Research Square

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

Self Reported Practical Innovation and Its Associations with Critical Thinking Problem Solving and Teamwork among Indonesian Biology Undergraduates

Luthpi Safahi

luthpi_safahi@uhamka.ac.id

Universitas Muhammadiyah Prof Dr Hamka

Gunawan Suryoputro

Universitas Muhammadiyah Prof Dr Hamka

Samsul Maarif

Universitas Muhammadiyah Prof Dr Hamka

Zulherman Zulherman

Universitas Muhammadiyah Prof Dr Hamka

Research Article

Keywords: practical innovation, critical thinking, problem solving, teamwork, biology education

Posted Date: September 29th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10717162/v1

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

Additional Declarations: No competing interests reported.

Abstract

Higher-order and collaborative competencies are often compressed into a single score, obscuring whether self-reported critical thinking, problem solving, teamwork, and practical innovation show the same measurement and association patterns. We conducted an exploratory cross-sectional secondary analysis of 100 biology undergraduates who completed 113 biology-contextualized Likert-type items. PSS14–PSS21, which describe maladaptive emotional and avoidant responses to problems, were reverse-scored at the respondent level so that higher values consistently indicated more effective problem solving. Teamwork had the highest mean (3.898), followed by practical innovation (3.811), critical thinking (3.799), and problem solving (3.495). Internal consistency was very high for critical thinking, teamwork, and practical innovation (alpha = .967–.973; omega = .968–.973), but lower and structurally less coherent for problem solving (alpha = .815; omega = .727). Practical innovation correlated strongly with teamwork (r = .879) and critical thinking (r = .863), and moderately with problem solving (r = .468). In an HC3-robust associative model, critical thinking (beta = .450) and teamwork (beta = .550) retained positive concurrent associations with practical innovation, whereas problem solving did not (beta = − .048, p = .269); model R^2 was .850. The findings describe concurrent self-report score relations rather than demonstrated competence or causal mechanisms. They also show that scoring direction materially changes substantive conclusions and that construct-level reporting is preferable to an unvalidated overall HOTS total.

1. Introduction

Higher-order thinking skills (HOTS) are increasingly important educational outcomes because learners are expected to move beyond disciplinary recall toward evidence evaluation, knowledge transfer, explanation, and defensible action. These demands are particularly relevant in biology, where students must connect processes across levels of biological organization, interpret probabilistic evidence, and reason about problems embedded in environmental and social systems. The classical progression from knowledge and comprehension to analysis, synthesis, and evaluation provided an enduring foundation for these expectations [14]. Subsequent accounts have further emphasized that higher-order thinking encompasses critical, creative, reflective, and problem-oriented reasoning rather than performance on a single difficult test or composite score [38, 28].

Biology learning therefore requires more than simply exposing students to inquiry-oriented activities. Students need opportunities to interpret graphs and laboratory observations, formulate and compare explanations, design procedures, evaluate evidence, and judge the consequences of scientific decisions. Inquiry can create conditions for such reasoning, but its effectiveness depends on deliberate scaffolding, task quality, and alignment between instruction and assessment [43, 44, 53]. Accordingly, a defensible HOTS assessment should clearly specify the construct being measured and provide evidence that resulting scores are coherent, interpretable, and meaningfully related to educational behaviour [5, 47].

Recent assessment research reinforces the need for this construct-level precision. Scientific-reasoning instruments have been developed partly because conventional assessments frequently emphasize factual recall rather than reasoning processes [1]. In biology, rubrics aligned with Bloom's taxonomy can clarify cognitive expectations and reveal misconceptions [13], while embedded assessment can make question generation, argumentation, and critical evaluation integral to learning [11]. Nevertheless, cognitively demanding formats do not automatically produce higher-order performance. Higher-order clicker questions have increased engagement

and improved some subsequent outcomes without consistently improving all higher-order examination responses [8]. Similarly, open-note biology examinations appear most productive when students prepare for broad conceptual understanding rather than simple information retrieval [17]. Contemporary analyses of science assessments continue to identify an overrepresentation of remembering and understanding [2], while even teachers who receive item-development training may encounter difficulty constructing questions that genuinely elicit higher-order thinking [15].

The four measured domains are related but should not be treated as interchangeable components of a single latent HOTS trait. Critical thinking and problem solving are higher-order cognitive processes involving interpretation, evaluation, inference, problem representation, alternative generation, decision making, and monitoring [19, 20, 48, 18]. Teamwork is better conceptualized as a collaborative or socio-cognitive competence through which ideas, evidence, roles, and responsibilities are coordinated [22]. Practical innovation is an applied competence concerned with translating knowledge and ideas into feasible improvements, procedures, products, or communicable solutions; it overlaps with creativity but adds constraints of usefulness, evidence, and implementation [49, 45]. Accordingly, this article uses the broader phrase higher-order and collaborative competencies and reserves HOTS for cognitive processes rather than assuming a validated higher-order common factor.

The analytical framework is explicitly relational. Practical innovation is modelled as the focal self-report score because its items require respondents to describe the translation of biological knowledge, data, digital tools, and collaborative work into feasible action. Critical thinking, problem solving, and teamwork are examined as concurrent correlates, not temporally prior causes. The same respondents completed all four scales at one time point, so directional arrows would overstate the design. Figure 1 therefore depicts non-directional conceptual links and the regression is interpreted only as a conditional description of covariance among observed scores.

Pedagogical evidence provides support for this integrated interpretation while also indicating that HOTS dimensions do not develop uniformly. Inquiry-based learning shows a positive average association with critical thinking, although its effects vary substantially according to context and implementation features [9]. Creative pedagogy incorporated into problem-based learning can strengthen analysis, evaluation, and workable solution generation [3], while explicit HOTS-oriented instruction can improve both critical-thinking skills and dispositions [10]. Biology laboratory practice has also been associated with stronger HOTS and experiment-planning performance [32], and relatively short scientific-inquiry interventions can strengthen preservice teachers' scientific reasoning [35]. Project-based learning likewise shows potential for developing creative thinking [26], whereas socioscientific-issue teaching illustrates the value of discussion around complex and evidence-dependent decisions [31]. Virtual laboratories may further extend opportunities for higher-order engagement, although infrastructure and teacher preparation remain important constraints [6].

Technology-supported and collaborative learning environments further demonstrate that twenty-first-century competencies may emerge at different rates. Reviews of mobile learning in science identify collaborative, inquiry-based, project-based, problem-based, game-based, and flipped approaches as recurring pathways for developing HOTS [4]. Mobile-supported inquiry may make gains in collaboration and communication more visible than changes in critical thinking, problem-solving, or creativity [34]. In a virtual ecosystem task, collaboratively constructed reasoning maps were associated with inquiry performance [16]. Within biology teacher education, metacognition has been linked to critical thinking [7], integrated problem-based learning and digital mind maps

have connected critical thinking with creativity [29], and blended project-based approaches have strengthened evaluating and creating [30]. Comparable improvements in analytical, creative, and metacognitive outcomes have been reported in animal physiology [40], although classroom evidence suggests that professional development focused on HOTS does not necessarily translate into explicit metacognitive teaching [36]. Authentic citizen-science activities provide another practical pathway by requiring students to formulate predictions, visualize data, examine limitations, and communicate scientific interpretations [39]. Sustained exposure to active learning may also be more consequential than isolated learning episodes [12].

Three gaps motivate the present analysis. First, HOTS research remains fragmented across constructs and interventions, while assessment studies often aggregate heterogeneous competencies into a single score [42, 5]. Second, reliability is frequently reported without examining scoring polarity, average inter-item association, item-rest behaviour, and the possibility that extremely high internal-consistency coefficients reflect repetitive item content rather than construct validity. Third, practical innovation is commonly invoked as an educational aspiration but its empirical relations with critical thinking, problem solving, and teamwork are less often examined in the same biology-contextualized self-report instrument. The growth of generative artificial intelligence in biology assessment further increases the need to distinguish what a score actually measures from what is merely assumed from task labels [41].

