Jurnal Penelitian dan Pengkajian Ilmu Pendidikan: e-Saintika
https://journal-center.litpam.com/index.php/e-Saintika/index

July 2026 Vol. 10, No. 2
e-ISSN: 2615-6881
pp. 868-902

Psychometric Validation of a Two-Factor Self-Report Instrument for Science Process Skills-Related Behaviors in Indonesian High School Students

Yuli Astutik, * Siti Zubaidah, Bagus Priambodo

Biology Education Department, Faculty of Mathematics and Science Education, Universitas Negeri Malang. Jl. Semarang No. 5, Malang 65145, Indonesia.

*Corresponding Author e-mail: siti.zubaidah.fmipa@um.ac.id

Received: May 2026; Revised: June 2026; Published: July 2026

Abstract

Science process skills (SPS) are central to scientific inquiry, yet instruments designed to assess SPS-related constructs require empirical evidence that their measurement structure corresponds to their intended interpretation. In Indonesia, SPS instruments have often emphasized expert-based content validation, while evidence concerning the latent structure of self-reported SPS-related behaviors remains limited. This study developed and psychometrically evaluated a two-dimensional self-report instrument assessing senior high school students' reported engagement in behaviors associated with Basic Science Process Skills (BSPS) and Integrated Science Process Skills (ISPS). A total of 830 tenth-grade students from East Java, Indonesia, were allocated to three independent samples: preliminary analysis ($n = 166$), exploratory factor analysis (EFA; $n = 339$), and confirmatory factor analysis (CFA; $n = 325$). The EFA data were suitable for factor analysis (KMO = 0.964; Bartlett's $\chi^2(595) = 7{,}345.301$, $p < .001$). Parallel analysis, eigenvalue examination, scree-plot inspection, and theoretical interpretability supported a two-factor solution. Seven items were removed, yielding a 28-item instrument that explained 49.8% of the total variance. Because Mardia's tests indicated multivariate nonnormality, CFA was estimated using robust maximum likelihood. The correlated two-factor model demonstrated acceptable fit, $\chi^2(349) = 768.321$, $p < .001$, CFI = 0.925, TLI = 0.919, RMSEA = 0.061, 90% CI [0.055, 0.067], and SRMR = 0.045. AVE values were 0.510 for BSPS and 0.540 for ISPS, while HTMT was 0.805. Cronbach's alpha and McDonald's omega ranged from 0.928 to 0.962, with composite reliability ranging from 0.979 to 0.993. The findings support two closely related yet empirically distinguishable dimensions of self-reported SPS-related behavior. The instrument should therefore be interpreted as assessing reported behavioral engagement rather than demonstrated SPS competence. Generalization beyond comparable student populations should remain cautious because participants were recruited purposively from East Java.

Keywords: Science Process Skills; Self-Report; Psychometric Validation; Exploratory Factor Analysis; Confirmatory Factor Analysis; Senior High School Students

How to Cite: Astutik, Y., Zubaidah, S., & Priambodo, B. (2026). Psychometric Validation of a Two-Factor Self-Report Instrument for Science Process Skills-Related Behaviors in Indonesian High School Students. *Jurnal Penelitian Dan Pengkajian Ilmu Pendidikan: E-Saintika, 10*(2), 868-902. https://doi.org/10.36312/e-saintika.v10i2.6140

10.36312/e-saintika.v10i2.6140

INTRODUCTION

Science Process Skills (SPS) refer to the competencies that enable individuals to acquire, evaluate, and apply scientific knowledge through activities such as observing, measuring, classifying, formulating hypotheses, designing experiments, interpreting data, and communicating findings (Rezba et al., 2022). These competencies contribute to rational thinking, responsiveness to evidence, and the ability to address complex problems (Dolapcioglu & Subasi, 2022; Mustofa et al., 2021). Their importance is consistent with contemporary science education, which expects students not only to acquire scientific knowledge but also to investigate phenomena, evaluate evidence, and apply scientific reasoning to real-world problems (Fitriyana et al., 2025; Wan et al., 2024; Amina et al., 2022). SPS therefore provide an important connection between conceptual scientific knowledge and its application through systematic and contextually meaningful scientific activity (Krell et al., 2022; Idris et al., 2022).

The SPS framework distinguishes two theoretically related dimensions: Basic Science Process Skills (BSPS) and Integrated Science Process Skills (ISPS) (Rezba et al., 2022). BSPS comprise observing, predicting, classifying, measuring, inferring, and communicating, whereas ISPS involve controlling variables, defining variables operationally, formulating hypotheses, interpreting data, and conducting experiments. The integrated processes require students to coordinate several foundational processes when planning, conducting, and interpreting scientific investigations. Accordingly, BSPS and ISPS are better conceptualized as related dimensions that differ in the complexity and coordination of the scientific behaviors they involve rather than as completely independent categories. This view is consistent with the proposition that proficiency in foundational process skills supports engagement in more complex integrated scientific activity (Gizaw & Sota, 2023).

This distinction is particularly relevant at the senior high school level, where students are increasingly expected to formulate hypotheses, identify and control variables, analyze empirical evidence, and draw conclusions from investigations. Empirical studies have indicated that competence in BSPS is associated with the development of more complex integrated process skills (Jardinico & Linaugo, 2023; Kamarudin et al., 2022). At the same time, students’ demonstrated SPS performance remains uneven in several educational contexts. Pujiyanto et al. (2025), for example, reported very low performance in observing, predicting, and measuring and low performance in communication among senior high school students. Such findings underscore the importance of valid SPS assessment, but they also highlight the need to distinguish clearly between what different forms of assessment actually measure.

SPS-related assessment may target different aspects of students’ engagement with scientific activity, including knowledge, demonstrated performance, confidence, and behavioral frequency. SPS knowledge refers to conceptual understanding of scientific procedures and principles, whereas demonstrated SPS performance concerns students’ ability to execute scientific processes successfully through tests, investigations, or laboratory-based tasks (Avan & Aydınlı, 2026). Confidence concerns students’ perceived capability or ease in carrying out a task (Liu et al., 2024), while behavioral frequency concerns how often students report engaging in SPS-related activities. These measurement targets may be related, but they are not interchangeable because self-report and performance-based measures capture different forms of evidence about scientific activity (Borgonovi, Pokropek, et al., 2023). Performance-oriented instruments assess whether students can demonstrate particular scientific processes, whereas self-report instruments capture students’ reported tendencies, experiences, or engagement in behaviors associated with those processes (Borgonovi, Ferrara, et al., 2023; Ulfah & Supahar, 2026).

This distinction has direct implications for item design and score interpretation. Statements such as “I find it easy” or “I can do this without difficulty” primarily reflect confidence or perceived competence rather than behavioral frequency. In contrast, frequency-based self-report items should describe specific SPS-related actions that can be meaningfully answered using response categories such as Always, Often, Sometimes, Rarely, and Never. The instrument developed in the present study therefore does not assess whether students can successfully demonstrate SPS under controlled testing or laboratory conditions. Instead, it targets students' self-reported frequency of engagement in behaviors associated with BSPS and ISPS during science learning. Scores should consequently be interpreted as indicators of self-reported SPS-related behavioral tendencies rather than as direct measures of SPS knowledge or demonstrated competence.

The need for rigorous measurement is especially relevant to SPS instrument development in Indonesia. Several instruments have been developed for specific science topics, but some have relied predominantly on expert judgment and content-oriented validation. Habibah and Fuadiyah (2025), for example, developed an SPS assessment instrument for senior high school biology using self-evaluation, expert review, and one-to-one evaluation without empirically examining its latent structure. Comparable development procedures were reported for SPS instruments addressing the respiratory system (Yanti & Fuadiyah, 2025) and circulatory system (Junianda & Fuadiyah, 2025). Although expert evaluation is essential for assessing item relevance and content representation, it cannot by itself determine whether the theoretically proposed dimensional structure is reproduced in responses from the target population. Consequently, evidence regarding the empirical BSPS–ISPS structure of SPS-related self-report behavior remains limited in the Indonesian context.

Content validity and construct validity therefore provide complementary but distinct forms of measurement evidence. Content validity concerns whether instrument items adequately represent the intended domain, whereas construct validity concerns whether relationships among item responses are consistent with the proposed theoretical structure (Obilor & Miwari, 2022). Empirical examination of dimensionality is therefore necessary in addition to expert evaluation. Exploratory Factor Analysis (EFA) can be used to examine how items organize empirically without imposing a fully specified measurement model, while Confirmatory Factor Analysis (CFA) can subsequently evaluate whether the resulting structure is reproduced in an independent sample. Marar et al. (2023), for example, applied EFA and CFA to separate samples when developing and validating an instrument, illustrating the value of distinguishing exploratory structure identification from confirmatory model evaluation.

Recent SPS-related instruments also demonstrate considerable diversity in both measurement format and psychometric approach. García-Carmona et al. (2024) validated a performance-oriented instrument assessing basic SPS through expert review, pilot testing, and empirical administration. Karim et al. (2024) employed the Rasch Measurement Model to evaluate an SPS competency instrument based on students' reported perceptions, whereas Avan and Aydınlı (2026) used open-ended performance tasks to assess science-related cognitive skills. These approaches illustrate that SPS-related constructs may be assessed using performance tests, constructed-response tasks, self-perception instruments, and item-level psychometric models. However, these measurement formats address different inferential targets and do not directly answer whether self-reported behavioral tendencies related to BSPS and ISPS form a reproducible correlated two-factor structure.

A similar gap remains within Indonesian psychometric research. Komisia et al. (2025) developed an SPS assessment instrument for senior high school chemistry and reported evidence of content validity, but the instrument focused on test-based skill attainment and did not examine a latent BSPS–ISPS structure through sequential EFA and CFA. Studies employing Rasch analysis or item-response approaches provide valuable evidence regarding item fit, difficulty, discrimination, and measurement precision, but such analyses address different psychometric questions from those involved in testing whether two theoretically proposed domains are represented as correlated latent dimensions. Thus, the empirical dimensionality of self-reported BSPS- and ISPS-related behaviors remains insufficiently examined.

The contribution of the present study therefore lies in three related aspects. First, it operationalizes SPS as self-reported behavioral engagement rather than demonstrated performance, allowing the resulting scores to be interpreted according to the frequency of reported SPS-related actions. Second, it empirically examines whether BSPS- and ISPS-related behavioral tendencies emerge as distinct but correlated dimensions rather than assuming that the theoretical classification will necessarily be reproduced in student responses. Third, it evaluates the resulting factor structure through sequential EFA and CFA using independent samples, thereby separating exploratory structure identification from confirmatory cross-validation. The novelty of the study therefore lies not in the use of EFA or CFA themselves, but in applying these procedures to test the dimensional organization of a behaviorally framed SPS self-report instrument.

Based on the theoretical distinction between BSPS and ISPS and their expected functional interrelationship, the study hypothesized that students' responses would be represented by two positively correlated but empirically distinguishable latent dimensions broadly corresponding to BSPS- and ISPS-related behavioral tendencies.

Accordingly, this study aimed to develop and examine the psychometric properties of a two-dimensional self-report instrument assessing SPS-related behavioral tendencies among senior high school students in East Java, Indonesia. The study addressed the following research questions:

1. What evidence supports the factorial validity, convergent validity, discriminant validity, and internal consistency reliability of the developed self-report SPS instrument?
2. To what extent do EFA and CFA support a correlated two-factor structure representing self-reported BSPS- and ISPS-related behavioral tendencies?

