RUDN Journal of Informatization in Education
Вестник РУДН. Серия: Информатизация образования
ISSN 2312-8631 (Print); ISSN 2312-864X (Online)
2026 Vol. 23 No. 1 25–38
http://journals.rudn.ru/nformatization-education

DOI: 10.22363/2312-8631-2026-23-1-25-38
EDN: ZMZGIO
UDC 37.02:004.9

Research article / Научная статья

Multimodal verification of pedagogical features using SAR analysis

Timur M. Bosenko, Albina R. Sadykova

Moscow City University, Moscow, Russian Federation
sadykovaar@mgpu.ru

Abstract. *Problem statement.* Traditional methods of verifying pedagogical interaction features (hereinafter referred to as pedagogical features) are limited to qualitative observation and correlation analysis, which do not allow to establish causal relationships between the actions of the teacher and learning outcomes. The objective of the study is to develop a scientific and methodological approach to the experimental verification of pedagogical features based on SAR analysis and the formation of reference datasets for training SAR agents. *Methodology.* A scientific and methodological approach based on SAR analysis (Socially Assistive Robotics Analysis) is proposed, including the formation of a context-enriched reference dataset of multimodal data for training SAR agents and experimental verification of features through their translation into behavioral modules for automated analysis. *Results.* A comparative analysis with the CLASS video analysis methodology showed a reduction in verification labor costs during scaling by more than 79% (from 808 to 167 person-hours per 500 videos) while increasing the objectivity of the assessment and obtaining causal knowledge instead of correlational knowledge. *Conclusion.* The methodology creates a basis for building evidence-based pedagogy and developing a new generation of intelligent learning systems with automatic recognition of effective pedagogical practices based on SAR agents.

Keywords: evidence-based pedagogy, social robotics, causal inference, expert annotation, reference dataset, CLASS methodology, educational data

Authors' contribution. *Timur M. Bosenko* – conceptualization, methodology, software, validation, formal analysis, data curation, project administration, visualization, writing – original draft. *Albina R. Sadykova* – supervision, writing – review and editing. All authors have read and approved the final version of the manuscript.

Conflicts of interest. The authors declare that there is no conflict of interest.

Article history: received 13 June 2025; revised 10 September 2025; accepted 28 September 2025.

For citation: Bosenko TM, Sadykova AR. Multimodal verification of pedagogical features using SAR analysis. *RUDN Journal of Informatization in Education.* 2026;23(1):25–38. http://doi.org/10.22363/2312-8631-2026-23-1-25-38

Мультимодальная верификация педагогических признаков с использованием SAR-анализа

Т.М. Босенко, А.Р. Садыкова✉

Московский городской педагогический университет, Москва, Российская Федерация
✉sadykovaar@mgpu.ru

Аннотация. *Постановка проблемы.* Традиционные методы верификации признаков педагогических взаимодействий (педагогические признаки) ограничиваются качественным наблюдением и корреляционным анализом, не позволяя установить каузальные связи между действиями педагога и образовательными результатами. Цель исследования – разработка научно-методологического подхода к экспериментальной верификации педагогических признаков на основе SAR-анализа и формирования эталонных датасетов для обучения SAR-агентов. *Методология.* Предложен научно-методологический подход на основе SAR-анализа (Socially Assistive Robotics Analysis), включающий формирование контекстно-обогащенного эталонного датасета мультимодальных данных для обучения SAR-агентов и экспериментальную верификацию признаков через их трансляцию в поведенческие модули для автоматизированного анализа. *Результаты.* Сравнительный анализ с методологией видеоанализа CLASS показал снижение трудозатрат на верификацию при масштабировании более чем на 79 % (с 808 до 167 человеко-часов на 500 видео) при повышении объективности оценки и получении каузального знания вместо корреляционного. *Заключение.* Методология создает основу для построения доказательной педагогики и разработки интеллектуальных обучающих систем нового поколения с автоматическим распознаванием эффективных педагогических практик на примере SAR-агентов.

Ключевые слова: доказательная педагогика, социальная робототехника, причинно-следственный вывод, экспертная аннотация, эталонный датасет, методология CLASS, образовательные данные

Вклад авторов. *Т.М. Босенко* – концепция и дизайн исследования, разработка методологии, программного обеспечения, верификация данных, анализ, синтез, администрирование и визуализация данных, написание рукописи. *А.Р. Садыкова* – надзор и руководство за планированием и выполнением исследования, редактирование рукописи. Все авторы прочли и одобрили окончательную версию рукописи.

Заявление о конфликте интересов. Авторы заявляют об отсутствии конфликта интересов.

История статьи: поступила в редакцию 13 июня 2025 г.; доработана после рецензирования 10 сентября 2025 г.; принята к публикации 28 сентября 2025 г.

