

School-based physical activity support among Chinese adolescents: development and initial psychometric evaluation of a five-domain student-report scale

Hu Shaojuan

Hengyang Normal University

Xiong Jie

Hengyang Normal University

Su Yi

Hengyang Normal University

Li Liting

Hengyang Normal University

Han Dong

Shenzhen No. 3 Senior High School, Shenzhen, Guangdong Province, China

Li Xinyuan

lixinyuan@hynu.edu.cn

Hengyang Normal University

Research Article

Keywords: Adolescent health, China, Physical activity, Physical education, Psychometrics, School environment, Scale development

Posted Date: September 29th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10834196/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

Background

Schools shape adolescents' opportunities, social experiences and institutional conditions for physical activity, but existing measures often combine policy, environment and lived experience into broad scores. This study developed and initially evaluated the 38-item School-Based Physical Activity Support Scale–Student Report (SBPASS-SR38), which distinguishes curricular, extracurricular activity, peer, cultural and institutional support.

Methods

Three sequential studies were conducted in China. Study 1 used review by five experts and cognitive interviews with nine adolescents (18 item-focused records) to evaluate content and response processes. A version audit restricted all analyses to 38 items with identical wording across study materials. Study 2 included 706 adolescents from eight schools and used item analysis, parallel analysis and polychoric exploratory factor analysis. Study 3 included 724 adolescents from six independent schools; 718 complete cases from 30 classes were used to compare seven class-cluster-robust ordinal models. Reliability, 14-day stability in 178 students and correlations with reports from 19 physical education teachers covering 30 classes were also examined.

Results

Scale-level content validity indices ranged from 0.953 to 0.989, and Study 2 recovered the proposed five factors. In Study 3, correlated five-factor confirmatory factor analysis fitted well (comparative fit index [CFI] = 0.957; root mean square error of approximation [RMSEA] = 0.018; standardised root mean square residual [SRMR] = 0.042), and five-factor target exploratory structural equation modelling fitted better (CFI = 0.983; RMSEA = 0.012; SRMR = 0.026). Second-order and bifactor confirmatory models did not converge. Bifactor exploratory structural equation modelling fitted slightly better, but the general factor was concentrated in extracurricular and peer items (omega hierarchical = 0.484; explained common variance = 0.449). Subscale omega ranged from 0.667 to 0.912 and absolute-agreement intraclass correlations from 0.677 to 0.880. All five means changed over 14 days, and no student–teacher correlation was statistically significant.

Conclusions

Initial evidence supports five student-reported support domains, but not a total score or interchangeability of student and teacher reports. The SBPASS-SR38 may support school-based public

health research and programme evaluation after further testing of subgroup invariance, predictive validity and intervention sensitivity.

Clinical trial registration

Clinical trial number: not applicable.

Background

Insufficient physical activity remains widespread among adolescents and is patterned by social and institutional conditions rather than by individual choice alone [1–3]. Schools are important public health settings because they repeatedly expose most adolescents to physical education, organised and informal activity opportunities, peers, norms, facilities and participation rules. Reviews and observational evidence nevertheless indicate substantial variation in the extent to which school environments enable or constrain activity [4, 5]. Distinguishing the mechanisms through which schools provide support may therefore improve needs assessment, implementation research and evaluation of school-based physical activity programmes.

Ecological models locate health behaviour within interacting interpersonal, organisational, community and policy systems [6, 7]. In schools, support may operate through physical education, opportunities beyond lessons, peer relationships, school-wide meanings attached to activity and formal implementation arrangements. Self-determination theory adds a motivational perspective: choice, appropriately challenging instruction, informational feedback, belonging and inclusion can support autonomy, competence and relatedness in physical education and physical activity [8–10]. Peer invitations, encouragement, non-ridicule and active modelling are also associated with adolescent activity [11]. These processes are likely to be related, but they are not conceptually interchangeable.

Conceptualisation of school-based physical activity support

We define school-based physical activity support as students' perception that school practices, opportunities, social relationships, shared meanings, and formal arrangements enable equitable, safe, sustained, and worthwhile participation in physical activity. Curricular support concerns student-centred physical education: choice, adaptation, feedback, safety, persistence, inclusion, transferable learning, meaningfulness, and appropriate exercise load. Extracurricular activity support concerns variety, scheduled opportunities, ability-inclusive access, timely information, low selection barriers, and continuity. Peer support concerns invitations, co-participation, encouragement, inclusive group membership, non-ridicule, cooperation, gender equity, and active-peer modelling. School physical activity culture concerns whether routine participation, persistence, enjoyment, belonging, broad participation, and values beyond competition rankings are recognised and enacted. Institutional support concerns

protection of scheduled provision, implementation of extracurricular opportunities, equitable facility access, accommodations, and channels for addressing participation problems.

This organisation distinguishes immediate pedagogical and interpersonal experiences from the school-wide culture and implementation arrangements that sustain them. It also yields competing score interpretations. If support is broadly unitary, a general or second-order factor should account for most reliable variance. If it is differentiated, five correlated domains should provide a better representation than one- or two-factor alternatives, and meaningful domain-specific variance should remain after any general factor is modelled.

Measurement context and study objectives

Existing instruments address related but distinct targets. The original student-report Questionnaire Assessing School Physical Activity Environment focused on physical and social environments, and its recent Chinese redevelopment added school policy and examined measurement invariance and associations with physical activity [12–14]. The School Physical Activity Policy Assessment is an organisational audit of physical education and other activity opportunities [15], consistent with comprehensive school physical activity programming [16]. Other recent Chinese instruments assess the implementation of physical education programmes in junior secondary schools [17] or perceived school influences on the formation of physical activity habits [18]. These measures are useful for their intended purposes, but they do not separately characterise how adolescents experience supportive pedagogy, extracurricular accessibility, peer inclusion, enacted school culture and institutional implementation within one scoring framework.

The present study therefore developed a student-report measure that integrates five theoretically distinct and potentially modifiable support mechanisms while testing whether their covariance justifies a total score. Teacher reports were included as an external perspective rather than as an assumed criterion because discrepancies between informants may reflect differences in access to information and frames of reference [19, 20]. Consistent with current validity theory, conclusions were framed in relation to proposed score uses and interpretations rather than treating validity as an inherent property of an instrument [21].

