Are Prospective Biology Teachers Ready for Competency-Based Assessment? Evidence from Achievement Test Construction

Namitha P. M.

Senior Prof. (Dr.) Abdul Gafoor K. Department of Education, University of Calicut

DOI: https://doi.org/10.47772/IJRISS.2026.102000012

Received: 15 October 2026; Accepted: 20 October 2026; Published: 19 November 2026

ABSTRACT

The National Education Policy (NEP) 2020 advocates for competency-based, constructivist evaluation. However, prospective teachers often struggle to translate assessment theory into high-quality test design. This study critically evaluated the structural, cognitive, and contextual quality of biology achievement tests constructed by prospective teachers. In this study, 31 biology achievement tests developed by 4th-semester B.Ed. Natural Science students during their school internship were analyzed using an expert-validated 5-point holistic rubric across seven core dimensions. This study reveals notable gaps between prospective teachers' content knowledge and their operational literacy of test construction. Pre-service teachers consistently over-relied on lower-order cognitive recall (LOTS: Remembering and Understanding) according to Revised Bloom's Taxonomy (RBT), showing a marked deficit in formulating higher-order, contextualized problem-solving questions (HOTS). Most achievement tests exhibited only superficial real-world context, weak integration of non-verbal stimuli, and inconsistent structural formatting. Inter-criterion correlation analysis confirmed the evaluation rubric's reliability as a diagnostic benchmark for teacher education. The findings highlight an urgent need to reform assessment coursework in B.Ed. program's curricula to explicitly train prospective biology teachers in crafting context-rich, competency-aligned evaluation tools.

Keywords: Teacher Assessment Literacy, Achievement Tests, Revised Bloom's Taxonomy, Natural Science Education, Constructivist Evaluation.

INTRODUCTION

The current policy framework emphasizes that modern teachers' roles extend beyond conventional teaching to include action research and evaluation competencies. Consequently, educators must be highly literate in diverse assessment strategies and continuous formative evaluation. Assessment is a foundational and indispensable element of the teaching-learning process, providing crucial feedback that guides instruction and monitors student progress. The transition from traditional, rote-based learning to student-centred, constructivist approaches demands a corresponding transformation in assessment practices. The National Curriculum Framework (NCF) 2005 advocates for assessments that enable students to think critically, solve problems individually or collectively, and demonstrate creativity. A constructivist mode of assessment requires a dialogue between teacher and student, incorporating higher-order thinking skills such as analyzing, creating, predicting, and evaluating. It is situation-based, non-judgmental, and relies on continuous feedback to assess the true potential of learners (Sheeja, 2023).

To meet the demands of constructivist assessment, educators must employ context-rich items, moving away from recall and simple memorization. Instead of asking for a direct definition, a context-rich item embeds core concepts within realistic, engaging, and meaningful real-world scenarios. The skill of context-rich item writing marks an operational shift from simple, definitional questions ("What is X?") to complex, narrative-based problems ("How do you use X to solve a real-world problem?"). This skill is particularly vital in the subject area of Biology, which holds an inherent connection to constructivist principles. Many biological concepts are best understood and assessed through inquiry-based situations and real-world problem-solving. Despite its documented importance, a critical gap exists between policy formulation and ground-level execution. Current teacher education programs often overemphasize theoretical frameworks and content delivery, resulting in a

demonstrable deficit in pre-service teachers' (Bachelor of Education) practical ability to design high-quality, constructivist-based assessments. This deficit perpetuates outdated rote-assessment methods in schools. While literature acknowledges this deficiency, there is an urgent need to empirically audit the actual assessment tools drafted by prospective teachers during their field placements.

To address this gap, the present study undertakes a systematic analysis of achievement tests developed by the B.Ed. candidates during their school internships. By breaking down item-writing skills into distinct, measurable criteria such as context authenticity, cognitive demand, and structural distractors, the study demonstrates how teacher education institutions can objectively audit internship artifacts (Brookhart, 2011; Haladyna et al., 2002; Xu & Brown, 2016). This study is designed as a rubric-based evaluation to analyze teacher-developed achievement items.

