iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: XII Month of publication: December 2025

DOI: https://doi.org/10.22214/ijraset.2025.75457

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XII Dec 2025- Available at www.ijraset.com

Evaluex Student Gradient System: A Survey of AI-Driven Academic Evaluation and Feedback Models

Prof. G.H. Wani, Yash Pansare,Vaibhav Parabhane, Atharva Patil
Department of Artificial Intelligence and Data Science AISSMS Institute of Information Technology, Pune, India

Abstract: Artificial Intelligence has brought transformative shifts in nearly every domain, and education is no exception.Yet,despitetherapidadoptionofdigitallearningplatforms and online assessments, the core evaluation processes in most institutions remain strikingly traditional—manual, delayed, and often lacking consistency [2]. Students wait days or weeks for feedback, instructors struggle to maintain fairness and pace, and institutions are left grappling with inefficiency.

To address these longstanding challenges, this survey explores the evolution of AI-driven academic assessment tools and intro- duces the EvalueX Student Gradient System, a next-generation conceptual framework built around gradient-based performance tracking and LLM-powered evaluation [4]. Unlike conventional assessmentsystemsthatjudgestudentsatasinglepointin time,EvalueXemphasizesunderstandinghowstudentslearnover time—capturing growth, stagnation, and learning momentum. Through analysis of modern research across automated grad- ing, feedback generation, explainability, and human-AI collabo- ration, the survey identifies critical gaps in existing approaches and highlights why the future of academic evaluation must shift towardcontinuous,transparent,andstudent-centeredevaluation ecosystems. EvalueX represents that shift—a move from static evaluations to dynamic, personalized learning trajectories aug- mented by AI [3].

I. INTRODUCTION

Educationhasalwaysreliedonassessmenttoguideteachingandmeasurelearning,buttoday'sacademicenvironmentsexposethelimitationsoftraditionalevaluationmethods.Instructorsoftenmanagelargeclasses,handlehundredsofresponses,andstruggletoprovidetimely,consistent,andmeaningfulfeedback[1].Students,ontheotherhand,waitdaysorweekstoreceivecommentsthatmaynolongerberelevanttotheirlearningmoment.Thediversityofmodernclassroomswithvaryinglearningspeeds,linguisticbackgrounds,andexpressionstylesmakesfairandconsistentevaluationevenmorechallenging.Asaresult,assessment oftenbecomesabottleneckratherthanafacilitatoroflearning.

Amid this complexity, Artificial Intelligence is emerging as a transformative force. Modern AI models, particularly large language models (LLMs), can analyze written responses with surprising sensitivity to context, reasoning, and conceptual clarity [5]. They offer significant advantages: instant grading, consistent evaluation, and the ability to generate personalized, constructive feedback at scale. More importantly, AI enables somethingthattraditionalsystemscannot—trackinglearning patternsovertime. Insteadofevaluatingstudentsbasedonisolatedtestscores,AIcanobservegrowth,identifyimprovementtrends,anddetectareaswhereearlyinterventionisneeded. Thisshifttowardcontinuous,trajectory-basedassessmentrep- resents a more holistic understanding of student learning[7].Thesepossibilitieshighlighttheneedforaunified,AIpoweredevaluationframework.TheEvalueXStudentGradientSystemaddressesthisbycombininggradientbasedperformancetracking,LLMdrivengrading,andtransparentexplainableAImechanisms[9].Ratherthanfunctioningasseparatetoolsforscoring,feedback,oranalytics,EvalueXintegratesthesecapabilitiesintoasingleplatformdesignedtosupportinstructorsandempowerstudents.Thissurveyexaminesexistingresearchinautomatedgrading,feedbackgeneration,explainability,andhumanAIcollaboration,andidentifieshowEvalueXbuildsupontheseadvancementstocreatethenextgenerationofintelligenteducationalevaluationsystems[11].

A. Background and Context

Imagineateacherhandling120papersafteraunittest.Each answer must be interpreted, scored, justified, and commented on. Even the most dedicated educator faces difficulty in main- taining consistency across so many submissions, and delaysin feedback are almost inevitable [2].

Research consistently shows that instructors spend a sub- stantial portion of their professional time grading, cross- checking, and documenting assessments. These tasks, though essential, do not directly contribute to deeper learning. Stu- dents struggle with feedback that arrives too late or lacks personalization [10].