The contribution of this study is therefore methodological and relational rather than a claim of instrument validation. We use comparable per-item scale means, verify item polarity directly from the archived item wording, reverse-score the negatively keyed problem-solving block at the respondent level, report alpha together with one-factor omega, mean inter-item correlations and corrected item-rest correlations, and estimate concurrent associations with heteroscedasticity-robust uncertainty. These procedures provide score-level evidence but do not substitute for convergent, discriminant, factor-structural, or invariance evidence [21, 27, 47, 52]. Prediction-oriented and robustness frameworks may inform future latent-variable work when a substantially larger validation sample is available [24, 23, 37, 46, 51].

Accordingly, the study addresses four questions: (1) What are the score distributions of self-reported critical thinking, problem solving, teamwork, and practical innovation? (2) How internally coherent are the four scales after directionally consistent scoring, and what do item-level diagnostics reveal? (3) How strongly are the four scale scores associated under Pearson and rank-based analyses? and (4) What concurrent associations with practical innovation remain when critical thinking, problem solving, and teamwork are entered together? We expected positive zero-order associations among the four scores, with particularly strong relations of practical innovation to teamwork and critical thinking; the unique problem-solving association was treated as uncertain because its items combine rational problem-solving content with negatively phrased emotional and avoidant responses.

2. Materials and methods

2.1 Research design and analytical scope

This study used an exploratory cross-sectional descriptive-correlational secondary-analysis design. The individual respondent was the unit of analysis. Descriptive procedures mapped score distributions and uncertainty; measurement diagnostics examined internal covariance coherence; and relational procedures

examined concurrent associations among the four observed scale means. The study was not preregistered and the present analysis should be interpreted as exploratory. Because all variables were measured once through the same self-report instrument, temporal ordering, causal effects, and latent construct validity cannot be inferred.

The framework in Fig. 1 guided the relational analysis. Critical thinking, problem solving, and teamwork were entered simultaneously as observed explanatory scores associated with practical innovation. This regression specification does not establish that practical innovation is an outcome caused by the other domains. Its purpose is to determine whether pairwise associations remain after shared variance among the three concurrent scores is taken into account.

2.2 Data source, setting, and participants

Data were collected through a structured online questionnaire administered using Google Forms from 14 to 30 December 2024. The target population comprised undergraduate students enrolled in Biology Education or Biology programmes at several public and private higher education institutions in Indonesia.

Participant recruitment was facilitated through formal research access procedures. Research Permission Letter No. 2463/B.04.02/2024 was submitted to the respective Vice Dean for Academic Affairs or Head of the relevant Study Programme at the participating institutions. Institutional representatives then facilitated dissemination of the study information and questionnaire invitation to eligible students. Recruitment was therefore conducted through institutional channels rather than through a centrally maintained list of individual participants.

Eligibility was defined as current enrolment in an undergraduate Biology Education or Biology programme at one of the participating institutions during the data-collection period. Students who received the study invitation were free to decide whether to participate. Participation was voluntary, and no participant was required to complete the questionnaire as a condition of academic participation or assessment.

Before accessing the questionnaire, prospective participants received information about the study purpose, the nature of participation, the voluntary character of participation, and the handling of their responses. Informed consent to participate was obtained electronically from all participants before questionnaire completion.

The data contained 100 complete response records. Demographic fields were gender, semester, academic programme, and teaching status. Institution information was entered in free-text fields rather than a standardized institutional identifier. Conservative normalization identified 12 named institutions for 95 respondents; five records contained location or organization descriptions that could not be classified reliably. Because clustering variables, invitation denominators, and response-rate data were not systematically retained, institution-level clustering and population response rates could not be modelled. The sample is therefore described as a voluntary multi-institutional sample rather than as statistically representative of Indonesian biology undergraduates.

Table 1
Respondent profile in the archived dataset.

Variable	Category	n	Percent
Gender	Male	11	11.0
Gender	Female	89	89.0
Academic programme	Biology Education	92	92.0
Academic programme	Biology/Science/Other	8	8.0
Semester	1	17	17.0
Semester	3	16	16.0
Semester	5	31	31.0
Semester	7	31	31.0
Semester	9	5	5.0
Teaching status	Not teaching/other	93	93.0
Teaching status	Teaching	7	7.0

Note. Percentages are based on n = 100. Category labels follow the coding documented in the archived analytical file.

2.3 Instrument structure

The instrument contained 113 items recorded as integers from 1 to 5 and was administered in Indonesian. Critical thinking comprised 40 items (CT1–CT40), problem solving 21 items (PSS1–PSS21), teamwork 26 items (TS1–TS26), and practical innovation 26 items (PIS1–PIS26). The archived item wording is explicitly contextualized to biology, including biological concepts, laboratory experiments, experiment design, ecological consequences, scientific discussion, biological data, and digital tools used in biology learning. Representative items include CT1, “I am able to raise critical questions about biological concepts in face-to-face discussions”; PSS1, “I can integrate biological theory learned online with laboratory experiments to solve biological problems”; TS7, “I listen carefully to team members’ views and provide constructive feedback during scientific project discussions”; and PIS8, “I can translate community needs into innovative biology-based solutions.” The export retained numeric responses but not a separate document defining the verbal anchors for 1–5; therefore, no verbal anchor labels are reconstructed here. The full original item wording, scoring direction, and item diagnostics are provided in Supplementary Information ESM 1.

Raw item totals were divided by the number of items in each domain to create comparable 1–5 scale means. An overall HOTS total was not used as a substantive outcome because the unequal item counts would implicitly weight the domains by scale length and because a higher-order common factor has not been established. Cronbach’s alpha, one-factor McDonald’s omega, mean inter-item correlations, and corrected item-rest correlations were treated as score-coherence diagnostics rather than evidence of validated dimensionality. The sample of 100 respondents is not sufficient for stable confirmatory modelling of a 113-item structure; factor-structural and measurement-invariance analyses are therefore deferred to independent validation research.

2.4 Scoring procedure and item-polarity sensitivity

PSS14–PSS21 describe maladaptive responses to biological problems, including loss of patience, anger, explosive or uncooperative behaviour, coercive reactions, social withdrawal, avoidance, and disengagement. Because endorsement of these items indicates poorer rather than better problem regulation, they were reverse-scored at the respondent level as 6 minus the observed response so that higher values consistently represented more effective problem solving. This direction is consistent with the distinction between constructive and dysfunctional problem orientations in the social problem-solving framework [18]. The archived materials did not include a standalone scoring-key document, so this study does not claim documentary verification of the original developer's item key; instead, the scoring rule is reported transparently and the archived raw-scoring results are retained as a sensitivity comparison in Table 3.

2.5 Ethical approval and consent to participate

No new participants were recruited and no new data were collected for this secondary analysis beyond the questionnaire responses collected during the original study period. Human-related procedures were conducted through institutional research access channels, and participation was voluntary.

The original study received full ethical approval from the Research Ethic Committee of the Institute of Research, Community Service, and Publication (LPPMP), Universitas Muhammadiyah Prof. DR. HAMKA (UHAMKA), Jakarta, Indonesia (Reference No. 758/LPPMP/LT/2025; approval date 12 December 2024; approved research period 10 December 2024 to 10 December 2025). The reported data-collection period, 14–30 December 2024, therefore occurred after the stated approval date and within the approved research period. Human-related procedures were conducted in accordance with Ministry of Research and Technology Decree No. 25/M/Kp/III/2013 on Research Ethics, Indonesian Institute of Science Decree No. 06/E/2013 on Research Ethics, and UHAMKA Research Ethics and Publication requirements.

Electronic informed consent to participate was obtained from all participants before questionnaire completion. The original Google Forms export contained name and email fields in addition to study responses. For the present secondary analysis, those direct identifier columns were excluded from the analytical dataset and no participant-level identifiers are reported or shared. The analysis used only the questionnaire responses and non-identifying demographic fields required for the analyses described here.

2.6 Statistical analysis

For each domain, respondent-level means, standard deviations, observed ranges, standard errors, and 95% confidence intervals were calculated. Confidence intervals used the t distribution with 99 degrees of freedom. The midpoint of 3.0 is shown in Fig. 3 only as a response-scale reference and is not treated as a proficiency or mastery threshold. No percentage-of-maximum statistic is used as an achievement criterion.