Study 1 developed the item content and examined expert and adolescent response-process evidence. Study 2 explored dimensionality in an independent adolescent sample. Study 3 compared seven ordinal structural representations, including confirmatory factor analysis (CFA), exploratory structural equation modelling (ESEM) and bifactor alternatives. We expected correlated five-factor CFA and five-factor ESEM to represent the data better than simpler models. Bifactor models were used to examine total-score interpretability rather than treating superior global fit as sufficient evidence for aggregation. We also assessed subscale reliability, 14-day test–retest stability and student–teacher convergence.

Methods

Study design, ethics and consent

A sequential, three-study instrument-development design combined content and response-process evidence, exploratory calibration, independent ordinal structural analysis, temporal stability and comparison with an external teacher perspective. The study was conducted in accordance with the Declaration of Helsinki and all relevant institutional guidelines and regulations. The protocol was approved before data collection by the Research Ethics Committee of Hengyang Normal University (approval no. 2026lllw023; 20 March 2026). All adult participants, including the expert reviewers and physical education teachers, provided electronically signed informed consent. For each adolescent participant, a parent or legal guardian signed an electronic informed-consent form, and the adolescent signed an electronic assent form. Permission was also obtained from all participating schools. Paper surveys were administered during school-organised sessions. Identification codes rather than names were used for analysis and retest linkage.

Students rated each statement from 1 (strongly disagree) to 5 (strongly agree), and all items were positively keyed. The Chinese wording was administered. The English wording in Additional file 1 is a reporting translation and has not been psychometrically validated. Subscale scores are calculated as the mean of the constituent items when all required items are available. A 38-item total score was not calculated because the structural evidence did not support that interpretation.

For Study 2, the operational target was at least 600 analysable students, exceeding 15 observations per field-test item. Eight hundred questionnaires were distributed to allow for non-return and unusable responses. Study 3 targeted approximately 700 students in an independent multi-school sample and at least 150 matched retest records. These pragmatic targets were intended to stabilise ordinal item correlations and multifactor models and were consistent with general scale-development guidance [22, 23]. No effect-size-based Monte Carlo power analysis was conducted before data collection [24, 25].

Common-form version audit

The Study 1 record described an intended 39-item field-test pool rather than a final scoring form. Before the final Mplus analyses, comparison of the locked item record with the administered forms showed that item 39 did not have the same content across materials. The intended curricular candidate had not been administered and therefore had no response data; it had been replaced by an institutional item concerning advance publication of facility opening times. That substituted item had previously failed the Study 1 relevance criterion (item-level content validity index [I-CVI] = 0.60). Items 1–38 had identical wording in the content record and both empirical studies. A common-form rule based on item identity and content was therefore applied: both versions of item 39 were excluded, no item was relabelled and no loading or fit statistic informed the decision. The resulting School-Based Physical Activity Support Scale–Student Report, 38-item form (SBPASS-SR38), comprises the 38 wording-identical items for which comparable evidence was available across studies (Fig. 1; Additional file 1: Table S9).

Study 1: content and response-process development

Item generation and expert review

A source pool of 53 student-facing statements was assembled from the provisional five-domain content map, targeted literature mapping and reformulation of school-support indicators into experiences that adolescents could judge [22, 26]. Five experts spanning school physical activity policy, physical education curriculum and pedagogy, frontline teaching and implementation, and school governance reviewed the construct definitions and candidate items. They proposed four additional curricular statements, producing 57 candidates before iterative deletion, merging and rewriting. Experts independently rated relevance, clarity and age appropriateness on four-point scales. Ratings of 3 or 4 were counted as content valid; I-CVI and scale-level average content validity index (S-CVI/Ave) were calculated [27]. Experts also assigned each item to a domain, identified redundancy and suggested missing content. Items that underwent substantive revision were re-rated using the exact wording proposed for field administration.

Cognitive interviewing and coding

Cognitive interviewing examined comprehension, retrieval, judgement and mapping to the response options [28, 29]. Nine adolescents completed item-focused interviews, generating 18 student-by-item records. The sample comprised four junior- and five senior-secondary students, four girls and five boys. Physical activity strata were low ($n = 2$), moderate ($n = 4$) and high ($n = 3$); perceived-competence strata were low ($n = 3$), moderate ($n = 2$) and high ($n = 4$). Participants paraphrased selected items, recalled relevant school experiences, explained the basis for their judgement, selected a response from 1 to 5 and distinguished adjacent response options.

Item assignment prioritised wording that remained under-tested after expert review. Three previously under-tested items were each assessed by five students. One was the revised curricular item, 'Physical education teachers arrange an appropriate amount of physical activity', which was assessed using the exact final Chinese wording and then re-rated by all five experts. Two researchers independently coded the 18 records for intended comprehension, language or construct problems and the resulting retain-or-revise decision. Initial judgements were recorded before discussion; the coders agreed on 16 records (88.9%), and the two disagreements were resolved by reviewing the full response context.

Study 2: exploratory calibration

Participants and procedure

Eight anonymised secondary-school sampling sites spanning Grades 7–11 participated. Only site-coded identifiers were retained in the verified analysis file; school names and geographical descriptors are therefore not reported, and no claim of regional representativeness is made. Teachers distributed 800 paper questionnaires; 735 were returned (91.9%). After removal of 29 incomplete or ineligible records, 706 were analysed (88.2% of those distributed and 96.1% of those returned). The sample comprised 378 boys (53.5%) and 328 girls (46.5%). Grade representation was 30 in Grade 7 (4.2%), 49 in Grade 8 (6.9%), 47 in Grade 9 (6.7%), 302 in Grade 10 (42.8%) and 278 in Grade 11 (39.4%). Self-reported activity level was low for 154 participants (21.8%), moderate for 265 (37.5%) and high for 287 (40.7%); 266 participants (37.7%) reported left-behind status. The 38 common items contained no missing or out-of-range values (Table 1; Additional file 1: Table S8).

Statistical analysis

Study 2 analyses were implemented in Python 3 using reproducible scripts. Item distributions, floor and ceiling percentages, corrected item–subscale rank correlations and alpha if an item was deleted were examined. Factor number was assessed by a 500-replication parallel analysis based on eigenvalues from the Spearman item-correlation matrix and the 95th percentile of random-data eigenvalues [30]. A five-factor principal-axis solution was then fitted to the polychoric correlation matrix and subjected to oblique Promax rotation. The Study 1 item–domain map was used to align and interpret the rotated factors, not as a rotation target; all secondary loadings were freely estimated [31, 32]. Internal consistency was summarised using coefficient alpha, ordinal alpha and McDonald’s omega [33, 34]. Heterotrait–monotrait ratios (HTMT) were used to assess factor distinctiveness [35].

