International Journal of Contemporary Sciences (IJCS)
Vol. 4, No. 9, 2026: 2249-2260

Measuring Pedagogical Quality: Validity, Reliability, Practicality, and Effectiveness Parameters of the SISSCA Learning Model

Siska Nerita[1*], Lufri[2], Helendra[3]
[1]Universitas PGRI Sumatera Barat
[2,3]Universitas Negeri Padang
Corresponding Author: Siska Nerita1 siskabio84@gmail.com

ARTICLEINFO

Keywords: Validity Index, Reliability, Practicality Index, SISSCA Model

Received : 15 July
Revised : 20 August
Accepted: 29 September

©2026 Nerita, Lufri, Helendra: This is an open-access article distributed under the terms of the Creative Commons Atribusi 4.0 Internasional.

ABSTRACT

This article analyzes the quality parameters of the SISSCA e-module learning model for pre-service biology teachers. The study focuses on four measurement parameters: validity, reliability, practicality, and effectiveness. Data were collected from five expert validators and sixty-eight fifth-semester students in West Sumatra, divided into experimental and control groups. The results show a mean validity index of 0.89, reliability of 0.983, practicality index of 0.99, and a statistically significant learning gain favoring the experimental group (t=7.643; p=0.001; Cohen' sd=1.02). These findings demonstrate the convergence of the four parameters, offering a comprehensive triangulation template for evaluating pedagogical innovations

INTRODUCTION

In educational design research (EDR), a model is not judged by its conceptual elegance alone but by whether it satisfies a defined set of quality parameters. Nieveen (1999) frames product quality around three interlocking criteria — validity, practicality, and effectiveness — each operationalized into measurable indicators before a design can be considered ready for wider adoption, and Plomp (2013) situates these criteria within a cyclical Preliminary Research–Prototyping–Assessment sequence that treats measurement as inseparable from design itself. Validity asks whether a design is theoretically sound and internally consistent; practicality asks whether practitioners can actually use it under ordinary conditions; effectiveness asks whether it produces the intended learning outcomes. A fourth, often under-reported parameter — reliability, or the consistency of the measurement process itself (Cronbach, 1951; Koo & Li, 2016) — determines how much confidence can be placed in the first three.

The SISSCA model (Stimulating, Identifying Problems, Solving Problems, Strategizing, Creating Instruments, Appraising) was developed to strengthen higher-order thinking skills (HOTS) and assessment-design competence among pre-service biology teachers, a competency global research consistently flags as under-developed (Anderson & Krathwohl, 2001; Barak & Dori, 2005; Ibrahim & Suparno, 2021; Liu & Li, 2022). SISSCA integrates three pedagogical traditions whose individual evidence bases are well established but rarely combined and jointly measured: Project-Based Learning, shown across meta-analytic and discipline-specific studies to strengthen critical thinking, creativity, and evaluation literacy (Chen, Wang, & Zhang, 2022; Genc, 2015; Hidayat & Rohman, 2023; Rahmawati & Taylor, 2022; Savery, 2006; Smyrnaiou, Georgakopoulou, & Sotiriou, 2020; Thomas, Mergendoller, & Michaelson, 1999); drill-based reinforcement, grounded in behaviorist learning theory (Aqib & Rohmadi, 2016; Skinner, 1953); and a digital e-module, consistent with evidence that interactive digital scaffolds raise engagement and self-regulation (Kaur & Nadarajah, 2022; Laili, Rohman, & Diah, 2019; Wibowo & Kurniawan, 2022).

This article deliberately narrows its scope to the measurement layer of the SISSCA study: the specific parameters, instruments, and thresholds used to certify the model as valid, reliable, practical, and effective. Readers who intend to evaluate their own instructional designs need a concrete, reusable account of which parameters to measure and how to interpret them, independent of the substantive content of any one model — a need documented across the problem-based and project-based learning literature, where inconsistent measurement has made cross-study comparison difficult (Hmelo-Silver, 2004; Jonassen, 2011; Kirschner, Sweller, & Clark, 2006). Isolating the parameters also allows a clearer view of how the four quality dimensions relate to one another, a distinction increasingly emphasized in active-learning research generally (Freeman et al., 2014; Prince, 2004). Accordingly, this article aims to (a) describe the parameters and instruments used to evaluate SISSCA, situating each in the measurement literature it draws on; (b) report the resulting data for each parameter family; and (c) discuss how the four parameter sets jointly substantiate the model's quality.