By addressing these questions, the study evaluates whether the theoretically proposed BSPS–ISPS distinction is reflected in students' self-reported behavioral response patterns while maintaining a clear distinction between reported engagement and demonstrated SPS performance.

METHOD

Research Design and Psychometric Validation Framework

This study employed a quantitative instrument-development and psychometric-validation design following the scale-development framework proposed by Boateng et al. (2018) and recent methodological recommendations for instrument validation (Stefana et al., 2025; Lim et al., 2026; Kline, 2023). The instrument was grounded in the two-dimensional Science Process Skills (SPS) framework proposed by Rezba et al.

(2022), comprising Basic Science Process Skills (BSPS) and Integrated Science Process Skills (ISPS).

As illustrated in Figure 1, the development and validation process proceeded sequentially through six stages: (1) literature review and theoretical specification, (2) item development based on BSPS and ISPS indicators, (3) expert-based content validation and item refinement, (4) preliminary item analysis, (5) exploratory factor analysis (EFA), and (6) confirmatory factor analysis (CFA). This sequential design enabled the proposed measurement structure to be examined progressively, beginning with theoretical and content-based evidence and followed by empirical exploration and independent cross-validation.

Preliminary analyses were conducted using IBM SPSS Statistics version 27 to examine data quality and initial item characteristics. EFA and CFA were subsequently conducted in JASP using independent samples. The finalized CFA measurement model was additionally visualized using AMOS to display standardized factor loadings, error terms, and relationships between latent variables; AMOS was used for graphical representation of the finalized model rather than as a separate model-development stage. EFA and CFA therefore served complementary but distinct purposes: EFA identified the empirical latent structure, whereas CFA evaluated whether the resulting structure could be reproduced in an independent sample (Flora & Flake, 2017; Hair et al., 2019; Marar et al., 2023). The CFA stage further evaluated model fit, standardized factor loadings, convergent validity, discriminant validity, and reliability in accordance with contemporary psychometric reporting recommendations (Cheung et al., 2024).

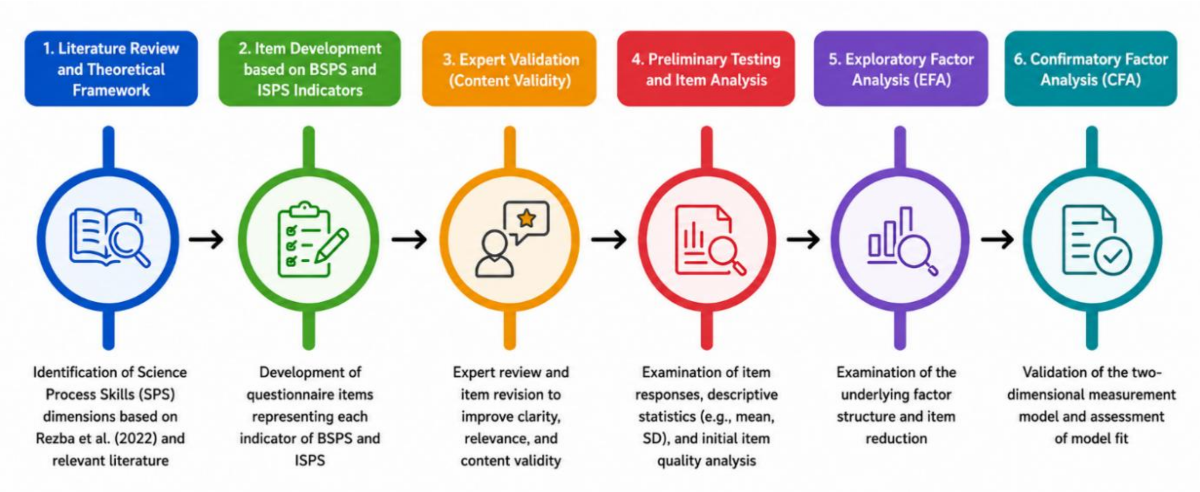

Figure 1. Stages of the Development and Psychometric Validation of the Self-Report Science Process Skills Instrument

Participants and Sample Allocation

The study involved 830 tenth-grade students recruited through purposive sampling from senior high schools in East Java, Indonesia. To preserve independence across the development and validation stages, participants were allocated to three non-overlapping analytical samples. The preliminary-analysis sample comprised 166 students and was used to examine descriptive response characteristics and initial item performance. A separate sample of 339 students was used to conduct EFA on the initial 35-item instrument. Following EFA-based item refinement, the resulting 28-item structure was subsequently evaluated through CFA using an independent sample of 325 students. No participant contributed data to more than one analytical stage.

This sequence is important because the 28-item structure was derived from the EFA of the initial 35-item pool rather than being specified before the EFA stage. Accordingly, the EFA sample corresponded to approximately 9.69 participants per initial item and 12.11 participants per item in the final retained EFA solution, whereas the CFA sample corresponded to approximately 11.61 participants per retained item. These ratios fall within commonly used practical ranges for factor analysis. Nevertheless, sample adequacy was not judged solely from participant-to-item ratios because contemporary methodological evidence emphasizes that factor stability also depends on factor loadings, communalities, and model complexity (Lorenzo-Seva & Ferrando, 2024; Stefana et al., 2025).

Although no formal a priori power analysis was conducted, the use of separate samples for exploratory structure identification and confirmatory evaluation reduced the risk of evaluating the measurement structure in the same observations from which it was derived. Participant characteristics across the three analytical samples, including gender, age range, grade level, and school type, are summarized in Table 1.

Table 1. Demographic Characteristics of Participants

Characteristic	Preliminary Testing (n = 166)	EFA (n = 339)	CFA (n = 325)
Schools, n	2	3	3
Gender, n (%)			
Male	74 (44.6%)	136 (40.1%)	146 (44.9%)
Female	92 (55.4%)	203 (59.9%)	179 (55.1%)
Age range	14–16	14–16	14–16
Grade	Grade 10	Grade 10	Grade 10
Type of school, n (%)			
Islamic Senior High School (MA)	40 (24.1%)	131 (38.6%)	114 (35.1%)
Public Senior High School (SMA)	126 (75.9%)	208 (61.4%)	211 (64.9%)

Instrument Development

Construct Specification and Initial Item Pool

The SPS instrument was designed as a self-report questionnaire assessing students' reported frequency of engagement in behaviors associated with science process skills during science-learning activities. Thus, the intended measurement target was self-reported SPS-related behavioral tendencies rather than students' demonstrated ability to perform SPS under controlled testing or laboratory conditions.

Instrument development followed the two-dimensional framework proposed by Rezba et al. (2022). The BSPS dimension comprised six indicators: observing, predicting, classifying, measuring, inferring, and communicating. The ISPS dimension comprised five indicators: controlling variables, defining variables operationally, formulating hypotheses, interpreting data, and conducting experiments. The two theoretical dimensions were operationalized into 11 indicators represented by an initial pool of 35 self-report items.

Because the complete construct blueprint and item wording are important for transparency but occupy substantial space, the initial 35-item specification is provided in Appendix A. Table A reports each dimension, indicator, indicator description, item number, and original item statement.

Response Format and Score Interpretation

All items were formulated as behavioral statements reflecting activities that students could encounter during science learning. Responses were recorded using a five-point Likert-type frequency scale: Always, Often, Sometimes, Rarely, and Never. The use of frequency-based categories was aligned with the intended measurement target because the instrument was designed to capture how often students reported engaging in SPS-related behaviors rather than whether they could successfully execute the corresponding scientific skills in a direct performance assessment.

Accordingly, responses represented reported engagement in activities such as observing phenomena, measuring, identifying variables, interpreting experimental data, formulating hypotheses, communicating evidence, and conducting investigations. The five-category format also provided sufficient response variability for differentiating students' reported behavioral engagement and supporting subsequent psychometric analyses (Tondong et al., 2024).

Content Validation and Item Refinement

Expert Evaluation

The initial 35-item instrument was evaluated by three experts with expertise in science education, biology education, curriculum, and educational assessment. The purpose of this stage was to assess whether each item adequately represented its intended SPS indicator and whether the wording was understandable and contextually appropriate for senior high school students. The experts evaluated the items in terms of relevance, clarity, appropriateness of terminology, and suitability for science-learning contexts. The characteristics of the expert panel are presented in Table 2.

Table 2. Characteristics of Content-Validity Experts

Expert	Designation	Expertise	Organization	Experience
Expert 1	Assessment Expert	Biology Education	University A	6 years
Expert 2	Assessment Expert	Science Education	University B	7 years
Expert 3	Curriculum Specialist	Science Education	University C	11 years

Content Validity Index

Expert judgments of item relevance were quantified using the Content Validity Index (CVI), a commonly used approach for summarizing expert agreement regarding the relevance of scale items (Polit et al., 2007). Each item was rated on a five-point scale ranging from 1 (not relevant) to 5 (highly relevant). For CVI calculation, ratings of 4 or 5 were classified as relevant and coded as 1, whereas ratings of 1, 2, or 3 were classified as not relevant and coded as 0.

The Item-level Content Validity Index (I-CVI) was calculated as the proportion of participating experts who classified each item as relevant. Scale-level content validity was summarized using both the average method (S-CVI/Ave) and universal-agreement method (S-CVI/UA). S-CVI/Ave represented the mean I-CVI across the 35 items, whereas S-CVI/UA represented the proportion of items for which all three experts agreed that the item was relevant.

All 35 items received relevant ratings from all three experts, resulting in I-CVI values of 1.00 for every item and S-CVI/Ave and S-CVI/UA values of 1.00. These results indicate complete agreement within the participating expert panel and provide favorable content-validity evidence for the initial item pool. Because the item-level matrix contains identical agreement patterns across all 35 items, the complete CVI results are provided in Appendix B (Table B) rather than reproduced in the main text.

In addition to the quantitative ratings, the experts provided qualitative feedback concerning item wording, terminology, and contextual suitability. Their recommendations were used to simplify selected statements, improve clarity, adjust terminology to the senior high school science context, and better align item wording with students' learning experiences without changing the intended indicator content. A summary of the qualitative feedback and corresponding revision actions is provided in Appendix C (Table C).

Data Analysis

Preliminary Data Screening and Item Analysis

Preliminary data screening was conducted using IBM SPSS Statistics version 27. Screening included examination of missing responses, response patterns, potential outliers, item means, standard deviations, and corrected item–total correlations. The purpose of this preliminary stage was to identify problematic response patterns and determine whether the initial items demonstrated sufficient response variability and association with the broader item pool to proceed to factor analysis.

The preliminary item analysis was treated as a screening procedure rather than as evidence establishing the dimensional structure of the instrument. Consequently, the final determination of item retention and factor membership was based on the subsequent EFA and CFA stages.

Exploratory Factor Analysis

EFA was conducted in JASP using responses from the independent EFA sample ($n = 339$). The analysis began with the initial 35-item pool. The adequacy of the data for factor analysis was evaluated using the Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy and Bartlett's Test of Sphericity.

Promax oblique rotation was used because the theoretical framework does not assume BSPS and ISPS to be statistically independent; rather, foundational SPS-related behaviors are expected to be associated with more integrated scientific processes (Rezba et al., 2022). Factor retention was determined using multiple complementary sources of evidence: parallel analysis, eigenvalue examination, visual inspection of the scree plot, and theoretical interpretability of the resulting factors. This approach avoided reliance on a single numerical factor-retention rule and allowed the empirical solution to be evaluated in relation to the proposed SPS framework.