Study 3: independent structural, temporal and multisource evaluation

Participants and procedure

An independent sample was recruited from six schools in Hunan, Henan and Shaanxi. The schools represented urban and county settings, and the 30 participating classes covered junior- and senior-secondary education. At Time 1, 724 students participated, of whom 718 had complete data for all 38 common items. The full sample comprised 364 boys and 360 girls, 346 junior- and 378 senior-secondary students, 467 students from urban and 257 from rural family backgrounds, 83 left-behind students and 132 boarders. Grades 9 and 12 were not sampled because students were preparing for entrance examinations. Grade counts were 188, 158, 189 and 189 for Grades 7, 8, 10 and 11, respectively. Chronological age was not recorded; assigning the reporting ages specified for these grades (13, 14, 16 and 17 years) yielded an approximate mean of 15.05 years (SD = 1.60; range 13–17). At two senior schools, locally entered codes 11 and 12 denoted Senior 1 and Senior 2; these were harmonised to

Grades 10 and 11 for cross-school reporting. Grade was not used in the structural models (Table 1; Additional file 1: Table S8).

For temporal stability, 180 students completed a second paper questionnaire 14 days after Time 1. All identification codes matched those used at Time 1, and 178 participants had complete item data at both assessments. For the external-informant comparison, 19 physical education teachers completed audit reports. Each teacher taught between one and three of the 30 participating classes; classes taught by the same teacher were combined as one teacher-defined instructional unit. Student subscale scores were averaged within these 19 units and linked to the corresponding teacher report (22–80 students per unit; mean = 37.79; median = 29).

Statistical analysis

All 38 items were treated as five-category ordered variables. Models were estimated in Mplus 8.3 using weighted least squares mean- and variance-adjusted estimation (WLSMV), DELTA parameterisation and a maximum of 20 000 iterations [36, 37]. TYPE = COMPLEX with CLUSTER = clusid adjusted standard errors and model tests for non-independence within the 30 classes. Seven models were evaluated: (M1) one-factor CFA; (M2) an empirical two-factor comparator contrasting the 10 curricular items with the remaining 28 items; (M3) five correlated CFA factors; (M4) five first-order factors with a second-order factor; (M5) orthogonal bifactor CFA; (M6) five-factor target ESEM; and (M7) target bifactor ESEM. M2 was an empirical sensitivity comparator rather than a prespecified theoretical model. In M6, non-target loadings were estimated while the rotation targeted the Study 1 content map. In M7, a general factor was specified as orthogonal to the five domain-specific factors [38, 39]. M6 and M7 used 100 random starts; target rotation was oblique in M6 and orthogonal in M7.

Model evaluation considered normal termination, admissibility, the comparative fit index (CFI), Tucker–Lewis index (TLI), root mean square error of approximation (RMSEA) with its 90% confidence interval (CI), standardised root mean square residual (SRMR), primary and cross-loadings, factor correlations and substantive interpretability. Conventional fit thresholds were used as descriptive guides rather than automatic decision rules [40, 41]. Mplus input and output files and a parameter-level results workbook were retained in the protected analysis archive; manuscript values were checked against the final DELTA output files.

For the converged bifactor ESEM, omega total (ω), omega hierarchical (ω_H), explained common variance (ECV), percentage of uncontaminated correlations (PUC), item explained common variance (IECV) and omega hierarchical subscale (ω_{HS}) were examined to assess whether a total score or residual specific-factor scores could be interpreted [42, 43]. Reliability and HTMT were calculated from polychoric correlations. Fourteen-day stability was estimated using the single-measure absolute-agreement intraclass correlation coefficient, ICC(A,1), with 5 000 participant-bootstrap samples, alongside the consistency ICC(C,1), Pearson correlation, mean change, standardised paired change (dz), standard error of measurement and minimum detectable change [44, 45]. Convergence with teacher

reports was assessed using Spearman correlations, 5 000 bootstrap samples and Holm-adjusted p values.

Use of generative AI-assisted tools

During manuscript preparation, the authors used OpenAI ChatGPT/Codex for English-language editing, consistency checks between archived Mplus syntax and native output and the reported results, and preparation of Fig. 1 from author-specified content. The tool did not select models, alter data or make inferential decisions. The authors checked all numerical reporting and figure content against the study records and take full responsibility for the manuscript.

Results

Participant flow and common-form evidence

Table 1 summarises the staged design and participant flow. The version audit identified 38 wording-identical items across the locked content record and empirical forms. All subsequent item, score and structural results therefore refer to the SBPASS-SR38 rather than to the intended 39-item field-test pool.

Table 1. Study design and participant flow

Study	Purpose	Sampling frame	Analysed evidence
Study 1	Content and response-process development	Five experts; item-focused cognitive interviews across school stage, sex, activity and competence strata	38-item common form: S-CVI/Ave 0.953/0.989/0.979; 18 interview records
Study 2	Exploratory calibration	Eight anonymised secondary schools; Grades 7–11	800 distributed; 735 returned; 706 analysed
Study 3, Time 1	Independent structure and reliability	Six schools; three provinces; 30 classes; Grades 7, 8, 10 and 11	724 participants; 718 complete structural cases
Study 3, Time 2	14-day stability	Subsample of the Study 3 cohort	180 returned; 178 complete matched pairs
Teachers	External-informant comparison	19 PE teachers covering 30 classes; 19 instructional units	19 teacher reports linked to student unit means

Note. S-CVI/Ave values are relevance/clarity/age appropriateness. PE, physical education.

Study 1: content and response-process evidence

Iterative expert and student review yielded an intended 39-item field-test pool comprising 11 curricular, six extracurricular activity, eight peer, nine cultural and five institutional candidates. The cross-study

comparison showed that only 38 intended candidates had been administered consistently. For this 38-item common form, S-CVI/Ave was 0.953 for relevance, 0.989 for clarity and 0.979 for age appropriateness. Every retained item had a relevance I-CVI of at least 0.80. The intended curricular item 39 had no response data, whereas the different institutional item printed in its place had a relevance I-CVI of 0.60 and was excluded from scoring and modelling.

Students distinguished activities open to all pupils from selectively constituted school teams; interpreted recognition beyond competitive rankings in terms of participation, persistence, health and enjoyment; and distinguished the availability of a feedback channel from a guarantee that the school would resolve every problem. These findings supported field testing of the exact wording but did not, on their own, establish factor structure or score use.