Internal consistency was evaluated using Cronbach's alpha, one-factor McDonald's omega, mean inter-item correlations, and corrected item-rest correlations. Omega was estimated from a standardized one-factor model for each scale and is reported only as an additional coherence index; it is not presented as evidence that the scale is unidimensional. Item means, standard deviations, item-rest correlations, alpha if item deleted, and omega if item deleted are reported in ESM 1.

Pearson correlations with 95% Fisher-z confidence intervals were calculated among the four domain means. Spearman rank correlations were used as a monotonic sensitivity analysis. Practical innovation was then

regressed simultaneously on critical thinking, problem solving, and teamwork. HC3 heteroscedasticity-robust standard errors and confidence intervals were used for inference, and standardized coefficients were obtained from z-standardized scores. Variance inflation factors described predictor overlap. Unique increments in R^2 were calculated by comparing the full model with reduced models that omitted one predictor at a time; these increments are descriptive and are not interpreted as causal importance.

Model diagnostics included residual inspection, the Breusch–Pagan test, a quadratic Ramsey RESET check, leverage, externally studentized residuals, and Cook's distance. The Breusch–Pagan statistic was LM = 7.305 ($p = 0.063$) and the RESET test was $F = 3.333$ ($p = 0.071$). HC3 inference was retained because it is conservative in a modest sample and less sensitive to leverage. The largest Cook's distance was 0.182, the largest leverage value was 0.176, and the largest absolute externally studentized residual was 2.940. Five records showed a uniform response across all 113 items. Because no response-time or attention-check exclusion rule had been prespecified, the primary analysis retained all complete cases; a sensitivity model excluding those five records yielded a similar coefficient pattern and $R^2 = 0.834$. No duplicate email or name records were present in the source export. Analyses were performed in Python 3.13 using pandas, NumPy, SciPy, statsmodels, and scikit-learn. All tests were two-sided with alpha = .05.

Demographic comparisons were not used for inferential claims because several groups were very small and measurement invariance has not been established. Gender, programme, teaching-status, and semester distributions are therefore reported descriptively in Table 1, with group-specific domain means supplied in ESM 1 for transparency. No p-values are interpreted for demographic differences.

3. Results

3.1 Data completeness and participant profile

All 100 records were complete across the 113 item responses, and every item value fell within the intended 1–5 numeric range. The source export contained no duplicate email records and no duplicate normalized name records. Five respondents used a single response category across all 113 items; these were retained in the primary analysis because no prospective attention-check or response-time exclusion criterion existed, and excluding them did not materially change the regression pattern. The sample was predominantly female (89%), enrolled in Biology Education (92% when the five Biology/Science and three "other" programme records are combined as the non-Biology-Education category), and not currently teaching (93% when "not yet teaching" and "other" are grouped). Twelve named institutions could be conservatively normalized for 95 respondents, while five institution entries remained ambiguous.

No overall HOTS score is reported as a substantive result. The analysis preserves the four domain means because they have unequal item counts and because no higher-order measurement model has established that they can be interpreted as a single common construct. This construct-level approach also prevents strong teamwork or critical-thinking scores from obscuring a distinct problem-solving profile.

3.2 Dimension profile, reliability, and scoring sensitivity

After reverse-scoring PSS14–PSS21, teamwork had the highest mean item score ($M = 3.898$, $SD = 0.629$, 95% CI [3.773, 4.023]), followed by practical innovation ($M = 3.811$, $SD = 0.600$) and critical thinking ($M = 3.799$, $SD =$

0.543). Problem solving remained lower (M = 3.495, SD = 0.463, 95% CI [3.403, 3.587]). The response-scale midpoint shown in Fig. 3 is descriptive only and does not denote competence.

Table 2
Descriptive statistics and internal-consistency diagnostics for the four self-report domains.

Construct	Items	Mean	SD	95% CI	Alpha	Omega	Mean inter-item r	Item-rest r
Critical thinking	40	3.799	0.543	3.691–3.906	0.968	0.968	0.430	0.496–0.777
Problem solving	21	3.495	0.463	3.403–3.587	0.815	0.727	0.190	0.126–0.525
Teamwork	26	3.898	0.629	3.773–4.023	0.973	0.973	0.581	0.598–0.819
Practical innovation	26	3.811	0.600	3.692–3.930	0.967	0.968	0.532	0.579–0.784

Note. Means are per-item scores on the 1–5 response scale after reverse-scoring PSS14–PSS21. Omega is one-factor McDonald's omega; mean inter-item r is the average pairwise item correlation; CI = confidence interval. Very high alpha or omega values may reflect item redundancy and should not be interpreted as construct validation.

Respondent-level reverse scoring materially changed the interpretation of the problem-solving scale. Under the archived raw scoring, the problem-solving mean was 3.415, alpha was 0.929, omega was 0.928, and the correlation with practical innovation was r = 0.614; the standardized problem-solving coefficient in the three-predictor model was beta = 0.128 (p = 0.014). After reversing PSS14–PSS21, the mean increased modestly to 3.495, but internal consistency decreased (alpha = 0.815; omega = 0.727), the correlation with practical innovation decreased to r = 0.468, and the standardized regression coefficient became beta = -0.048 (p = 0.269). Table 3 makes this scoring dependence explicit rather than treating item polarity as a minor limitation.

Table 3
Sensitivity of problem-solving results to archived versus directionally consistent scoring.

Scoring scenario	PSS mean	Alpha	Omega	PSS–PI r	PSS beta (HC3)	p	Model R^2
Data raw scoring	3.415	0.929	0.928	0.614	0.128	0.014	0.857
Reverse PSS14–PSS21	3.495	0.815	0.727	0.468	-0.048	0.269	0.850

Note. Corrected scoring reverse-scores PSS14–PSS21 as 6 minus the observed response. Regression coefficients are standardized and use HC3 robust inference. The comparison demonstrates that the scoring rule changes substantive conclusions.

Critical thinking, teamwork, and practical innovation showed very high internal consistency: alpha/omega were 0.968/0.968, 0.973/0.973, and 0.967/0.968, respectively. Their mean inter-item correlations ranged from 0.430 to 0.581, values high enough to raise a redundancy concern for long scales. Problem solving was less internally coherent after reverse scoring (alpha = 0.815, omega = 0.727, mean inter-item r = 0.190); its corrected item-rest

correlations ranged from 0.126 to 0.525. Thus, reverse scoring aligned score direction but also exposed heterogeneity within the problem-solving item set.

Item-level patterns reinforced this interpretation. Critical-thinking item means ranged from 3.54 to 4.18, teamwork from 3.73 to 4.10, and practical innovation from 3.58 to 3.95. PSS1–PSS13 ranged from 3.63 to 3.84, whereas the reverse-scored PSS14–PSS21 block ranged from 2.69 to 3.37. The diagnostic map in Fig. 4 shows that several problem-solving items also had weaker item-rest relations than items in the other domains. Complete item-level diagnostics, including alpha and omega if an item is deleted, are provided in ESM 1.

3.3 Associations among self-report competency scores

All six Pearson correlations were positive. Practical innovation was strongly correlated with teamwork (r = 0.879, 95% CI [0.826, 0.917]) and critical thinking (r = 0.863, 95% CI [0.802, 0.906]). Problem solving showed a more moderate association with practical innovation (r = 0.468, 95% CI [0.299, 0.609]). Spearman correlations produced the same qualitative ordering, indicating that the main zero-order pattern was not dependent on Pearson linearity alone. These observed-score correlations should not be treated as evidence of discriminant validity or as proof that the four domains represent a single higher-order construct.

Table 4
Pairwise Pearson correlations with 95% confidence intervals and Spearman sensitivity estimates.