LITERATURE REVIEW

Educational Design Research Quality Criteria

Nieveen (1999) argues that an instructional product can only be judged ready for wider use once it clears three interlocking criteria — validity, practicality, and effectiveness — each translated into concrete, measurable indicators. Plomp (2013) extends this into a cyclical Preliminary Research–Prototyping–Assessment sequence, treating measurement as inseparable from the design process itself rather than a final check performed after development ends.

H1: The SISSCA model's design products (blueprint, guidebook, e-module) achieve a mean validity index at or above the 0.70 threshold recommended by Polit and Beck (2006).

Reliability of Expert Judgment

Because validity ratings are only as trustworthy as the consistency of the raters producing them, Cronbach (1951) and Koo and Li (2016) argue that reliability should be reported as an independent parameter rather than assumed from validity alone. Cronbach's Alpha assesses internal consistency of the rating instrument, while the Intraclass Correlation Coefficient assesses agreement among multiple raters, with values above 0.90 considered excellent.

H2: The SISSCA validity rubric achieves Cronbach's Alpha and Intraclass Correlation Coefficient values at or above their respective 0.70 and 0.75 thresholds.

After the hypothesis section, the conceptual framework below situates SISSCA's three integrated learning paradigms — Project-Based Learning, behaviorist drill-based reinforcement, and connectivist digital scaffolding — against the measurement parameters used to evaluate the resulting model.

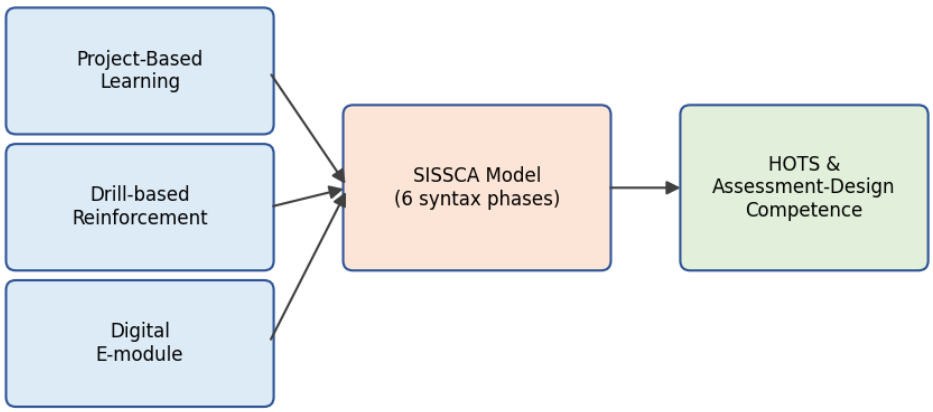

Figure 1. Conceptual Framework linking SISSCA's three pedagogical traditions to its measured outcomes

METHODOLOGY

This study re-examines the measurement component of a Research and Development (R&D) study that followed Plomp's (2013) three-phase design — Preliminary Research, Prototyping, and Assessment. The analysis here is restricted to the Assessment phase, in which four parameter families were operationalized.

Participants comprised two groups. Five expert validators — specialists in biology education, learning assessment, and educational technology — rated the model's design products (model blueprint, lecturer's guidebook, student e-module) for validity. Sixty-eight fifth-semester students enrolled in the Evaluation of Biological Learning Processes and Outcomes course were divided into an experimental group (n = 34), taught using SISSCA, and a control group (n = 34), taught through conventional lecture and discussion; both groups were assessed for practicality and effectiveness parameters. Four parameter families and their instruments were used, summarized in Table 1.

Table 1. Parameter families, instruments, and acceptance thresholds used to evaluate the SISSCA model.