Study 2: exploratory structure and reliability

The first five observed eigenvalues (10.24, 6.76, 4.06, 2.23 and 2.15) exceeded their random 95th-percentile values (1.52, 1.46, 1.40, 1.37 and 1.34); the sixth did not (0.86 versus 1.31). In the Promax-rotated five-factor solution, every item had its largest absolute loading on its intended factor. Intended-domain primary loadings ranged from 0.655 to 0.975, the largest absolute secondary loading was 0.166 and no secondary loading reached 0.30. Factor correlations ranged from -0.208 to 0.464, and HTMT values from 0.052 to 0.499. Subscale omega ranged from 0.929 to 0.960 (Table 2; Additional file 1: Tables S2–S3).

Table 2. Study 2 subscale descriptives and reliability

Subscale	Items	Mean	SD	Alpha	Ordinal alpha	Omega
Curricular support	10	3.78	0.74	0.931	0.948	0.948
Extracurricular activity support	6	3.15	0.93	0.947	0.960	0.960
Peer support	8	3.21	0.81	0.934	0.954	0.954
School physical activity culture	9	2.47	0.68	0.907	0.929	0.929
Institutional support	5	3.12	0.90	0.940	0.960	0.960

Note. N = 706. Scores range from 1 to 5. HTMT values ranged from 0.052 to 0.499. Full loadings and item statistics are reported in Additional file 1.

Study 3: structural models

The one- and two-factor models showed poor overall fit despite RMSEA values below 0.05 (M1: $\chi^2(665) = 1797.23$, $p < 0.001$, CFI = 0.665, TLI = 0.646, RMSEA = 0.049, 90% CI 0.046–0.051, SRMR = 0.091; M2: $\chi^2(664) = 1486.42$, $p < 0.001$, CFI = 0.757, TLI = 0.743, RMSEA = 0.042, 90% CI 0.039–0.044, SRMR =

0.081). Correlated five-factor CFA fitted well, $\chi^2(655) = 800.89$, $p < 0.001$, CFI = 0.957, TLI = 0.954, RMSEA = 0.018 (90% CI 0.013–0.022) and SRMR = 0.042. The second-order model (M4) and orthogonal bifactor CFA (M5) did not terminate normally after 20 000 iterations and produced near-zero variances and extreme loadings; their fit and parameters were not interpreted. Five-factor target ESEM fitted very well, $\chi^2(523) = 579.95$, $p = 0.043$, CFI = 0.983, TLI = 0.977, RMSEA = 0.012 (90% CI 0.003–0.018) and SRMR = 0.026. Target bifactor ESEM fitted slightly better, $\chi^2(490) = 535.52$, $p = 0.076$, CFI = 0.987, TLI = 0.981, RMSEA = 0.011 (90% CI 0.000–0.017) and SRMR = 0.024 (Table 3).

Table 3. Study 3 class-cluster-robust ordinal model comparisons

Model	Specification	χ^2	df	p	CFI	TLI	RMSEA (90% CI)	SRMR
M1	One-factor CFA	1797.23	665	<0.001	0.665	0.646	0.049 (0.046–0.051)	0.091
M2	Empirical two-factor CFA	1486.42	664	<0.001	0.757	0.743	0.042 (0.039–0.044)	0.081
M3	Correlated five-factor CFA	800.89	655	<0.001	0.957	0.954	0.018 (0.013–0.022)	0.042
M4	Second-order five-factor model	—	—	—	—	—	Did not converge	—
M5	Orthogonal bifactor CFA	—	—	—	—	—	Did not converge	—
M6	Five-factor target ESEM	579.95	523	0.043	0.983	0.977	0.012 (0.003–0.018)	0.026
M7	Target bifactor ESEM	535.52	490	0.076	0.987	0.981	0.011 (0.000–0.017)	0.024

Note. N = 718 across 30 classes. Values are from final Mplus 8.3 WLSMV runs using DELTA parameterisation, TYPE = COMPLEX, CLUSTER = clusid and 20 000 iterations. M4 and M5 did not terminate normally; their fit and parameters are not reported. M6 was retained for interpretation because M7's general factor did not span all five domains.

In M6, all 38 target loadings were statistically significant and ranged from 0.397 to 0.835. Across 152 non-target loadings, the mean absolute loading was 0.055 (median = 0.041; maximum = 0.213); 24 had an absolute magnitude of at least 0.10, two reached 0.20 and none reached 0.30. Factor correlations ranged from –0.325 to 0.755, with the strongest association between extracurricular activity and peer support ($r = 0.755$). In the correlated CFA, primary loadings ranged from 0.419 to 0.777 and the corresponding extracurricular–peer correlation was 0.810. HTMT values ranged from 0.094 to 0.810.

The overlap was concentrated in one theoretically plausible pair rather than distributed as a uniform general factor (Additional file 1: Table S4).

Although M7 showed slightly better global fit, its loadings did not support a broadly defined general factor. Mean general-factor loadings were -0.145 for curricular items, 0.609 for extracurricular items, 0.704 for peer items, 0.040 for cultural items and 0.168 for institutional items; only 13 of 38 IECV values were at least 0.50 . Global indices were $\omega = 0.863$, $\omega_H = 0.484$, $ECV = 0.449$ and $PUC = 0.809$. Domain-specific ω_{HS} values were 0.649 , 0.234 , 0.118 , 0.775 and 0.568 for curricular, extracurricular, peer, cultural and institutional support, respectively. The general factor therefore primarily represented covariance between extracurricular and peer items rather than balanced school-wide support. On the basis of convergence, global fit, loading pattern and substantive interpretability, M6 was retained for structural interpretation, and five subscale scores rather than a 38-item total were recommended (Additional file 1: Table S5).

Study 3: reliability, temporal stability and teacher comparison

At Time 1, omega was 0.736 for curricular support, 0.861 for extracurricular activity support, 0.912 for peer support, 0.779 for school physical activity culture and 0.667 for institutional support (Table 4). Institutional support therefore had comparatively limited precision and should be interpreted cautiously. The corresponding means were 4.11 , 3.87 , 3.72 , 2.96 and 2.36 , indicating substantial differences in endorsement across domains.

