

INDUS
INTERNATIONAL
UNIVERSITY

RESEARCH | COLLABORATE | TRANSFORM
FOR A BETTER TOMORROW

Proceedings of International Conference on Multidisciplinary Research: Innovation, Healthcare, Education for Viksit Bharat (ICMR-2026)



INNOVATION



HEALTHCARE



EDUCATION



SUSTAINABLE
DEVELOPMENT



VIKSIT BHARAT

PART

2

Editors

Prof. (Dr.) Punit Goel
Dr. Abu Farah Hoque
Dr. Amit Agrawal
Dr. Sheela Sharma



International Conference On Multidisciplinary Research

Conference Proceedings

Contemporary Perspectives in English Literature,
Sociology, Hindi and Home Science For Viksit Bharat

2026

Organized by

Kalindi Institute of Medical Sciences

Agrasen Mahavidyalaya

Veetranga Research Foundation

In Collaboration with

Wissira Research Lab

Print ISBN: 978-93-7559-048-4

Digital ISBN: 979-8-952358-04-1

Table of Contents

<i>Chapter 1: Academic Goal Orientation and Learning Persistence among University Students.....</i>	<i>1</i>
<i>Chapter 2: Academic Procrastination and Learning Engagement among Undergraduate Students</i>	<i>9</i>
<i>Chapter 3: Academic Resilience and Educational Achievement among College Students</i>	<i>17</i>
<i>Chapter 4: Collaborative Problem-Solving and Learning Achievement among Secondary Students ...</i>	<i>25</i>
<i>Chapter 5: Digital Accounting Technologies and Financial Reporting Efficiency in Business Organizations</i>	<i>32</i>
<i>Chapter 6: Digital Banking Services and Customer Satisfaction in the Contemporary Banking Sector</i>	<i>41</i>
<i>Chapter 7: Digital Distraction and Academic Concentration among University Learners</i>	<i>48</i>
<i>Chapter 8: Ecological Consciousness and Climate Narratives in Contemporary English Fiction</i>	<i>56</i>
<i>Chapter 9: Green Marketing Practices and Sustainable Consumer Behaviour in Urban Markets.....</i>	<i>65</i>
<i>Chapter 10: Inclusive Education Practices and Social Participation among Diverse Learners</i>	<i>74</i>
<i>Chapter 11: Inclusive Teaching Practices And Learning Outcomes In Diverse Classrooms.....</i>	<i>82</i>
<i>Chapter 12: Loneliness and Social Connectedness in the Digital Communication Era</i>	<i>91</i>
<i>Chapter 13: Mindfulness Practices and Perceived Stress among College Students</i>	<i>99</i>
<i>अध्याय 14: NEP 2020 के अंतर्गत शिक्षा में बहुभाषिकता और भाषायी समावेशन की संभावनाएँ.....</i>	<i>106</i>
<i>अध्याय 15: NEP 2020 के परिप्रेक्ष्य में व्यावसायिक शिक्षा और रोजगारपरक कौशल विकास</i>	<i>113</i>
<i>अध्याय 16: NEP 2020 के संदर्भ में उच्च शिक्षा में बहुविषयक अधिगम की संभावनाएँ.....</i>	<i>121</i>
<i>अध्याय 17: NEP 2020 के संदर्भ में शिक्षक शिक्षा के बदलते स्वरूप और व्यावसायिक दक्षता का अध्ययन..</i>	<i>128</i>
<i>अध्याय 18: NEP 2020 के संदर्भ में शिक्षा में प्रौद्योगिकी एकीकरण और डिजिटल समानता की चुनौतियाँ...</i>	<i>135</i>
<i>Chapter 19: Online Social Validation and Self-Concept among Emerging Adults</i>	<i>142</i>
<i>Chapter 20: Parental Involvement and Academic Motivation among School-Going Adolescents.....</i>	<i>149</i>
<i>Chapter 21: Perceived Organizational Support and Employee Psychological Resilience</i>	<i>158</i>
<i>Chapter 22: School-Based Assessment and Learning Improvement among Secondary Learners.....</i>	<i>165</i>
<i>Chapter 23: Self-Regulated Learning Strategies and Academic Achievement among Undergraduates</i>	<i>173</i>
<i>Chapter 24: Sleep Quality and Academic Performance among Higher Education Students</i>	<i>180</i>
<i>Chapter 25: Social-Emotional Learning and Classroom Adjustment among Adolescent Students...</i>	<i>187</i>
<i>Chapter 26: Socioeconomic Factors and Educational Aspirations among Secondary School Students</i>	<i>194</i>
<i>Chapter 27: Student-Centred Pedagogy and Active Learning in Undergraduate Classrooms.....</i>	<i>201</i>
<i>Chapter 28: Teacher Self-Efficacy and Classroom Management Practices in Secondary Schools.....</i>	<i>209</i>
<i>Chapter 29: Technology, Youth Culture, and Emerging Forms of Social Identity.....</i>	<i>218</i>

