Research on the Construction of a Dynamic Teaching Evaluation System Based on Generative Artificial Intelligence

Siyi Wang[1*]

[1] Guangdong Polytechnic Normal University, Guangzhou, 510665, China

* 2505595980@qq.com

https://doi.org/10.70695/IAAI202602A12

Abstract

Digital technology is moving so fast that we're running into a real shortage of top-tier innovative talent. This shift is making it hard for traditional teaching assessment models to keep up, and it's also bringing some deep-rooted, long-standing issues to the surface. Generative Artificial Intelligence (GenAI) is great at handling massive datasets, which gives us a fresh way to look at dynamic teaching assessment. In this paper, we're going to explore how we can use this technology to build a more dynamic evaluation framework for education. The introduction of Artificial Intelligence (AI) has changed the underlying logic of the entire evaluation system. Old-school evaluation systems used to be straightforward and top-down, but that just doesn't work anymore. We now need more voices involved. The focus has shifted away from just testing surface-level knowledge toward actually digging into how students think. What's more, those old, slow grading methods are being replaced by systems that offer real-time feedback. Because of this, assessment shouldn't just be an end-of-term event. It needs to be woven into every part of the learning journey—before, during, and after class—to keep the evaluation process ongoing. Of course, teachers have to stay in the driver's seat throughout the teaching process. This is non-negotiable. We really need to be careful about the risks and ethical baggage that can come with using new tech. Overall, this new system aims to break past the limits of traditional quantitative assessment. It brings intelligent tech back to its core purpose: supporting human growth. By blending practical tech with the human side of education, we're proposing an evaluation model that is both more scientific and genuinely effective for training talent in this digital age.

Keywords Generative Artificial Intelligence; Dynamic Teaching Evaluation; Formative Assessment

1 Drivers of the Transformation of Teaching Evaluation Systems in the Era of GenAI

Every major technological breakthrough has profoundly reshaped the form of productivity and, in turn, redefined the educational ecosystem. From mechanization and electrification to informatization, technology has consistently acted as a key driving force behind the transformation of education [1]. Educational evaluation is not only an essential means of guiding teaching practices, fostering students' core competencies, and improving the quality of classroom instruction, but also an intrinsic driving force behind educational reform [2]. Against the backdrop of rapid technological advancement and profound social transformation, teaching evaluation models are undergoing a significant and dynamic shift.

1.1 The Demand for Highly Qualified Personnel Calls for a Reform of Evaluation Methods

As GenAI technology matures, the global educational focus has shifted from traditional knowledge retention to the development of higher-order thinking skills. This is highly emphasized by the OECD's "Education and Skills 2030" and UNESCO's "Skills for the 21st Century" initiative. Consequently, major developed countries have elevated educational reform to a national strategic level, with initiatives like the US's "STEM Education 2026" plan and the EU's "Digital Education Action Plan" actively exploring new paths for talent cultivation in the digital age.

Faced with this global transformation, China is at a crucial turning point in building a strong educational nation. General Secretary Xi Jinping emphasized that "China attaches great importance to the profound impact of AI on education, actively promotes the deep integration of AI and education, and encourages educational reform and innovation" [3]. To meet current requirements, the Master Plan for

Building a Strong Educational Focus Country (2024-2035) also specified the need to "promote AI to support educational reform and open new avenues of development and shape new strengths through the digitalization of education" [4].

However, the current assessment methods used in schools clearly lag behind these rising demands. Traditional evaluations are too rigid and overly focused on rote memorization, making them incapable of measuring the critical thinking and higher-order skills required today. This stark mismatch between modern educational needs and outdated assessment standards means we must fundamentally rethink and rebuild the entire educational evaluation system, breaking away from the old mold to create a dynamic model that truly works for the new era.

1.2 Traditional Teaching Evaluation Models Highlight Inherent Limitations

The standards for innovative and well-rounded talents in today's society are constantly rising, and traditional teaching and assessment models have exposed a number of unavoidable shortcomings in the process of actual design and implementation. At the level of evaluation subjects, previous classroom assessments were basically completed by teachers alone, without the joint participation and collaborative evaluation of multiple subjects. The single observation perspective has obvious limitations, and it is difficult to fully capture students' true learning status and growth. In terms of evaluation content, the existing assessment content is generally superficial, mostly focusing on students' memorization and accumulation of textbook knowledge. It cannot effectively explore students' logical reasoning ability in the process of thinking in the humanities and social sciences, nor can it identify students' implicit core competencies such as learning attitude and value orientation.