Table 4. Study 3 subscale scores, reliability and 14-day stability

Subscale	Items	Time 1 mean	Time 1 SD	Time 1 omega	ICC(A,1) (95% CI)	Time 2 – Time 1	dz
Curricular support	10	4.11	0.48	0.736	0.730 (0.639–0.792)	–0.100	–0.347
Extracurricular activity support	6	3.87	0.61	0.861	0.677 (0.628–0.718)	–0.404	–1.203
Peer support	8	3.72	0.80	0.912	0.803 (0.765–0.833)	–0.167	–0.413
School physical activity culture	9	2.96	0.53	0.779	0.880 (0.843–0.907)	+0.042	+0.168
Institutional support	5	2.36	0.64	0.667	0.765 (0.703–0.811)	–0.119	–0.308

Note. Time 1 reliability N = 718; the retest included n = 178 complete matched pairs. ICC confidence intervals used 5 000 participant-bootstrap samples. All five paired changes remained statistically significant after Holm correction; magnitude and direction should be interpreted alongside ICC.

Absolute-agreement test–retest ICCs were 0.730, 0.677, 0.803, 0.880 and 0.765 for curricular, extracurricular, peer, cultural and institutional support. Their confidence intervals were compatible with at least moderate stability. However, every domain showed a statistically detectable mean change after Holm adjustment. The largest change was for extracurricular activity support (mean change = −0.404; dz = −1.203); curricular, peer and institutional scores also declined, whereas cultural support increased slightly (mean change = 0.042; dz = 0.168). The coefficients should therefore be interpreted alongside the occasion effects and not as evidence of score invariance over 14 days. Time 2 alpha values were also lower (0.424–0.743) (Additional file 1: Table S6).

Student and teacher reports did not converge at the instructional-unit level. Spearman correlations ranged from −0.313 to 0.171; every bootstrap confidence interval included zero, and all Holm-adjusted p values were at least 0.962. The findings do not support interchangeability of the two informants. Different observational access and reference frames may contribute, although the small number of teacher-linked units produced wide confidence intervals. The teacher audit is therefore best treated as an external perspective rather than a parallel form of the student scale (Table 5; Additional file 1: Table S7).

Table 5. Student–teacher correlations by support domain

Domain	Teacher units	Spearman ρ	Bootstrap 95% CI	Holm-adjusted p
Curricular support	19	0.140	−0.347 to 0.607	1.000
Extracurricular activity support	19	−0.313	−0.693 to 0.155	0.962
Peer support	19	−0.279	−0.622 to 0.185	0.993
School physical activity culture	19	0.046	−0.363 to 0.447	1.000
Institutional support	19	0.171	−0.340 to 0.567	1.000

Note. Correlations compare each PE teacher’s audit score with mean student scores across that teacher’s class or classes (19 teacher-defined units covering 30 classes; 22–80 students per unit). No correlation supported student–teacher convergence.

Discussion

Principal findings

Across three studies, the SBPASS-SR38 yielded initial evidence for five domains of adolescents' perceived school-based physical activity support. Expert review and cognitive interviewing supported content coverage and comprehensibility; an independent exploratory sample recovered the proposed five-factor structure with small secondary loadings; and ordinal analysis in a second sample supported correlated five-factor CFA and, more strongly, target ESEM. These results support reporting five subscale scores. They do not support a total score, temporal invariance over 14 days or interchangeability of student and teacher reports.

Public health and theoretical interpretation

The five-domain profile links ecological and motivational accounts of adolescent physical activity to potentially modifiable features of school practice. Curricular support reflects whether physical education provides choice, appropriate challenge, safety, feedback, inclusion, meaning and transfer. Extracurricular support concerns whether opportunities are sustained, visible and accessible. Peer support represents the interpersonal conditions that can make participation socially safe and attractive. School physical activity culture concerns shared recognition, belonging, enjoyment and values beyond competitive status. Institutional support reflects whether timetables, facilities, accommodations and feedback mechanisms make provision reliable in practice. Considering these domains together may help public health researchers identify where support is strong or weak without treating facilities, written policy, enacted practice and student experience as equivalent.

The SBPASS-SR38 is complementary to, rather than a replacement for, established instruments. Q-SPACE characterises physical, social and policy environments [12–14]; administrator- or teacher-facing instruments assess policy and programme implementation [15–17]; and the Chinese school-environment scale focuses on influences on habit formation [18]. The present scale instead separates five support mechanisms as experienced by students and explicitly tests whether they can be combined. This distinction is relevant to implementation evaluation: a school may report that a programme or facility exists while students vary in whether they perceive it as visible, accessible, inclusive or sustained.

The strong association between extracurricular and peer support also illustrates why target ESEM was informative. Adolescents commonly experience organised activity through their peers; fixing all cross-loadings to zero in CFA may therefore inflate the latent correlation between these domains. Target ESEM retained the proposed content map while allowing modest contextual overlap. The bifactor results nevertheless showed that covariance was not distributed evenly across all five domains. The slightly better fit of M7 did not justify a total score because its general factor largely represented extracurricular and peer items and ω_H was only 0.484.

Score use and multisource interpretation

Researchers should calculate five subscale means using the scoring key in Additional file 1: Table S1. A high curricular score indicates supportive physical education experiences; it should not be taken as evidence that extracurricular access or institutional implementation is also strong. Conversely, low

institutional support may identify problems involving scheduling, access, accommodation or feedback channels even when peer relationships are supportive. At present, the measure is most appropriate for group-level research, school needs assessment and programme evaluation. It should not be used for high-stakes decisions about individual students or as a national surveillance indicator without further evidence.

The lack of student–teacher convergence indicates that students’ experience should not be used as a proxy for policy compliance, nor should teacher reports be treated as a criterion that necessarily supersedes student perceptions. Teachers may describe arrangements that are formally available, whereas students report whether those arrangements are visible, accessible and experienced in practice. Future evaluations may retain both perspectives and model discrepancies rather than assuming high agreement [19, 20]. Larger teacher samples and direct indicators from school records are needed before strong conclusions about cross-informant correspondence can be drawn.

Strengths and limitations

Strengths include the complementary expertise of the content reviewers, item-focused cognitive evidence, independent exploratory and confirmatory samples, variation across schools and grades, ordinal modelling with class-cluster correction, ESEM and bifactor score diagnostics, and explicit reporting of non-convergence and null teacher correlations. The version audit also ensured that the non-equivalent item 39 was excluded consistently before scoring and structural analysis.