The diversity of modern classrooms further complicates evaluation. Traditional systems, designed decades ago, cannot fully support contemporary learning environments with vary- ing student profiles and learning speeds.

B. The Imperative for AI-Based Solutions

The growing complexity of academic environments makes the case for AI-driven evaluation not just appealing, but necessary. Traditional assessment systems were built for a time when classrooms were smaller, learning resources were limited,andevaluationwasprimarilyaclericaltask.Today,students engage with dynamic content, multimodal learning platforms, and diverse academic activities—yet their assess- ments are still measured using static, manual methods thatfail to capture the full picture of learning. This mismatch creates delays, inconsistencies, and an overall fragmented understanding of student performance [3].

AI-based evaluation systems address these limitations by bringing speed, precision, and depth that are difficult for humans to maintain at scale. Large Language Models (LLMs) such as GPT-4 can interpret long-form text, recognize rea- soning patterns, detect misconceptions, and generate high- qualityfeedbackwithinseconds[4].Theyevaluatenotjustthe correctness of an answer but also the clarity of explanation, logical structure, and conceptual depth. This capability aligns directly with the goals of the EvalueX Student Gradient System, which aims to understand how students learn—not just what they write on a single exam [5].

C. Scope of This Survey

This paper provides a comprehensive survey based on significant research contributions in the field of AI-driven academic evaluation. Care has been taken to ensure that no sentenceisreproducedverbatimfrompriorliterature,and any such duplication is strictly avoided to maintain academic integrity. The scope of this survey covers the major domains where AI is making measurable impact in educational assess- ment, including:

- Automatedgradingandintelligentanswerevaluation,
- Smart performance risk assessment and learning- trajectory monitoring,
- Predictive analytics for early intervention and retention improvement,
- The use of advanced AI/ML models such as NLP, Large Language Models (LLMs), and Retrieval-Augmented Generation (RAG).

We introduce a practical application to ground our analysis—**EvalueX**, a conceptual AI-powered evaluation and feedbacksystem.EvalueXservesasaunifyingframeworkthat synthesizes insights from the surveyed literature, illustrating how modern technological advancements and research trends can converge to produce a usable, scalable, and future-ready solution for academic institutions [10].

D. OrganizationofthePaper

Theremainderofthispaperisorganizedasfollows.

- TheliteraturesurveyispresentedinSectionII,wherekey research works and comparative findings are discussed.
- Section III highlights areas that require further research and introduces the EvalueX concept as a unified frame- work.
- ThesurveyconcludeswithasummaryofinsightsinSectionIV.
- SectionVlistsallreferencesusedinthispaper.

II. LITERATURE SURVEY

A. Individual Paper Summaries

1) *Flode´n (2025):* A large-scale comparison of AI-driven and human grading showed ChatGPT achieving accuracyclose to experienced instructors, demonstrating the practical reliability of LLM-based assessment [1].
2) *Burrows et al. (2015):* A historical overview high- lighting the limitations of early rule-based scoring systemsand emphasizing the need for deeper linguistic and semantic modeling in automated assessment [2].
3) *Ramesh and Sanampudi (2022):* A systematic review confirming that neural network architectures and end-to-end deep models consistently outperform feature-based and statis- tical approaches in essay scoring [3].
4) *Johnson et al. (2024):*An empirical evaluation showing GPT-4 achieving nearly 85% agreement with expert graders, often exceeding human inter-rater consistency on complex university entrance assessments [4].
5) *Bewersdorff et al. (2023) and Dai et al. (2024):*Studies demonstrating that LLM-generated feedback can match or exceed the pedagogical richness, specificity, and clarity of instructor-generated comments [5], [6].
6) *Xu and Ouyang (2022):* A meta-analysis emphasizing the critical role of timely, formative, and iterative feedback in improving student learning outcomes and supporting continu- ous academic growth [7].