Items were evaluated using the factor-loading and cross-loading criteria applied in the EFA. Seven items that did not contribute sufficiently to a clear and interpretable factor structure were removed during refinement, yielding the final 28-item solution. Detailed factor-retention results, eigenvalues, explained variance, retained-item loadings, and the scree plot are reported in the Results section because they represent empirical findings rather than analytical procedures.

The use of EFA for empirical structure identification followed by independent CFA for cross-validation is consistent with recommendations for scale development and validation (Flora & Flake, 2017; Hair et al., 2019; Marar et al., 2023).

Confirmatory Factor Analysis

CFA was conducted in JASP using the independent CFA sample (n = 325). The CFA model was specified from the finalized 28-item EFA solution before evaluating the CFA dataset. This procedure preserved the distinction between exploratory structure identification and confirmatory model testing and prevented the CFA stage from functioning as an extension of exploratory model modification. CFA was used to evaluate whether the hypothesized measurement structure adequately reproduced the observed relationships among the retained indicators (Li & Zhang, 2022).

Before CFA estimation, multivariate normality was assessed using Mardia's tests of multivariate skewness and kurtosis. Because the CFA data showed significant multivariate nonnormality, robust maximum likelihood (MLR) estimation was used to reduce the sensitivity of parameter estimates and model-fit statistics to departures from multivariate normality. The Mardia statistics themselves are reported in the Results section because they are empirical diagnostic findings.

The CFA model was evaluated using multiple goodness-of-fit indices, including the chi-square statistic (χ^2), chi-square-to-degrees-of-freedom ratio (χ^2/df), Comparative Fit Index (CFI), Tucker–Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Square Residual (SRMR). Model adequacy was interpreted by considering these indices jointly rather than treating any single cutoff as definitive. The study used χ^2/df values below 3, RMSEA values within approximately 0.05–0.10, SRMR values below approximately 0.05–0.08, and CFI and TLI values approaching or exceeding 0.90 as practical indicators of acceptable model fit (Marsh et al., 2004; Byrne, 2010; Kline, 2023).

After estimation in JASP, the finalized CFA measurement model was visualized in AMOS to provide a graphical representation of standardized factor loadings, residual terms, and the correlation between the two latent factors.

Convergent Validity, Discriminant Validity, and Reliability

Construct-related measurement evidence was evaluated beyond global model fit. Standardized factor loadings were examined to determine the association of each retained item with its specified latent factor. Convergent validity was assessed using Average Variance Extracted (AVE), with values above 0.50 interpreted as evidence that a latent construct accounted for an adequate proportion of variance in its indicators (Hair et al., 2022; Henseler et al., 2015).

Discriminant validity between the two correlated dimensions was assessed using the Heterotrait–Monotrait ratio (HTMT). The HTMT results were interpreted in conjunction with the factor correlation and the overall CFA solution rather than as evidence that BSPS and ISPS were independent constructs. The conservative criterion of HTMT < 0.85 was used when evaluating whether the two dimensions remained empirically distinguishable (Henseler et al., 2015).

Internal consistency reliability was evaluated using Cronbach's alpha (α), McDonald's omega (ω), and composite reliability (CR). These coefficients were examined primarily at the dimensional level because the factor-analytic model specified two correlated constructs. Reliability estimates for the complete 28-item set were treated as descriptive indicators of overall internal consistency and were not interpreted as evidence that the instrument was unidimensional. Reliability, convergent validity, and discriminant validity were evaluated in accordance with contemporary psychometric reporting recommendations (Cheung et al., 2024).

Ethical Considerations

The study received an Ethical Exemption from the Research Ethics Committee of Universitas Negeri Malang (Approval No. 27.04.16/UN32.14.2.8/LT/2026). Before data collection, participants were informed about the study objectives, procedures, anticipated benefits, voluntary nature of participation, and their rights as research participants. Participation proceeded only after informed consent had been obtained.

All data were collected anonymously and managed confidentially using respondent identification codes. These procedures protected the identities of individual participants and participating institutions throughout data analysis and dissemination of the findings.

RESULTS AND DISCUSSION

Preliminary Item Analysis and Descriptive Statistics

Preliminary item analysis was conducted using responses from 166 students to examine the response characteristics of the initial 35-item self-report SPS instrument before factor analysis. The analysis focused on item means, standard deviations, and corrected item–total correlations.

Across the 35 items, mean scores ranged from 2.79 to 3.75, while standard deviations ranged from 1.21 to 1.55. Corrected item–total correlations ranged from 0.637 to 0.773, and all correlations were statistically significant ($p < .001$). These results indicated that all initial items showed sufficient response variability and were adequately associated with the broader item pool to proceed to factor-analytic evaluation.

The complete item-level descriptive statistics are provided in Appendix D (Table D) rather than the main text because the 35-row table is primarily supplementary to the factor-analytic findings.

The preliminary analysis was treated as an initial screening stage rather than as evidence of the dimensional structure of the instrument. Accordingly, final decisions concerning item retention and factor membership were based on the subsequent EFA and CFA analyses.

Multivariate Normality of the CFA Data

Before CFA estimation, multivariate normality was examined using Mardia's tests of multivariate skewness and kurtosis. As shown in Table 3, both the multivariate skewness and kurtosis statistics were statistically significant. The skewness statistic was $\chi^2(7770) = 14{,}735.589$, $p < .001$, while the small-sample adjusted skewness statistic was $\chi^2(7770) = 14{,}873.283$, $p < .001$. The kurtosis statistic was 54.214, $p < .001$.These results indicated substantial departure from multivariate normality and supported the use of robust maximum likelihood estimation for CFA.

Table 3. Multivariate Normality Assessed Using Mardia's Test

Test	Value	Statistic	df	p
Skewness	260.807	14,735.589	7770	< .001
Small-Sample Skewness	260.807	14,873.283	7770	< .001
Kurtosis	1594.704	54.214	—	< .001

Exploratory Factor Analysis

Sampling Adequacy and Factor Retention

EFA was conducted on the initial 35-item instrument using the independent EFA sample of 339 students. The Kaiser–Meyer–Olkin value was 0.964, indicating a high degree of sampling adequacy. Bartlett's Test of Sphericity was statistically significant, $\chi^2(595) = 7,345.301$, $p < .001$, indicating that the item correlation matrix was appropriate for factor analysis. The factor-model chi-square was also statistically significant, $\chi^2(526) = 1,051.861$, $p < .001$.

Factor retention was evaluated using parallel analysis, eigenvalues, the scree plot, and theoretical interpretability. In the parallel analysis of the initial 35-item pool, the observed eigenvalues of the first and second factors were 16.297 and 2.109, respectively. Both exceeded their corresponding 95% random eigenvalues of 1.745 and 1.640. The third and subsequent observed eigenvalues did not exceed the corresponding random eigenvalues.

Consistent with the parallel-analysis results, the scree plot displayed a clear inflection after the second factor, as shown in Figure 2. Together, these results supported retention of two factors.

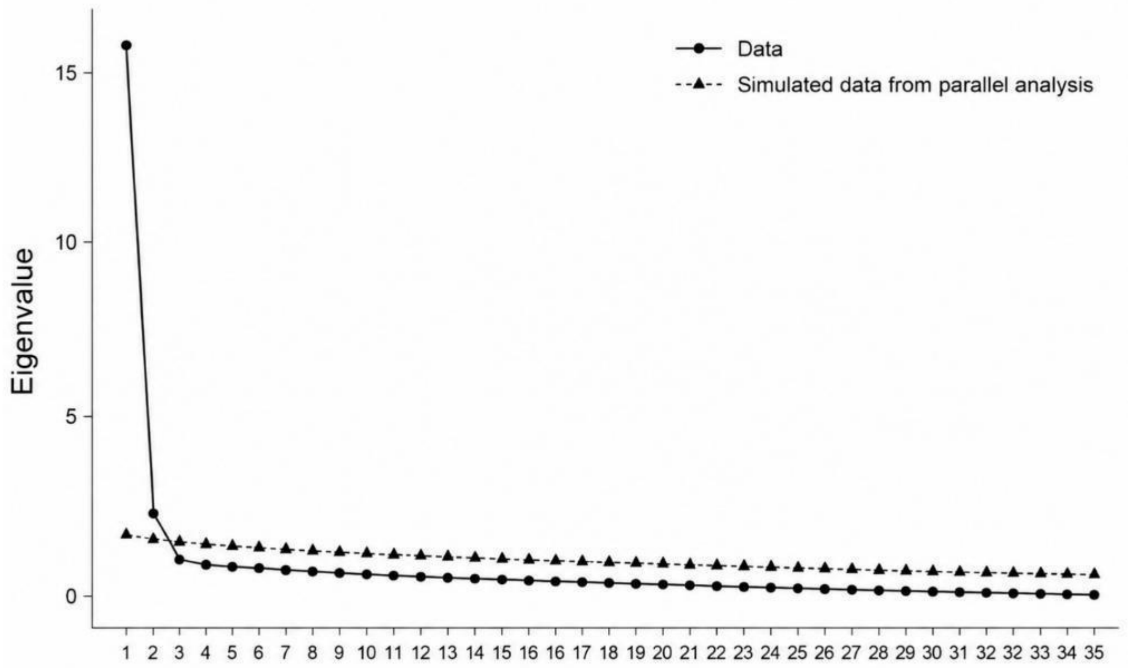

Figure 2. Scree Plot for Factor Retention in Exploratory Factor Analysis

Final EFA Factor Structure

Following factor extraction and item refinement, seven items were removed because they did not contribute sufficiently to a clear and interpretable two-factor solution according to the factor-loading and cross-loading criteria applied in the analysis. The resulting structure consisted of 28 retained items.

As shown in Table 4, Factor 1 contained 17 items with loadings ranging from 0.506 to 0.886, whereas Factor 2 contained 11 items with loadings ranging from 0.564 to 0.884. All retained items therefore met the minimum loading criterion applied in the study (Kline, 2023; DeVellis & Thorpe, 2021).

In the final 28-item solution, Factor 1 had an eigenvalue of 15.80 and accounted for 29.70% of the variance. Factor 2 had an eigenvalue of 1.639 and accounted for an additional 20.10%. Together, the two factors explained 49.80% of the total variance. The difference between these eigenvalues and those reported during parallel analysis reflects the use of different item sets: the parallel-analysis values correspond to the initial 35-item pool, whereas the values in Table 4 correspond to the final 28-item solution.

The correlation between the two extracted factors was 0.754. Based on the predominant item content and the original SPS framework (Rezba et al., 2022), Factor 1 was labeled Basic Science Process Skills (BSPS) and Factor 2 Integrated Science Process Skills (ISPS). The theoretical implications of this factor correlation and the interpretation of items located near the conceptual boundary between the two factors are addressed in the Discussion section.

Table 4. Final EFA Results for the 28-Item Science Process Skills Instrument

Dimension	Item	Factor 1	Factor 2
Factor 1: BSPS	1	0.886	
	2	0.871	
	8	0.793	
	3	0.786	
	7	0.719	
	11	0.718	
	5	0.699	
	14	0.678	
	4	0.673	
	15	0.640	
	18	0.588	
	13	0.582	
	23	0.556	
	17	0.547	
	10	0.537	
	21	0.506	
	12	0.571	
Factor 2: ISPS	34		0.884
	35		0.862
	33		0.785
	28		0.739
	27		0.704
	32		0.702
	31		0.663
	25		0.623
	30		0.608
	26		0.590
	29		0.564
Eigenvalue		15.80	1.639
Variance explained (%)		29.70	20.10
Cumulative variance (%)		29.70	49.80

The complete wording and empirical factor assignment of the 28 retained items are provided in Appendix E (Table E). This table is placed in the appendix because the full item wording is important for instrument transparency but duplicates information already summarized in Table 4.