Для цитирования: *Bosenko T.M., Sadykova A.R.* Multimodal verification of pedagogical features using SAR analysis // Вестник Российского университета дружбы народов. Серия: Информатизация образования. 2026. Т. 23. № 1. С. 25–38. http://doi.org/10.22363/2312-8631-2026-23-1-25-38

Problem statement. Modern pedagogical science faces the need to form an evidence base for the effectiveness of pedagogical practices [1]. There is a significant methodological gap between the accumulation of empirical data on pedagogical *features* (specific manifestations of pedagogical interactions at the behavioral level) and the establishment of causal links between the actions of the teacher and learning outcomes [2]. Specific, measurable behavioral features of teachers (gestures, intonation, phrases) are traditionally identified through qualitative observation, which does not allow correlations to be separated from causal relationships.

Pedagogical science is dominated by an expert approach to evaluating the effectiveness of pedagogical actions. Studies demonstrate the priority of qualitative methods based on reflective analysis and expert evaluations [3]. However, the subjectivity of interpretation and the inability to reproduce results limit the applicability of such methods for the formation of evidence-based pedagogy.

International practice demonstrates more systematic approaches to verification. The CLASS (Classroom Assessment Scoring System) protocol, developed by Pianta et al. [4], has become the standard for observational evaluation of the quality of classroom interactions. However, even CLASS remains a correlational tool: it records the relationship between pedagogical behavior and outcomes, but does not prove causality.

The development of multimodal learning analytics has opened up new opportunities for automated *feature* analysis [5; 6]. Blikstein [7] showed that joint processing of verbal, nonverbal, and paraverbal (voice features) data allows identifying patterns of effective learning. Prieto et al. [8] developed a methodology for the automatic extraction of orchestration graphs from wearable sensor data. Worsley and D'Angelo [9] identified key areas for the development of MMLA in a systematic review, emphasizing the need to move from descriptive to causal analytics.

Schlotterbeck et al. [10] demonstrated a cost-effective approach to detecting pedagogical practices through spectral analysis of classroom audio, but their method is also limited to identifying correlations. Martinez-Maldonado et al. [11] emphasize the critical importance of contextual interpretation of automatically detected patterns, which remains an unresolved issue in most MMLA studies.

A critical analysis of current approaches reveals a fundamental limitation: all existing methodologies – from expert observation to advanced multimodal analytics – only allow for the identification of correlations between the *features* under consideration and learning outcomes. Controlled experiments are required to prove causality [2], but conducting such experiments in real educational settings is fraught with ethical and organizational constraints.

An alternative solution could be the use of social assistive robotics (SAR). Belpaeme et al. [12] showed in their review of the use of social robots in education that SAR systems can reproduce actions in controlled conditions. Scassellati et al. [13] substantiated the potential of SAR for creating standardized experimental protocols. However, there are no methodologies in the literature for using SAR specifically to verify *features* identified through multimodal analysis of real-world practice.

The objective of the study is to develop and substantiate a scientific and methodological approach to the experimental verification of pedagogical *features*, which is intended to overcome the limitations of correlation methods through the integration of multimodal analysis and SAR technologies.

Research tasks: to develop a procedure for forming a context-enriched dataset of pedagogical *features* based on multimodal analysis of educational interactions; to create a methodology for translating the identified features into behavioral modules of a social assistance robot for experimental verification; to conduct a comparative analysis of the effectiveness of the proposed approach with existing methods of verifying pedagogical practices; to justify the prospects for scaling the methodology through agent systems.

Methodology. The methodological basis of the study is a comprehensive approach that combines methods of multimodal analysis of educational data, experimental verification of pedagogical features, and comparative evaluation of the effectiveness of methodologies.

Multimodal analysis of pedagogical *features* was based on expert annotation of video data from training sessions (50 video recordings, ~18.3 hours) using Label Studio tools and the author's Pedagogical Pattern Infrastructure (PPI)[1] platform [14; 15]. The procedure included the identification of pedagogical events (specific pedagogically significant episodes of interaction) and their classification into two main categories of features with the recording of verbal, nonverbal, and paraverbal components. To assess the reliability of the labeling, a statistical method for calculating inter-expert agreement (Cohen's kappa coefficient) was used with the participation of 12 expert methodologists divided into two teams, which made it possible to quantitatively assess the subjectivity of expert assessments ($k = 0.68$, 95% CI: 0.61–0.75) [16].

SAR analysis (Socially Assistive Robotics Analysis) was implemented through a three-step procedure of translating the identified features into SAR agent training parameters, followed by an experimental evaluation of their causal impact on learning outcomes. This method ensured the transition from correlation analysis to proof of causal relationships between pedagogical *features* and learning outcomes.