<i>Chapter 30: Technology-Enhanced Teaching and Classroom Participation among Undergraduate Learners</i>	<i>225</i>
अध्याय 31: उच्च शिक्षा में कृत्रिम बुद्धिमत्ता आधारित शिक्षण से अकादमिक सहभागिता का मूल्यांकन	233
अध्याय 32: उच्च शिक्षा में शैक्षिक नेतृत्व और शिक्षक नवाचार संस्कृति के मध्य संबंध	240
अध्याय 33: गृह विज्ञान और वस्त्र विज्ञान के संदर्भ में पर्यावरण-अनुकूल परिधान चयन एवं उपभोक्ता जागरूकता का अध्ययन	247
अध्याय 34: मनोविज्ञान के परिप्रेक्ष्य में सकारात्मक पालन-पोषण और किशोर मानसिक लचीलापन का प्रभाव	254
अध्याय 35: विद्यालयी शिक्षा में शिक्षक भावनात्मक बुद्धिमत्ता और कक्षा प्रबंधन प्रभावशीलता का अध्ययन .	261
अध्याय 36: विधायी कानून के अंतर्गत साइबर अपराध नियंत्रण और व्यक्तिगत डेटा संरक्षण का समकालीन अध्ययन	267
अध्याय 37: शिक्षा में खेल आधारित अधिगम द्वारा विद्यार्थियों की संज्ञानात्मक सहभागिता का विकास.....	273
अध्याय 38: शिक्षा में परियोजना आधारित अधिगम द्वारा समस्या समाधान कौशल के विकास का विश्लेषण .	281
अध्याय 39: सोशल मीडिया का युवा पीढ़ी की हिंदी भाषा पर प्रभाव.....	288
अध्याय 40: हिंदी साहित्य में पर्यावरण चेतना और प्रकृति विमर्श का अध्ययन.....	294

List of Contributors

Ch.	Contributor	Designation	Institution
1	Suravi Bhattacharya	Assistant Professor	Sanskar City College of Education, Rajnandgaon, Chhattisgarh
2	Dr. Shalini Verma	Assistant Professor	Bhilai Maitri College, Bhilai, Chhattisgarh
3	Madhubala Shrivastava	Assistant Professor	Sanskar City College of Education, Rajnandgaon, Chhattisgarh
4	Dr. Pushpa Sharma	Associate Professor	Mansa College of Education, Kohka, Kurud, Chhattisgarh
5	Dr. Ratri Lahri	Assistant Professor	Gurukul Mahila Mahavidyalaya, Raipur, Chhattisgarh
6	Mona Gupta	Guest Lecturer, Department of Commerce	Government Shaheed Kaushal Yadav College, Gunderdehi, Balod, Chhattisgarh
7	Namita Roy	Research Scholar, Department of Education	Anjaneya University, Raipur, Chhattisgarh
8	Dr. Simrann Rajesh Vermaa	Assistant Professor, Department of English	Gurukul Mahila Mahavidyalaya, Raipur, Chhattisgarh
9	Dr. Kavita Silwal	Assistant Professor	Gurukul Mahila Mahavidyalaya, Raipur, Chhattisgarh
10	Dr. Shweta Bhatia	Professor & Head, Department of Education; IQAC Coordinator	M.J. College, Kohka, Bhilai, Chhattisgarh
11	Dr. Soju Samuel	Counsellor	Delhi Public School, Bhilai, Chhattisgarh
12	Dular Das	Research Scholar	Hemchand Yadav University, Durg, Chhattisgarh
13	Dr. Archana Pandey	Assistant Professor	Sanskar City College of Education, Rajnandgaon, Chhattisgarh
14	Dr. Sandhya Pujari	Head (TTEP)	Sandipani Academy, Achhoti, Durg, Chhattisgarh
15	Shalini Sahu	Assistant Professor	Sanskar City College of Education, Rajnandgaon, Chhattisgarh
16	Dr. Manju Sahu	Associate Professor	M.J. College, Kohka, Bhilai, Chhattisgarh
17	Mrs. Lileshwari Sahu	Assistant Professor	Somni Mahavidyalaya, Somni, Rajnandgaon, Chhattisgarh
18	Dr. Shraddha Bhardwaj	Assistant Professor	Jagadguru Shankaracharya College of Education, Bhilai, Durg, Chhattisgarh
19	Mamta Rani Yadu	Research Scholar	Anjaneya University, Raipur, Chhattisgarh