Parameter family	Instrument	Statistic / index	Acceptance threshold
Validity	Analytic expert rubric across eight design dimensions	Content Validity Index (CVI), mean validation score (0-1)	CVI / mean ≥ 0.70 (Polit & Beck, 2006)
Reliability	Same expert rubric, repeated across five raters	Cronbach's Alpha (Cronbach, 1951); Intraclass Correlation Coefficient (Koo & Li, 2016)	$\alpha \geq 0.70$; ICC ≥ 0.75
Practicality	Observation checklists (six syntax phases) and student-response questionnaires	Normalized practicality index (0-1) (Nieveen, 1999)	Very high: 0.81-1.00
Effectiveness	Pre-test / post-test HOTS instrument (C4 analysis, C5 evaluation, C6 creation)	Independent-samples t-test; Cohen's d (Cohen, 1988)	$p < 0.05$; $d \geq 0.8$ = large effect

For the validity parameter, the CVI for each dimension was computed by dividing the number of experts rating an item as valid (score of 3 or 4 on a four-point scale) by the total number of experts, following the calculation and reporting conventions clarified by Polit and Beck (2006). For the reliability parameter, Cronbach's Alpha (Cronbach, 1951) assessed internal consistency of the rubric, while the Intraclass Correlation Coefficient — interpreted using the benchmarks of Koo and Li (2016) — assessed agreement among the five raters. For the practicality parameter, two trained observers scored implementation fidelity across SISSCA's six syntax phases on a four-point Likert scale, normalized to a 0-1 index consistent with Nieveen's (1999) practicality criterion, and students completed a response questionnaire covering usability, engagement, clarity, and facilitation. For the effectiveness parameter, both groups completed pre-test and post-test HOTS instruments constructed around Anderson and Krathwohl's (2001) revision of Bloom's taxonomy; the difference in gain scores was tested with an independent-samples t-test after normality (Shapiro & Wilk, 1965) and homogeneity assumptions were confirmed, and effect

size was quantified with Cohen's d (Cohen, 1988), where 0.2-0.5 indicates a small effect, 0.5-0.8 a medium effect, and ≥ 0.8 a large effect. All statistical analyses were performed in SPSS version 26.0.

RESEARCH RESULT

Validity Parameters

Five experts rated the SISSCA design products across eight rubric dimensions: model construction, rationale, supporting theory, syntax, reaction principles, social system, support system, and instructional/nurturant effects. Every dimension exceeded the 0.70 acceptance threshold recommended by Polit and Beck (2006), and the overall mean validation index reached 0.89, categorized as 'very valid.' Scores are reported in Table 2 and Figure 2.

Table 2. Validity parameter scores by rubric dimension.

Validity dimension	Mean score
Social system	0.92
Model construction	0.90
Support system	0.90
Instructional and nurturant effects	0.90
Model rationale	0.89
Syntax	0.89
Reaction principle	0.88
Supporting theory	0.85
Overall mean validity index	0.89

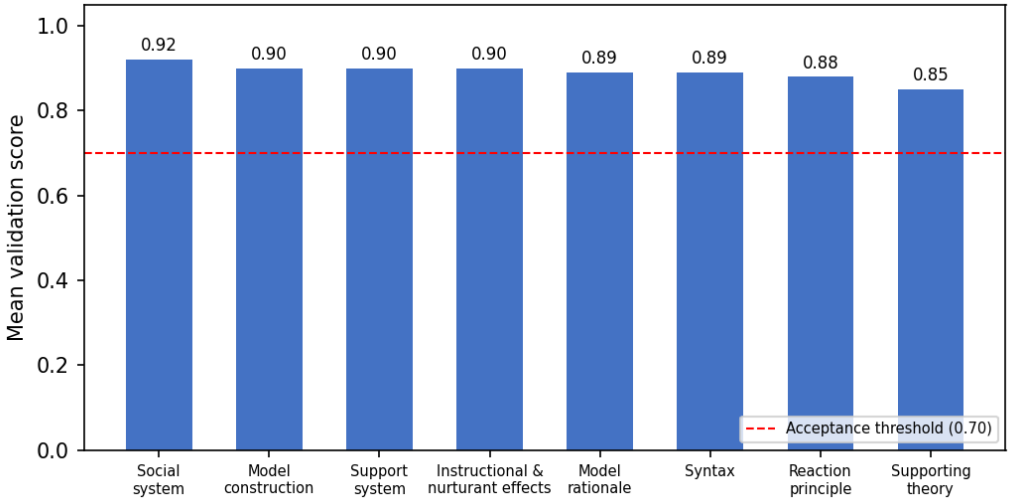

Figure 2. Validity parameter scores across eight rubric dimensions.

