BG Pulse Journal USA
ISSN:1947-2536 http://www.bgpulseusa.com/index.php/bg

# A MULTI-INSTRUMENT SYSTEM FOR ASSESSING THE AI-BASED PEDAGOGICAL COMPETENCE OF PRE-SERVICE PRIMARY SCHOOL TEACHERS

**Maxsudova Shohsanam Muzaffarxo‘jayevna**
Senior Lecturer, Department of Intellectual Sciences and Information Technologies,
Namangan State Pedagogical Institute
Phone: +998 97 132-32-22;
e-mail: shohsanammaxsudova@gmail.com
doi:https://doi.org/10.5281/zenodo.24028593

**Abstract.** The article presents a system for assessing the artificial intelligence (AI)-based pedagogical competence of pre-service primary school teachers. The system follows the chain “component → criterion → indicator → diagnostic instrument → level” and assigns a separate instrument to each component: a questionnaire, a theoretical test, rubric-scored practical tasks, cases and a reflection sheet. The procedure for converting raw scores to a 100-point scale, the score ranges of competence levels, the diagnostic points, the method of ensuring the content validity of the instruments and the pedagogical conditions of the assessment are described.
**Keywords:** artificial intelligence, pedagogical competence, assessment system, diagnostic instrument, rubric, competence levels, content validity, pre-service primary school teacher.

**Annotatsiya.** Maqolada bo‘lajak boshlang‘ich sinf o‘qituvchilarining sun’iy intellektga asoslangan pedagogik kompetensiyasini baholash tizimi yoritilgan. Tizim “komponent → mezon → ko‘rsatkich → diagnostik vosita → daraja” zanjiriga asoslanadi va har bir komponent uchun alohida diagnostik vositani nazarda tutadi: so‘rovnoma, nazariy test, rubrika asosida baholanadigan amaliy topshiriqlar, keyslar va refleksiya varaqasi. Birlamchi ballarni 100 ballik shkalaga keltirish tartibi, kompetensiya darajalarining ball oraliqlari, diagnostika kesimlari, vositalarning mazmuniy validligini ta’minlash usuli hamda baholashning pedagogik shart-sharoitlari bayon qilingan.
**Kalit so‘zlar:** sun’iy intellekt, pedagogik kompetensiya, baholash tizimi, diagnostik vosita, rubrika, kompetensiya darajalari, mazmuniy validlik, bo‘lajak boshlang‘ich sinf o‘qituvchisi.

**Аннотация.** В статье представлена система оценки педагогической компетенции будущих учителей начальных классов, основанной на искусственном интеллекте (ИИ). Система построена по цепочке «компонент → критерий → показатель → диагностический инструмент → уровень» и предусматривает отдельный инструмент для каждого компонента: анкету, теоретический тест, практические задания, оцениваемые по рубрике, кейсы и лист рефлексии. Изложены порядок перевода первичных баллов в 100-балльную шкалу, интервалы уровней компетенции, диагностические срезы, способ обеспечения содержательной валидности инструментов и педагогические условия оценки.
**Ключевые слова:** искусственный интеллект, педагогическая компетенция, система оценки, диагностический инструмент, рубрика, уровни компетенции, содержательная валидность, будущий учитель начальных классов.

## Introduction

Over the past few years, generative artificial intelligence (AI) tools are increasingly becoming part of teachers’ everyday work. Researchers describe AI literacy as a combination of knowing and understanding AI, using and applying it, evaluating and creating with it, and addressing its ethical

issues [1]. For a teacher, these abilities cannot be separated from pedagogy: as the TPACK framework shows, technological knowledge becomes professionally meaningful only when it is integrated with content and pedagogical knowledge [2]. However, many existing AI literacy instruments, such as the scale developed by B. Wang and colleagues, are self-report scales [3]: they show what students believe about their abilities rather than what they can actually do. This article presents an assessment system developed within a doctoral study for measuring the AI-based pedagogical competence of pre-service primary school teachers through several instruments, together with the pedagogical conditions under which the assessment is carried out.

**Structure of the assessment system**

In this study, competence is viewed as an integrated quality consisting of four components: motivational-value, cognitive, activity-methodological and ethical-reflective. The assessment system follows a five-link chain: component → criterion → indicator → diagnostic instrument → level. The first three links were theoretically grounded in the first chapter of the dissertation; the second chapter assigns a specific instrument, a scoring procedure and level boundaries to each indicator (Table 1).