When it comes to assessment, our current system is clearly lopsided. It leans too heavily on final results and relies entirely on standardized tests at fixed intervals. This completely misses the student's actual learning journey—things like how they navigate cognitive biases, adjust their thinking, or self-correct—none of these crucial growth markers ever show up in the final grade. When it comes to feedback, the biggest issue with traditional methods is how slow they are. The feedback is often outdated and repetitive, which just trains students to chase standard answers rather than actually learning. Because of this, it fails to provide the real, timely guidance students need to improve their skills or adjust their study methods. The old, cookie-cutter approach to assessment has largely held students back, preventing them from developing the full range of skills and qualities they actually need. Bista's critique of the "measurement culture" in contemporary education is incisive, pointing out that this evaluation paradigm alienates teachers into producers of data, while students are objectified as measurable indicators [5]. Given all these issues, building a new assessment system that actually matches modern teaching goals isn't just an option—it's an urgent priority for education reform right now.

1.3 GenAI Offers an Opportunity for a Transformation in Evaluation

GenAI, as a new factor of production in human society, has successfully transcended the scope of traditional factors of production with its unparalleled intelligence and efficiency, reshaping the face and boundaries of education [6]. Currently, traditional educational assessment models are no longer sufficient to meet the needs of individualized and innovative talent development in the digital age. GenAI, with its remarkable ability to adapt to different situations, provide personalized feedback, and actually generate knowledge, offers a fresh way to solve the long-standing challenges we've faced in reforming our assessment models. Dynamic evaluation advocates using high-frequency interaction during the teaching process to measure and promote the improvement of students' comprehensive qualities.

However, traditional IT has consistently struggled to keep pace in handling unstructured data. It is simply incapable of handling the heavy processing required for this level of data analysis. This is why dynamic assessment has always been difficult to implement in large-scale classrooms in the real world—data collection is too difficult, analysis is too costly, and feedback always seems to fail to arrive in time when students really need it. The rise of GenAI has been a complete game-changer. Finally, we have the tools to push past those old technical hurdles and shift toward a smarter, more effective way of evaluating student learning. The diagnostic and generative functions of assessment have been truly activated by GenAI, which provides a key impetus for reshaping teaching assessment methods; GenAI has strong language understanding and multi-information processing capabilities. It can analyze a large

amount of process data, such as classroom discussions and exploration notes, in real time. Moreover, it is very sensitive and can track changes in students' logical thinking and immediately provide personalized feedback to help them make up for their shortcomings in thinking.

2 Dynamic Reconstruction of Teaching Evaluation Dimensions Driven by GenAI

In the era of AI, establishing a more equitable, educationally sound, diversified, and scientific talent evaluation system is beneficial to both students' growth and professional development and teachers' professional growth, taking into account all aspects [7]. The deep involvement of GenAI has completely broken the closed pattern of traditional evaluation and promoted a fundamental shift in the evaluation system in four dimensions: subject, content, method, and feedback, thus constructing a new dynamic evaluation paradigm that meets the requirements of talent cultivation in the era.

2.1 Evaluation Subject: Shifting from a Single Authority to Multi-Party Collaboration

Building a more scientific and dynamic teaching evaluation model often requires taking multi-party collaborative evaluation methods as the foundation. Looking back at the traditional teaching evaluation system, it does have many limitations. In the past, teachers were basically the sole implementers of evaluation. Teachers have limited time and energy, and their personal cognitive perspectives also have limits. It is difficult for them to conduct frequent and detailed learning observations and learning situation diagnoses for every student in the class. The results from the traditional education evaluation model are actually difficult to be completely comprehensive and objective. After adding AI to the education scene, this old-fashioned industry pain point finally has a more tangible solution. It can become the core driving force for the reform of education evaluation, in the final analysis, because the technology itself has the characteristics of independent adaptation that traditional tools cannot match [8]. A large number of complicated and scattered information generated in the usual teaching, the system now relies on the monitoring system throughout the whole process to capture, and then take the mature massive data parallel processing ability to run, and finally turn into a standardized text to sort out and integrate. As a result, a multi-angle and high-precision evaluation feedback channel is built up. With the support of the system, the large number of repetitive and mechanical evaluation and correction work can be taken over by AI. The burden of the first-line teachers is much lighter, just to focus the free energy on the students' emotional state research and judgment, and to do the targeted value guidance work. At the same time, the evaluation participation dimension of the whole education has also been broadened. Students themselves log on to the online application portal and complete the personalized self-assessment at any time, and can also conduct anonymous mutual evaluation among students. The system takes the feedback content given by different evaluation subjects to do in-depth combing, which is equivalent to pulling the power of schools, society, and other aspects into the education evaluation link. When this multi-party collaborative evaluation model really turns up, the previous fixed evaluation thinking that relies solely on scores is naturally broken. This not only jumps out of the limitations of the old traditional evaluation, but also truly constructs a dynamic education evaluation system that just meets the actual needs of high-quality personnel training in the new era.