Social system (0.92) — the quality of prescribed lecturer-student and peer-peer roles — scored highest, consistent with problem-based and project-based learning research showing that clearly defined social roles are a strong predictor of successful implementation (Hmelo-Silver, 2004; Jonassen, 2011; Savery, 2006). Supporting theory (0.85), while still comfortably above threshold, was the weakest dimension, echoing a broader critique in the PjBL literature that hybrid models often integrate practices more successfully than they integrate the theories behind them (Kirschner, Sweller, & Clark, 2006).

Reliability Parameters

Because validity scores are only as trustworthy as the consistency of the raters producing them, reliability was assessed alongside validity rather than treated as a formality. Cronbach's Alpha reached 0.983 (Cronbach, 1951), and the Intraclass Correlation Coefficient reached 0.983 — both far above their respective 0.70 and 0.75 thresholds, and within the 'excellent' band described by Koo and Li (2016). Practically, this means that items targeting the same construct moved together, and that the five experts agreed with each other to a degree that leaves little room for idiosyncratic bias.

Practicality Parameters

Practicality was measured across eight implementation sessions of the Evaluation of Biological Learning Processes and Outcomes course, combining observation checklists, student-response questionnaires, and lecturer reflection logs, consistent with Nieveen's (1999) recommendation that practicality be triangulated across observed use and perceived use rather than inferred from either alone. The composite practicality index reached 0.99, categorized as 'very high.' Phase-level and respondent-level scores are shown in Table 3 and Figure 3.

Table 3. Practicality parameter scores by implementation phase and respondent measure.

Practicality parameter	Score
Stimulating (syntax 1)	1.00
Identifying Problems (syntax 2)	1.00
Solving Problems (syntax 3)	0.95
Strategizing (syntax 4)	>0.98
Creating Instruments (syntax 5)	>0.98
Appraising (syntax 6)	>0.98
Student response (overall)	0.90
– Facilitation	0.99
– Instructional clarity	0.96
Composite practicality index	0.99

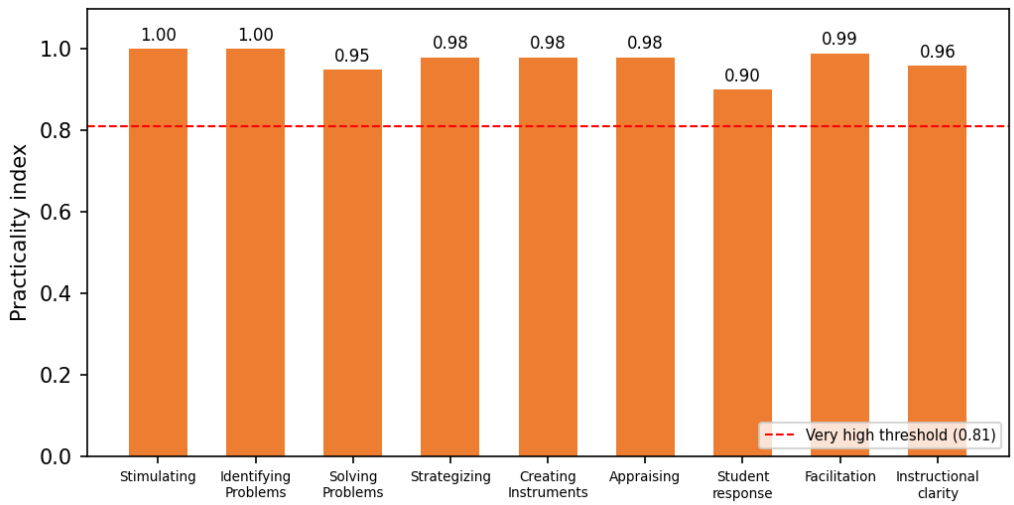

Figure 3. Practicality parameter scores across implementation phases, student response, and student perception.