**Table 1. The assessment system of AI-based pedagogical competence**

| Component | Main indicator | Diagnostic instrument | Maximum raw score |
|---|---|---|---|
| Motivational-value | Interest in and responsible attitude toward using AI for pedagogical purposes (M1–M5) | Questionnaire, 20 Likert-type items | 100 (20 × 5) |
| Cognitive | Knowledge of AI capabilities and limitations (K1–K6) | Theoretical test, 30 items | 30 |
| Activity-methodological | Creating didactic products with AI (F1–F7) | 3 practical tasks scored with a rubric | 63 (3 × 21) |
| Ethical-reflective | Critical and ethical evaluation of AI output (E1–E5) | 2 cases and a reflection sheet | 30 (2 × 12 + 6) |

Only one of the five instruments, the questionnaire, is fully based on self-assessment. It uses a five-point Likert scale [4] and contains four subscales of five items each: interest in AI, awareness of its pedagogical value, motivation for independent learning, and a responsible attitude and balanced position. Four items are reverse-worded to detect mechanical responding. The test, the rubric-scored tasks and the cases measure students' actual performance. The reflection sheet, although written by the student, is scored by the instructor against set criteria.

**Scoring procedure and competence levels**

To make the results of different instruments comparable, the raw score of each component is converted to a 100-point scale. For the questionnaire, the four reverse-worded items are recoded first (score = 6 − response), and the total score S, which ranges from 20 to 100, is transformed as $M = (S - 20) / 80 \times 100$. For the other components, the score is the proportion of the maximum: $K = b / 30 \times 100$, $F = b / 63 \times 100$ and $E = b / 30 \times 100$, where b is the raw score. The composite score is

the arithmetic mean of the four component scores: I = (M + K + F + E) / 4. Equal weights reflect the systemic view of competence: a weakness in one component cannot be fully compensated by high scores in another.

**Table 2. Score ranges of competence levels**

| Level | Score range (components and composite score) | Additional condition |
|---|---|---|
| Low (acquisition) | 0–55 | — |
| Medium (deepening) | 56–85 | — |
| High (creation) | 86–100 | For the composite score: no component below 56 |

The level boundaries are aligned with the 100-point grading system used in higher education institutions. The additional rule for the high level reflects the integrative nature of competence. A student whose composite score is 86 or above but who has a component below 56 is placed at the medium level and is assigned individual corrective work on the weak component.

**Diagnostic points and validity**

Diagnostics are carried out at three points. The baseline assessment, conducted before the module, establishes students' initial level. The ongoing assessment takes place at the end of each stage of the pedagogical technology and triggers corrective measures. The final assessment records the level achieved at the end of the module. Parallel versions of equal difficulty are used in the baseline and final assessments, which allows a direct comparison of results for each component.

The content validity of the instruments is checked by a panel of ten experts in pedagogy, primary education methodology and information technology, who rate how well each item corresponds to its indicator. The ratings are processed using C.H. Lawshe's content validity ratio (CVR) [5]. For ten experts, the critical value of CVR is 0.62; items below this value are revised or removed. A preliminary version of the instruments was piloted with students in April 2026, and its results were used only to refine the instruments.

**Pedagogical conditions**

The methodology specifies five pedagogical conditions, each with concrete measures and implementation indicators. Instructor preparation includes an 8-hour training seminar for instructors who teach the module. The other conditions are technological infrastructure (access to AI tools for every student), teaching and methodological support, an ethical and legal basis (rules for using AI and an ethical checklist) and a motivational environment in which students are not afraid of making mistakes. Instructor preparation is decisive, since all other conditions are implemented through the instructor.

**Conclusion**

The proposed system makes it possible to measure AI-based pedagogical competence not only through what students report about themselves but also through what they actually produce and decide. Its main features are a separate instrument for each component, a common 100-point scale, a level rule that does not allow a weak component to be hidden by a high composite score, and three diagnostic points linked to corrective measures.

**References**

1. Ng D.T.K., Leung J.K.L., Chu S.K.W., Qiao M.S. Conceptualizing AI literacy: An exploratory review // Computers and Education: Artificial Intelligence. – 2021. – Vol. 2. – Article 100041.
2. Mishra P., Koehler M.J. Technological pedagogical content knowledge: A framework for teacher knowledge // Teachers College Record. – 2006. – Vol. 108, No. 6. – P. 1017–1054.
3. Wang B., Rau P.-L.P., Yuan T. Measuring user competence in using artificial intelligence: validity and reliability of artificial intelligence literacy scale // Behaviour & Information Technology. – 2023. – Vol. 42, No. 9. – P. 1324–1337.
4. Likert R. A technique for the measurement of attitudes // Archives of Psychology. – 1932. – No. 140. – P. 1–55.
5. Lawshe C.H. A quantitative approach to content validity // Personnel Psychology. – 1975. – Vol. 28, No. 4. – P. 563–575.