Confirmatory Factor Analysis

Model Fit

The final 28-item two-factor structure obtained from EFA was evaluated using an independent CFA sample of 325 students. The measurement model specified two correlated latent factors, with 17 observed indicators assigned to the BSPS-related factor and 11 indicators assigned to the ISPS-related factor.

The model demonstrated acceptable fit to the observed data: $\chi^2(349) = 768.321$, $p < .001$, $\chi^2/df = 2.20$, CFI = 0.925, TLI = 0.919, RMSEA = 0.061, 90% CI [0.055, 0.067], and SRMR = 0.045. These indices were evaluated using the criteria specified in the Method section (Marsh et al., 2004; Byrne, 2010; Kline, 2023). The standardized measurement model is presented in Figure 3.

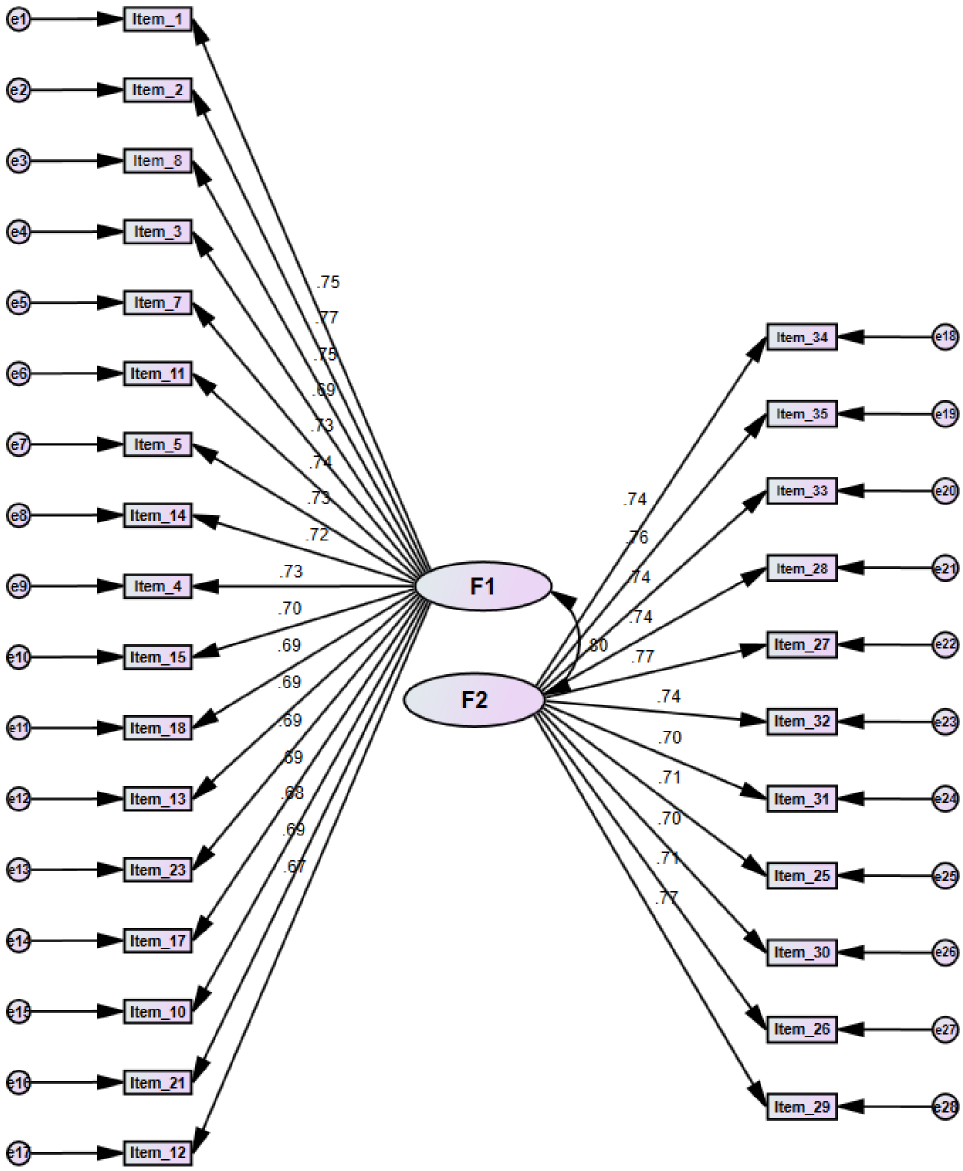

Figure 3. Path Diagram of the Correlated Two-Factor Model of Science Process Skills

Standardized Factor Loadings and Factor Correlation

As displayed in Figure 3 and summarized in Table 5, all retained items loaded significantly onto their specified latent factors ($p < .001$). Standardized loadings ranged from 0.66 to 0.77 for the BSPS-related factor and from 0.70 to 0.80 for the ISPS-related factor. Across the full measurement model, standardized factor loadings therefore ranged from 0.66 to 0.80, exceeding the minimum loading criterion used in the study (Hair et al., 2019; Brown, 2015; Kline, 2023).

The standardized latent correlation between BSPS and ISPS was 0.800 ($p < .001$), indicating a strong positive association between the two factors. The conceptual interpretation of this correlation is reserved for the Discussion section.

Convergent and Discriminant Validity

Convergent validity was examined using Average Variance Extracted (AVE). As reported in Table 5, the AVE value was 0.510 for BSPS and 0.540 for ISPS. Both values exceeded the 0.50 criterion applied in this study, providing evidence of adequate convergence among the indicators within each latent factor (Henseler et al., 2015; Hair et al., 2022).

Discriminant validity was evaluated using the Heterotrait–Monotrait ratio. The HTMT value between BSPS and ISPS was 0.805, which remained below the conservative 0.85 criterion used in the analysis (Henseler et al., 2015). Thus, despite their substantial latent correlation, the two factors remained empirically distinguishable according to the HTMT criterion.

Internal Consistency and Composite Reliability

Reliability coefficients for the final measurement model are presented in Table 5. For BSPS, composite reliability was 0.979, Cronbach's alpha was 0.946, and McDonald's omega was 0.947. For ISPS, composite reliability was 0.993, Cronbach's alpha was 0.928, and McDonald's omega was 0.928.

For the complete 28-item instrument, composite reliability was 0.992, Cronbach's alpha was 0.961, and McDonald's omega was 0.962. These results indicated high internal consistency within both dimensions and across the complete item set (McNeish, 2018; Roco-Videla et al., 2024; Kalkbrenner, 2024). The substantive interpretation of the very high reliability coefficients, including their possible relationship to overlapping item content, is considered in the Discussion section.

Table 5. Standardized Factor Loadings, Composite Reliability, Average Variance Extracted, and Internal Consistency of the Final Science Process Skills Instrument

Dimension	Standardized Factor Loadings (λ)	Composite Reliability (CR)	Average Variance Extracted (AVE)	Cronbach's Alpha (α)	McDonald's Omega (ω)
Basic Science Process Skills (17 items)	0.66–0.77	0.979	0.510	0.946	0.947
Integrated Science Process Skills (11 items)	0.70–0.80	0.993	0.540	0.928	0.928
Overall Instrument (28 items)	0.66–0.80	0.992	—	0.961	0.962

Summary of Psychometric Results

Across the sequential analyses, the initial 35-item pool was reduced to a 28-item instrument comprising two correlated factors. EFA supported a two-factor solution explaining 49.80% of the total variance, with factor loadings ranging from 0.506 to

0.886. The resulting structure was subsequently evaluated in an independent CFA sample and demonstrated acceptable model fit, standardized factor loadings from 0.66 to 0.80, AVE values above 0.50, an HTMT value of 0.805, and high internal consistency and composite reliability. These empirical findings provide the basis for the theoretical interpretation presented in the Discussion section.

Discussion

Psychometric Structure of the Self-Report SPS Instrument

The principal finding of this study is that students' responses to the self-report Science Process Skills (SPS) instrument were represented by two strongly related yet empirically distinguishable dimensions that broadly corresponded to Basic Science Process Skills (BSPS) and Integrated Science Process Skills (ISPS). This interpretation is supported by converging evidence across the sequential psychometric analyses. The preliminary item analysis showed adequate response variability and consistently positive corrected item–total correlations, indicating that the initial items were sufficiently associated with the broader item pool to proceed to structural evaluation. Although such preliminary analyses are useful for identifying potentially problematic items, they do not independently establish dimensionality (Lim et al., 2026; Mohammadi et al., 2025; Küçükaydın & Ayaz, 2025). The interpretation of the measurement structure therefore rests primarily on the EFA and its subsequent evaluation through CFA using an independent sample.

The separation of EFA and CFA samples strengthens the structural evidence because the model evaluated through CFA was not tested using the same observations from which the exploratory solution had been derived. This procedure reduces the likelihood that the confirmed structure merely reflects sample-specific optimization and provides a more rigorous assessment of whether the exploratory structure can be reproduced in an independent dataset (Worthington & Whittaker, 2006; Brown, 2015). In the present study, parallel analysis, eigenvalue examination, scree-plot inspection, and theoretical interpretability converged on a two-factor EFA solution, while the corresponding CFA model demonstrated acceptable fit. This convergence of exploratory and confirmatory evidence provides stronger support for the proposed measurement structure than reliance on either analysis alone (Kline, 2023; Küçükaydın & Ayaz, 2025; Romine et al., 2025).

The final EFA solution accounted for 49.80% of the variance, and all retained items met the loading criterion applied in the study. This proportion indicates that the two factors captured a meaningful component of variation in students' self-reported SPS-related behaviors while leaving room for individual, contextual, and measurement-specific influences that are expected in educational self-report data. Accordingly, loading magnitude and explained variance are most appropriately interpreted together with theoretical coherence and the reproducibility of the factor structure rather than as isolated indicators of construct validity (DeVellis & Thorpe, 2021; Asmaningrum et al., 2025; McChesney et al., 2025).

Relationship Between BSPS- and ISPS-Related Behavioral Tendencies

The substantial relationship between the two factors is an important characteristic of the measurement structure rather than evidence that the dimensions are indistinguishable. The EFA factor correlation of 0.754 and CFA latent correlation of approximately 0.800 indicate that students who reported more frequent engagement in foundational SPS-related behaviors also tended to report more frequent engagement in integrated inquiry-related behaviors. At the same time, the HTMT value of 0.805 remained below the conservative criterion of 0.85 applied in this study, and the correlated two-factor CFA model demonstrated acceptable fit. Taken together, these findings indicate strong conceptual proximity without empirical collapse of the two dimensions (Henseler et al., 2015; Brown, 2015; Kline, 2023).

This pattern broadly corresponds to the SPS framework proposed by Rezba et al. (2022), in which basic scientific processes such as observing, predicting, classifying, measuring, inferring, and communicating provide resources that students draw upon when engaging in more integrated processes such as formulating hypotheses, working with variables, interpreting data, and conducting experiments. However, the present results should not be interpreted as statistical evidence of a hierarchical measurement structure because no second-order or formally hierarchical CFA model was evaluated. The findings instead support a correlated two-factor representation in which BSPS- and ISPS-related behaviors differ in their predominant cognitive and behavioral complexity while remaining strongly interconnected.