The only phase to score below 0.98 was Solving Problems (0.95), where groups spent additional time reconciling competing explanations before converging on a solution — a pattern consistent with Jonassen's (2011) observation that the solution-generation stage of problem-based designs is where cognitive load peaks, even when overall implementation runs smoothly. Among student-response sub-parameters, facilitation (0.99) scored higher than instructional clarity (0.96), suggesting that the human-facilitation component of the model is, at present, slightly outperforming the e-module's self-explanatory design.

Effectiveness Parameters

Effectiveness was measured through pre-test and post-test HOTS instruments covering analysis (C4), evaluation (C5), and creation (C6), following Anderson and Krathwohl's (2001) revised taxonomy. Table 4 reports the resulting parameters.

Table 4. Effectiveness parameters: pre-test/post-test HOTS scores and statistical indices.

Effectiveness parameter	**Experimental group (n = 34)**	**Control group (n = 34)**
Pre-test mean ± SD	45.20 ± 5.34	44.85 ± 6.02
Post-test mean ± SD	79.47 ± 4.88	66.25 ± 5.71
Gain score	34.27	21.40
t-statistic (independent samples)	7.643	—
p-value	0.001	—
Cohen's d	1.02 (large effect)	—

The experimental group's gain score (34.27 points) exceeded the control group's (21.40 points) by a considerable margin. The independent-samples t-test confirmed the difference was statistically significant ($p = 0.001$, below the 0.05 threshold), and Cohen's d of 1.02 exceeded the 0.8 cut-off Cohen (1988) sets for a large effect. In interpretive terms, the average SISSCA-taught student outperformed roughly 79 percent of conventionally-taught peers. All formulas used to derive gain scores follow the standard difference formula:

$$\text{Gain} = \text{Post-test score} - \text{Pre-test score} \quad \ldots\ldots\ldots\ldots\ldots (1)$$

DISCUSSION

Considered individually, each parameter family clears its respective threshold: validity ($0.89 \geq 0.70$), reliability (α and ICC $= 0.983 \geq 0.70/0.75$), practicality (0.99, 'very high'), and effectiveness ($p = 0.001$, $d = 1.02$). Considered together, the four families tell a coherent story rather than four disconnected facts. The high reliability of the validity ratings gives confidence that the validity scores reflect genuine design quality rather than one rater's opinion. The strength of the social-system validity dimension (0.92) is echoed in the practicality data, where facilitation — the human-interaction component the social-system dimension was meant to capture — scored highest among student-response sub-parameters (0.99). And the model's classroom practicality plausibly underpins its effectiveness: a model that is difficult to implement with fidelity would be unlikely to produce the learning gains observed here, regardless of how well it scores on paper — a caution raised more generally in the problem-based learning

literature, where designs that validate well on paper have sometimes failed to translate into classroom gains when practicality was not separately verified (Hmelo-Silver, 2004; Kirschner, Sweller, & Clark, 2006).

This magnitude of effect is broadly consistent with meta-analytic evidence that active, project-based approaches outperform lecture-based instruction, though typically with smaller effects than reported here (Freeman et al., 2014; Prince, 2004), suggesting the combined PjBL–drill–e-module design may compound the benefits each component produces individually (Chen, Wang, & Zhang, 2022; Smyrnaiou, Georgakopoulou, & Sotiriou, 2020). This convergence pattern is the central methodological takeaway of a parameter-level analysis: no single parameter family is sufficient evidence of a model's quality on its own, but when validity, reliability, practicality, and effectiveness parameters are measured independently and still point in the same direction, the resulting claim of quality is considerably stronger than any one figure could support alone. For researchers designing their own evaluation protocols — particularly in teacher-education contexts where assessment literacy itself is the target competency (Liu & Li, 2022; Rahmawati & Taylor, 2022; Setiawan & Nasrullah, 2023) — this suggests that resources are better spent measuring several parameter families with modest precision than a single family with maximal precision.