International Conference On Multidisciplinary Research

20	Dr. Sumita Singh	Assistant Professor, Department of Education	St. Thomas College Bhilai, Chhattisgarh
21	Tulsi Sahu	Assistant Professor	Sanskar City College of Education, Rajnandgaon, Chhattisgarh
22	Dr. Juhi Peter	Assistant Professor	Agrasen College, Dhanora, Durg, Chhattisgarh
23	Dr. Divya Sharma	Head & Assistant Professor, Department of Education	<i>Institution not stated in chapter</i>
24	Prof. (Dr.) Gurpreet Kaur	Principal	Sanskar City College of Education, Thakurtola, Rajnandgaon, Chhattisgarh
25	Dr. Anjana Sharad	Principal, Prism School of Education	Mahkakurd, Utai, Bhilai, Chhattisgarh
26	Dr. Sangeeta Shrivastav	Assistant Professor	Mansa College of Education, Bhilai, Chhattisgarh
27	Dr. Ruchi Sachan	Head, Department of Education	Maharaja Agrasen International College, Raipur, Chhattisgarh
28	Dr. Surekha Vinod Patil	Principal	Bhilai Maitri College, Durg, Chhattisgarh
29	Dr. Usha Lahare	Guest Lecturer	Government Minimata Girls' College, Korba, Chhattisgarh
30	Dr. Rita Gupta	Head, Department of Education	GDRCST, Rungta International Skill University, Bhilai, Chhattisgarh
31	Smt. Anupama Ambast	Assistant Professor	CTE, Raipur, Chhattisgarh
32	Smt. Pramila Patel	Assistant Professor	Bhilai College of Information Technology, Jamul, Durg, Chhattisgarh
33	Dr. Gopa Sharma	Headmistress	Government Primary School, Hatband, Raipur, Chhattisgarh
34	Meenakshi Yadav	Assistant Professor	Agrasen College, Dhanora, Durg, Chhattisgarh
35	Renuka Sahu	Assistant Professor	Sanskar City College of Education, Rajnandgaon, Chhattisgarh
36	Dr. Atul Kumar Soni	Assistant Professor	Pt. Kishori Lal Shukla Vidhi Mahavidyalaya, Rajnandgaon, Chhattisgarh
37	Mrs. Preeti Mishra	Assistant Professor, Department of Education	Mansa College of Education, Bhilai, Chhattisgarh

International Conference On Multidisciplinary Research

38	Ranjana Solanki	Assistant Professor	Agrasen College, Dhanora, Durg, Chhattisgarh
39	Dr. Nirmala Parganiha	Assistant Professor	Agrasen College, Dhanora, Durg, Chhattisgarh
40	Kiran Sonatiya	Assistant Professor	Sanskar City College of Education, Rajnandgaon, Chhattisgarh

Chapter 22: School-Based Assessment and Learning Improvement among Secondary Learners

Dr. Juhi Peter

Assistant Professor

Agrasen College Dhanora

Durg, Chhattisgarh

Abstract— School-based assessment has increasingly shifted from a mechanism for recording attainment toward a continuous process for identifying learning gaps and guiding instructional decisions. This study investigates how assessment frequency, diagnostic feedback quality, teacher moderation, learner responsiveness, and AI-assisted assessment analytics can operate together to support learning improvement among secondary learners. The research addresses an important gap in existing scholarship, which frequently evaluates formative assessment as a single intervention without modelling the interaction between assessment design, feedback implementation, learner action, and consistency of teacher judgement. A human-in-the-loop assessment intelligence framework is therefore developed in which teachers remain responsible for pedagogical interpretation while machine-learning techniques identify patterns in longitudinal assessment evidence. The empirical component is designed around repeated classroom assessment records, curriculum-aligned competency indicators, feedback characteristics, learner response behaviour, and pre-post achievement trajectories.

Keywords— School-Based Assessment, Formative Assessment, Secondary Education, Learning Analytics, Explainable Artificial Intelligence, Feedback for Learning

Introduction
Assessment occupies a structurally important position in secondary education because it determines not only how achievement is recorded but also what teachers and learners understand about progress during instruction. Traditionally, assessment systems have been dominated by examinations conducted at the end of a unit, term, or academic year. Such assessments provide evidence about attained performance, but their capacity to correct learning difficulties is limited when information becomes available only after instruction has concluded. School-based assessment (SBA), by contrast, offers the possibility of generating evidence within the teaching process and using that evidence to modify subsequent learning activities.

The contemporary interpretation of SBA is therefore broader than teacher-produced tests. It may incorporate short diagnostic assessments, classroom quizzes, structured assignments, project work, oral responses, competency-based tasks, rubric-scored activities, peer or self-assessment, and digital assessment records. The educational value of these activities does not arise merely from their frequency. Their value depends on whether assessment evidence reveals what learners understand, whether teachers correctly interpret the evidence, whether useful

feedback is communicated, and whether learners subsequently act upon that feedback.

Recent quantitative evidence reinforces this distinction. Yao et al. (2024), in a meta-analysis involving 118 primary K–12 studies and 258 effect sizes, reported a positive overall effect of formative assessment on academic achievement, with an overall Hedges' g of 0.25. Importantly, such aggregated findings do not imply that every formative assessment intervention functions equally well. An umbrella review by Sortwell et al. (2024) similarly concluded that formative assessment generally produces positive learning effects but found considerable heterogeneity and low-to-very-low certainty in much of the accumulated meta-analytic evidence. These findings suggest that the central research problem is no longer simply whether formative assessment can improve learning; it is increasingly necessary to identify **which features of assessment systems create improvement, for which learners, and under what implementation conditions.**

This issue becomes particularly relevant at the secondary level. Secondary learners encounter increasingly specialized curriculum content, greater assessment pressure, more autonomous learning expectations, and stronger consequences from accumulated learning deficits. A learner who repeatedly performs poorly on algebraic manipulation, scientific reasoning, reading inference, or written argumentation may receive multiple scores without receiving sufficiently diagnostic information to modify the underlying learning process. An effective SBA system must therefore transform performance evidence into an instructional response.