2.2 Evaluation Content: Shift from Surface Knowledge to Tacit Thinking

The previous static education evaluation basically only focuses on written knowledge, and students' real high-level thinking and innovative literacy are actually difficult to measure. In order to cultivate high-quality talents, it is not enough to see how basic knowledge is mastered. The focus must be on logical reasoning, critical thinking, and a deep value standpoint. After all, these implicit qualities are the key to assessment. This kind of comprehensive ability often lacks a fixed measurement structure. If we still take the past standard answer mode to set, it is not accurate at all.

However, with the application of GenAI, this problem has a better solution, and the academic community has finally found a new way to judge students' implicit thinking. The knowledge base and advanced natural language processing capabilities of GenAI can efficiently analyze massive educational data such as classroom dialogues and teaching materials [9], that kind of complex topic exploration scenario, the system itself can grab useful conceptual information from multimodal data, and can also

dig deep into the logical rigor of students' argumentation problems to see whether their thinking perspectives are rich or not. After such a round of innovation, the focus of the evaluation system has also undergone an essential change. It is no longer just a shallow knowledge point test, but a real focus on the excavation of interdisciplinary thinking and core literacy.

In addition, the platform can also use the implicit thinking data captured in the process of inquiry and then combine the specific academic situation of students to push basic guidance materials or advanced expansion resources in a targeted manner, just to the rhythm of cognitive development of different students. In this way, the evaluation content can be updated and dynamically generated in real time along with the students' cognitive growth, and this method can be said to have completely achieved the original intention of educational evaluation.

2.3 Evaluation and Feedback: Shifting from Delayed Scoring to Immediate Support

The old evaluation feedback usually only gives a simple score or a grade. This practice of scoring afterwards has long been separated from the specific situation of learning at that time, and there is no substantive guidance and advice at all. So, students often do not know where their weaknesses are after they get the results, let alone how to improve them, which leads to the whole evaluation can easily become a one-way, direct conclusion. However, after the introduction of GenAI, this old mechanism has been changed. It can turn the original rigid evaluation feedback into a specific and easy-to-use learning aid. After all, the real value of introducing these technologies is to stimulate students' thinking quality and innovation ability, rather than improving efficiency in traditional rote learning [1]. Therefore, when the intelligent system analyzes the data left by the learning process and finds that students have logical faults or cognitive blind spots.It will no longer give a simple right and wrong judgment as before, but will directly provide a very clear optimization and improvement path. For example, when students encounter deviations in exploring problems, the system will immediately generate some questions that can guide thinking. Or to supplement the necessary background literature, or even to disassemble the original complex learning tasks and plan a step-by-step compensation training ; this kind of immediate and very real feedback can make students clearly see where they are wrong and know how to continue to learn in the next step, so that they can take the initiative to fill in the short board and adjust their state. This not only provides a very intuitive basis for teachers to improve the follow-up teaching arrangements, but also truly implements the concept of Promoting learning through evaluation. Let the evaluation feedback completely become a practical tool that can effectively guide action and promote the continuous development of the ability.

3 Implementation Path of Dynamic Teaching Evaluation Empowered by GenAI

3.1 Introduce a Subject-Specific Knowledge Base to Train a Dedicated Evaluation Model

Sure, big language models like ChatGPT are pretty impressive when it comes to generating text and handling logical analysis. But the reality is, once you try to drop these general models into the actual teaching of specific subjects, their lack of deep professional knowledge becomes a real issue. If we want to genuinely improve intelligent teaching evaluation systems, we have to upgrade and reshape these general tools using specialized knowledge. They need to be fully adapted to the real teaching situations we face across various subjects. Getting these targeted large language models off the ground really requires educators and technical staff to team up and focus on joint research.

When we build these new models, they absolutely have to cover complete disciplinary knowledge frameworks. Beyond just the content, they need to account for the actual ways students think and incorporate positive educational ideas. In terms of making it work on the ground, we can run dynamic verification and continuous optimization by anchoring the models in national curriculum standards, classic academic documents, and high-quality teaching cases. Pushing through these practical optimization measures is what helps the intelligent models handle the messiness of diverse and complex classroom scenarios. It allows the system to map out exactly how professional knowledge points develop and connect with each other, ultimately giving us a set of scientific and reliable intelligent scoring standards that actually fit the real situation of classroom teaching.

