. https://doi.org/10.3390/ijerph18083985
13. Fine, S. L., Pinandari, A. W., Muzir, S. M., Agnesia, L., Novitasari, P. I., Bass, J. K., Blum, R. W., van Reeuwijk, M., Wilopo, S. A., & Mmari, K. (2023). "If it's really excessive, it can enter your heart": A Mixed Methods Investigation of Bullying Among Early Adolescents in Semarang, Indonesia. *Journal of Interpersonal Violence*, *38*(3–4), 4088–4113. https://doi.org/10.1177/08862605221111422

14. Giuliani, A., Petruccelli, I., & D'Urso, G. (2025). Longitudinal Evidence on Peer Victimisation and Persistent Mental Health Outcomes in Youth: A Systematic Review. *Behavioral Sciences*, *15*(12), 1734. https://doi.org/10.3390/bs15121734
15. Graves, K. A., Mirielli, L. G., Ousley, C., & Rose, C. A. (2025). Literature Review and Content Analysis of Bullying Assessments: Are We Measuring What We Intend to? *International Journal of Environmental Research and Public Health*, *22*(1), 29. https://doi.org/10.3390/ijerph22010029
16. Grothaus, C. (2024). Coping strategies and social representations of bullies among bullying victims from individualistic and collectivist societies. *Culture & Psychology*, *30*(1), 115–136. https://doi.org/10.1177/1354067X231164796
17. Halliday, S., Gregory, T., Taylor, A., Digenis, C., & Turnbull, D. (2021). The Impact of Bullying Victimization in Early Adolescence on Subsequent Psychosocial and Academic Outcomes across the Adolescent Period: A Systematic Review. *Journal of School Violence*, *20*(3), 351–373. https://doi.org/10.1080/15388220.2021.1913598
18. Hasan, M. T., Hossain, M. A. E., Mukta, M. S. H., Akter, A., Ahmed, M., & Islam, S. (2023). A Review on Deep-Learning-Based Cyberbullying Detection. *Future Internet*, *15*(5), 179. https://doi.org/10.3390/fi15050179
19. Hu, X., Hu, X., & Xie, S. (2025). Unraveling the multidimensional association of covert bullying, abandonment, and adolescent depression: A critical review and conceptual framework. *Acta Psychologica*, *261*, 105726. https://doi.org/10.1016/j.actpsy.2025.105726
20. Hysing, M., Askeland, K. G., La Greca, A. M., Solberg, M. E., Breivik, K., & Sivertsen, B. (2021). Bullying Involvement in Adolescence: Implications for Sleep, Mental Health, and Academic Outcomes. *Journal of Interpersonal Violence*, *36*(17–18), NP8992–NP9014. https://doi.org/10.1177/0886260519853409
21. Iurino, C., Townsend, S., Yoon, J., Bauman, S., & Yee, C. (2025). Adolescents' Reporting of Bullying: The Role of Minority Status, Bullying Type, Outcome Expectations, and Affect. *International Journal of Bullying Prevention*. https://doi.org/10.1007/s42380-025-00327-z
22. Johansson, S., & Englund, G. (2021). Cyberbullying and its relationship with physical, verbal, and relational bullying: A structural equation modelling approach. *Educational Psychology*, *41*(3), 320–337. https://doi.org/10.1080/01443410.2020.1769033
23. Jukes, M. C. H., Ahmed, I., Baker, S., Draper, C. E., Howard, S. J., McCoy, D. C., Obradović, J., & Wolf, S. (2024). Principles for Adapting Assessments of Executive Function across Cultural Contexts. *Brain Sciences*, *14*(4), 318. https://doi.org/10.3390/brainsci14040318
24. Juli, R., Sarfika, R., Basmanelly, Saifudin, I. M. Moh. Y., & Abdullah, K. L. (2025). Predictors of bullying victimization among early adolescents in junior high schools: A cross-sectional study. *Belitung Nursing Journal*, *11*(5), 547–559. https://doi.org/10.33546/bnj.3950
25. Kaufman, T. M. L., Huitsing, G., Bloemberg, R., & Veenstra, R. (2020). The Systematic Application of Network Diagnostics to Monitor and Tackle Bullying and Victimization in Schools. *International Journal of Bullying Prevention*, *3*(1), 75–87. https://doi.org/10.1007/s42380-020-00064-5
26. Koutsoumpis, A. (2025). Psychometric properties of personality assessment using machine learning. *Current Opinion in Psychology*, *65*, 102093. https://doi.org/10.1016/j.copsyc.2025.102093
27. Lechner, V., Crăciun, I. C., & Scheithauer, H. (2023). Barriers, resources, and attitudes towards (cyber-)bullying prevention/intervention in schools from the perspective of school staff: Results from focus group discussions. *Teaching and Teacher Education*, *135*, 104358. https://doi.org/10.1016/j.tate.2023.104358