Several limitations constrain interpretation. First, schools were recruited through existing contacts rather than probability sampling; school selection, unverified geographical descriptors in Study 2, grade representation and left-behind status limit national generalisation. Second, the studies were not preregistered and no effect-size-based a priori Monte Carlo power analysis was undertaken. Third, Study 1 provided expert and item-focused cognitive-interview evidence but was not a saturation-based qualitative concept-elicitation study. Fourth, measurement invariance across sex, grade, region, urban–rural background, activity level and left-behind status was not examined. Fifth, institutional support had marginal omega in Study 3, and loading-based average variance extracted from M3 was below 0.50 for curricular, cultural and institutional support (0.219–0.282). Sixth, the 14-day mean changes, particularly the decline in extracurricular support, and lower Time 2 alpha values indicate occasion-related variation. Seventh, only 19 teacher-defined units were available, and no cross-informant convergence was detected. Eighth, because the administered item 39 differed from the intended item, evidence applies only to the 38 wording-identical common items. Ninth, the class-level correction did not account for classes nested within only six schools, leaving school-level dependence as a source of uncertainty.

Future studies should preregister multilevel measurement models, test subgroup and longitudinal invariance, evaluate sensitivity to school-based interventions and examine associations with objectively measured and self-reported physical activity, autonomous motivation, habit and mental health. Larger teacher samples and direct policy or facility indicators are needed to establish how students’ experiential profiles correspond to organisational conditions. Cross-cultural adaptation will require new translation

and cognitive evidence; the English reporting translation should not be administered as though it were a validated English form.

Conclusions

The SBPASS-SR38 provides an initial student-centred framework for profiling curricular, extracurricular activity, peer, cultural and institutional support for physical activity at school. Its structure was recovered in an exploratory sample and supported independently by ordinal modelling, with five-factor target ESEM providing the best balance of fit, loading pattern and domain-level interpretability. Reliability was generally acceptable but weaker for institutional support; systematic 14-day mean changes and the absence of student–teacher convergence further limit score interpretation. On the available evidence, the five subscales rather than a total score are the appropriate outputs. Further work is required before the scale is used for subgroup comparisons, prediction, intervention evaluation or surveillance.

Abbreviations

CFA: confirmatory factor analysis;

CFI: comparative fit index;

CI: confidence interval;

ECV: explained common variance;

EFA: exploratory factor analysis;

ESEM: exploratory structural equation modelling;

HTMT: heterotrait–monotrait ratio;

ICC: intraclass correlation coefficient;

IECV: item explained common variance;

I. CVI: item-level content validity index;

II. PE: physical education;

III. PUC: percentage of uncontaminated correlations;

IV. RMSEA: root mean square error of approximation;

V. S-CVI/Ave: scale-level average content validity index;

VI. SBPASS-SR38: School-Based Physical Activity Support Scale–Student Report, 38-item form;

VII. SRMR: standardised root mean square residual;

VIII. TLI: Tucker–Lewis index; WLSMV: weighted least squares mean- and variance-adjusted estimation.

Declarations

Ethics approval and consent to participate

The study was conducted in accordance with the Declaration of Helsinki and all relevant institutional guidelines and regulations. The protocol was approved before data collection by the Research Ethics Committee of Hengyang Normal University (approval no. 2026lllw023; 20 March 2026). All adult participants, including the expert reviewers and physical education teachers, provided electronically signed informed consent. For each adolescent participant, a parent or legal guardian signed an electronic informed-consent form, and the adolescent signed an electronic assent form. Permission was also obtained from all participating schools.

Clinical trial registration

Clinical trial number: not applicable.

Consent for publication

Not applicable. The manuscript contains no identifiable individual participant data.

Availability of data and materials

The de-identified participant data and analysis materials are not publicly available because access is constrained by the ethics protocol, participating-school agreements and participant-protection requirements, but they are available from the corresponding author on reasonable request and subject to the required approvals. The protected project archive includes the scoring key, version-audit record, reproducible analysis scripts and final Mplus 8.3 input and output files.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the Third Hunan Provincial Basic Education Teaching Reform Research Key Project (25JGZD0091), the Hunan Provincial University Ideological and Political Education Research General Project (24SZ37), the Key Scientific Research Project of the Hunan Provincial Department of Education (25A0483), and an Externally Commissioned Research Project administered by Hengyang

Normal University (HXKY2025028). The funders had no role in study design, data collection, analysis, interpretation, manuscript preparation or the decision to submit the work for publication.

Authors' contributions

HS and LX conceived and designed the study. HS, XJ, SY, LL and HD contributed to instrument development, fieldwork and data management. HS and LX performed the analyses and drafted the manuscript. All authors interpreted the findings, critically revised the manuscript, read and approved the final version, and agree to be accountable for the work.

Acknowledgements

The authors thank the participating students, parents and guardians, schools, physical education teachers and expert reviewers for their contributions to the research.

References

1. Bull FC, Al-Ansari SS, Biddle S, Borodulin K, Buman MP, Cardon G, et al. World Health Organization 2020 guidelines on physical activity and sedentary behaviour. *Br J Sports Med*. 2020;54(24):1451–62. <https://doi.org/10.1136/bjsports-2020-102955>.
2. Guthold R, Stevens GA, Riley LM, Bull FC. Global trends in insufficient physical activity among adolescents: a pooled analysis of 298 population-based surveys with 1.6 million participants. *Lancet Child Adolesc Health*. 2020;4(1):23–35. [https://doi.org/10.1016/S2352-4642\(19\)30323-2](https://doi.org/10.1016/S2352-4642(19)30323-2).
3. van Sluijs EMF, Ekelund U, Crochemore-Silva I, Guthold R, Ha A, Lubans D, et al. Physical activity behaviours in adolescence: current evidence and opportunities for intervention. *Lancet*. 2021;398(10298):429–42. [https://doi.org/10.1016/S0140-6736\(21\)01259-9](https://doi.org/10.1016/S0140-6736(21)01259-9).
4. Morton KL, Atkin AJ, Corder K, Suhrcke M, van Sluijs EMF. The school environment and adolescent physical activity and sedentary behaviour: a mixed-studies systematic review. *Obes Rev*. 2016;17(2):142–58. <https://doi.org/10.1111/obr.12352>.
5. Rajbhandari-Thapa J, Metzger I, Ingels J, Thapa K, Chiang K. School climate-related determinants of physical activity among high school girls and boys. *J Adolesc*. 2022;94:642–55. <https://doi.org/10.1002/jad.12052>.
6. McLeroy KR, Bibeau D, Steckler A, Glanz K. An ecological perspective on health promotion programs. *Health Educ Q*. 1988;15(4):351–77. <https://doi.org/10.1177/109019818801500401>.
7. Sallis JF, Cervero RB, Ascher W, Henderson KA, Kraft MK, Kerr J. An ecological approach to creating active living communities. *Annu Rev Public Health*. 2006;27:297–322. <https://doi.org/10.1146/annurev.publhealth.27.021405.102100>.