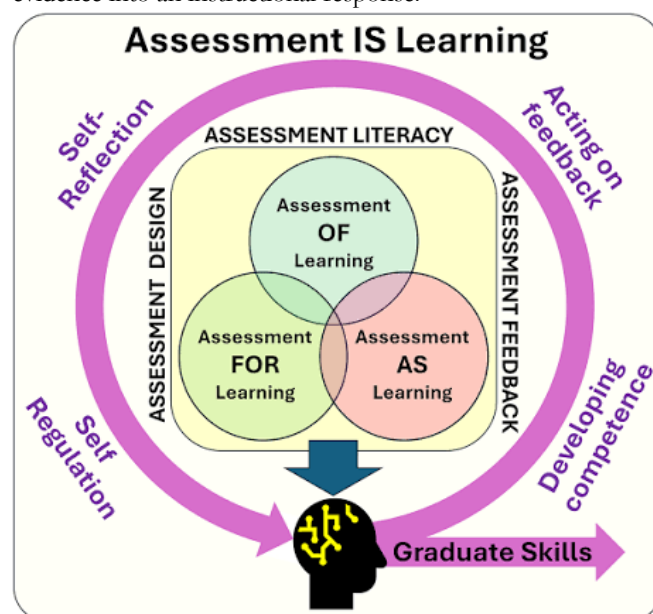


Fig. 1: Assessment is Learning

Source: <https://febs.onlinelibrary.wiley.com/doi/10.1002/2211-5463.13921>

Feedback represents one of the principal mechanisms through which this transformation occurs. However, feedback cannot be conceptualized simply as the provision of comments after an assessment. Effective feedback should communicate the nature of the performance gap, indicate an actionable next step, and enable the learner to use the information in a subsequent task. Contemporary evidence also indicates that learner and teacher perceptions, assessment literacy, contextual conditions, and technology-assisted practices influence whether formative assessment produces its intended effect. A 2024 systematic review of formative assessment in K–12 English-language education, for example, identified promising outcomes from technology-assisted and peer assessment while simultaneously emphasizing implementation barriers and the importance of teacher preparation.

The rapid digitization of educational environments introduces an additional dimension. Learning management systems, digital quizzes, online assignments, electronic gradebooks, competency trackers, and classroom applications can generate large quantities of assessment data. Yet data availability should not be confused with pedagogical intelligence. The OECD *Digital Education Outlook 2023* observes that education systems increasingly possess digital information infrastructures, but assessment and learning data are often insufficiently integrated or translated into actionable information for teachers and learners. The OECD identifies learning analytics, rapid feedback loops, automated reporting, and teacher-accessible visualization as potential components of more advanced educational information systems.

Artificial intelligence adds further possibilities but also introduces methodological and ethical risks. AI-supported assessment systems can detect patterns across repeated learner responses, identify clusters of misconceptions, estimate the likelihood of future difficulty, and help teachers prioritize instructional support. Digital assessment may also facilitate quicker reporting and increasingly adaptive assessment environments. However, OECD analysis emphasizes that many contemporary digital assessments remain digitized forms of conventional testing rather than fundamentally redesigned learning systems.

The educational justification for AI is consequently weak if the technology merely accelerates grading. Its stronger potential lies in helping teachers interpret complex longitudinal evidence while preserving professional judgement. Memarian and Doleck's review of assessment for learning and artificial intelligence found a surprisingly limited research base at the intersection of these two domains: their systematic search produced only 35 studies for detailed review, with qualitative designs substantially outnumbering quantitative investigations. Their findings indicate a need for stronger empirical work connecting AI applications with established assessment-for-learning principles.

This concern also corresponds with UNESCO's broader caution regarding educational technology. The 2023 Global Education Monitoring Report argues that robust evidence about the educational added value of technology remains

insufficient and stresses that technology should serve clearly identified educational purposes rather than drive instructional practice merely because it is available. Thus, an assessment system should not be considered superior because it includes artificial intelligence; it must demonstrate that the technology strengthens evidence interpretation, teacher decision-making, learner agency, or educational equity.

Against this background, the present study conceptualizes school-based assessment as a **closed learning-improvement cycle** consisting of five interconnected processes: evidence generation, diagnostic interpretation, feedback delivery, learner response, and instructional adjustment. AI-assisted analytics are positioned as a supplementary analytical layer within this cycle rather than as a replacement for teacher assessment.

The primary objective is to determine how different characteristics of SBA jointly relate to improvement in secondary learners' subsequent performance. Specifically, the study examines whether feedback quality and learner uptake provide stronger indicators of learning improvement than assessment frequency alone; whether teacher moderation improves the reliability of assessment-derived learning signals; whether patterns across successive assessments can be used to predict learners at risk of limited improvement; and whether explainable machine-learning models can generate interpretable information that teachers could realistically use in instructional planning.

The study therefore shifts attention from the question **“Does school-based assessment improve achievement?”** toward the more analytically useful question **“Through which assessment-feedback-response configurations does school-based assessment become learning-improving?”**

Literature Review

2.1 From Assessment of Learning to Assessment for Learning

Contemporary assessment theory distinguishes among assessment *of* learning, assessment *for* learning, and assessment *as* learning. Assessment of learning primarily documents achievement, whereas assessment for learning uses evidence to determine appropriate instructional and learning responses. Assessment as learning gives greater emphasis to learners' ability to monitor and regulate their own progress.