This interconnectedness is consistent with contemporary views of scientific inquiry, in which students rarely perform scientific processes as isolated operations. Observation, measurement, manipulation of variables, data interpretation, experimental design, and communication are typically coordinated within the same investigative activity (Furtak & Penuel, 2023; OECD, 2023). Similarly, Vo and Simmie (2025) identified hypothesis formulation, experimental design, manipulation and measurement of variables, control of confounding variables, data interpretation, and conclusion drawing as recurring components of scientific-inquiry assessment. Such evidence supports viewing scientific inquiry as an organization of interacting processes rather than as a collection of completely independent skills.

The present factor structure is therefore more appropriately interpreted as distinguishing the predominant orientation of students' reported SPS-related behaviors—foundational or integrated—rather than establishing an absolute conceptual boundary between BSPS and ISPS. This distinction is particularly relevant to the empirical placement of Items 21 and 23.

Items at the Conceptual Boundary Between BSPS and ISPS

Item 21: Distinguishing Experimental Variables

Item 21, "I distinguish between independent, dependent, and control variables in an experiment," was initially developed under the controlling-variables indicator of ISPS but was empirically associated with the BSPS-related factor, with an EFA loading of 0.506. This result does not imply that controlling variables should theoretically be reclassified as a basic science process skill. Rather, the wording of the item appears to emphasize a cognitive operation located at the interface between foundational classification and integrated experimental reasoning.

The principal action expressed in Item 21 is to distinguish among categories of variables. Responding to this statement requires students to recognize differences in the functions of independent, dependent, and control variables and to assign them to appropriate conceptual categories. This process resembles classification because it involves recognizing defining characteristics and differentiating among categories. However, these categories acquire their functional meaning within experimental contexts, where identification of variable roles is necessary for designing and evaluating controlled investigations.

The scientific-reasoning literature provides support for this interpretation. Schwichow et al. (2016) characterized the control-of-variables strategy as a central component of scientific inquiry and experimental reasoning. At the same time, the specific cognitive demands elicited by an assessment task influence how scientific reasoning is expressed. In the present item, the wording emphasizes recognition and differentiation of variable types rather than the active manipulation or systematic control of variables during experimental design.

Accordingly, the empirical placement of Item 21 is best understood as a characteristic of the present response structure rather than as a revision of the theoretical SPS taxonomy. The item appears to function as a boundary indicator connecting classification-related processes with integrated experimental reasoning (Rezba et al., 2022; García-Carmona et al., 2024).

Item 23: Measurement and Operational Definition

A comparable pattern was observed for Item 23, "I identify the method for measuring each variable in an experiment." The item was initially developed within the defining-variables domain of ISPS but loaded on the BSPS-related factor at 0.556. As with Item 21, this empirical placement can be understood more appropriately by considering the dominant action represented in the item rather than by redefining its original theoretical domain.

The wording of Item 23 primarily emphasizes identifying how a variable should be measured. Measurement is generally considered a foundational science process because students must recognize relevant quantities, select suitable measurement approaches, and obtain empirical information from observations or experimental activities (Karim et al., 2024; Pertiwi & Aini, 2025). At the same time, specifying how a variable will be measured is closely connected to operational definition because operationalization requires linking an abstract or conceptual variable to an observable or measurable procedure within an investigation.

This conceptual overlap is consistent with broader models of scientific inquiry in which measurement is embedded within experimental design rather than performed as an isolated activity. Vo and Simmie (2025), for example, identified measuring dependent variables, manipulating independent variables, controlling confounding variables, designing experiments, and interpreting data as interacting components of scientific-inquiry assessment. Consequently, the same observable behavior may contribute to multiple stages of an investigation depending on its functional purpose.

The loading of Item 23 on the BSPS-related factor therefore does not indicate that operational definition itself constitutes a basic science process skill. Instead, the factor placement suggests that students' responses to the item may be dominated by its measurement component. The item nevertheless retains conceptual relevance to ISPS because measurement decisions become operationally meaningful when they are applied to variables within experimental design (OECD, 2023; Furtak & Penuel, 2023).

Implications of Boundary Items for Construct Interpretation

The empirical placement of Items 21 and 23 illustrates why factor membership should not necessarily reproduce an initial theoretical blueprint perfectly. Theoretical frameworks guide item generation and establish intended content representation, whereas factor analysis examines how responses to those items are organized empirically within the target population. When an item combines cognitive processes associated with adjacent theoretical domains, its empirical association may reflect the component that is most salient to respondents rather than the category under which the item was originally generated.

Construct interpretation therefore requires statistical evidence to be considered alongside theoretical coherence rather than allowing either source of evidence to operate independently. Similar instrument-development research has emphasized that theoretically proposed structures need empirical examination because item relationships may differ when applied to actual learners and specific educational contexts (Küçükaydın & Ayaz, 2025; Romine et al., 2025; Asmaningrum et al., 2025; McChesney et al., 2025).

Retention of Items 21 and 23 is also consistent with their adequate EFA loadings and with the performance of the complete 28-item model in the independent CFA sample. Neither item prevented the model from demonstrating acceptable fit, adequate convergent validity, or discriminant validity. Their factor placement is therefore most appropriately treated as evidence of conceptual overlap within the observed response structure rather than as evidence that the original SPS taxonomy should be redefined.

Factorial, Convergent, and Discriminant Validity

The CFA provided complementary evidence that the factor structure identified through EFA remained viable in an independent sample. Standardized factor loadings ranged from 0.66 to 0.77 for BSPS and from 0.70 to 0.80 for ISPS, indicating consistent associations between the retained items and their specified latent factors (Hair et al., 2019; Brown, 2015; Kline, 2023). The acceptable overall model fit further indicates that the proposed correlated two-factor structure provided a reasonable representation of the covariance pattern observed in the CFA sample.

Evidence for convergent validity was also adequate. AVE values of 0.510 for BSPS and 0.540 for ISPS exceeded the criterion applied in the study, indicating that the retained indicators shared sufficient variance with their corresponding latent dimensions (Henseler et al., 2015; Hair et al., 2022). Together with the standardized loading pattern, these results indicate that items assigned to each dimension converged on their specified latent factor.

Discriminant validity requires a more qualified interpretation because BSPS and ISPS were strongly correlated. Nevertheless, the HTMT value of 0.805 remained below the conservative criterion of 0.85, indicating that the dimensions retained empirical differentiation despite their close association (Henseler et al., 2015). Thus, the findings do not support interpreting BSPS and ISPS as independent constructs, but neither do they justify collapsing the two dimensions automatically into a single undifferentiated factor. The evidence instead supports two closely related but empirically distinguishable dimensions of self-reported SPS-related behavior.

Reliability and Potential Item Redundancy

The final instrument demonstrated high internal consistency. Cronbach's alpha and McDonald's omega were 0.946 and 0.947 for BSPS and 0.928 and 0.928 for ISPS, respectively, while composite reliability was also high for both dimensions. The convergence of alpha, omega, and composite-reliability evidence indicates that the retained items function consistently within their respective factors (McNeish, 2018; Roco-Videla et al., 2024; Kalkbrenner, 2024).

The very high reliability coefficients should nevertheless be interpreted cautiously. Values approaching unity may indicate that several items capture closely overlapping manifestations of the same behavioral domain rather than contributing entirely distinct information (Kalkbrenner, 2024). This possibility is particularly relevant to an instrument containing multiple statements representing related classroom and laboratory activities. Thus, high reliability constitutes evidence of strong internal consistency, but it does not necessarily indicate that every retained item provides unique measurement information.

Similarly, the high reliability coefficient for the complete 28-item set should not be interpreted as evidence that the instrument is unidimensional. High overall consistency is plausible because the two dimensions themselves are strongly correlated. Dimensionality is determined by the factor-analytic evidence, which supported two correlated factors, whereas reliability concerns the consistency of responses within the measured score structure. The present reliability pattern is therefore most appropriately interpreted as strong internal consistency accompanied by the possibility that some retained items represent closely overlapping behavioral content.

Theoretical Implications

The findings provide initial empirical support for applying the BSPS–ISPS distinction to students' self-reported behavioral tendencies in the Indonesian senior high school context. The theoretical contribution lies not in establishing definitive boundaries between basic and integrated SPS, but in showing that students' reported SPS-related behaviors can be represented by two correlated dimensions that broadly correspond to foundational and integrated scientific processes.

This interpretation is consistent with perspectives in science education that conceptualize scientific proficiency as involving increasingly coordinated forms of engagement. Foundational processes such as observation, prediction, classification, measurement, inference, and communication provide resources that students can combine when formulating hypotheses, planning investigations, working with experimental variables, interpreting data, and evaluating evidence (Rezba et al., 2022; García-Carmona et al., 2024; Wan Hussain et al., 2026). Scientific inquiry assessment likewise commonly involves multiple interacting processes rather than isolated skills (Tauhidah & Rofi'ah, 2023; Vo & Simmie, 2025).

The findings also reinforce the importance of evaluating theoretically derived measurement structures empirically. Expert judgment provides evidence that items represent intended content, whereas factor analysis examines whether responses from the target population reproduce the proposed organization. In this study, the empirical structure broadly corresponded to the theoretical BSPS–ISPS distinction while also revealing meaningful item-level overlap. This combination of theoretical correspondence and empirical qualification provides a more defensible representation of SPS-related behaviors than either forcing every item to reproduce its original theoretical classification or allowing statistical results alone to redefine the construct (Karim et al., 2024; García-Carmona et al., 2024).

The substantial relationship between BSPS and ISPS further indicates that foundational and integrated SPS-related behaviors are better understood as interconnected dimensions than as independent domains. However, this interpretation is limited to the correlated two-factor structure tested in the present study and does not constitute evidence of a formal hierarchical or higher-order SPS structure.

Practical Implications

From a practical perspective, the instrument provides researchers and educators with a structured means of examining students' reported engagement in foundational and integrated SPS-related behaviors. Because BSPS and ISPS are represented as related but distinguishable dimensions, dimension-level scores can provide more differentiated information than a single undivided SPS score. Such patterns may characterize students' reported engagement in activities involving observation, measurement, communication, hypothesis formulation, variable-related reasoning, data interpretation, and experimentation.

The practical interpretation of these scores must nevertheless remain consistent with the self-report format. The instrument captures how frequently students report engaging in SPS-related behaviors rather than how successfully they can demonstrate those processes during direct observation or performance-based assessment. Its practical value therefore lies primarily in research, educational evaluation, needs assessment, and characterization of students' reported behavioral engagement rather than in certifying or diagnosing actual scientific competence.

Limitations of the Study

Several limitations should be considered when interpreting the findings. First, participants were recruited purposively from senior high schools in East Java, Indonesia. Although the use of independent EFA and CFA samples strengthened the internal evaluation of the measurement structure, recruitment from a single geographical region limits the extent to which the findings can be generalized to students in other provinces, educational systems, or sociocultural contexts. The present psychometric evidence should therefore be interpreted primarily in relation to populations reasonably similar to the sampled students.

Second, the instrument relies on students' self-reported frequency of SPS-related behaviors. Responses may consequently be influenced by self-perception, memory accuracy, response tendencies, or social-desirability effects. More importantly, reported frequency of engagement is conceptually distinct from demonstrated SPS performance. The present evidence therefore supports interpretation of the instrument as a measure of self-reported behavioral engagement rather than as direct evidence of students' competence in performing scientific investigations.

Third, the empirical placement of Items 21 and 23 differed from their original theoretical classifications. Although these patterns can be interpreted as reflecting conceptual overlap between classification and variable-related reasoning, and between measurement and operational definition, the results indicate that the empirical boundary between BSPS- and ISPS-related behaviors is not absolute. Consequently, correspondence between the theoretical blueprint and observed factor structure should be understood as broad rather than exact.