CONCLUSIONS AND RECOMMENDATIONS

This article isolated and reported the measurement parameters used to evaluate the SISSCA learning model: validity (mean 0.89 across eight rubric dimensions), reliability ($\alpha = 0.983$; ICC = 0.983), practicality (composite index 0.99), and effectiveness ($p = 0.001$; Cohen's $d = 1.02$). Each parameter family independently exceeded its established acceptance threshold, and, examined jointly, the four families corroborate one another rather than standing as isolated figures. Beyond certifying SISSCA specifically, the parameter set and thresholds documented here offer a transferable template — grounded in Plomp's (2013) and Nieveen's (1999) design-research framework and standard psychometric practice (Cohen, 1988; Cronbach, 1951; Koo & Li, 2016; Polit & Beck, 2006; Shapiro & Wilk, 1965) — for how validity, reliability, practicality, and effectiveness can be operationalized, measured, and triangulated when evaluating any instructional model, including those integrating project-based, behaviorist, and digitally mediated components.

It is recommended that teacher-education institutions adopting SISSCA or similar hybrid instructional designs measure all four parameter families independently rather than relying on a single aggregate quality score, and that facilitation support be maintained during the Solving Problems phase, where implementation fidelity was comparatively lower.

ADVANCED RESEARCH

This study's measurement scope was limited to a single institution, a single cohort of 68 students, and a single offering of the Evaluation of Biological Learning Processes and Outcomes course, which constrains the generalizability of the reported parameter values. Future research should replicate this parameter-level assessment across multiple institutions and cohorts, examine whether the validity-reliability-practicality-effectiveness convergence pattern holds for other hybrid instructional models, and investigate longer-term retention of the HOTS gains observed here beyond the immediate post-test.

ACKNOWLEDGMENT

The authors gratefully acknowledge the five expert validators and the fifth-semester students enrolled in the Evaluation of Biological Learning Processes and Outcomes course for their participation in this study, as well as the institution's Faculty of Teacher Training and Education for facilitating the research.

REFERENCES

Anderson, L. W., & Krathwohl, D. R. (2001). A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. Longman.

Aqib, Z., & Rohmadi, M. (2016). Metode pembelajaran efektif. Yrama Widya.

Barak, M., & Dori, Y. J. (2005). Enhancing higher order thinking skills among inservice science teachers via embedded assessment. Journal of Science Teacher Education, 16(4), 293–310. https://doi.org/10.1007/s10972-005-1107-8

Buck Institute for Education. (2013). Project based learning handbook. BIE Press.

Chen, X., Wang, L., & Zhang, D. (2022). Project-based learning in STEM education: A meta-analysis. International Journal of STEM Education, 9(1), 45. https://doi.org/10.1186/s40594-022-00377-4

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Lawrence Erlbaum Associates.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. Psychometrika, 16(3), 297–334. https://doi.org/10.1007/BF02310555

Darling-Hammond, L., Flook, L., Cook-Harvey, C., Barron, B., & Osher, D. (2020). Implications for educational practice of the science of learning and development. Applied Developmental Science, 24(2), 97–140. https://doi.org/10.1080/10888691.2018.1537791

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. Proceedings of the National Academy of Sciences, 111(23), 8410–8415. https://doi.org/10.1073/pnas.1319030111

Genc, M. (2015). The project-based learning approach in environmental education. International Research in Geographical and Environmental Education, 24(2), 105–117. https://doi.org/10.1080/10382046.2014.993169

Hidayat, A., & Rohman, M. (2023). Enhancing pre-service teachers' evaluation literacy through digital project-based learning. Asia-Pacific Education Researcher, 32(2), 177–190. https://doi.org/10.1007/s40299-022-00713-8

Hmelo-Silver, C. E. (2004). Problem-based learning: What and how do students learn? Educational Psychology Review, 16(3), 235–266. https://doi.org/10.1023/B:EDPR.0000034022.16470.f3