3.2 Implement ongoing dynamic assessment throughout the entire classroom process

If we really want to implement dynamic teaching assessment, a dedicated assessment model is definitely a core support. However, if we want to use this model in the classroom, we need to change our old habits first. The previous separation of teaching and assessment is no longer feasible. Now, there is multi-modal data technology at the bottom of the support, a set of evaluation system can just logically spread into all aspects of pre-class, in-class and after-class, and truly follow the whole process of teaching.

The pre-class stage is mainly to make a pre-diagnosis. Before the formal class, the system first sorts out the data left by the students before the class to see how their knowledge base is and where it is still relatively weak. By the way, select some content that is similar to their current cognitive level to record; this set of down, the basic learning situation of the whole class to understand, later want to see how much progress in the end of the students, at least there is a reference to the beginning of the data.

The main stage of the lesson is the accompanying evaluation. When teaching activities go forward, the system will always record the students' various learning behaviors in the background, such as text, voice, and image. It even covers the data generated by sensory interactions such as vision and hearing. The system collects this messy information to generate real-time learning situations, and then sends out guidance suggestions; At this time, the evaluation is not only to see whether the question is right, but to stare at the students' logical derivation when they encounter cognitive conflicts that are not strict, and the critical thinking is not deep. Including the effect of teamwork, the original rigid scoring standard has naturally transformed into a process diagnosis that is monitored at any time.

The after-school stage is used to generate a comprehensive development evaluation. When a class is completed, the system will integrate the previous process data again. Take out a multi-dimensional assessment result, so that people can fully see the performance of students in this class; it can also according to the evaluation results, to different levels of students to push the difficulty of the appropriate expansion of the material and personalized work; As for how these evaluation data are presented, the data analysis system can directly use text or pictures to show the trend of teaching development and the pre-judgment results, which provides a more reliable data reference for teachers to adjust the teaching rhythm and do accurate counseling.

3.3 Establishing Teacher Leadership and Mitigating Technological and Ethical Risks

In the process of gradually applying GenAI technology in the field of educational evaluation, front-line teachers and relevant scientific researchers need to maintain a rational and prudent application attitude. First of all, users need to rationalize the application of technology; intelligent technology can indeed simplify the evaluation process of teachers and effectively improve the overall teaching efficiency, but the problem of technology dependence needs to be avoided. The dynamic learning situation and internal thinking analysis results of students sorted out by an AI system can only be used as a reference for teachers to carry out teaching evaluation. Teachers cannot directly apply the data generated by the platform as the final conclusion to judge students' learning situation.

Secondly, teachers need to always adhere to their own dominant position in teaching evaluation; At present, the update and iteration speed of AI technology is extremely fast, and some intelligent computing and analysis capabilities have surpassed the manual level. Human-machine competition and human-machine collaboration have become the normal development in the field of education. In the face of the AI era of 'machine humanization' change and man-machine competition, emotion has become the key weight of human victory [6]. In daily teaching and evaluation practice, teachers should focus on the core work of emotional communication and value guidance and give full play to their leading role. The application of technology is only an auxiliary means, and the operation logic of intelligent tools cannot be allowed to influence the evaluation criteria of educational value. Especially in key links such as the assessment of students' ideological and political literacy, the screening of complex moral and ethical issues, and the judgment of students ' emotional identity, teachers must strictly control them. Teachers can rely on the temperature of humanistic education to make up for the inherent drawbacks of intelligent algorithms, so as to adhere to the original intention of educating people and implement the core educational goals of awakening students' minds and cultivating students' good moral character.

Furthermore, technology applications need to adhere to the bottom line of compliance and do a good job in the prevention and control of data security and ethical risks. In order to realize the long-term and stable application of the dynamic teaching evaluation model, we must pay attention to the overall safety construction of the education scene. In the whole process of student data collection and information

analysis, schools and intelligent platforms need to build a perfect security protection mechanism to effectively protect students' personal privacy and daily learning data and avoid security risks such as data leakage from the source.

Finally, teachers should do a good job in guiding students' education and helping students to objectively recognize the application scope and functional boundary of intelligent tools. From the theoretical level of knowledge construction, the characteristics of real-time feedback of AI will enable students to skip the necessary learning links of independent review, self-reflection, and optimization of learning strategies, and greatly reduce the exercise space of students' critical thinking [10]. In the long run, it will lead to deviations in students' knowledge, cognition, and thinking construction. This requires teachers to adopt scientific guidance methods to guide students to standardize the use of intelligent tools and prevent students from over-reliance on AI technology.