Fourth, the very high internal-consistency coefficients indicate strong coherence among the retained items but also raise the possibility that some items represent closely overlapping behavioral content. The present analyses establish that the items function consistently within the identified factors, but they do not determine whether every retained item contributes unique measurement information.

Fifth, the validation evidence examined in this study focused primarily on internal structure, convergent validity, discriminant validity, and internal consistency. Criterion-related validity, predictive validity, temporal stability, and measurement invariance across student groups were not examined. The present results therefore provide an initial psychometric foundation for interpretation of the instrument but do not encompass all possible sources of evidence relevant to score validity and comparability.

Finally, the CFA evaluated a correlated two-factor structure rather than alternative structural representations such as a second-order or bifactor model. Accordingly, the strong association between BSPS and ISPS should not be interpreted as empirical evidence for a general higher-order SPS factor or a formally hierarchical measurement structure.

CONCLUSIONS

This study developed and psychometrically evaluated a self-report Science Process Skills (SPS) instrument for senior high school students in East Java, Indonesia, using sequential exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) with independent samples. The final instrument comprised 28 items organized into two correlated dimensions: Basic Science Process Skills (BSPS), represented by 17 items, and Integrated Science Process Skills (ISPS), represented by 11 items. The final measurement model demonstrated acceptable factorial fit, adequate convergent and discriminant validity, and high internal consistency within the examined samples.

The resulting structure broadly corresponded to the theoretical distinction between BSPS and ISPS while also indicating that the two dimensions are closely interconnected rather than completely independent. The empirical placement of Items 21 and 23 further showed that some SPS-related behaviors may lie near the conceptual boundary between foundational and integrated scientific processes. These findings are therefore best interpreted as supporting two related but empirically distinguishable dimensions of self-reported SPS-related behavior, rather than as establishing rigid conceptual boundaries between BSPS and ISPS.

Importantly, the instrument measures students' reported frequency of engagement in SPS-related behaviors rather than their demonstrated competence in performing scientific investigations. Within this interpretive boundary, the instrument provides initial psychometric support for examining patterns of students' reported engagement in foundational and integrated scientific practices. However, the findings should be generalized cautiously because the current evidence was derived from purposively sampled senior high school students in East Java.

RECOMMENDATIONS

Further validation should examine the stability of the two-factor structure across broader and more diverse educational contexts. Replication with students from different geographical regions, school systems, and demographic groups is particularly important for determining whether the BSPS–ISPS structure and the empirical placement of boundary items remain consistent across populations.

Measurement-invariance analyses are recommended to evaluate whether the instrument functions equivalently across relevant student groups, such as gender, school type, and geographical region. Establishing invariance would provide a stronger basis for comparing BSPS- and ISPS-related scores across groups.

Additional validity evidence should also be obtained by examining relationships between self-reported SPS-related behavioral tendencies and external indicators. Performance-based assessments, laboratory tasks, structured observations, academic outcomes, and other inquiry-related measures may be used to determine the extent to which students' reported engagement corresponds to their demonstrated scientific performance.

Given the high reliability coefficients observed in this study, further item-level evaluation is also warranted to determine whether some retained items provide overlapping information. Approaches such as item response theory, Rasch modeling, or related item-functioning analyses may help determine whether a more parsimonious version can retain adequate construct coverage and psychometric quality while improving measurement efficiency.

For practical use, the instrument is most appropriately applied as a research, educational evaluation, or needs-assessment tool for examining students' reported engagement in BSPS- and ISPS-related activities. The resulting information may be used to characterize patterns of engagement across foundational and integrated scientific practices. However, the instrument should not be used as a stand-alone diagnostic or certification measure of students' actual SPS competence until stronger evidence is available from direct performance-based validation.

Acknowledgment

The authors would like to express their sincere gratitude to Dr. Kristin Sangur, M.Pd., and Ikramina Yusti Amina, M.Pd., for their valuable contributions through constructive feedback, critical suggestions, and insightful comments during the preparation and refinement of this manuscript. Their academic perspectives and recommendations greatly contributed to improving the clarity, rigor, and quality of this study. The authors also extend their appreciation to the participating schools, students, and expert reviewers who contributed to the data collection process and content validation of the instrument. During the manuscript preparation process, the authors used NotebookLM as a supporting tool for literature organization and exploration. The tool assisted the authors in managing relevant literature and facilitating the review process. However, the authors' interpretation, critical analysis, academic judgment, and final decisions regarding the content of the manuscript remained entirely their own responsibility.

Funding Information

This research was funded by Universitas Negeri Malang through the internal PNBP research grant scheme, under the Thesis Grant Scheme (14.04.1102/UN32.14.1/LT/2026).

Conflict of Interest Statement

The authors declare no conflict of interest.

Informed Consent

The authors obtained informed consent from all individuals included in this study.

Ethical Approval

This study received ethical exemption approval from the Research Ethics Committee of Universitas Negeri Malang (Approval No. 27.04.16/UN32.14.2.8/LT/2026). All research procedures complied with relevant institutional regulations and ethical principles for research involving human participants.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request. The data are not publicly available because they contain information that could compromise the privacy of research participants.

Author Contributions Statement

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Yuli Astutik	✓	✓	✓		✓	✓			✓	✓	✓			
Siti Zubaidah	✓								✓	✓		✓	✓	
Bagus Priambodo	✓	✓					✓			✓		✓		

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

REFERENCES

Amina, I. Y., Zubaidah, S., Gofur, A., & Wicaksono, A. G. C. (2022). Remap-GI: A learning model to enhance students' science process skills and biology cognitive learning outcomes. *Jurnal Pendidikan: Teori, Penelitian, dan Pengembangan, 7*(11), 515–525. https://doi.org/10.17977/2502-471X.2486

Angco, R. J. N., Cortes, S. T., Castañeda, W. D. Jr., & Selle, J. D. (2026). Psychometric validation of the student assessment survey for teachers scale (SAST) for higher education institutions. *Frontiers in Education, 11*, 1768305. https://doi.org/10.3389/feduc.2026.1768305

Asmaningrum, H. P., Sathasivam, R. V., & Shahali, E. H. M. (2025). Development and validation of cultural intelligence, cultural diversity climate, and intention to implement Ethno-STEM scales for Indonesian novice science teachers: An exploratory factor analysis approach. *Journal of Cultural Analysis and Social Change, 10*(4), 2902–2911. https://doi.org/10.64753/jcasc.v10i4.3360

Avan, Ç., & Aydınlı, B. (2026). Developing and implementing a STEM-based assessment tool to measure higher-order cognitive skills in science education. *Research in Science & Technological Education*.

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health, 6*, Article 149. https://doi.org/10.3389/fpubh.2018.00149

Borgonovi, F., Ferrara, A., & Piacentini, M. (2023). From asking to observing: Behavioural measures of socio-emotional and motivational skills in large-scale assessments. *Social Science Research, 112*, 102874.

Borgonovi, F., Pokropek, A., & Jakubowski, M. (2023). The relationship between self-reported and performance-based measures of students' skills: Evidence from large-scale assessments. *Educational Assessment, Evaluation and Accountability, 35*, 1–25.

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). Guilford Press.

Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (2nd ed.). Routledge.

Chao, C.-T., & Liang, J.-C. (2025). The development and validation of a multi-dimensional medical students' learning self-efficacy questionnaire for clinical education. *Medical Education Online, 30*(1), 2534053. https://doi.org/10.1080/10872981.2025.2534053

Chen, C., & Zhang, L. J. (2025). Development and validation of a scale for metacognitive regulation of interactive metadiscourse strategies in L2 argumentative writing. *Metacognition and Learning, 20*(1). https://doi.org/10.1007/s11409-025-09435-z

Chengere, A. M., Bono, B. D., Zinabu, S. A., & Jilo, K. W. (2025). Enhancing secondary school students' science process skills through guided inquiry-based laboratory activities in biology. *PLOS ONE, 20*(4), e0320692. https://doi.org/10.1371/journal.pone.0320692

Cheung, G. W., Cooper-Thomas, H. D., Lau, R. S., & Wang, L. C. (2024). Reporting reliability, convergent and discriminant validity with structural equation modeling: A review and best-practice recommendations. *Asia Pacific Journal of Management, 41*(2), 745–783.

DeVellis, R. F., & Thorpe, C. T. (2021). *Scale development: Theory and applications* (5th ed.). SAGE Publications.

Dolapcioglu, S., & Subasi, M. (2022). The relationship between scientific process skills and science achievement: A meta-analysis study. *Journal of Science Learning, 5*(2), 363–372.

Elfeky, A. I. M., Najmi, A. H., & Elbyaly, M. Y. H. (2024). The impact of advance organizers in virtual classrooms on the development of integrated science process skills. *PeerJ Computer Science, 10*, e1989. https://doi.org/10.7717/peerj-cs.1989

Fitriyana, S., Farhan, A., Susanna, Hamid, A., & Mahzum, E. (2025). The effect of guided inquiry learning model on students' science process skills and learning outcomes in physics science lessons. *Jurnal Pendidikan Fisika, 13*(3), 555–571. https://doi.org/10.26618/0m4ny850

Flora, D. B., & Flake, J. K. (2017). The purpose and practice of exploratory and confirmatory factor analysis in psychological research: Decisions for scale development and validation. *Canadian Journal of Behavioural Science, 49*(2), 78–88. https://doi.org/10.1037/cbs0000069

Furtak, E. M., & Penuel, W. R. (2023). Learning through scientific practices: Connecting classroom inquiry with scientific reasoning. *Science Education, 107*(5), 1125–1147.

García-Carmona, A. (2025). Scientific thinking and critical thinking in science education: Two distinct but symbiotically related intellectual processes. *Science & Education, 34*, 227–245. https://doi.org/10.1007/s11191-023-00460-5

García-Carmona, A., Muñoz-Franco, G., Criado, A. M., & Cruz-Guzmán, M. (2024). Validation of an instrument for assessing basic science process skills in initial elementary teacher education. *International Journal of Science Education, 46*(4), 362–381. https://doi.org/10.1080/09500693.2023.2232936

Gizaw, G. G., & Sota, S. S. (2023). Improving science process skills of students: A review of literature. *Science Education International, 34*(3), 216–224. https://doi.org/10.33828/sei.v34.i3.5

Gültekin, S. B., & Altun, T. (2022). Investigating the impact of activities based on scientific process skills on 4th grade students' problem-solving skills. *International Electronic Journal of Elementary Education, 14*(4), 491–500. https://doi.org/10.26822/iejee.2022.258

Habibah, S. N., & Fuadiyah, S. (2025). Validitas instrumen tes keterampilan proses sains pada pembelajaran biologi materi sistem pencernaan kelas XI SMA/MA. *Jurnal Jendela Pendidikan, 5*(2), 353–360. https://doi.org/10.57008/jjp.v5i02.1359

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage Learning.

Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2022). *A primer on partial least squares structural equation modeling (PLS-SEM)* (3rd ed.). SAGE Publications.

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science, 43*(1), 115–135. https://doi.org/10.1007/s11747-014-0403-8

Idris, N., Talib, O., & Razali, F. (2022). Strategies in mastering science process skills in science experiments: A systematic literature review. *Jurnal Pendidikan IPA Indonesia, 11*(1), 155–170. https://doi.org/10.15294/jpii.v11i1.32969

Jardinico III, E. C., & Linaugo, J. D. (2023). Acquired basic and integrated science process skills and academic performance in Earth Science of Grade 11 students in a Philippine public high school. *Philippine Social Science Journal, 6*(1), 63–73.