Pair	Pearson r	95% CI	Spearman rho
Critical thinking – Problem solving	0.468	0.299–0.608	0.537
Critical thinking – Teamwork	0.791	0.704–0.855	0.784
Critical thinking – Practical innovation	0.863	0.802–0.906	0.857
Problem solving – Teamwork	0.556	0.403–0.678	0.581
Problem solving – Practical innovation	0.468	0.299–0.609	0.501
Teamwork – Practical innovation	0.879	0.826–0.917	0.847

Note. Pearson confidence intervals use Fisher's z transformation. Spearman coefficients are included as a monotonic sensitivity analysis. All estimates are based on n = 100.

The three-predictor associative model accounted for 85.0% of the observed variation in practical innovation (R^2 = 0.850, adjusted R^2 = 0.845); the HC3 robust joint test was F(3, 96) = 269.30, p < .001. Teamwork had a standardized coefficient of beta = 0.550 (95% HC3 CI [0.348, 0.752]), and critical thinking beta = 0.450 (95% HC3 CI [0.272, 0.629]). The corrected problem-solving coefficient was beta = -0.048 (95% HC3 CI [-0.135, 0.038], p = 0.269). Unique increments in R^2 were 0.100 for teamwork, 0.076 for critical thinking, and 0.002 for problem solving. Because the predictors are correlated, these statistics are descriptive decompositions rather than a ranking of causal importance.

Table 5
HC3-robust regression estimates for concurrent associations with practical innovation.

Explanatory score	B	HC3 SE	Standardized beta	95% CI for beta	p	VIF	Unique ΔR^2
Critical thinking	0.498	0.099	0.450	0.272–0.629	< .001	2.682	0.076
Problem solving	-0.063	0.056	-0.048	-0.135–0.038	0.269	1.452	0.002
Teamwork	0.525	0.097	0.550	0.348–0.752	< .001	3.032	0.100

Note. B = unstandardized coefficient; HC3 SE = heteroscedasticity-robust standard error; beta = standardized coefficient; VIF = variance inflation factor; unique ΔR^2 = reduction in model R^2 when the predictor is omitted. The model is associative and cross-sectional.

Variance inflation factors were 2.68 for critical thinking, 1.45 for problem solving, and 3.03 for teamwork. The Breusch–Pagan test was borderline but not conventionally significant (LM = 7.305, p = 0.063), and the quadratic RESET test did not provide strong evidence of nonlinearity (F = 3.333, p = 0.071). Eight observations exceeded the heuristic Cook's-distance value of 4/n, but the maximum was 0.182; five observations exceeded a leverage value of 2(k + 1)/n, with a maximum of 0.176. Excluding the five uniform-response records reduced the sample to 95 and yielded R^2 = 0.834; the coefficient signs and substantive interpretation were unchanged. The high R^2 nevertheless requires caution because all domains were measured concurrently by self-report within the same instrument.

3.4 Exploratory demographic descriptives

The demographic structure was strongly imbalanced: 89 respondents identified as female and 11 as male; 92 were enrolled in Biology Education; only seven reported already teaching; and semester 9 contained five respondents. These cell sizes do not support stable comparative inference, particularly because measurement invariance has not been evaluated. Inferential subgroup tests reported in the earlier draft were therefore removed from the main analysis.

For transparency, ESM 1 provides group-specific domain means by gender, programme, teaching status, and semester. These values are descriptive only and should not be interpreted as demographic effects. A future study designed for subgroup comparison would require planned sampling, adequate group sizes, standardized institution identifiers, and evidence that the measurement structure is invariant across the groups being compared.

4. Discussion

The revised analysis changes the central interpretation in an important way. Practical innovation was strongly concurrent with teamwork and critical thinking, but the apparent unique problem-solving association in the archived analysis disappeared after directionally consistent respondent-level scoring. The most defensible conclusion is therefore not that practical innovation is an "integrated HOTS outcome," but that self-reported innovation co-occurs closely with collaborative and evaluative self-perceptions in this sample while the problem-solving scale behaves differently. This distinction matters because it separates a substantive association pattern

from a scoring artefact and avoids treating correlated self-report scores as a causal mechanism or validated latent structure.

The strong teamwork–practical-innovation association is educationally plausible because biological inquiry is often socially organised. Learners may distribute laboratory roles, compare observations, negotiate explanations, coordinate resources, and communicate procedures, thereby making reasoning visible to peers [22]. Technology-supported inquiry can make collaboration and communication especially salient [34], and blended project structures can connect collaborative production with evaluating and creating [30]. Repeated active-learning exposure may also matter across students with different social dispositions [12]. The present data, however, show only that students who rated themselves highly on teamwork also tended to rate themselves highly on practical innovation; they do not demonstrate that teamwork causes innovation.

Critical thinking also retained a strong concurrent association with practical innovation. This pattern is compatible with accounts of critical thinking that emphasize interpretation, evaluation, inference, and reasoned judgment [19, 20]. Innovation in biology is constrained by evidence, feasibility, safety, and disciplinary coherence, so evaluative self-perceptions may reasonably covary with self-perceptions of practical innovation. The continuing dominance of lower-order demands in some science assessments [2] and the difficulty teachers experience when constructing genuinely higher-order items [15] underscore why evaluative reasoning should be measured explicitly rather than inferred from task difficulty.

Problem solving showed a different pattern. After PSS14–PSS21 were reversed, its zero-order correlation with practical innovation remained positive but moderate, while its unique regression coefficient was close to zero after critical thinking and teamwork were included. More importantly, alpha fell from .929 under archived raw scoring to .815 under directionally consistent scoring, omega fell from .928 to .727, and several item-rest correlations became weak. This does not imply that problem solving is educationally unimportant. Instead, it suggests that the 21-item set may combine rational problem-solving behaviours with emotional and avoidant response tendencies that do not operate as a single homogeneous scale. The distinction is consistent with social problem-solving theory, which separates problem orientation from problem-solving style [18].

The scoring audit therefore becomes part of the substantive result rather than a footnote. PSS14–PSS21 are unambiguously negative in behavioural direction: endorsing anger, coercion, avoidance, withdrawal, or fear cannot be interpreted as stronger effective problem solving when the remaining items are positively keyed. Reverse scoring is necessary for directional consistency, yet the reduced internal coherence after that correction shows that a scoring rule alone does not establish unidimensionality. Independent cognitive interviewing, item-development documentation, and larger-sample factor analysis are needed before this problem-solving score is used for high-stakes interpretation or as evidence that one competence is “weaker” than another.

The high model R^2 is likewise a reason for caution rather than a claim of explanatory power. All four domains were assessed at the same time, through the same response format, and within one long questionnaire. Shared method variance, acquiescence, self-confidence, semantic overlap, and common response tendencies can strengthen correlations among self-report scores. An item about proposing an improvement during group work, for example, may simultaneously cue teamwork, critical evaluation, and innovation. The regression therefore summarizes conditional covariance among observed scores; it does not show that practical innovation is determined by the other domains.

Reliability and validity must also remain conceptually separate. Very high alpha and omega coefficients for critical thinking, teamwork, and practical innovation indicate strong covariance coherence but may also reflect repetitive content in long scales. Future validation should compare plausible correlated-factor, higher-order, and bifactor structures, examine convergent and discriminant validity, and establish measurement invariance before subgroup means are compared [21, 27, 24, 23, 37, 46, 51]. External performance criteria are equally important because self-reported dispositions and demonstrated reasoning are not equivalent [47, 48, 52].

A practical assessment implication follows directly from the analysis: the four domains should be reported separately rather than collapsed into an overall HOTS total. Construct-level reporting preserves information that would otherwise be hidden by unequal item counts and by an unverified higher-order structure. School-based HOTS assessment is most useful when it informs instructional design rather than merely ranking learners [50, 33]. Biology rubrics that score content and reasoning together can clarify cognitive expectations [13], while embedded assessment can make argument construction part of learning [11]. Higher-order clicker questions and open-note examinations further show that cognitively demanding formats require explicit guidance and appropriate preparation [8, 17].