THEORETICAL OVERVIEW

Constructivism posits that knowledge is an evolving, subjective framework actively constructed by the learner through experience, social interaction, and reflection, rather than a static product passively received. This epistemological view shifts classroom dynamics from traditional, teacher-centred instruction toward student-centred environments characterized by active inquiry and personalized learning (Alam, 2016; Allen, 2022; Pandey, 2019). Within this framework, teachers act as facilitators and scaffolders, assisting learners in connecting new information to existing mental schemas (Priyamvada, 2018). Consequently, classrooms utilize experiential frameworks such as the 5E model to prioritize authentic engagement, critical thinking, and collaborative problem-solving over rote memorization (Singh & Yaduvanshi, 2015; Sharma, 2025).

To sustain this pedagogical shift, assessment practices must undergo a parallel transformation from traditional behaviourist models toward process-oriented, constructivist approaches (Alam, 2016; Allen, 2022). Traditional measurement relies heavily on decontextualized selection tasks and summative sorting mechanisms that evaluate passive recall (Allen, 2022). In contrast, constructivist assessment relies on authentic, situation-based tasks that elicit higher-order thinking skills—such as analyzing, evaluating, and creating—within real-world problem scenarios (Alam, 2016; Priyamvada, 2018). By embedding opportunities for student reflection and immediate feedback, constructivist evaluation functions as a diagnostic and formative tool that tracks changes in a learner’s conceptual schema over time (Pandey, 2019; Sheeja, 2023).

Effective implementation of these practices relies heavily on an educator's assessment literacy, a construct initially defined by Stiggins (1991, 1995) as the capacity to distinguish between sound and unsound measurement. Popham (2020) further emphasizes that assessment-literate teachers master fundamental measurement principles to guide instructional decisions. Modern frameworks expand assessment literacy beyond discrete technical skills into an iterative, multi-dimensional competence encompassing conceptual knowledge, practical skills, and socio-emotional beliefs (Xu & Brown, 2016; Pastore, 2019, 2023). In discipline-specific contexts like science education, Abell and Siegel (2011) highlight that teachers require specialized Science Teacher Assessment Literacy (STAL) to evaluate complex domain goals, including scientific inquiry, experimental reasoning, and conceptual modelling (Hung, 2024).

Despite these theoretical advancements, empirical literature repeatedly reveals a persistent gap between assessment theory and actual classroom practice (Koksal, 2004; Pastore, 2023). Teachers frequently exhibit low competence in test construction, resulting in teacher-made tests that suffer from poor content validity, misaligned instruction, and a heavy concentration on lower-level cognitive domain items (Koksal, 2004; Hamafyelto, 2015). While existing research addresses general teacher conceptions and training frameworks (Fulcher, 2012; Meijer et al., 2023; Hull, 2025), there remains a critical lack of empirical investigation specifically dissecting how teachers construct classroom tests. This highlights an urgent need for a detailed analysis of achievement test item construction patterns, particularly among pre-service teachers, to identify systemic item-writing deficiencies and inform targeted interventions that foster higher-order, context-rich assessment literacy.

Objectives of the Study

1. To assess the competence in achievement test item writing in total among 4th semester B.Ed. students of Natural Science at Calicut University
2. To identify specific strengths and weaknesses in the achievement tests constructed by the B.Ed. Natural Science students.

METHODOLOGY

This study is designed as a rubric-based evaluation to analyze teacher-made achievement test items, where the performance of prospective secondary school biology teachers on the achievement test was analyzed. It is carried out in a representative sample of 31 students in the fourth semester of B.Ed. natural science option from colleges within the Palakkad and Malappuram revenue districts of Kerala affiliated to the University of Calicut. The achievement test papers developed by prospective biology teachers during their school internships were collected via a Google Form. These test papers were analyzed using a five-point scale quantitative holistic rubric (Rubric for evaluation of achievement test created by prospective secondary-school Biology teachers). It is designed as a comprehensive evaluative framework for assessing the quality and pedagogical soundness of achievement tests, with a specific focus on contextualization. It has seven criteria listed below

1. Learning outcome alignment
2. Cognitive level distribution
3. Contextualization depth
4. Structural integrity
5. Scientific accuracy and content
6. Quality of nonverbal context
7. Quality of instruction

To establish content validity and pedagogical soundness, the draft rubrics were subjected to a formal expert review process involving two experienced biology educators with over five years of specialization in secondary school science and educational assessment. Reliability is confirmed using inter-criterion correlation analysis of the rubric, which effectively measures core pillars of assessment design. Strong links were found between Learning Outcomes and Cognitive Levels (r=0.68), validating that alignment with objectives helps balance cognitive demands. For the scope of this study, formal statistical reliability coefficients such as Inter-Rater Reliability (IRR) via Cohen's Kappa or Intraclass Correlation Coefficients (ICC)–were not empirically computed. Because these rubrics were designed as qualitative-diagnostic instruments to map baseline error patterns and drive the structural development of the Constructivist Assessment Literacy Workshop Module rather than to assign high-stakes, summative student grades the determination of face, content, and construct validity via intensive expert vetting was deemed methodologically sufficient for this investigation.