7) *Alkafaweenetal.(2024):*Anempiricalstudycomparing instructor-designed and LLM-generated test suites for pro- gramming assignments, demonstrating that LLM-created tests can identify hidden edge cases and complement traditional autograding workflows [8].
8) *Lietal.(2024):*AsystemdemonstrationofAERAChat, an explainable assessment platform that pairs automated scor- ing with rationale generation, highlight-based explanations, andaudittoolstoenhancetransparencyandeducatortrust[9].
9) *Katuka et al. (2024):* An experimental investigation showing that PEFT-trained and 4-bit quantized LLaMA-2 models can accurately predict grades and generate expert-like feedback, enabling cost-efficient, scalable automated assess- ment [10].
10) *Xie et al. (2024):*A multi-agent grading frame-work—Grade Like a Human—that divides rubric generation, grading,andscorereviewacrossspecializedagents,improving grading consistency and reducing evaluation errors [11].

B. Comparative Summary

III. FUTURE WORK AND PRACTICAL APPLICATION

While the surveyed literature demonstrates significant ad- vancements in the application of Artificial Intelligence within academicevaluationandfeedbacksystems,thestudyuncovers a number of ongoing gaps and practical challenges. A cohe- sive, end-to-end solution that fully incorporates these diverse advancementshasyettoberealized.Thissectionidentifiesthe keylimitationsofcurrentresearch,exploresthepotentialof emerging technologies, and introduces EvalueX as a practical approachtofillthesegaps—anapplicationthatbuildsuponthe surveyed works to form a next-generation evaluation solution [9].

TABLEI
COMPARATIVESUMMARYOFSURVEYEDLITERATURE

Paper#	Authors(Year)	FocusArea	Methodology	KeyContribution	RelevancetoEvalueX
[1]	Flode´n(2025)	AIvsHuman Grading	ComparativeStudy	DemonstratesLLMgrad- ingaccuracycomparable tohumanevaluators	SupportsLLM-basedevalu- ation
[2]	Burrowsetal.(2015)	EvolutionofAES	LiteratureReview	Identifieslimitationsof rule-basedscoringsys- temsandtheneedfor deeperlinguisticmodels	Motivates neural/LLM adoption
[3]	Ramesh & Sanampudi (2022)	EssayScoring	SystematicReview	Neural models consistently outperform feature-based scoring systems	FoundationforEvalueX scoringmodel
[4]	Johnsonetal.(2024)	GPT-4 Assessment	EmpiricalStudy	GPT-4achieves 85% agreementwithhuman ex- pertgraders	ValidatesLLMreliability forEvalueX
[5]	Bewersdorffetal.(2023)	AutomatedFeed- back	ExperimentalStudy	LLM-generatedfeedback matchestherichnessand clarityofteacherfeedback	SupportsEvalueXfeedback generation
[6]	Xu&Ouyang(2022)	FormativeFeed- back	Meta-analysis	Timely,iterativefeedback significantlyimprovesstu- dentlearningoutcomes	Justifies trajectory-based evaluationinEvalueX

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XII Dec 2025- Available at www.ijraset.com

[7]	Alkafaweenetal.(2024)	Program Autograding	EmpiricalStudy	LLM-generatedtestsuites findhidden	
[8]	Lietal.(2024)	ExplainableAs- sessment	SystemDemonstration	AERA Chat provides rationale generatio n, highlight-based explanations, and auditability	GuidesEvalueXexplain- abilityandtransparencyde - sign
[9]	Katukaetal.(2024)	Grading&Feed- back(PEFT)	ExperimentalStudy	PEFT-trained 4-bit LLaMA-2 models accuratelypredictgrades andgenerateexpert-like feedbackatlowcost	Enablescost- efficientfine- tuningandscalablefeed- backinEvalueX
[10]	Xieetal.(2024)	Multi-Agent Grading	System/Empirical	Multi-agent rubric–grading–review pipelineimprovesgrading consistencyandreduces errors	SupportsEvalueXmulti- agentworkflowforrubric+ reviewevaluation

A. Identified Research Gaps and Limitations

- Lack of Temporal Learning Interpretation: Current systems evaluate students based on isolated responses, lacking the ability to interpret long-term learning trajec- tories or performance gradients [7].
- OpaqueEvaluationLogic:ManyAIgradingmodelsare difficult to interpret, making it challenging for educators to justify or audit evaluation decisions [9].
- Inconsistent Feedback Quality: Feedback generated by AI varies due to prompt sensitivity and model behavior, highlighting the need for more controlled and pedagogi- cally aligned feedback generation [10].
- Scalability and Deployment Issues: Systems that per- formwellinresearchsettingsoftenfacelatency,through- put, or cost challenges in real-world institutional deploy- ments [12].
- BiasandFairnessConcerns:AImayunintentionallyfa- vor certain writing styles or linguistic patterns, requiring strong fairness monitoring and bias mitigation strategies [1].