8. Deci EL, Ryan RM. The 'what' and 'why' of goal pursuits: human needs and the self-determination of behavior. *Psychol Inq.* 2000;11(4):227–68. https://doi.org/10.1207/S15327965PLI1104_01.
9. Vasconcellos D, Parker PD, Hilland T, Cinelli R, Owen KB, Kapsal N, et al. Self-determination theory applied to physical education: a systematic review and meta-analysis. *J Educ Psychol.* 2020;112(7):1444–69. <https://doi.org/10.1037/edu0000420>.
10. White RL, Bennie A, Vasconcellos D, Cinelli R, Hilland T, Owen KB, et al. Self-determination theory in physical education: a systematic review of qualitative studies. *Teach Teach Educ.* 2021;99:103247. <https://doi.org/10.1016/j.tate.2020.103247>.
11. Fitzgerald A, Fitzgerald N, Aherne C. Do peers matter? A review of peer and/or friends' influence on physical activity among American adolescents. *J Adolesc.* 2012;35(4):941–58. <https://doi.org/10.1016/j.adolescence.2012.01.002>.
12. Robertson-Wilson JE, Lévesque L, Holden RR. Development of a questionnaire assessing school physical activity environment. *Meas Phys Educ Exerc Sci.* 2007;11(2):93–107. <https://doi.org/10.1080/10913670701294088>.
13. Martin JJ, McCaughy N, Flory S, Murphy A, Wisdom K. Validity and reliability of the School Physical Activity Environment Questionnaire. *Meas Phys Educ Exerc Sci.* 2011;15(4):274–82. <https://doi.org/10.1080/1091367X.2011.611462>.
14. Leung KM. Redevelopment and examination of the psychometric properties of school physical activity environment in Chinese. *BMC Public Health.* 2025;25:2348. <https://doi.org/10.1186/s12889-025-23571-w>.
15. Lounsbery MAF, McKenzie TL, Morrow JR, Holt KA, Budnar RG. School Physical Activity Policy Assessment. *J Phys Act Health.* 2013;10(4):496–503. <https://doi.org/10.1123/jpah.10.4.496>.
16. Webster CA, Rink JE, Carson RL, Moon J, Gaudreault KL. The Comprehensive School Physical Activity Program model: a proposed illustrative supplement to help move the needle on youth physical activity. *Kinesiol Rev.* 2020;9(2):112–21. <https://doi.org/10.1123/kr.2019-0048>.
17. He L, Ayub AFBM, Amri SB. Development and validation of a questionnaire to assess the implementation of physical education programs in Chinese junior high schools. *BMC Public Health.* 2024;24:2387. <https://doi.org/10.1186/s12889-024-19844-5>.
18. Dai S, Wu Y, Yan J, Zeng L. Development and validation of the Chinese school environment influencing physical activity habits scale. *Front Educ.* 2025;10:1534843. <https://doi.org/10.3389/feduc.2025.1534843>.
19. De Los Reyes A, Augenstein TM, Wang M, Thomas SA, Drabick DAG, Burgers DE, et al. The validity of the multi-informant approach to assessing child and adolescent mental health. *Psychol Bull.* 2015;141(4):858–900. <https://doi.org/10.1037/a0038498>.
20. Kraemer HC, Measelle JR, Ablow JC, Essex MJ, Boyce WT, Kupfer DJ. A new approach to integrating data from multiple informants in psychiatric assessment and research: mixing and matching contexts and perspectives. *Am J Psychiatry.* 2003;160(9):1566–77. <https://doi.org/10.1176/appi.ajp.160.9.1566>.

21. American Educational Research Association, American Psychological Association. National Council on Measurement in Education. Standards for educational and psychological testing. Washington, DC: American Educational Research Association; 2014.
22. Boateng GO, Neilands TB, Frongillo EA, Melgar-Quinonez HR, Young SL. Best practices for developing and validating scales for health, social, and behavioral research: a primer. *Front Public Health*. 2018;6:149. <https://doi.org/10.3389/fpubh.2018.00149>.
23. Worthington RL, Whittaker TA. Scale development research: a content analysis and recommendations for best practices. *Couns Psychol*. 2006;34(6):806–38. <https://doi.org/10.1177/0011000006288127>.
24. Muthén LK, Muthén BO. How to use a Monte Carlo study to decide on sample size and determine power. *Struct Equ Model*. 2002;9(4):599–620. https://doi.org/10.1207/S15328007SEM0904_8.
25. Wolf EJ, Harrington KM, Clark SL, Miller MW. Sample size requirements for structural equation models: an evaluation of power, bias, and solution propriety. *Educ Psychol Meas*. 2013;73(6):913–34. <https://doi.org/10.1177/0013164413495237>.
26. DeVellis RF, Thorpe CT. Scale development: theory and applications. 5th ed. Thousand Oaks: Sage; 2021.
27. Polit DF, Beck CT, Owen SV. Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Res Nurs Health*. 2007;30(4):459–67. <https://doi.org/10.1002/nur.20199>.
28. Beatty PC, Willis GB. Research synthesis: the practice of cognitive interviewing. *Public Opin Q*. 2007;71(2):287–311. <https://doi.org/10.1093/poq/nfm006>.
29. Willis GB. Cognitive interviewing: a tool for improving questionnaire design. Thousand Oaks: Sage; 2005.
30. Horn JL. A rationale and test for the number of factors in factor analysis. *Psychometrika*. 1965;30(2):179–85. <https://doi.org/10.1007/BF02289447>.
31. Fabrigar LR, Wegener DT, MacCallum RC, Strahan EJ. Evaluating the use of exploratory factor analysis in psychological research. *Psychol Methods*. 1999;4(3):272–99. <https://doi.org/10.1037/1082-989X.4.3.272>.
32. Schmitt TA. Current methodological considerations in exploratory and confirmatory factor analysis. *J Psychoeduc Assess*. 2011;29(4):304–21. <https://doi.org/10.1177/0734282911406653>.
33. Dunn TJ, Baguley T, Brunsden V. From alpha to omega: a practical solution to the pervasive problem of internal consistency estimation. *Br J Psychol*. 2014;105(3):399–412. <https://doi.org/10.1111/bjop.12046>.
34. McNeish D. Thanks coefficient alpha, we'll take it from here. *Psychol Methods*. 2018;23(3):412–33. <https://doi.org/10.1037/met0000144>.
35. Henseler J, Ringle CM, Sarstedt M. A new criterion for assessing discriminant validity in variance-based structural equation modeling. *J Acad Mark Sci*. 2015;43:115–35. <https://doi.org/10.1007/s11747-014-0403-8>.