ANALYSIS AND FINDINGS

This study evaluated the assessment literacy of 31 prospective secondary-school biology teachers during their school internships using an expert-validated five-point holistic rubric across seven core dimensions. The strengths and weaknesses of achievement test construction were identified using basic statistical procedures such as mean, median, mode and percentage analysis

Table 1 presents the component-wise performance on competence in achievement test item writing among student teachers (N = 31). Performance was assessed across seven criteria using a 5-point performance scale (1 = Needs Improvement, 2 = Satisfactory, 3 = Good, 4 = Very Good, 5 = Excellent), yielding a maximum

potential score of 35 across all seven criteria.

Table 1 Component-Wise Performance on Competence in Achievement Test Item Writing Among Student Teachers

Criteria	Needs Improvement	Satisfactory	Good	Very Good	Excellent	M	Mean%	MEDIAN	MODE	SD
Instructional Clarity	12.9	9.7	22.6	22.6	32.26	3.52	70.32	4.0	5	1.39
Scientific Accuracy & Content	0.0	9.7	41.9	41.9	6.45	3.45	69.03	3.0	3	0.77
Structural integrity	3.2	3.2	51.6	32.3	9.68	3.42	68.39	3.0	3	0.85
Learning Outcome Alignment	0.0	19.4	45.2	25.8	9.68	3.26	65.16	3.0	3	0.89
Cognitive Level (RBT)	0.0	32.3	54.8	9.7	3.23	2.84	56.77	3.0	3	0.73
Quality of Non-verbal Context	45.2	35.5	16.1	3.2	0.0	1.77	35.48	2.0	1	0.84
Contextualisation Depth	58.1	25.8	12.9	3.2	0.0	1.61	32.26	1.0	1	0.84
Total	-	-	-	-	-	19.87	56.77	20.0	16	3.69

Table 1 shows that across the seven criteria, the Mean score of Competence in Achievement Test Item Writing is 19.87 (*SD* = 3.69, Mdn = 20.00, *Mode*= 16 out of 35.00) representing an overall student teacher's performance efficiency on *Achievement Test Item Writing* of 56.77%. This composite average is equivalent to "Satisfactory to Good" on overall assessment literacy.

However, a detailed analysis of individual criteria reveals a divergent performance profile: student teachers demonstrated strong performance on mechanical, structural, and linguistic components, but exhibited severe, performance deficits in constructivist design components (contextualization, non-verbal visual integration, and higher-order cognitive balancing) of test item writing competence.

A. **High-Performing Criteria**

1. Instructional Clarity

With a mean score of 3.52 (*SD =1.39, Mdn of 4.00, and Mode of 5.00*) Instructional Clarity emerged as the highest-performing domain with 70.32 percent. A substantial majority of the prospective teachers (77.42%) met or exceeded the "Good" benchmark, with 32.26% reaching the "Excellent" level (22.58% Very Good, 22.58% Good). Conversely, 9.68% were rated Satisfactory and 12.90 percent fell into Needs Improvement; indicating that student teachers possess strong phrasing and communication skills for crafting direct, comprehensible directions. However, the relatively large standard deviation (*SD =1.39*) reflects a notable variance between high-performing and struggling student teachers.

2. Scientific Accuracy and Content

With a mean score of 3.45 (*SD =0.77, Mdn =3.00, and Mode=3.00*) Scientific Accuracy ranked second highest overall perming test item writing component among student teachers with a 69.03 percent efficiency.

Student teachers demonstrated strong subject-matter knowledge, with 90.33% of participants achieving "Good" (41.94%), "Very Good" (41.94%), or "Excellent" (6.45%) ratings. Only 9.68% scored at the "Satisfactory" level, and 0.00% fell into "Needs Improvement." The low standard deviation (*SD =0.77*) confirm that student teachers reliably construct factually sound items that match standard secondary biology terminology.