28. Magaz, A. M., Chorot, P., Santed, M. A., Valiente, R. M., & Sandín, B. (2016). Evaluación del bullying como victimización: Estructura, fiabilidad y validez del Cuestionario de Acoso entre Iguales (CAI). *Revista de Psicopatología y Psicología Clínica*, *21*(2), 77–95. https://doi.org/10.5944/rppc.vol.21.num.2.2016.16990
29. Messman, E., Heinze, J., Hsieh, H.-F., Hockley, N., Pomerantz, N., Grodzinski, A., Scott, B., Goldstein, N., & Zimmerman, M. (2024). Anonymous Reporting Systems for School-Based Violence Prevention: A Systematic Review. *Health Education & Behavior*, *51*(1), 62–70. https://doi.org/10.1177/10901981211073734
30. Nurhalimah, N., Haryati, O., Wartonah, W., Adelia, A., Bellavista, C., Dzagonga, T., Chipuriro, I., Chikwasha, V., Matsvimbo, P. F., Yinka-Banjo, C., Ajayi, O., Akinyemi, M., Tresner-Kirsch, D., Omotubora, A., Cherie, K. T., Alemayehu, G. M., Getie, E. T., Krisnana, I., Rachmawati, P. D., … Adelia, A. (2025). The impact of verbal bullying on the mental health of students at a senior high school in Bekasi, Indonesia. *Healthcare in Low-Resource Settings*, *13*(s1). https://doi.org/10.4081/hls.2025.13092
31. Orrù, G., Galli, A., Gattulli, V., Gravina, M., Micheletto, M., Marrone, S., Nocerino, W., Procaccino, A., Terrone, G., Curtotti, D., Impedovo, D., Marcialis, G. L., & Sansone, C. (2023). Development of Technologies for the Detection of (Cyber)Bullying Actions: The BullyBuster Project. *Information*, *14*(8), 430. https://doi.org/10.3390/info14080430
32. Ortega Ruiz, R., Rey Alamillo, R. del, & Casas, J. A. (2016). Evaluar el bullying y el cyberbullying validación española del EBIP-Q y del ECIP-Q. *https://journals.copmadrid.org/psed*. https://doi.org/10.1016/j.pse.2016.01.004
33. Polanin, J. R., Espelage, D. L., Grotpeter, J. K., Spinney, E., Ingram, K. M., Valido, A., El Sheikh, A., Torgal, C., & Robinson, L. (2021). A meta-analysis of longitudinal partial correlations between school violence and mental health, school performance, and criminal or delinquent acts. *Psychological Bulletin*, *147*(2), 115–133. https://doi.org/10.1037/bul0000314
34. Rachman, A., Putro, H. Y. S., Rusandi, M. A., & Situmorang, D. D. B. (2024). The development and validation of the "Kuesioner Tema Proyek Penguatan Profil Pelajar Pancasila" (KT P5): A new tool for strengthening the Pancasila Student Profile in Indonesian pioneer schools. *Heliyon*, *10*(16), e35912. https://doi.org/10.1016/j.heliyon.2024.e35912
35. Samara, M., Da Silva Nascimento, B., El-Asam, A., Hammuda, S., & Khattab, N. (2021). How Can Bullying Victimisation Lead to Lower Academic Achievement? A Systematic Review and Meta-Analysis of the Mediating Role of Cognitive-Motivational Factors. *International Journal of Environmental Research and Public Health*, *18*(5), 2209. https://doi.org/10.3390/ijerph18052209
36. Useche, S. A., Valle, E., Valle-Escolano, R., & Colomer-Pérez, N. (2021). Psychometric properties, validity and insights of the School Bullying Questionnaire (CIE-A) in secondary schools of the Valencian Community (Spain). *PLOS ONE*, *16*(11), e0259392. https://doi.org/10.1371/journal.pone.0259392
37. Wójcik, M., & Rzeńca, K. (2021). Disclosing or Hiding Bullying Victimization: A Grounded Theory Study From Former Victims' Point of View. *School Mental Health*, *13*(4), 808–818. https://doi.org/10.1007/s12310-021-09447-5
38. Xie, Z., Man, W., Liu, C., & Fu, X. (2023). A PRISMA-Based Systematic Review of Measurements for School Bullying. *Adolescent Research Review*, *8*(2), 219–259. https://doi.org/10.1007/s40894-022-00194-5
39. Zaneva, M., Minnick, E., Nahar, Ginting, V., Aryani, F., Ohan, F., Haryanti, R. H., Winarna, S., Arsianto, Y., Budiyawati, H., Widowati, E., Saraswati, R., Kristianto, Y., Suryani, Y. E., Ulum, D. F., & Bowes, L. (2023). Social Norms Predict Bullying: Evidence from an Anti-Bullying Intervention Trial in Indonesia. *International Journal of Bullying Prevention*, *7*(4), 253–265. https://doi.org/10.1007/s42380-023-00176-8

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- INSTRUMENT.pdf
- BullyTRACEinstrumentanalysisfinal.xlsx
- Supplementarynature.docx