36. Flora DB, Curran PJ. An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychol Methods*. 2004;9(4):466–91. <https://doi.org/10.1037/1082-989X.9.4.466>.
37. Li CH. Confirmatory factor analysis with ordinal data: comparing robust maximum likelihood and diagonally weighted least squares. *Behav Res Methods*. 2016;48:936–49. <https://doi.org/10.3758/s13428-015-0619-7>.
38. Asparouhov T, Muthén B. Exploratory structural equation modeling. *Struct Equ Model*. 2009;16(3):397–438. <https://doi.org/10.1080/10705510903008204>.
39. Marsh HW, Morin AJS, Parker PD, Kaur G. Exploratory structural equation modeling: an integration of the best features of exploratory and confirmatory factor analysis. *Annu Rev Clin Psychol*. 2014;10:85–110. <https://doi.org/10.1146/annurev-clinpsy-032813-153700>.
40. Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Model*. 1999;6(1):1–55. <https://doi.org/10.1080/10705519909540118>.
41. Marsh HW, Hau KT, Wen Z. In search of golden rules: comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's findings. *Struct Equ Model*. 2004;11(3):320–41. https://doi.org/10.1207/S15328007SEM1103_2.
42. Rodriguez A, Reise SP, Haviland MG. Applying bifactor statistical indices in the evaluation of psychological measures. *J Pers Assess*. 2016;98(3):223–37. <https://doi.org/10.1080/00223891.2015.1089249>.
43. Rodriguez A, Reise SP, Haviland MG. Evaluating bifactor models: calculating and interpreting statistical indices. *Psychol Methods*. 2016;21(2):137–50. <https://doi.org/10.1037/met0000045>.
44. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15(2):155–63. <https://doi.org/10.1016/j.jcm.2016.02.012>.
45. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86(2):420–8. <https://doi.org/10.1037/0033-2909.86.2.420>.

Figures

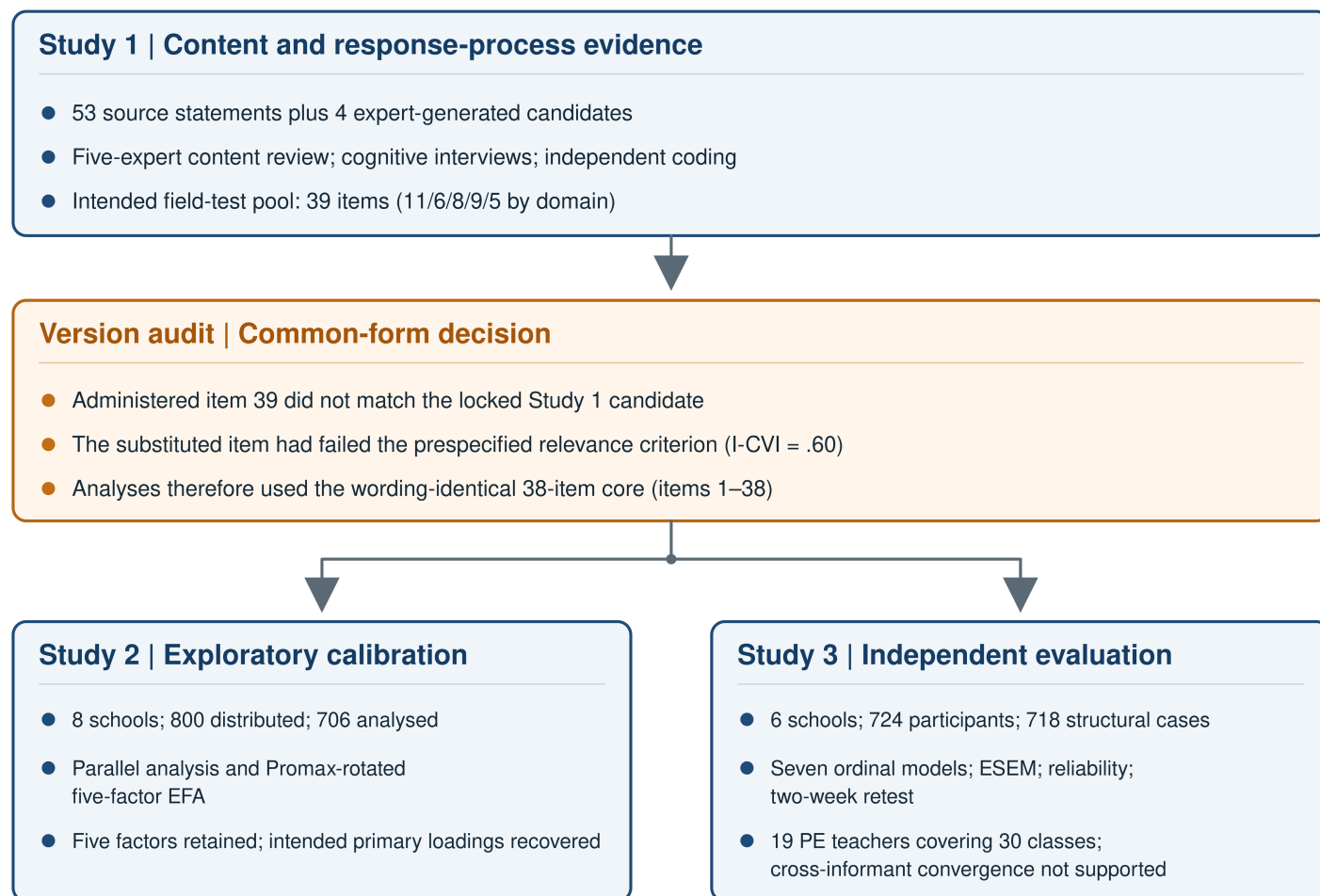


Figure 1

Development, version audit and three-study evaluation sequence

The 38-item common form was identified before structural modelling by comparing the locked Study 1 item record with the forms administered in Studies 2 and 3. The diagram was prepared with OpenAI ChatGPT/Codex from author-specified content and verified against the study records. EFA, exploratory factor analysis; ESEM, exploratory structural equation modelling; I-CVI, item-level content validity index; PE, physical education.

Supplementary Files

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

- [BMCPublicHealthSupplementSBPASSSR38REVISED CLEAN.docx](#)