The biology-specific wording of the archived instrument is an important strength relative to the earlier draft's description, but it does not convert self-report into performance evidence. The items refer explicitly to biological concepts, laboratory experiments, ecological consequences, design problems, data analysis, and communication of biological findings. These contexts align conceptually with inquiry, problem-based, project-based, and citizen-science practices reported in biology and science education [9, 32, 35, 39, 29, 3, 26]. A future performance study could test whether the present self-report domains predict the quality of actual experimental designs, data interpretations, or innovation products.

Metacognitive scaffolding is one plausible intervention hypothesis for that future work, particularly for the heterogeneous problem-solving domain. Learners could be asked to define a problem, distinguish causes from symptoms, identify constraints, generate alternatives, specify decision criteria, implement a response, and state what evidence would justify revision. Reviews of HOTS-oriented teaching emphasize explicit regulation of such processes [25]. Associations between metacognition and critical thinking in biology-related settings [7], incomplete transfer of HOTS professional development into explicit metacognitive teaching [36], and gains in analytical, creative, and metacognitive outcomes under PjB-HOTS instruction [40] justify testing this proposition experimentally rather than assuming it from the present correlations.

The demographic data do not support subgroup conclusions. Small and imbalanced groups can produce unstable means and effect estimates, and the present instrument has not been shown to function equivalently across gender, programme, teaching status, semester, or institution. Descriptive subgroup values are therefore reported only for transparency. Longitudinal or stratified replication would be needed to distinguish curricular development from cohort composition, selective participation, institutional context, or changes in self-perception.

The study's contribution is consequently diagnostic and relational. It demonstrates how a long biology-contextualized self-report matrix can be audited transparently by separating construct scores, checking item direction, reporting multiple internal-consistency diagnostics, and examining concurrent associations with robust uncertainty. The most consequential result is that the problem-solving conclusion changes after respondent-level

reverse scoring, whereas the strong practical-innovation associations with teamwork and critical thinking remain. This is a reproducibility lesson as much as an educational one: score construction is part of the scientific claim.

Several limitations constrain interpretation. The design is cross-sectional and self-reported, the sample is modest and nonprobabilistic, and several demographic cells are small. Institution fields were free-text, preventing reliable multilevel modelling; invitation denominators and response rates were not retained. The source export contained names and email addresses, although these fields were removed from the analytical copy used here. Five uniform response patterns were retained because no prespecified attention-check or response-time rule existed, and Google Forms did not provide completion-duration data. The archived materials did not include a standalone authoritative scoring-key document or a full instrument-development dossier. Factor structure, discriminant validity, measurement invariance, and external criterion validity were not tested. These constraints mean that the findings should be interpreted as associations among observed self-report scores in this voluntary sample, not as national estimates or demonstrated cognitive performance.

5. Conclusion

This exploratory secondary analysis mapped self-reported critical thinking, problem solving, teamwork, and practical innovation in 100 biology undergraduates using a 113-item biology-contextualized questionnaire. Directionally consistent scoring of PSS14–PSS21 modestly increased the problem-solving mean but substantially changed its psychometric and relational profile: alpha and omega decreased, its correlation with practical innovation weakened, and its adjusted regression coefficient was no longer distinguishable from zero. Critical thinking, teamwork, and practical innovation remained highly internally consistent, although coefficients of this magnitude also raise the possibility of item redundancy.

Practical innovation showed strong concurrent associations with teamwork and critical thinking, while the full three-predictor model explained a large proportion of observed score variation. These results should be read as covariance among self-perceptions rather than evidence of causal influence, demonstrated skill, or a validated integrated HOTS construct. The immediate measurement priority is to document the instrument's development and scoring key, evaluate its factor structure in a larger independent sample, and link self-report scores to authentic biology performance. For current reporting, separate construct scores are more defensible than an overall HOTS total.

Declarations

Funding

This study was self-funded and received no external funding or grant support.

Data availability

The data supporting the findings of this study are based on an archived, de-identified assessment dataset. Due to institutional and ethical restrictions related to participant privacy, the full dataset is not publicly available. De-identified data or analysis materials may be made available from the corresponding author upon reasonable request and subject to institutional approval.

Health and safety

No laboratory, animal, or hazardous experimental procedures were performed in this secondary analysis.

Declaration of generative AI use

OpenAI's ChatGPT was used only for language polishing, organizational refinement, and preparation of visual materials. All substantive aspects of the study, including data analysis, interpretation of results, scientific judgment, and the final conclusions, were determined by the authors, who take full responsibility for the accuracy and integrity of the manuscript.

Author accountability statement

All authors made substantial contributions to the conception and design of the study, data interpretation, manuscript drafting, critical revision, and approval of the final version. All authors agree to be accountable for the integrity of the work and confirm that the manuscript has not been published elsewhere.

Author Contribution

LS: Conceptualization, methodology, data curation, formal analysis, writing—original draft, visualization, and project administration.GS: Methodology, validation, supervision, writing—review and editing.SM: Conceptualization, literature review, validation, and writing—review and editing.Zul: Methodology, visualization, validation, and writing—review and editing.

Acknowledgement

The authors would like to acknowledge Universitas Muhammadiyah Prof. DR. HAMKA for academic support during the preparation of this manuscript.

Data Availability

The data supporting the findings of this study are based on an archived, de-identified assessment dataset. Due to institutional and ethical restrictions related to participant privacy, the full dataset is not publicly available. De-identified data or analysis materials may be made available from the corresponding author upon reasonable request and subject to institutional approval.

References

1. Abate T, Michael K, Angell C. Assessment of scientific reasoning: Development and validation of scientific reasoning assessment tool. Eurasia J Math Sci Technol Educ. 2020;16(12):1–15. https://doi.org/10.29333/ejmste/9353.
2. Adofo S, Kärkkäinen S, Keinonen T. Examining Ghanaian lower secondary school science test through revised Bloom's taxonomy. Sci Educ Int. 2024;35(4):382–93. https://doi.org/10.33828/sei.v35.i4.9.

3. Affandy H, Sunarno W, Suryana R. Integrating creative pedagogy into problem-based learning: The effects on higher order thinking skills in science education. Think Skills Creativity. 2024;53:101575. https://doi.org/10.1016/j.tsc.2024.101575.
4. Afikah A, Astuti SRD, Suyanta S, Jumadi J, Rohaeti E. Mobile learning in science education to improve higher-order thinking skills (HOTS) and communication skills: A systematic review. Int J Adv Comput Sci Appl. 2022;13(7):698–704. https://doi.org/10.14569/IJACSA.2022.0130782.
5. Ahmad S, Prahmana RCI, Kenedi AK, Helsa Y, Arianil Y, Zainil M. (2018). The instruments of higher order thinking skills. Journal of Physics: Conference Series, 943(1). https://doi.org/10.1088/1742-6596/943/1/012053
6. Alshaikh AAN. (2022). The reality of using virtual labs in teaching advanced biology curricula in developing higher-order thinking skills (HOTS) among female teachers at secondary level in Al-Kharj. Education Research International, 2022, 8605202. https://doi.org/10.1155/2022/8605202
7. Amin AM, Corebima AD, Zubaidah S, Mahanal S. The correlation between metacognitive skills and critical thinking skills at the implementation of four different learning strategies in animal physiology lectures. Eur J Educational Res. 2020;9(1):143–63. https://doi.org/10.12973/eu-jer.9.1.143.
8. Anderson DK, Schoenleber M, Korshavn S. Higher-order clicker questions engage students and prepare them for higher-order thinking activities. J Microbiol Biology Educ. 2023;24(1). https://doi.org/10.1128/jmbe.00151-22.
9. Arifın Z, Saputro S, Kamari A. The effect of inquiry-based learning on students' critical thinking skills in science education: A systematic review and meta-analysis. Eurasia J Math Sci Technol Educ. 2025;21(3). https://doi.org/10.29333/ejmste/15988.
10. Arviani FP, Wahyudin D, Dewi L. Role of teaching strategies in promoting students' higher order thinking skills and critical thinking dispositions. Int J Learn Teach Educational Res. 2023;22(9):332–49. https://doi.org/10.26803/ijlter.22.9.19.
11. Barak M, Dori YJ. Enhancing higher order thinking skills among inservice science teachers via embedded assessment. J Sci Teach Educ. 2009;20(5):459–74. https://doi.org/10.1007/s10972-009-9141-z.
12. Beckerson WC, Anderson J, Kulkarni S, Perpich J, Yoder-Himes DR. It’s about time: Exploring the dose-dependent effects of active learning on students of different social personalities in an upper-level biology course. J Coll Sci Teach. 2024;53(2):177–86. https://doi.org/10.1080/0047231X.2024.2316378.
13. Bissell AN, Lemons PP. A new method for assessing critical thinking in the classroom. Bioscience. 2006;56(1):66–72. https://doi.org/10.1641/0006-3568. 056%5B0066:ANMFAC%5D2.0.CO;2.
14. Bloom BS, editor. (1956). Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain. David McKay. https://books.google.com/books?id=hos6AAAAIAAJ
15. Bolat A, Karamustafaoğlu S. Investigation of the items prepared by teachers who received training on item preparation measuring higher-order thinking skills. Participatory Educational Res. 2024;11(6):189–215. https://doi.org/10.17275/per.24.86.11.6.
16. Chen J, Wang M, Dede C, Grotzer TA. Analyzing student thinking reflected in self-constructed cognitive maps and its influence on inquiry task performance. Instr Sci. 2021;49(3):287–312. https://doi.org/10.1007/s11251-021-09543-8.
17. Driessen EP, Beatty AE, Ballen CJ. Evaluating open-note exams: Student perceptions and preparation methods in an undergraduate biology class. PLoS ONE. 2022;17(8):e0273185.