B. New Technology and Future Directions

Promising directions include advanced LLMs for nuanced reasoning interpretation, multimodal evaluation models, Ex- plainable AI (XAI) for transparent justification of scores, gradient-based learning analytics for trajectory modeling, and cloud-native RAG architectures for reliable real-time evalua- tion. Privacy-preserving training and fairness-controlled scor- ing frameworks are also important future research areas [3], [9].

C. Practical Application: The Evalue XPlatform

EvalueX is an integrated AI-powered academic evaluation platform that incorporates the technologies examined in this study into a unified solution for educational institutions. Key functionalities include:

- AI-drivengradingandlearninginsightsusingLLM-based analysis of descriptive and analytical responses [4].
- NLP and RAG-based personalized feedback generation aligned with curriculum requirements [5].
- Learning gradient tracking for monitoring performance trends, improvement patterns, and early risk prediction [7].
- Adaptive recommendations for learning materials, revi- sion tasks, and intervention strategies.
- A centralized platform that supports instructor review, score adjustment, bias checking, and transparent reason- ing visualization.

II. CONCLUSION

AI is reshaping academic evaluation, offering consistency, insight, and scalability. However, existing systems lack holis- tic design and temporal awareness. EvalueX addresses these limitations by introducing a gradient-based, transparent, and student-centered model. The future of assessment lies not in static scoring but in dynamic, AI-augmented learning trajec- tories.

III. ACKNOWLEDGMENT

We thank our guide Prof. G.H.Wani,Department of Ar- tificial Intelligence and Data Science,AISSMS Institute of InformationTechnology,Puneforhisguidance,motivationand constant support incompleting this survey paper.Thi swork benefitedtremendouslyfromhisvaluableinputandcomments which helped shape the outcome.

REFERENCES

[1] J. Flode´n, "Grading Exams Using Large Language Models: A Com-parison Between Human and AI Grading in Higher Education UsingChatGPT," British Educational Research Journal, 2025.
[2] S. Burrows, I. Gurevych, and B. Stein, "The Eras and Trends in Auto-maticEssayAssessment,"InternationalJournalofArtificialIntelligencein Education, vol. 25, no. 1, pp. 60–117, 2015.
[3] R. Ramesh and V. Sanampudi, "Automated Essay Scoring Systems: ASystematicLiteratureReview,"ArtificialIntelligenceReview,vol.55,
[4] pp.2495–2527,2022.
[5] A. Johnson et al., "Evaluating GPT-4's Vision-and-Language Capabili-ties on Brazilian University Admission Exams," arXiv preprint, 2024.
[6] P.Bewersdorffetal.,"LargeLanguageModelsforAutomatedFeedbackonStudentWriting,"IEEETransactionsonLearningTechnologies,vol.17, pp. 234–248, 2024.
[7] Y. Dai et al., "Pedagogical Feedback with LLMs," Conference onLearning Technologies, 2024.
[8] K. Xu and M. Ouyang, "Effectiveness of Automatic Feedback andTeacher Feedback on Chinese Learners' English Writing," Computersand Education: Artificial Intelligence, vol. 3, p. 100059, 2022.
[9] A. Alkafaween et al., "LLM-Generated Test Suites for ProgrammingAutograding," unpublished manuscript, 2024.
[10] J. Li, A. Bobrov, D. West, C. Aloisi, and Y. He, "An AutomatedExplainable Educational Assessment System Built on LLMs (AERAChat)," demo paper, 2024.
[11] G.A.Katuka,A.Gain,Y.-Y.Yu,"InvestigatingAutomaticScoringandFeedback using Large Language Models," arXiv preprint, 2024.
[12] W. Xie, J. Niu, C. J. Xue, and N. Guan, "Grade Like a Human:RethinkingAutomatedAssessmentwithLargeLanguageModels,"arXivpreprint, 2024.
[13] Gloria A.Katuka etal., "InvestigatingAutomatic Scoringand Feedbackusing Large Language Models," 2024.