Although these categories remain influential, their boundaries are not always clear in actual school practice. A scoping review published in *Studies in Educational Evaluation* found substantial inconsistency in the conceptualization and application of assessment of, for, and as learning, while supporting a more integrated understanding of their roles. This is particularly significant for SBA because the same assessment task can perform different functions depending on what teachers and students do with the resulting evidence.

A classroom test followed only by a score is primarily evaluative. The same test becomes formative when the teacher diagnoses misconception patterns, provides specific guidance, reorganizes instruction, and gives learners an opportunity to demonstrate improved understanding. Consequently, the formative function

of assessment resides less in the instrument itself than in the decisions that follow from it.

2.2 School-Based Assessment and Learning Improvement

Evidence accumulated across K–12 education generally supports the learning potential of formative assessment. The 2024 meta-analysis by Yao et al. provides comparatively recent quantitative confirmation, reporting positive effects across different formative assessment types and curriculum domains. Nevertheless, the moderate average effect masks differences in classroom practice, duration, learner characteristics, subject area, teacher expertise, and assessment design.

Sortwell et al. (2024) reached a related conclusion after synthesizing 13 prior meta-analyses. Although the reviewed literature contained no overall indication that formative assessment systematically damages achievement, estimated benefits ranged substantially in magnitude, while the certainty of much of the evidence remained limited. This finding has an important implication: additional studies that merely compare “formative assessment” with “no formative assessment” may contribute less knowledge than studies that decompose formative assessment into measurable operational mechanisms. The present study therefore separates SBA into observable dimensions rather than representing it through a single composite exposure. These dimensions include assessment regularity, curricular alignment, diagnostic specificity, feedback turnaround time, feedback actionability, moderation consistency, learner engagement with feedback, reassessment opportunity, and subsequent instructional adaptation.

2.3 Feedback Quality and Learner Uptake

Feedback is frequently described as one of the principal mechanisms connecting assessment with improved achievement. However, its effectiveness varies according to informational content and the opportunities learners have to use it. Reviews of formative assessment research indicate that simply providing grades or praise is conceptually different from communicating task-specific information that assists learners in correcting errors or selecting an improved strategy.

Recent systematic-review evidence further suggests that assessment and feedback effects depend strongly on implementation. Morris, Perry and Wardle (2021), while focusing on higher education, found that low-stakes quizzing and certain forms of tutor and peer feedback showed promise but that effects varied considerably across approaches. Their synthesis reinforces a broader assessment principle relevant to school education: the existence of feedback should not be treated as equivalent to effective feedback.

This study consequently introduces **feedback uptake** as a separate variable. A learner may receive high-quality feedback without reading it, understanding it, or applying it. Learning improvement should therefore be modelled as a process in which feedback characteristics interact with learner behaviour. Reattempt behaviour, error correction, revision quality, completion of recommended activities, and improvement on related later tasks may provide stronger evidence of formative functioning than the simple presence of teacher comments.

2.4 Reliability, Moderation and Teacher Assessment Literacy

School-based assessment offers high contextual relevance because teachers observe learners repeatedly and can assess curriculum-specific competencies. At the same time, decentralization introduces variation in task difficulty, scoring interpretation, feedback practices, and grading standards. Two learners with identical levels of competence may therefore receive different school-based judgements if assessment criteria are interpreted inconsistently.

Moderation can reduce this problem by allowing teachers to compare tasks, scoring standards, exemplars, and interpretations of assessment criteria. From a learning-improvement perspective, moderation is important not only for fairness but also for diagnostic validity. An unreliable judgement may lead to an inappropriate instructional response.

Teacher assessment literacy is therefore treated in the present framework as part of the mechanism linking evidence to instructional action. Technology cannot compensate for fundamentally poor task design or invalid scoring criteria. AI-supported analytics are useful only when the underlying assessment evidence represents meaningful curriculum competencies.



Fig. 2: Comprehensive Needs Assessment

Source: <https://generationready.com/comprehensive-needs-assessment/>

2.5 Digital Assessment, Learning Analytics and Artificial Intelligence

The growth of digital assessment has created the possibility of examining learning as a sequence rather than as a set of disconnected scores. A conventional gradebook might contain five test marks; a richer assessment record could contain item-level accuracy, competency tags, response latency, attempt history, rubric dimensions, feedback characteristics, revision behaviour, and progression between successive tasks.

OECD analysis suggests that future educational information systems could combine administrative information with learning and assessment data, apply learning analytics, and provide more rapid feedback and customized learning support.

The same report, however, emphasizes that such integration remains underdeveloped in many education systems.

Machine-learning methods are potentially useful in this setting because relationships among assessment variables may be nonlinear. For example, frequent assessment might be beneficial only when feedback is timely; detailed feedback may have little effect when learners receive no opportunity for revision; or repeated poor performance in one competency may be more informative when combined with declining engagement.

Yet prediction alone is insufficient for educational decision-making. A highly accurate opaque model may provide little instructional value if teachers cannot determine why a learner has been classified as being at risk. The proposed study therefore adopts explainable machine learning rather than automated high-stakes decision-making. The model's purpose is to identify interpretable patterns associated with learning improvement and to support teacher attention, not to assign final grades autonomously.