https://doi.org/10.1371/journal.pone.0273185.

18. D'Zurilla TJ, Nezu AM. Development and preliminary evaluation of the Social Problem-Solving Inventory. Psychol Assess. 1990;2(2):156–63. https://doi.org/10.1037/1040-3590.2.2.156.
19. Ennis RH. (2011). The nature of critical thinking: An outline of critical thinking dispositions and abilities. University of Illinois. https://education.illinois.edu/docs/default-source/faculty-documents/robert-ennis/thenatureofcriticalthinking_51711_000.pdf?sfvrsn=7bb51288_0
20. Facione PA. (2015). Critical thinking: What it is and why it counts. Insight Assess. https://insightassessment.com/iaresource/critical-thinking-what-it-is-and-why-it-counts/
21. Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. J Mark Res. 1981;18(1):39–50. https://doi.org/10.1177/002224378101800104.
22. Gokhale AA. Collaborative learning enhances critical thinking. J Technol Educ. 1995;7(1):22–30. https://doi.org/10.21061/jte.v7i1.a.2.
23. Hair JF, Hult GTM, Ringle CM, Sarstedt M. (2022). A primer on partial least squares structural equation modeling (PLS-SEM). SAGE Publications. https://us.sagepub.com/en-us/nam/a-primer-on-partial-least-squares-structural-equation-modeling-pls-sem/book270548
24. Hair JF, Risher JJ, Sarstedt M, Ringle CM. When to use and how to report the results of PLS-SEM. Eur Bus Rev. 2019;31(1):2–24. https://doi.org/10.1108/EBR-11-2018-0203.
25. Hamzah H, Hamzah MI, Zulkifli H. Systematic literature review on the elements of metacognition-based higher order thinking skills (HOTS) teaching and learning modules. Sustainability. 2022;14(2):813. https://doi.org/10.3390/su14020813.
26. Hasibuan MP, Sunarno W. Creative thinking enhancement through project-based learning in science: A meta-analytic review. Int J Evaluation Res Educ. 2025;14(5):3994–4005. https://doi.org/10.11591/ijere.v14i5.34521.
27. Henseler J, Ringle CM, Sarstedt M. A new criterion for assessing discriminant validity in variance-based structural equation modeling. J Acad Mark Sci. 2015;43:115–35. https://doi.org/10.1007/s11747-014-0403-8.
28. Heong YM, Othman WB, Yunos JBM, Kiong TT, Hassan RB, Mohammad MMB. The level of Marzano higher order thinking skills among technical education students. Int J Social Sci Humanity. 2011;1(2):121–5. https://doi.org/10.7763/IJSSH.2011.V1.20.
29. Hidayati N, Zubaidah S, Suarsini E, Praherdhiono H. Examining the relationship between creativity and critical thinking through integrated problem-based learning and digital mind maps. Univers J Educational Res. 2019;7(9A):171–9. https://doi.org/10.13189/ujer.2019.071620.
30. Hujjatusnaini N, Corebima AD, Prawiro SR, Gofur A. The effect of blended project-based learning integrated with 21st-century skills on pre-service biology teachers' higher-order thinking skills. Jurnal Pendidikan IPA Indonesia. 2022;11(1):104–18. https://doi.org/10.15294/jpii.v11i1.27148.
31. Högström P, Gericke N, Wallin J, Bergman E. Teaching socioscientific issues: A systematic review. Sci Educ. 2025;34(5):3079–122. https://doi.org/10.1007/s11191-024-00542-y.
32. Kamarudin N, Lee TJ, Talib O, Hassan A. Enhancing higher order thinking skills via inquiry-based laboratory practices among science students. Man India. 2017;97(17):185–94. https://www.researchgate.net/profile/Othman-Talib-2/publication/309291204_INQUIRY-

BASED_LEARNING_IN_LABORATORY_SCIENCE/links/5b14e96ba6fdcc4611e20404/INQUIRY-BASED-LEARNING-IN-LABORATORY-SCIENCE.pdf.

33. Kosasih A, Supriyadi T, Firmansyah MI, Rahminawati N. Higher-order thinking skills in primary school: Teachers' perceptions of Islamic education. J Ethnic Cult Stud. 2022;9(1):56–76. https://doi.org/10.29333/ejecs/994.
34. Kousloglou M, Petridou E, Molohidis A, Hatzikraniotis E. Assessing students' awareness of 4Cs skills after mobile-technology-supported inquiry-based learning. Sustainability. 2023;15(8):6725. https://doi.org/10.3390/su15086725.
35. Koyunlu Ünlü Z, Babayiğit Ö, Ünlü V. Development of preservice elementary teachers' scientific reasoning skills through scientific inquiry. Res Sci Technological Educ. 2025;43(2):656–73. https://doi.org/10.1080/02635143.2024.2332720.
36. Labak I, Maruncek I, Blazetic S. Biology teaching that develops the metacognitive aspect of learning: How to learn competence. J Educ E-Learning Res. 2024;11(1):113–27. https://doi.org/10.20448/jeelr.v11i1.5395.
37. Legate AE, Hair JF, Chretien JL, Risher JJ. PLS-SEM: Prediction-oriented solutions for HRD researchers. Hum Res Dev Q. 2023;34(1):91–109. https://doi.org/10.1002/hrdq.21466.
38. Lewis A, Smith D. Teaching for higher order thinking. Theory Into Pract. 1993;32(3):131–7. https://doi.org/10.1080/00405849309543588.
39. Lichti D, Mosley P, Callis-Duehl K. Learning from the trees: Using Project Budburst to enhance data literacy and scientific writing skills in an introductory biology laboratory during remote learning. Citiz Science: Theory Pract. 2021;6(1). https://doi.org/10.5334/CSTP.432.
40. Liline S, Tomhisa A, Rumahlatu D, Sangur K. The effect of the PjB-HOTS learning model on cognitive learning, analytical thinking skills, creative thinking skills, and metacognitive skills of biology education students. J Turkish Sci Educ. 2024;21(1):175–95. https://doi.org/10.36681/tused.2024.010.
41. Linnebur LA, Syed SB, Girolamo N, Reid J. Can ChatGPT pass introductory biology? An analysis of ChatGPT responses to introductory biology exam questions. J Coll Sci Teach. 2025;54(6):463–76. https://doi.org/10.1080/0047231X.2025.2535962.
42. Lismaya L, Hartono H, Subali B, Sumarni W, Ridlo S, Nuswowati M. Higher order thinking skills research trends: A bibliometric analysis in selected journals (2014 to 2023). Int J Evaluation Res Educ. 2025;14(1):695–707. https://doi.org/10.11591/ijere.v14i1.29590.
43. Marshall JC, Horton RM. The relationship of teacher-facilitated, inquiry-based instruction to student higher-order thinking. School Sci Math. 2011;111(3):93–101. https://doi.org/10.1111/j.1949-8594.2010.00066.x.
44. Miri B, David BC, Uri Z. Purposely teaching for the promotion of higher-order thinking skills: A case of critical thinking. Res Sci Educ. 2007;37(4):353–69. https://doi.org/10.1007/s11165-006-9029-2.
45. Ramirez RPB, Ganaden MS. (2008). Creative activities and students' higher order thinking skills. Education Quarterly, 66(1), 22–33. https://d1wqtxts1xzle7.cloudfront.net/92907336/1511-libre.pdf?1666534792=&response-content-disposition=inline%3B+filename%3DCreative_Activities_and_Students_Higher.pdf&Expires=1788002882&Signature=aol2M1NFrx2rmPt29~E~9qLFbJ8t54skzmaAvSta4ZMwRmvxTmwlccmVDWoBoT~K5CJu57FlxB18lA372v2sOtK7JF~EgeaNy4v~w18z3-ZuoqBq5WX4MjlvyCHhrL589tzoduKdg29qctVpFmbKsaPhhlphqm9F6PMRZuEoOGA38IJ-CDum5hj23u2Zmuio6pMZ-Y7qgRLx0InjCijXLx55L8v2rgvGap-whw2IVVLiHP9Q3N~xpxXH241VgVu4RjQ50H3CID3~JjaRbB9Iu5VPj-