Junianda, W., & Fuadiyah, S. (2025). Validitas instrumen tes keterampilan proses sains pada materi sistem peredaran darah kelas XI SMA/MA [Validity of a science process skills test instrument on the circulatory system topic for Grade XI senior high school students]. *Afeksi: Jurnal Penelitian dan Evaluasi Pendidikan, 6*(4), 737–748. https://doi.org/10.59698/afeksi.v6i4.517

Kalkbrenner, M. T. (2024). Choosing between Cronbach’s coefficient alpha, McDonald’s coefficient omega, and coefficient H: Confidence intervals and the advantages and drawbacks of interpretive guidelines. *Measurement and Evaluation in Counseling and Development, 57*(2), 93–105. https://doi.org/10.1080/07481756.2023.2283637

Kamarudin, N., Wahida, M., & Ahrari, S. (2022). Exploring basic and integrated science process skills and their impact on science achievement among university students. *Journal of Public Administration and Governance, 12*(4S), 74–90. https://doi.org/10.5296/jpag.v12i4S.20572

Karim, E., Ibrahim, S. I., Rasit, H. H., & Sumintono, B. (2024). Development and validation of students’ competency instrument on science process skills. *Proceedings of the Pacific-Rim Objective Measurement Symposium (PROMS 2023)*, 311–325. https://doi.org/10.2991/978-94-6463-494-5_20

Kline, R. B. (2023). *Principles and practice of structural equation modeling* (5th ed.). Guilford Press.

Komisia, F., Tukan, M. B., Fallo, D. Y. A., Hale, G. J. P., Meta, G. B., Lanmai, P. O., Eflani, F., & Leba, M. A. U. (2025). Validity analysis of a science process skills assessment instrument for high school students in chemistry. *Jurnal Penelitian Pendidikan IPA, 11*(12). https://doi.org/10.29303/jppipa.v11i12.1286

Krell, M., Vorholzer, A., & Nehring, A. (2022). Scientific reasoning in science education: From global measures to fine-grained descriptions of students’ competencies. *Education Sciences, 12*(2), 97. https://doi.org/10.3390/educsci12020097

Küçükaydın, M. A., & Ayaz, E. (2025). Validation of the Scientific Reasoning Competencies Instrument: Relationships with epistemological beliefs and analytical thinking. *International Journal of Science and Mathematics Education, 23*, 343–363. https://doi.org/10.1007/s10763-024-10482-2

Li, C., & Zhang, L. (2022). Confirmatory factor analysis and construct validity evaluation in educational measurement. *Educational Measurement: Issues and Practice, 41*(4), 25–36.

Lim, W. M., Khatri, P., Shiva, A., & Dabas, V. (2026). Guidelines for scale development and validation. *Journal of Consumer Behaviour, 25*(2), 953–973. https://doi.org/10.1002/cb.70096

Liu, R., Jong, C., & Fan, M. (2024). Reciprocal relationship between self-efficacy and achievement in mathematics among high school students. *Large-scale Assessments in Education, 12*, Article 14.

Lorenzo-Seva, U., & Ferrando, P. J. (2024). Determining sample size requirements in EFA solutions: A simple empirical proposal. *Multivariate Behavioral Research, 59*(5), 899–912. https://doi.org/10.1080/00273171.2024.2342324

Marar, S., Hamza, M. A., Ayyash, M., & Abu-Shaheen, A. (2023). Development and validation of an instrument to assess the knowledge and perceptions of predatory journals. *Heliyon, 9*(11), e22270. https://doi.org/10.1016/j.heliyon.2023.e22270

Marsh, H. W., Hau, K. T., & Wen, Z. (2004). 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. *Structural Equation Modeling: A Multidisciplinary Journal, 11*(3), 320–341. https://doi.org/10.1207/s15328007sem1103_2

McChesney, E. T., Schunn, C. D., Dorvè-Lewis, G., Godwin, A., & DeAngelo, L. (2025). Reconceptualizing learning engagement: Evidence for a context-sensitive structure in STEM education. *Frontiers in Psychology, 16*, Article 1649744.

McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods, 23*(3), 412–433. https://doi.org/10.1037/met0000144

Mohammadi, M. M., Sohrabi, P., & Mohammadi, M. (2025). Development and psychometric properties of Ethical Competence Questionnaire for Nursing Students. *Health Science Reports, 8*, e70750. https://doi.org/10.1002/hsr2.70750

Molefe, L., & Aubin, J.-B. (2023). A nexus of scientific investigations, science process skills, and sustainable development goals: Pre-service teachers' views concerning mangroves fieldwork. *Journal of Baltic Science Education, 22*(4), 682–700. https://doi.org/10.33225/jbse/23.22.682

Molefe, L., & O'Donoghue, R. (2026). Science process skills and education for sustainable development competences: A review through the lens of critical realism. *Eurasia Journal of Mathematics, Science and Technology Education, 22*(5), em2835. https://doi.org/10.29333/ejmste/18519

Mustofa, A., Zubaidah, S., & Kuswantoro, H. (2021). Pengembangan modul berbasis proyek berdasarkan analisis lintas karakter agronomi dan morfologi kedelai untuk meningkatkan keterampilan proses sains. *Jurnal Pendidikan: Teori, Penelitian, dan Pengembangan, 6*(1), 24–34. https://doi.org/10.17977/jptpp.v6i1.14373

Negerie, D., Habtamu, T., & Tafesse, T. (2026). Effects of the 5E guided inquiry-based learning model on science process skills of secondary school biology students in Mattu town, Ethiopia. *Discover Education, 5*, Article 320. https://doi.org/10.1007/s44217-026-01351-x

Ngo, H. M., Nguyen, T. T., & Tran, T. H. (2025). Assessment of convergent validity and measurement model evaluation in structural equation modeling. *Heliyon*.

Obilor, E. I., & Miwari, G. U. (2022). Content validity in educational assessment. *International Journal of Innovative Education Research, 10*(2), 57–69.

Ocy, D. R., Sarifah, I., & Riyadi. (2025). EFA and CFA analysis: Development and validation of a test instrument for mathematical abstraction skills. *Journal of Research and Advances in Mathematics Education, 10*(2).

OECD. (2023). *PISA 2022 assessment and analytical framework: Mathematics, reading, science and financial literacy*. OECD Publishing.

Pertiwi, N., & Aini, F. Q. (2025). Development of scientific process skill assessment instrument on stoichiometry using Rasch model in the Merdeka Curriculum. *Journal of Educational Sciences, 9*(1), 241–255.

Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing & Health, 30*(4), 459–467. https://doi.org/10.1002/nur.20199

Pujiyanto, Nuroso, H., & Sumarno. (2025). Profil keterampilan proses sains (KPS) siswa kelas XI SMA Negeri 1 Ngawen pada konsep gelombang bunyi [Profile of science process skills (SPS) among Grade XI students of SMA Negeri 1 Ngawen on the concept of sound waves]. *Jurnal Inovasi Pembelajaran di Sekolah, 6*(2), 61–65. https://doi.org/10.51874/jips.v6i2.317

Rattanachaithada, R., Kwangmuang, P., Vongtathum, P., Gamlunglert, R., & Srikoon, S. (2025). A confirmatory factor analysis of scientific critical thinking in secondary school. *Cogent Education, 12*(1), 2467508. https://doi.org/10.1080/2331186X.2025.2467508

Rezba, R. J., Sprague, C., McDonnough, J. T., & Matkins, J. J. (2022). *Learning and assessing science process skills* (5th ed.). Kendall/Hunt Publishing Company.

Roco-Videla, Á., Aguilera-Eguía, R., & Olguin-Barraza, M. (2024). Ventajas del uso del coeficiente de omega de McDonald frente al alfa de Cronbach. *Nutrición Hospitalaria, 41*(1), 233–234. https://doi.org/10.20960/nh.04879

Roemer, E., Schuberth, F., & Henseler, J. (2021). HTMT2—an improved criterion for assessing discriminant validity in structural equation modeling. *Industrial Management & Data Systems*.

Romine, W., Reed, E., Mahajan, R., & Todd, A. (2025). What is the Inventory of Student Evolution Acceptance (I-SEA) measuring? A factor analytic comparison of multiple competing hypotheses in practicing science teachers and undergraduate science non-majors. *International Journal of Science and Mathematics Education, 23*(7), 3047–3067. https://doi.org/10.1007/s10763-025-10570-x

Rönkkö, M., & Cho, E. (2022). An updated guideline for assessing discriminant validity. *Organizational Research Methods, 25*(1), 6–14. https://doi.org/10.1177/1094428120968614

Sadykova, B., Yesnazarova, U., & Tokbergenova, A. (2023). The geographic inquiry process skills scale: A validation study. *Eurasia Journal of Mathematics, Science and Technology Education, 19*(9), em2319. https://doi.org/10.29333/ejmste/13471

Schwichow, M., Croker, S., Zimmerman, C., Höffler, T., & Härtig, H. (2016). Teaching the control-of-variables strategy: A meta-analysis. *Developmental Review, 39*, 37–63. https://doi.org/10.1016/j.dr.2015.12.001

Šmida, D., Drozdíková, A., & Nechajová, R. (2026). A new perspective on the evaluation of pupils' inquiry skills using four-tier test. *Science & Education*. Advance online publication. https://doi.org/10.1007/s11191-026-00747-3

Stefana, A., Damiani, S., Granziol, U., Provenzani, U., Solmi, M., Youngstrom, E. A., & Fusar-Poli, P. (2025). Psychological, psychiatric, and behavioral sciences measurement scales: Best practice guidelines for their development and validation. *Frontiers in Psychology, 15*, 1494261. https://doi.org/10.3389/fpsyg.2024.1494261

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Pearson.

Tauhidah, D., & Rofi'ah, N. L. (2023). Validation of undergraduate science process skills tests: Rasch model analysis. *Research and Development in Education (RaDEn), 3*(1), 51–57. https://doi.org/10.22219/raden.v3i1.24585

Tondong, H. I., Apriyanto, A., Kusumastuti, S. Y., & Wahyuningsih, M. (2024). *Buku referensi pengantar statistik*. PT. Sonpedia Publishing Indonesia.