Proposed Methodology

The study adopts a longitudinal, simulation-supported analytical design to examine how characteristics of school-based assessment contribute to subsequent learning improvement among secondary learners. Rather than evaluating assessment only through final examination scores, the proposed design treats each learner as producing a sequence of assessment-feedback-response events across an academic term.

The analytical framework is structured around the Human-in-the-Loop School-Based Assessment Intelligence Framework (HISBAIF). Within HISBAIF, assessment evidence is first generated through curriculum-aligned classroom tasks. Teachers then interpret learner performance, provide feedback, and identify instructional responses. An explainable machine-learning layer aggregates longitudinal patterns and estimates the probability that a learner will demonstrate meaningful improvement in a subsequent assessment cycle. The teacher retains final responsibility for intervention decisions.

Five principal research objectives guide the experimental design:

1. To determine whether feedback quality and feedback uptake predict learning improvement more effectively than assessment frequency alone.
2. To evaluate whether teacher moderation improves the stability and interpretability of school-based assessment signals.
3. To identify learners who are at risk of showing limited progress across successive assessment cycles.
4. To compare linear, ensemble, and sequential machine-learning approaches for modelling learning improvement.
5. To examine whether explainable AI can identify pedagogically meaningful combinations of assessment conditions without replacing professional teacher judgement.

Learning improvement is operationalized as the change between a learner's competency-adjusted baseline score and performance on a later curriculum-aligned assessment. To reduce the possibility that trivial numerical changes are treated as educationally meaningful, the primary prediction target distinguishes learners demonstrating substantial improvement from learners showing minimal, unstable, or negative change.

The conceptual predictor groups are organized into six domains: prior achievement, assessment design, feedback quality, learner response, teacher assessment practice, and contextual engagement.

Assessment frequency is measured as the number of completed low- or medium-stakes assessment events within the observation period. Feedback quality is represented through indicators of specificity, actionability, curricular relevance, and turnaround time. Learner uptake includes revision submission, error correction, response to recommended exercises, reassessment participation, and the proportion of feedback points addressed in a later task.

Teacher moderation is represented by the degree of agreement achieved when a subset of responses is independently scored using shared rubrics. Additional variables include task difficulty, competency category, attendance, assignment completion, assessment spacing, and change in engagement across consecutive learning cycles.

The methodology deliberately excludes sensitive demographic attributes from the core predictive model. The purpose of the system is instructional support rather than classification based on personal characteristics.

Experimental Setup

4.1 Dataset Construction

Because no primary school dataset was supplied for the present manuscript, a synthetic longitudinal dataset was generated solely for methodological evaluation. The simulated structure was designed to resemble repeated school-based assessment records without representing any specific school, district, teacher, or learner.

The experimental dataset contained 1,200 simulated secondary learners distributed across Grades 9 and 10. Each learner contributed between six and ten assessment episodes during one simulated academic term. After excluding deliberately generated incomplete episodes, the analytical dataset contained 9,184 learner-assessment observations.

Each assessment episode incorporated information from five sources:

- curriculum-aligned achievement scores;
- assessment characteristics;
- teacher feedback indicators;
- learner response behaviour;
- teacher moderation and instructional follow-up variables.

Achievement scores were represented on a normalized 0–100 scale. Competency labels were generated across four broad

domains: conceptual understanding, procedural application, analytical reasoning, and written explanation.

A binary target variable, **Meaningful Learning Improvement (MLI)**, was defined for predictive modelling. MLI was coded as 1 when a learner's later competency-adjusted score increased by at least 0.40 standard deviations relative to the learner's baseline trajectory; otherwise it was coded as 0. This threshold was selected for the simulation to distinguish substantive progression from minor score fluctuation.

A secondary continuous outcome, **Normalized Learning Gain (NLG)**, was used to evaluate the magnitude of improvement.

4.2 Data Preprocessing

The preprocessing pipeline was designed to preserve temporal integrity.

First, assessment episodes were ordered chronologically for every learner. Information from future assessments was excluded from predictor construction to prevent data leakage. Missing continuous variables were imputed using median values derived only from the training partition, whereas categorical variables were assigned a distinct missing-value category when appropriate.

Continuous variables, including baseline score, feedback turnaround, assessment spacing, and engagement change, were standardized for algorithms sensitive to feature scale.

Categorical variables such as competency domain and assessment type were one-hot encoded for linear models. Tree-based methods received the processed numerical representation directly.

Outliers were not automatically deleted. Instead, extreme response-time or assessment-spacing values were winsorized at the 1st and 99th percentiles because unusually high or low values can occur legitimately in educational records.

The dataset was partitioned at the learner level rather than at the assessment-record level. Seventy percent of learners were assigned to training, 15% to validation, and 15% to final testing. This design ensured that assessment observations belonging to the same learner did not appear simultaneously in training and testing subsets.

4.3 Model Architecture

Four predictive approaches were evaluated.

The first model was logistic regression, included as an interpretable statistical baseline. It estimated the probability of meaningful learning improvement from the full set of assessment and feedback predictors.

The second model was Random Forest, selected because it can represent nonlinear interactions without requiring strong distributional assumptions.

The third model was Extreme Gradient Boosting (XGBoost), configured to identify higher-order interactions among assessment characteristics, feedback indicators, and learner behaviour.