3. Structural Integrity

With a mean score of 3.42 (*SD* =0.85, Mdn of 3.00, and Mode of 3.00) Structural Integrity of the test items student teachers write also is strong with 68.39 percent efficiency. High majority of the prospective teachers (93.55%) met or exceeded the "Good" benchmark, with 9.68% reaching the "Excellent" level (32.26% Very Good, 51.61% Good). Conversely, 3.23% were rated Satisfactory and 3.23% fell into Needs Improvement; indicating that student teachers can effectively format questions, define singular problems, and avoid obvious cues. Relatively low standard deviation (*SD* = 0.85) reflects a high consistency across the student teachers in ability to write test items with Structural Integrity.

4. Learning Outcome Alignment

With a mean score of 3.26 (*SD =0.89, Mdn =3.00, and Mode =3.00*) student teachers write items with Learning Outcome Alignment with 65.16 percent efficiency. A large majority of the prospective teachers (80.65%) met or exceeded the "Good" benchmark, with 9.68% reaching the "Excellent" level (25.81% Very Good, 45.16% Good). Only 19.35% were rated Satisfactory and none fell into Needs Improvement. Student teachers consistently possess strong capabilities in matching test items with prescribed curricular objectives (*SD =0.89*).

B. Moderate Competence on Cognitive Level / RBT Distribution

With a mean score of 2.84 (*SD =0.73, Mdn =3.00, and Mode =3.00*) Cognitive Level / RBT student teachers perform on writing test items with Cognitive Level / RBT Distribution moderately, with 56.77 percent efficacy. More than half of the prospective teachers (67.75%) met or exceeded the "Good" benchmark, with 3.23% reaching the "Excellent" level (9.68% Very Good, 54.84% Good). However, 32.26% were rated Satisfactory, but none fell into Needs Improvement; indicating that student teachers heavily rely on Lower-Order Thinking Skills and struggle to integrate Higher-Order questions. Standard deviation (0.73) reflects a consistent difficulty in balancing cognitive levels of their test items among student teachers.

C. Low-Performing Criteria

1. Quality of Non-Verbal Context

With a mean score of 1.77 (*SD =0.84, Mdn =2.00, and Mode =1.00*) Quality of Non-Verbal Context of the test items student teachers write is vulnerabile with they having only 35.48 percent efficacy on this task. Only a small minority of the prospective teachers (19.36%) met or exceeded the "Good" benchmark, with none reaching the "Excellent" level (3.23% Very Good, 16.13% Good). Majority of student teachers (35.48% Satisfactory and 45.16% Needs Improvement) consistently (*SD =0.84*) demonstrate a reliance on text-only formats and struggle to integrate visual contexts into their test items.

2. Contextualisation Depth

With a mean score of 1.61 (*SD =0.84, Mdn & Mode =1.00*) Contextualisation Depth of the items student teachers write is the lowest-rated ; with only 32.26 percent efficacy for them on this dimension. Only minority of the prospective teachers (16.13%) met or exceeded the "Good" benchmark, with none reaching the "Excellent" level (3.23% Very Good, 12.90% Good). Student teachers overwhelmingly write abstract, textbook-bound recall questions rather than real-world scenarios (25.81% Satisfactory; 58.06% Needs Improvement) indicating a uniformly poor capacity for authentic problem embedding for student teachers.

Figure 1 Strengths and weaknesses of achievement test constructed by Natural student teachers