2XG3BhqGtOkcsJuUzcvfyuMnDgYwR08Y9XxH1w-OM1R1L30MjjVYFuQHuOwQ__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA

46. Sarstedt M, Ringle CM, Cheah JH, Ting H, Moisescu OI, Radomir L. Structural model robustness checks in PLS-SEM. Tour Econ. 2020;26(4):531–54. https://doi.org/10.1177/1354816618823921.
47. Shaw A, Liu OL, Gu L, Kardonova E, Chirikov I, Li G, Hu S, Yu N, Ma L, Guo F, Su Q, Shi J, Shi H, Loyalka P. Thinking critically about critical thinking: Validating the Russian HEIghten critical thinking assessment. Stud High Educ. 2020;45(9):1933–48. https://doi.org/10.1080/03075079.2019.1672640.
48. Sosu EM. The development and psychometric validation of a Critical Thinking Disposition Scale. Think Skills Creativity. 2013;9:107–19. https://doi.org/10.1016/j.tsc.2012.09.002.
49. Torrance EP. (1974). Torrance Tests of Creative Thinking. Scholastic Testing Service. https://doi.org/10.1037/t05532-000
50. Wilson DM, Narasuman S. Investigating teachers' implementation and strategies on higher order thinking skills in school based assessment instruments. Asian J Univ Educ. 2020;16(1):70–84. https://doi.org/10.24191/ajue.v16i1.8991.
51. Yuan K-H, Zhang Z, Latan H, Hair JF, Noonan R. (2023). Statistical and psychometric properties of three weighting schemes of the PLS-SEM methodology. In H. Latan, J. F. Hair, & R. Noonan, editors, Partial least squares path modeling (2nd ed., pp. 81–112). Springer Nature. https://doi.org/10.1007/978-3-031-37772-3_4
52. Zhou Y, Ning Y, Chen J, Zhang W, Wijaya TT. Development and validation of Mathematical Higher-Order Thinking Scale for high school students. Psychol Sch. 2024;61(8):3160–92. https://doi.org/10.1002/pits.23213.
53. Zohar A, Dori YJ. Higher order thinking skills and low-achieving students: Are they mutually exclusive? J Learn Sci. 2003;12(2):145–81. https://doi.org/10.1207/S15327809JLS1202_1.

Figures

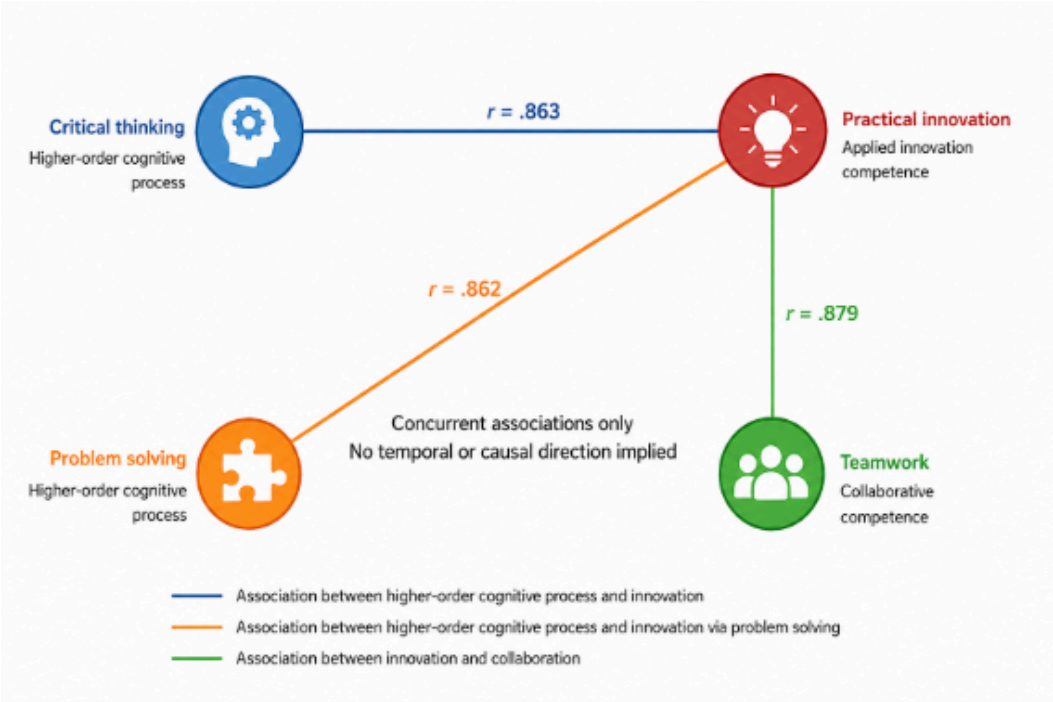


Figure 1

Conceptual framework for concurrent associations among higher-order and collaborative self-report competencies. Lines are non-directional and do not represent temporal or causal pathways.

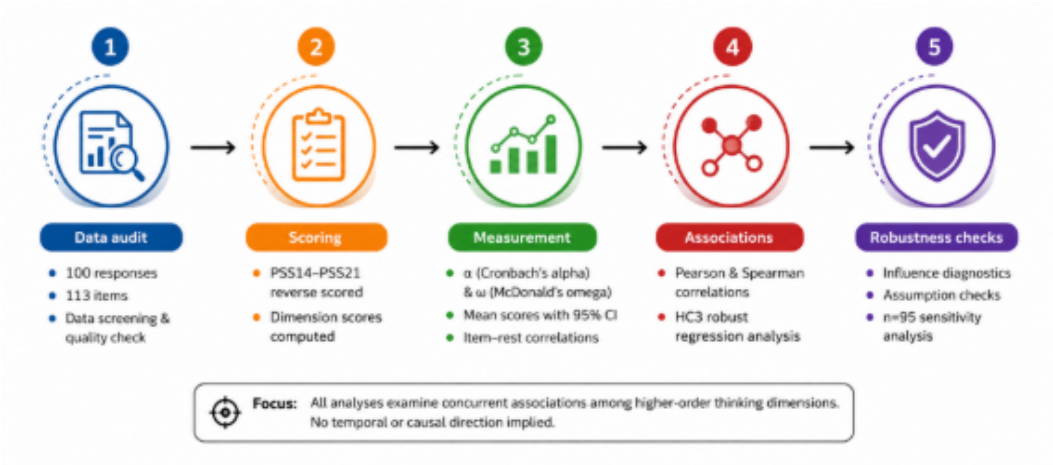


Figure 2

Analytical workflow from source-data audit and directionally consistent scoring to measurement diagnostics, robust associative modelling, and sensitivity analysis.