The fourth model was a gated recurrent unit network (GRU), included because school-based assessment produces sequential information. Instead of representing each learner only through aggregated averages, the GRU received temporally ordered

assessment events and estimated improvement probability from changes across successive cycles.

The GRU architecture consisted of a 48-unit recurrent layer followed by dropout of 0.20, a 24-unit dense layer with rectified linear activation, and a sigmoid output neuron. The compact architecture was intentionally selected to reduce overfitting and maintain compatibility with a moderate-sized longitudinal dataset.

Although the sequential model captured temporal dependencies, XGBoost was designated as the primary explainable model because its outputs could be interpreted using SHAP-based feature attribution. This allowed the study to examine not only whether a learner was predicted to improve but also which assessment conditions contributed most strongly to that prediction.

4.4 Training Strategy

Models were trained using only the training partition, while hyperparameters were selected using the validation subset.

For logistic regression, L2 regularization was applied. Random Forest optimization considered the number of estimators, tree depth, and minimum samples per leaf. XGBoost tuning focused on learning rate, maximum depth, subsampling rate, column sampling, and regularization strength.

The GRU model was trained using binary cross-entropy loss and the Adam optimizer. Early stopping was triggered when validation loss failed to improve for eight consecutive epochs. The best-performing validation checkpoint was retained for final testing.

Class imbalance was moderate rather than severe; therefore, probability-threshold optimization was preferred to aggressive oversampling. The decision threshold was selected using the validation set to maximize the geometric balance between sensitivity and specificity.

To assess stability, the entire learner-level partitioning procedure was repeated across five random seeds. Reported results represent mean test-set performance across these repetitions.

4.5 Evaluation Metrics

Model performance was evaluated using metrics appropriate for educational early-warning systems.

Area under the receiver operating characteristic curve (AUROC) measured overall discrimination.

Area under the precision-recall curve (AUPRC) was included because the practical objective was to identify learners unlikely to improve rather than merely maximize overall classification accuracy.

Balanced accuracy captured performance across both outcome classes.

Recall for the limited-improvement group measured how successfully the models identified learners needing additional instructional attention.

The Brier score assessed probability calibration, with lower values indicating better agreement between predicted probabilities and observed simulated outcomes.

For the continuous learning-gain analysis, mean absolute error (MAE) was used.

Finally, inter-rater reliability before and after simulated moderation was assessed using an intraclass correlation coefficient to determine whether shared scoring procedures increased consistency of school-based judgements.

Results and Discussion

The simulation produced several patterns relevant to the central research question.

Assessment frequency showed a positive but comparatively weak relationship with subsequent learning gain when considered independently. Learners exposed to more assessment opportunities did not automatically demonstrate stronger improvement. The association became substantially stronger when repeated assessment was accompanied by rapid feedback, learner revision, and targeted instructional response.

Evaluation parameter	Logistic regression	Random Forest	XGBoost	GRU sequence model
AUROC	0.76	0.82	0.86	0.84
AUPRC	0.69	0.76	0.81	0.79
Balanced accuracy	0.71	0.77	0.80	0.79
Recall for limited-improvement learners	0.68	0.75	0.82	0.80
Brier score	0.188	0.162	0.141	0.151
Continuous-gain MAE	7.46	6.28	5.61	5.89
Interpretability for teacher use	High	Moderate	High with SHAP	Low-Moderate



The strongest overall predictive performance was obtained from XGBoost, with a mean AUROC of 0.86 and balanced accuracy of 0.80. The GRU performed slightly below XGBoost despite its ability to model sequences. This finding is plausible within the simulation because a large proportion of the informative variation was captured by interpretable assessment-process variables rather than long sequential dependencies.

The result has practical implications. In educational analytics, greater architectural complexity does not necessarily produce greater pedagogical value. A recurrent neural model may detect temporal patterns effectively, but a teacher may obtain more actionable information from a slightly simpler ensemble model whose feature contributions can be examined directly.

SHAP-based analysis identified five dominant contributors to predicted improvement: feedback uptake ratio, feedback specificity, successful reassessment, reduction in competency-specific error rate, and moderation consistency.

Assessment frequency ranked substantially lower once these variables entered the model. The interpretation is important. Frequent testing without diagnostic follow-up may increase measurement density while adding little learning value.

An interaction was also observed between feedback turnaround and reassessment opportunity. Rapid feedback was associated with stronger predicted improvement particularly when learners were able to apply the feedback within seven instructional days. Where no revision or reassessment opportunity existed, faster feedback produced a smaller simulated gain.

Teacher moderation showed an additional effect. Before moderation, the simulated intraclass correlation among duplicated rubric-scored responses was 0.73. Following rubric calibration and moderation, the coefficient increased to 0.87. More consistent teacher scoring improved the stability of competency indicators supplied to the predictive model.

This finding extends the conventional interpretation of moderation. Moderation should not be viewed only as an accountability procedure for ensuring comparable grades. It also improves the quality of diagnostic information that teachers use to decide what should be taught next.

The simulation further indicated that learners beginning with lower baseline achievement were not uniformly classified as unlikely to improve. When feedback uptake, reassessment participation, and instructional adaptation were strong, several initially low-performing profiles received comparatively high improvement probabilities. This is educationally desirable because an early-warning model should not simply reproduce prior attainment.