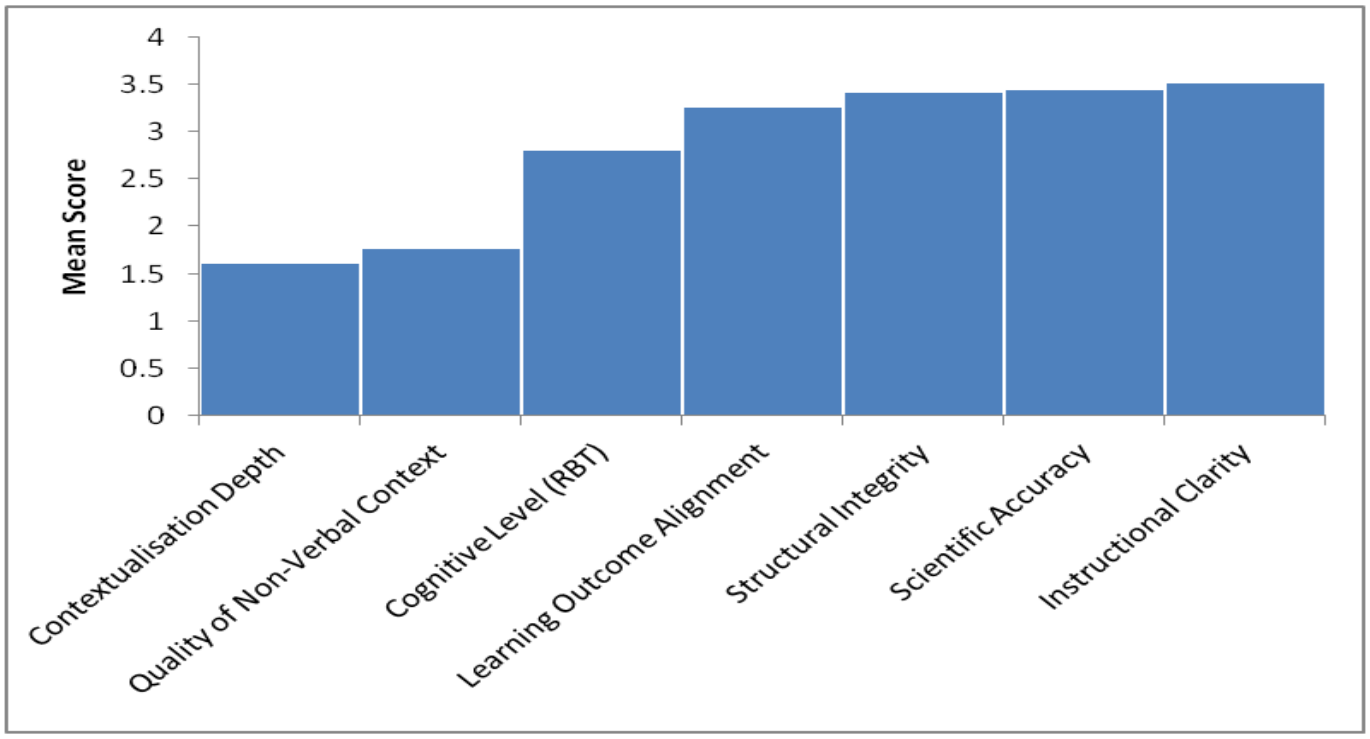

Figure 1 shows the major strengths and weaknesses of the achievement test constructed by prospective biology teachers. There is a struggle with advanced, constructivist design components.

The findings highlight a dual-profile pattern in pre-service biology teachers' test construction competencies. Whereas prospective teachers possess well-established technical baseline skills like writing clear instructions, maintaining factual scientific accuracy, structuring question layouts, and linking items to basic learning outcomes , they fail when required to apply constructivist, high-level assessment design. They struggle significantly to embed authentic real-world contexts, integrate non-verbal visual stimuli , and balance test papers with higher-order cognitive demands.

DISCUSSION AND IMPLICATIONS

While prospective teachers demonstrated a basic theoretical awareness of assessment alignment, a critical dimension of structural validity (Tractenberg et al., 2021), their test items were heavily skewed toward Lower-Order Thinking Skills (LOTS). Consistent with findings by Mpuangnan (2024) and Quansah et al. (2019), the assessments overwhelmingly prioritized the Remember and Understand levels of Bloom's Taxonomy. This suggests that while prospective teachers can successfully map items to learning objectives, they lack the technical competence in item writing needed to assess higher-order scientific inquiry beyond rote memorization.

While prospective teachers understand theoretical frameworks (e.g., Revised Bloom's Taxonomy), they lack the practical item-writing competence to translate this knowledge into constructivist assessments. The 'Contextualization Barrier' (the inability to design authentic, problem-based items) remains the primary obstacle to transitioning from traditional, rote-based evaluation to competency-based education. Current B.Ed. curricula overemphasize general evaluation theory at the expense of hands-on, practical skill development. Teacher Education Programmes (TEPs) must pivot toward:

a. Evidence-Based Intervention Modules: Using workshops to train student teachers to convert simple recall items into complex, scenario-based tasks.

b. Collaborative Peer Review: Integrating peer critique sessions to evaluate internship artifacts against standardized rubrics.

c. Practical Scaffolding: Implementing explicit training on visual literacy and real-world contextualization to meet the demands of modern education policies like the NEP 2020.

CONCLUSION

Prospective biology teachers possess "Satisfactory" baseline assessment literacy. However, to meet the goals of competency-based science education, teacher training must shift from passive rote instruction to intensive, practice-based training in deep contextualization, visual literacy, and higher-order item engineering. There is a need for the development of targeted professional development interventions such as a Constructivist Assessment Literacy Workshop Module to explicitly train prospective biology teachers in converting abstract, recall-based questions into context-rich, higher-order assessment tasks.