Ibrahim, M., & Suparno, P. (2021). Developing higher order thinking skills in biology education through inquiry-based instruction. Cakrawala Pendidikan, 40(3), 482–496. https://doi.org/10.21831/cp.v40i3.41945

Jonassen, D. H. (2011). Supporting problem solving in PBL. Interdisciplinary Journal of Problem-Based Learning, 5(2), 95–113. https://doi.org/10.7771/1541-5015.1256

Kaur, G., & Nadarajah, R. (2022). Students' perceptions of e-modules for blended learning in higher education. Education and Information Technologies, 27(6), 8417–8435. https://doi.org/10.1007/s10639-022-11098-y

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work. Educational Psychologist, 41(2), 75–86. https://doi.org/10.1207/s15326985ep4102_1

Kolb, D. A. (1984). Experiential learning: Experience as the source of learning and development. Prentice Hall.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. Journal of Chiropractic Medicine, 15(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Laili, E. N., Rohman, M., & Diah, M. (2019). The development of e-module based on problem-based learning for biology subject. Journal of Physics: Conference Series, 1227(1), 012003. https://doi.org/10.1088/1742-6596/1227/1/012003

Liu, S., & Li, X. (2022). The impact of digital assessment tools on pre-service teachers' competence in designing HOTS-based tests. Journal of Science Teacher Education, 33(5), 475–493. https://doi.org/10.1080/1046560X.2021.1976290

Nieveen, N. (1999). Prototyping to reach product quality. In J. van den Akker, R. M. Branch, K. Gustafson, N. Nieveen, & T. Plomp (Eds.), Design approaches and tools in education and training (pp. 125–135). Kluwer Academic Publishers. https://doi.org/10.1007/978-94-011-4255-7_10

Plomp, T. (2013). Educational design research: An introduction. In T. Plomp & N. Nieveen (Eds.), Educational design research (pp. 11–50). SLO.

Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? Critique and recommendations. Research in Nursing & Health, 29(5), 489–497. https://doi.org/10.1002/nur.20147

Prince, M. (2004). Does active learning work? A review of the research. Journal of Engineering Education, 93(3), 223–231. https://doi.org/10.1002/j.2168-9830.2004.tb00809.x

Rahmawati, Y., & Taylor, P. C. (2022). A transformative project-based learning approach for teacher education in Indonesia. Cultural Studies of Science Education, 17(2), 453–474. https://doi.org/10.1007/s11422-021-10079-6

Savery, J. R. (2006). Overview of problem-based learning: Definitions and distinctions. Interdisciplinary Journal of Problem-Based Learning, 1(1), 9–20. https://doi.org/10.7771/1541-5015.1002

Setiawan, A., & Nasrullah, M. (2023). Digital PjBL to foster HOTS among biology preservice teachers. Journal of Biological Education, 57(4), 433–445. https://doi.org/10.1080/00219266.2022.2035421

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). Biometrika, 52(3/4), 591–611. https://doi.org/10.1093/biomet/52.3-4.591

Siemens, G. (2005). Connectivism: A learning theory for the digital age. International Journal of Instructional Technology and Distance Learning, 2(1), 3–10.

Skinner, B. F. (1953). Science and human behavior. Macmillan.

Smyrnaiou, Z., Georgakopoulou, A., & Sotiriou, S. (2020). The effectiveness of a project-based learning approach in STEM education. Education Sciences, 10(8), 213. https://doi.org/10.3390/educsci10080213

Susanti, H., & Anwar, R. (2024). Blended project-based learning to improve students' creativity and HOTS in biology. Journal of Baltic Science Education, 23(1), 75–88. https://doi.org/10.33225/jbse/24.23.75

Thomas, J. W., Mergendoller, J. R., & Michaelson, A. (1999). Project-based learning: A handbook for middle and high school teachers. Buck Institute for Education.

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. Harvard University Press.

Wibowo, H., & Kurniawan, D. (2022). Evaluating the effectiveness of e-learning modules for developing HOTS in higher education. Heliyon, 8(9), e10735. https://doi.org/10.1016/j.heliyon.2022.e10735

World Economic Forum. (2022). The future of jobs report 2022. World Economic Forum.