Properly dealing with the above-mentioned application problems of various technologies can realize the two-way complementarity between human educational wisdom and AI technology. Promote the benign development of the human-machine collaborative education model and help the education evaluation work continue to optimize and upgrade.

4 Conclusion

The reform and upgrading of education evaluation systems go beyond simply bringing in intelligent teaching tools; it involves many aspects such as evaluation subjects, evaluation content, and feedback methods, and it drives a comprehensive transformation of the whole evaluation mechanism. GenAI has very strong abilities in collecting, integrating, and processing data, which can make up for various long-standing weaknesses in traditional teaching evaluation and allow many evaluation methods that were difficult to carry out before to be put into practice. For example, it can track students' learning throughout the entire process, analyze their implicit thinking, and offer timely and personalized teaching guidance even in large classes. Digital technology can help education advance quickly, but the core aim of using technology is still to cultivate students' higher-order thinking and innovative literacy. The education sector should not overlook the humanistic attributes that belong to education just because intelligent tools are widely used; teachers need to firmly hold on to their core and leading role in teaching evaluation, and bring together the technological strengths of GenAI with a people-oriented educational philosophy and positive value guidance. In this way, the dynamic teaching evaluation system can run stably over the long term, keep supporting the cultivation of high-quality talents, and truly achieve the key mission of reforming and making breakthroughs in the education evaluation system for the new era.

Acknowledgement

This work was supported without any funding.

Conflicts of Interest

The authors declare no conflicts of interest.

References

1. Song, Y. (2026) From Knowledge Transmission to Thinking Stimulation: Research on the Classroom Teaching Transformation Empowered by Generative Artificial Intelligence. China Educational Technology, (6): 103-109+124.
2. Xia, X.J. (2026) Logical Turn and Practical Direction of Classroom Teaching Evaluation in Primary and Secondary Schools under the New Curriculum Concept. Journal of Chinese Education, (3): 81-88.
3. China Youth Daily. (2019) Xi Jinping's Congratulatory Letter to the International Conference on Artificial Intelligence and Education. China Youth Daily, 01.
4. State Council Gazette. (2025) Outline of the Construction Plan for a Leading Country in Education (2024-2035). General Office of the State Council of the People's Republic of China.

5. Biesta, G. (2019) Good Education in an Age of Measurement: Ethics, Politics, Democracy. (Zhang, L.P., Han, Y.F., Trans.). Beijing Normal University Press, Beijing. pp. 11-25.
6. Wei, X., Lu, Y.S. (2026) Realization Mechanism of Generative Artificial Intelligence Empowering the Connotative Development of Ideological and Political Courses in Colleges and Universities. China Educational Technology, (6): 119-124.
7. Chen, J.E., Qiu, Y.J. (2023) Educational Evaluation Reform Boosts the Construction of New Liberal Arts in the Era of Artificial Intelligence. Education Research Monthly, (5): 106-112.
8. Li, J.L., Niu, Z.D. (2026) Exploration of the Practical Mechanism of Deep Integration of Artificial Intelligence into University Teaching Evaluation. China Higher Education, (Z1): 76-80.
9. Wang, Z.J., Long, S., Zhang, J. (2025) Construction of a Hierarchical Model for Human-Machine Collaborative Intelligent Classroom Teaching Evaluation. Distance Education Journal, 43(5): 32-40.
10. Wang, T.P., Qin, P. (2026) Thinking Dependence Risks and Avoidance of Generative Artificial Intelligence in Educational Applications. Modern Distance Education Research, 38(3): 39-46.

Biographies

1. **Siyi Wang** Master's student at Guangdong Polytechnic Normal University.

基於生成式人工智能的動態教學評估系統建構研究

王思懿[1]

[1]廣東技術師範大學，廣州，中國，510665

摘要：數智化時代對創新型高素質人才的訴求，凸顯了傳統教學評估的固有限制。生成式人工智能憑藉著強大的大量資料處理能力，為實現新型動態評估提供了歷史性契機。本文深入探討了基於生成式人工智能的動態教學評估體系建構：首先剖析了評估體系在智慧驅動下，從單一權威向多方協同、由表層知識考核向隱性思維深探、從滯後賦分向即時學習支架轉變的動態重構邏輯。其次，提出了透過引入學科知識庫訓練專屬評估模式、貫穿課前課中課後全流程落實伴隨式評估，以及在確立教師主導權的基礎上規避技術與倫理風險的具體實施路徑。此體系旨在打破傳統工具化評測的桎梏，實現技術理性與教育人性的深度平衡，為新時代人才培養提供科學的新型評價典範。

關鍵詞：生成式人工智能；動態教學評估；過程性評估

1. 王思懿，廣東技術師範大學在读碩士。