REFERENCE

1. Abell, S. K., & Siegel, M. A. (2011). Assessment literacy: What science teachers need to know and be able to do. In D. Corrigan, J. Dillon, & R. Gunstone (Eds.), The professional knowledge base of science teaching (pp. 205–221). Springer.
2. Alam, M. (2016). Constructivism: paradigm shift from teacher-centred to student-centred approach. The International Journal of Indian Psychology, 4(1), 79.
3. Allen, A. (2022). An introduction to constructivism: Its theoretical roots and impact on contemporary education. Journal of Learning Design and Leadership,1(1), 1-11.
4. Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. Educational Measurement: Issues and Practice, 30(1), 3–12.
5. Fulcher, G. (2012). Assessment literacy for the language classroom. Language Assessment Quarterly, 9(2), 113–132.https://doi.org/10.1080/15434303. 2011.642041.
6. Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. Applied Measurement in Education, 15(3), 309–334.
7. Hull, P., & Vígh, T. (2025). Teachers' assessment literacy: A descriptive literature review. Hungarian Educational Research Journal, 15(2), 210–225.https://doi.org/10.1556/063.2024.00317
8. Hung, C.-S., & Wu, H.-K. (2024). High school science teachers' assessment literacy for inquiry-based science instruction. International Journal of Science Education, 46(7), 621–642.https://doi.org/10.1080/09500693.2023.225165
9. Köksal, D. (2004). Assessing teachers' testing skills in ELT and enhancing their professional development through distance learning on the net. Turkish Online Journal of Distance Education-TOJDE, 5(1).
10. Meijer, K., Baartman, L., Vermeulen, M., & de Bruijn, E. (2023). Teachers' conceptions of assessment literacy. Teachers and Teaching, 29(7-8), 695–709.https://doi.org/10.1080.
11. Mpuangnan, K. N. (2024). Determining skills of basic schoolteachers in test item
12. construction. Review of Education, 12, e3478.
13. Pandey, A. (2019). Constructivism: a paradigm shift from teacher-centred to learner-centred education. JETIR, 6(3).
14. Pastore, S. (2023). Teacher assessment literacy: A systematic review. Frontiers in Education, 8, Article 1217167.https://doi.org/10.3389/feduc.2023.1217167.
15. Pastore, S. (2023). Teacher assessment literacy: A systematic review. Frontiers in Education, 8, Article 1217167.https://doi.org/10.3389/feduc.2023.1217167.
16. Pastore, S., & Andrade, H. L. (2019). Teacher assessment literacy: A three-dimensional model. Teaching and Teacher Education, 84, 128–138.https://doi.org/10.1016/j.tate.2019.05.003
17. Popham, W. J. (2020). Classroom assessment: What teachers need to know (9th ed.). Pearson.
18. Priyamvada. (2018). Exploring the constructivist approach in education: Theory, Practice, and Implications. International Journal of Research and Analytical Reviews (IJRAR), 5(2), 716-717.
19. Quansah, F., Amoako, I., & Ankomah, F. (2019). Teachers' test construction skills in senior high schools in Ghana: Document analysis. International Journal of Assessment Tools in Education, 6(1), 1–8.

20. Sharma, N. (2025). The impact of constructivist approach on achievement in science education: A comprehensive review. International Journal for Multidisciplinary Research (IJFMR).
21. Sheeja, T. (2023). Assessment and evaluation of learning in a constructivist paradigm. International Journal for Multidisciplinary Research IJFR, 5(2).
22. Singh, S., &Yaduvanshi, S. (2015). Constructivism in science classroom: Why and how. International Journal of Scientific and Research Publications, 5(3), 1–5.
23. Stiggins, R. J. (1991). Assessment literacy. Phi Delta Kappan, 72(7), 534–539.
24. Stiggins, R. J. (1995). Assessment literacy for the 21st century. Phi Delta Kappan, 77(3), 238–245.
25. Tractenberg, R. E. (2021). The assessment evaluation rubric: Promoting learning and learner-centred teaching through assessment in face-to-face or distanced higher education. Education Sciences, 11(8), 441. https://doi.org/10.3390/educsci11080441
26. Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. Teaching and Teacher Education, 58, 149–162.https://doi.org/10.1016/j.tate.2016.05.010