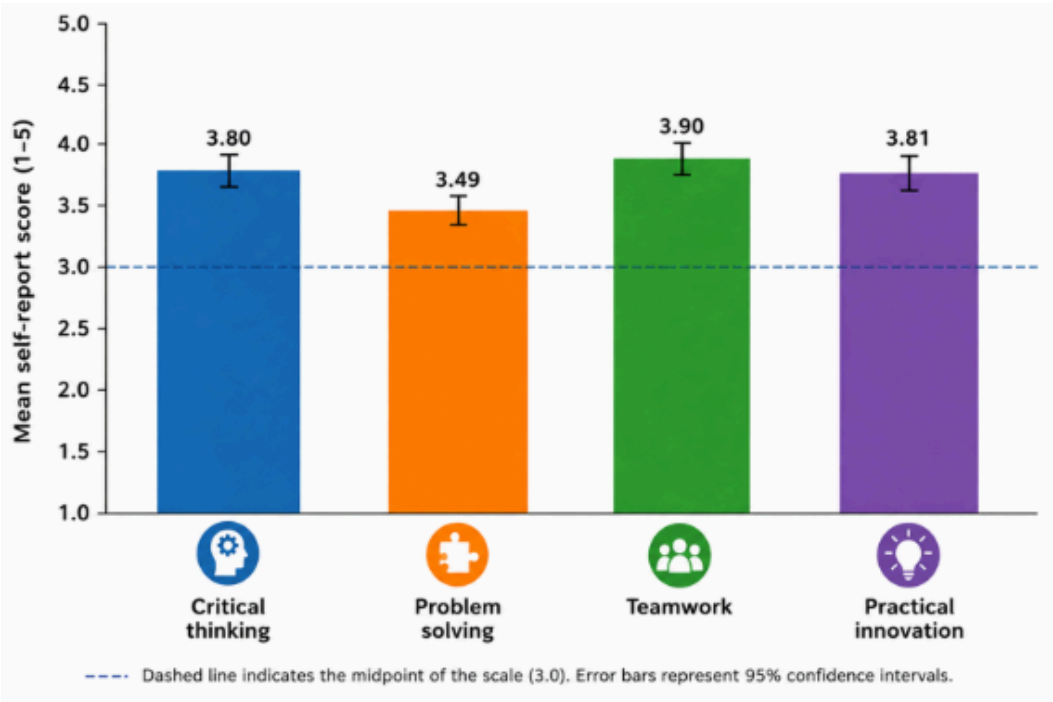


Figure 3

Mean item scores for the four self-report domains after directionally consistent scoring. Error bars are 95% confidence intervals; the dashed line marks the response-scale midpoint and is not a proficiency threshold.

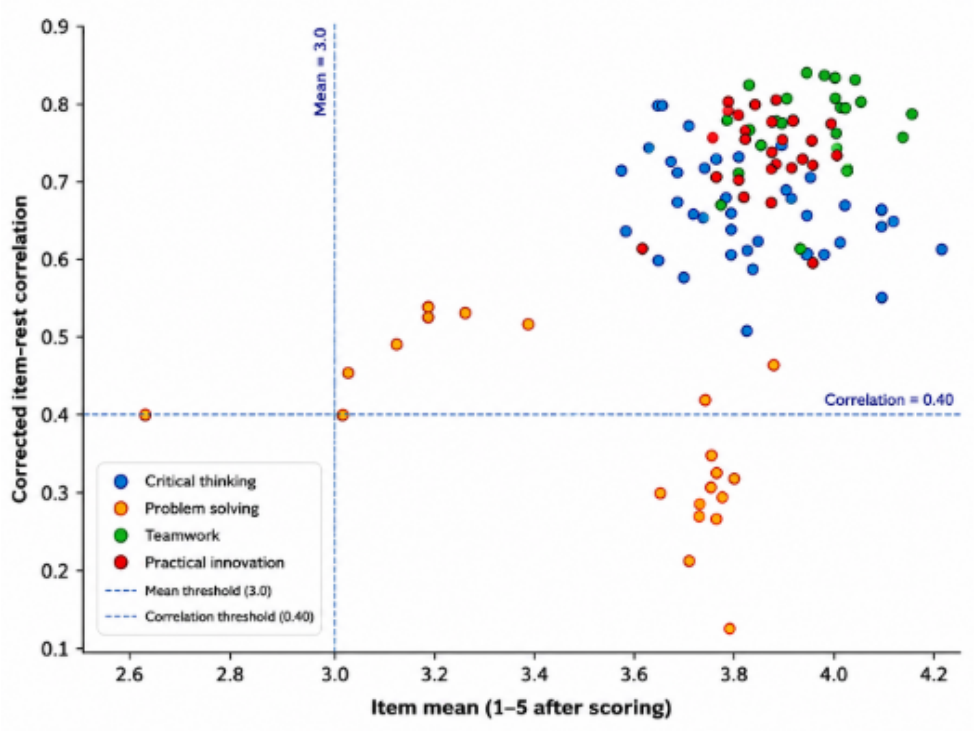


Figure 4

Item-level diagnostic map after directionally consistent scoring. Dashed lines mark an item mean of 3.0 and a corrected item-rest correlation of .40; neither line is a construct-validity threshold.

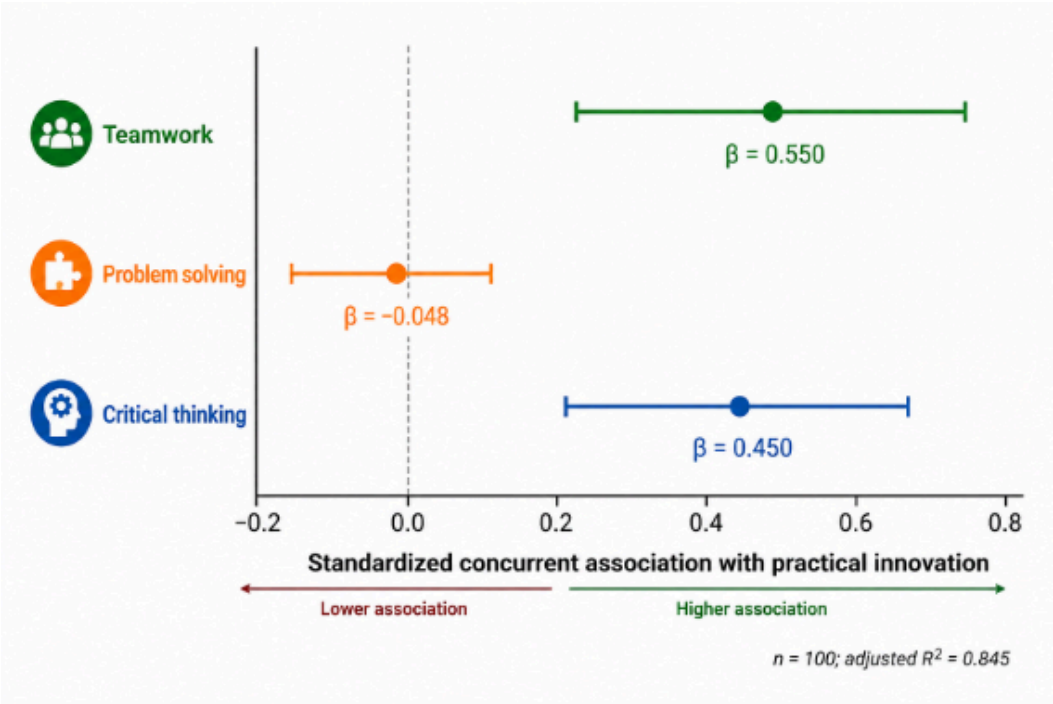


Figure 5

Standardized concurrent associations with practical innovation after reverse-scoring PSS14–PSS21. Error bars are 95% HC3 robust confidence intervals.

Supplementary Files

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

- ESM1ItemDiagnosticsandAnalysis.xlsx