Ulfah, A. H., & Supahar. (2026). Validity of performance assessment of Science Process Skills in senior high school biology subject. *Jurnal BIOEDUIN, 16*(1), 51–63. https://doi.org/10.15575/bioeduin.v16i1.44745

Vo, D. V., & Simmie, G. M. (2025). Assessing scientific inquiry: A systematic literature review of tasks, tools and techniques. *International Journal of Science and Mathematics Education, 23*, 871–906. https://doi.org/10.1007/s10763-024-10498-8

Wan Hussain, W. N. H., Shahali, E. H. M., & Fadzil, H. M. (2026). Assessing science process skills through conceptual frameworks: A scoping review. *International Journal on Studies in Education, 8*, 599–622. https://doi.org/10.46328/ijonse.6545

Wan, Z. H., Zhan, Y., & Zhang, Y. (2024). Positive or negative? The effects of scientific inquiry on science achievement via attitudes toward science. *Science Education, 108*(1), 3–24. https://doi.org/10.1002/sce.21825

Wang, S., Hu, A., Yuan, L., Tian, W., & Xin, T. (2026). Macro- and micro-level behavioral patterns in simulation-based scientific inquiry: Linking processes to performance among elementary students. *Journal of Intelligence, 14*(1), 6. https://doi.org/10.3390/jintelligence14010006

Wang, T., Zhou, Q., Yin, K., Hu, H., Xu, H., Huang, J., & Chen, S. (2026). Microsystem components and EFL achievement in adolescence: A network analysis. *Acta Psychologica, 263*, Article 106379. https://doi.org/10.1016/j.actpsy.2026.106379

Worthington, R. L., & Whittaker, T. A. (2006). Scale development research: A content analysis and recommendations for best practices. *The Counseling Psychologist, 34*(6), 806–838. https://doi.org/10.1177/0011000006288127

Yang, K.-L., Fang, S.-C., & Fan, S.-C. (2025). Development and validation of an instrument for assessing secondary students' transdisciplinary STEM practices. *International Journal of STEM Education, 12*, Article 5. https://doi.org/10.1186/s40594-025-00529-3

Yanti, S. R., & Fuadiyah, S. (2025). Validitas instrumen tes keterampilan proses sains pada materi sistem pernapasan kelas XI SMA/MA [Validity of a science process skills test instrument on the respiratory system topic for Grade XI senior high school students]. *Jurnal Pengabdian Masyarakat dan Riset Pendidikan, 3*(4), 4590–4597. https://doi.org/10.31004/jerkin.v3i4.1039

APPENDIX A

Initial Construct Blueprint and 35-Item Self-Report Instrument

Table A. Dimensions, Indicators, and Initial Items of the SPS Self-Report Instrument

Dimension	Indicator	Indicator Description	Item No.	Self-Report Item
Basic Science Process Skills	Observing	The process of gathering information about an object or an event.	1	I carefully observe objects or events during classroom learning activities, whether conducted in groups or individually.
			2	I carefully observed objects or events using my five senses during the practicum in the previous semester.
			3	I describe the characteristics of the observed object or phenomenon in a complete and systematic manner.
Basic Science Process Skills	Predicting	Predicting what will happen based on observations and data obtained from previous events.	4	I predict the outcome of an event or experiment based on previous data and observations.
			5	I predict the impact of an experiment based on a literature review I read before the practicum.
			6	I predict the results of the practicum based on the material I studied beforehand.
Basic Science Process Skills	Classifying	The process of grouping objects or events based on their properties and characteristics.	7	I group objects or events based on their similarities and differences.
			8	I classify the material being studied based on its similarities and differences.
			9	I determine the basis for grouping objects according to their observed characteristics.
Basic Science	Measuring	Conveying information based on the	10	I use measuring instruments correctly during laboratory practicum.

Dimension	Indicator	Indicator Description	Item No.	Self-Report Item
Process Skills		size of an object, such as length, width, weight, volume, and so on. This activity involves the use of measuring instruments.		
			11	I record measurement results using the appropriate units.
			12	I select the measuring instrument that is appropriate for measuring a specific quantity in an experiment.
Basic Science Process Skills	Concluding	Making a statement based on the observation or evaluation of the accuracy of data or facts, or explaining and interpreting something on the basis of an observation.	13	I draw conclusions based on data from observations or experiments that have previously been conducted.
			14	I draw conclusions based on data obtained from observations or experiments.
			15	I conclude the results of observations based on the data obtained.
Basic Science Process Skills	Communicating	Sharing information obtained from observations with others, whether in the form of tables, maps, graphs, or other means.	16	I present the results of observations or experiments clearly in the form of tables, graphs, or written explanations.
			17	I orally present the results of my observations to the teacher or classmates.
			18	I present experimental data in the form of tables, diagrams, or graphs.

Dimension	Indicator	Indicator Description	Item No.	Self-Report Item
Integrated Science Process Skills	Controlling Variables	Being able to identify and determine the variables that may affect the outcome of an experiment.	19	I identify the variables that influence the outcome of an experiment.
			20	I determine the variables to be used in an experiment.
			21	I distinguish between independent, dependent, and control variables in an experiment.
Integrated Science Process Skills	Defining Variables	Explaining, in operational terms, how each variable in an experiment is measured.	22	I explain how to measure each variable in an experiment.
			23	I identify the method for measuring each variable in an experiment.
			24	I formulate the operational definition of each research variable prior to conducting the experiment.
Integrated Science Process Skills	Formulating Hypotheses	Making a tentative conjecture about the expected experimental outcome, expressed as a statement that can be scientifically explained.	25	I formulate a hypothesis based on the problem given.
			26	I formulate a hypothesis before conducting an experiment.
			27	I align the hypothesis with the objective of the experiment to be conducted.
Integrated Science Process Skills	Interpreting Data	Organizing data and drawing conclusions on the basis of that data.	28	I process experimental data before drawing conclusions.

Dimension	Indicator	Indicator Description	Item No.	Self-Report Item
			29	I interpret experimental data based on the results obtained.
			30	I identify patterns or trends in experimental data.
			31	I relate experimental data to concepts or theories that have previously been studied.
Integrated Science Process Skills	Conducting Experiments	Carrying out the steps of an experiment in a systematic manner.	32	I carry out experimental activities according to the working steps, systematically and in an orderly manner.
			33	I carry out each practicum step in accordance with the available procedure.
			34	I use laboratory tools and materials in accordance with safety procedures throughout the experiment.
			35	I develop an experimental procedure to test the formulated hypothesis.

APPENDIX B

Item-Level Content Validity Index Results

Table B. Item-Level Content Validity Index Results

Item	Expert 1	Expert 2	Expert 3	Expert Agreement	I-CVI	UA
1	1	1	1	3	1.00	1
2	1	1	1	3	1.00	1
3	1	1	1	3	1.00	1
4	1	1	1	3	1.00	1
5	1	1	1	3	1.00	1
6	1	1	1	3	1.00	1
7	1	1	1	3	1.00	1
8	1	1	1	3	1.00	1
9	1	1	1	3	1.00	1
10	1	1	1	3	1.00	1
11	1	1	1	3	1.00	1
12	1	1	1	3	1.00	1
13	1	1	1	3	1.00	1
14	1	1	1	3	1.00	1
15	1	1	1	3	1.00	1
16	1	1	1	3	1.00	1
17	1	1	1	3	1.00	1
18	1	1	1	3	1.00	1
19	1	1	1	3	1.00	1
20	1	1	1	3	1.00	1
21	1	1	1	3	1.00	1
22	1	1	1	3	1.00	1

Item	Expert 1	Expert 2	Expert 3	Expert Agreement	I-CVI	UA
23	1	1	1	3	1.00	1
24	1	1	1	3	1.00	1
25	1	1	1	3	1.00	1
26	1	1	1	3	1.00	1
27	1	1	1	3	1.00	1
28	1	1	1	3	1.00	1
29	1	1	1	3	1.00	1
30	1	1	1	3	1.00	1
31	1	1	1	3	1.00	1
32	1	1	1	3	1.00	1
33	1	1	1	3	1.00	1
34	1	1	1	3	1.00	1
35	1	1	1	3	1.00	1
S-CVI/Ave					1.00	
S-CVI/UA						1.00

APPENDIX C

Summary of Expert Feedback and Item Revision

Table C. Summary of Expert Feedback and Item Revision

Aspect Evaluated	Expert Feedback	Revision Action
Item wording	Some item statements required minor refinement to improve clarity and simplify the wording for senior high school students.	Item statements were slightly revised by simplifying the wording and improving sentence clarity while maintaining the intended meaning of each indicator.
Terminology	Some terms needed adjustment to fit the senior high school science context.	Terms were modified according to expert recommendations.
Contextual suitability	Some items needed better alignment with science learning activities.	Items were revised to reflect students' actual learning experiences.

APPENDIX D

Descriptive Statistics for the Initial 35-Item Instrument

Table D. Descriptive Statistics for the Initial 35-Item Instrument

Item	Mean	SD	Corrected Item–Total Correlation (r)	p
1	3.66	1.21	0.750	< .001
2	3.24	1.52	0.773	< .001
3	3.75	1.23	0.700	< .001
4	3.41	1.31	0.743	< .001
5	3.13	1.45	0.689	< .001
6	3.12	1.27	0.637	< .001
7	3.48	1.32	0.748	< .001
8	3.28	1.43	0.748	< .001
9	3.41	1.23	0.663	< .001
10	3.25	1.30	0.719	< .001

Item	Mean	SD	Corrected Item–Total Correlation (r)	p
11	3.21	1.47	0.735	< .001
12	3.34	1.24	0.682	< .001
13	3.30	1.25	0.696	< .001
14	3.08	1.38	0.732	< .001
15	3.37	1.39	0.688	< .001
16	3.32	1.22	0.658	< .001
17	3.14	1.33	0.697	< .001
18	3.13	1.38	0.738	< .001
19	3.27	1.28	0.710	< .001
20	2.92	1.32	0.732	< .001
21	3.07	1.36	0.705	< .001
22	2.99	1.25	0.682	< .001
23	3.15	1.29	0.722	< .001
24	2.93	1.42	0.710	< .001
25	2.90	1.33	0.676	< .001
26	2.80	1.45	0.672	< .001
27	2.86	1.49	0.721	< .001
28	2.94	1.41	0.661	< .001
29	2.96	1.38	0.753	< .001
30	2.92	1.55	0.677	< .001
31	3.05	1.38	0.681	< .001
32	2.89	1.40	0.726	< .001
33	2.96	1.52	0.670	< .001
34	3.00	1.54	0.648	< .001
35	2.79	1.51	0.712	< .001

Note. SD = standard deviation; r = corrected item–total correlation. All item–total correlations were statistically significant at p < .001.

APPENDIX E

Dimensions and Items of the Final 28-Item SPS Instrument

Table E. Dimensions and Items of the Final 28-Item SPS Instrument

Dimension	Item	Self-Report Item
BSPS	1	I carefully observe objects or events during classroom learning activities, whether conducted in groups or individually.
	2	I carefully observed objects or events using my five senses during the practicum in the previous semester.
	8	I classify the material being studied based on its similarities and differences.
	3	I describe the characteristics of the observed object or phenomenon in a complete and systematic manner.
	7	I group objects or events based on their similarities and differences.
	11	I record measurement results using the appropriate units.
	5	I predict the impact of an experiment based on a literature review I read before the practicum.
	14	I draw conclusions based on data obtained from observations or experiments.
	4	I predict the outcome of an event or experiment based on previous data and observations.
	15	I conclude the results of observations based on the data obtained.

Dimension	Item	Self-Report Item
	18	I present experimental data in the form of tables, diagrams, or graphs.
	13	I draw conclusions based on data from observations or experiments that have previously been conducted.
	23	I identify the method for measuring each variable in an experiment.
	17	I orally present the results of my observations to the teacher or classmates.
	10	I use measuring instruments correctly during laboratory practicum.
	21	I distinguish between independent, dependent, and control variables in an experiment.
	12	I select the measuring instrument that is appropriate for measuring a specific quantity in an experiment.
ISPS	35	I develop an experimental procedure to test the formulated hypothesis.
	34	I use laboratory tools and materials in accordance with safety procedures throughout the experiment.
	33	I carry out each practicum step in accordance with the available procedure.
	28	I process experimental data before drawing conclusions.
	27	I align the hypothesis with the objective of the experiment to be conducted.
	32	I carry out experimental activities according to the working steps, systematically and in an orderly manner.
	31	I relate experimental data to concepts or theories that have previously been studied.
	25	I formulate a hypothesis based on the problem given.
	30	I identify patterns or trends in experimental data.
	26	I formulate a hypothesis before conducting an experiment.
	29	I interpret experimental data based on the results obtained.