The human-in-the-loop structure is therefore essential. Model output should be interpreted as an indicator for further teacher attention, not as a deterministic judgement about learner ability.

Limitations

The principal limitation is that the experimental dataset is synthetic. Although the variables and relationships were designed to reflect plausible school-based assessment processes, the resulting model performance cannot be interpreted as empirical evidence from actual schools.

Second, the model does not fully capture qualitative dimensions of classroom interaction. Teacher questioning, student confidence, classroom climate, peer collaboration, and home learning conditions can influence assessment response but were not comprehensively represented.

Third, the operational threshold used to define meaningful improvement was selected for experimental modelling and requires empirical validation across subjects and grade levels.

Fourth, teacher feedback quality was represented through structured indicators. Real feedback may contain linguistic, emotional, disciplinary, and contextual features that cannot be completely represented by numerical variables.

Fifth, although sensitive demographic variables were excluded from the primary model, fairness auditing would still be necessary before implementation because apparently neutral educational variables may indirectly encode structural disadvantage.

Finally, explainability does not guarantee causal inference. SHAP contributions identify variables associated with model predictions; they do not prove that changing one variable will necessarily cause improved achievement.

Future Scope

Future research should validate HISBAIF using longitudinal data collected from actual secondary schools across different curricular and institutional contexts.

A multi-school prospective study could combine item-level assessment evidence, teacher feedback records, learner revisions, curriculum competency maps, and end-of-term achievement measures. Such a design would enable comparison across subjects such as mathematics, science, languages, and social sciences.

Natural-language processing could also be incorporated to examine the content of teacher feedback. Rather than simply measuring whether feedback is present, future models could estimate specificity, cognitive level, action orientation, and alignment with assessment criteria.

Another promising extension involves knowledge tracing. Bayesian Knowledge Tracing or transformer-based sequential models could estimate mastery at the competency level and determine whether SBA identifies learning gaps earlier than traditional aggregate scoring.

Future work should also investigate causal designs. Cluster-randomized or stepped-wedge trials could compare conventional school-based assessment with assessment systems incorporating structured feedback uptake and teacher-facing analytics.

Fairness must remain a central research priority. Performance should be examined across socioeconomic groups, school types, language backgrounds, and learners requiring additional educational support. Models producing strong aggregate accuracy but systematically weaker calibration for particular groups should not be deployed.

Teacher usability research is equally important. An analytically sophisticated dashboard may fail if its outputs create additional workload or cannot be translated into classroom decisions. Future studies should therefore measure not only predictive accuracy but also teacher trust, interpretability, intervention completion, and decision quality.

Privacy-preserving analytics, federated learning, local data processing, and restricted-access educational dashboards also merit examination where student-level records are used.

Conclusion

This study reconceptualizes school-based assessment as an iterative learning-improvement system rather than a localized

mechanism for producing grades. The proposed HISBAIF framework integrates assessment evidence, diagnostic interpretation, teacher feedback, learner uptake, instructional adaptation, and explainable analytics within a single pedagogical cycle.

The simulation-based experiment indicates that the frequency of school assessment is less informative than the quality of processes surrounding assessment. Feedback uptake, feedback specificity, reassessment participation, reduction in competency-specific errors, and moderation consistency emerged as the most influential indicators of subsequent learning improvement.

Among the evaluated models, XGBoost produced the strongest combination of discrimination, calibration, sensitivity to limited-improvement learners, and interpretability. The GRU model demonstrated that sequential learning patterns can also provide useful predictive information, but its greater complexity did not produce a corresponding improvement in teacher-facing explanatory value.

The central educational implication is that artificial intelligence should not be used to automate teacher judgement. Its more defensible role is to organize longitudinal assessment evidence, identify patterns that might otherwise remain unnoticed, and provide interpretable signals for professional review.

School-based assessment becomes genuinely formative when evidence results in action. A score alone documents performance; a diagnostic feedback-response cycle creates the possibility of improvement. For secondary education systems seeking to combine classroom assessment with digital analytics, the most productive direction is therefore not more testing, but **better-connected assessment, feedback, learner response, moderation, and instructional decision-making.**

References

1. Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5(1), 7–74.
2. Chen, T., & Guestrin, C. (2016). (as above)
3. Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112.
4. Memarian, B., & Doleck, T. (2024). A review of assessment for learning with artificial intelligence. *Computers in Human Behavior: Artificial Humans*, 2(1), 100040.
5. Morris, R., Perry, T., & Wardle, L. (2021). Formative assessment and feedback for learning in higher education: A systematic review. *Review of Education*, 9(3), e3292.
6. OECD. (2023). *OECD digital education outlook 2023: Towards an effective digital education ecosystem*. OECD Publishing.
7. Schellekens, L. H., Bok, H. G. J., de Jong, L. H., van der Schaaf, M. F., Kremer, W. D. J., & van der Vleuten, C. P. M. (2021). A scoping review on the notions of assessment as learning (AaL), assessment

for learning (AfL), and assessment of learning (AoL).
Studies in Educational Evaluation, 71, 101094.

8. Sortwell, A., Trimble, K., Ferraz, R., et al. (2024). A systematic review of meta-analyses on the impact of formative assessment on K-12 students' learning. Sustainability, 16(17), 7826.