A comparative analysis of the effectiveness of the proposed approach was conducted in relation to the CLASS (Classroom Assessment Scoring System) methodology [4] based on quantitative indicators of labor costs, assessment objectivity, and scalability of the verification process for different volumes of analyzed material (50, 100, 200, 500 videos).

Architecture of the scientific and methodological approach. The proposed methodology consists of a sequence of seven logically connected stages (Figure 1), providing a complete research cycle from hypothesis to experimental verification.

[1] The PPI platform was developed as part of a scientific collaboration between the Russian-Chinese Center for Artificial Intelligence in Education at Moscow City University (MCU, Russia) and Central China Normal University (CCNU, China). Available from: https://github.com/BosenkoTM/ppi_mgpu (accessed: 10.12.2025).

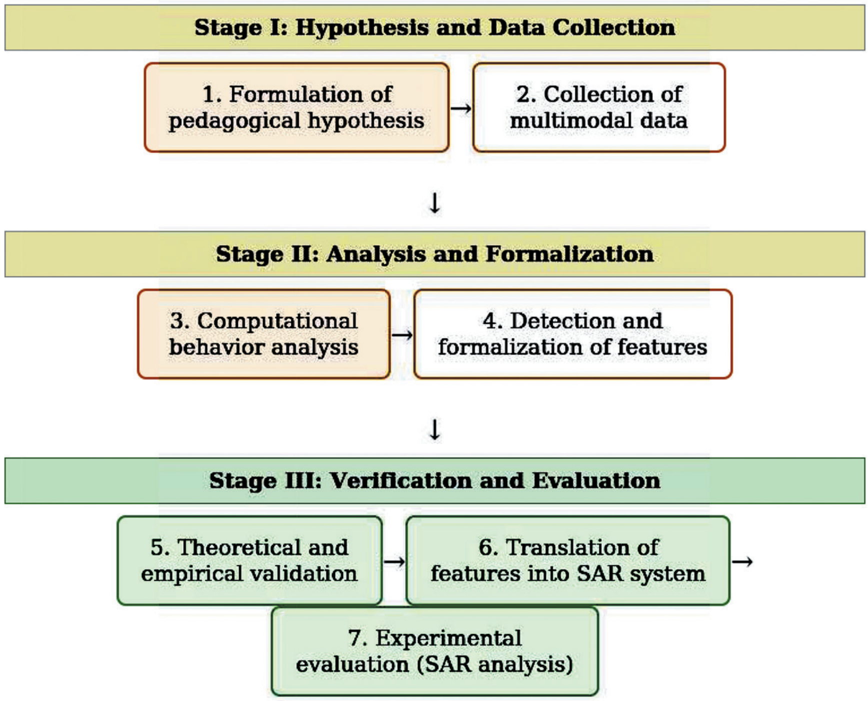

Figure 1. Architecture of the scientific and methodological approach to *feature* verification

Source: created by Timur M. Bosenko, Albina R. Sadykova.

Pedagogical hypothesis formulation. The following illustrative hypothesis was selected for testing the methodology: *Taking into account the pedagogical features that constitute growth-oriented feedback has a more significant positive impact on student engagement and academic self-sufficiency than evaluative feedback such as 'right/wrong'.*

This hypothesis was chosen based on Dweck's [17] theoretical propositions about the influence of feedback type on student motivation and the availability of preliminary correlation data [18].

Collection and analysis of multimodal data. To create a reference dataset, author-developed methodology based on expert annotation in Label Studio was used, followed by processing in the PPI web platform. Two teams of experts (6 people each) participated in the experiment, analyzing two categories of pedagogical *features*: 'Pedagogical Actions' and 'Communicative Modalities'. Each team worked in groups of three to ensure the reliability of the annotation.

A unified data collection template (Table 1) ensures that not only the observed features are recorded, but also the critically important pedagogical context.

For the category 'Pedagogical actions', experts identified the following subcategories: asking open/closed questions, explanations/clarifications, evaluating answers (reward/constructive criticism), organizing group work, redirecting attention, providing feedback, and using scaffolding.

For the category 'Communicative modalities', the following were recorded: verbal, nonverbal, and paraverbal components of interaction.

Table 1. **Template for data collection in the reference dataset**

Field	Source	Description
Event_ID	System	Unique event identifier
Subject	Expert	Action initiator (Teacher, Student, Group of students)
Category	Expert	Major category (Pedagogical actions, Communicative modalities)
Subcategory	Expert	Feature subcategory (multiple choice)
Action_description	Expert	Objective description of the action (what is happening)
Context and comments	Expert	**Key field:** context BEFORE; reaction AFTER; pedagogical meaning

Source: compiled by Timur M. Bosenko, Albina R. Sadykova.

Theoretical and empirical validation. The identified signs undergo two-level validation: compliance with pedagogical theory and the presence of a correlation with measurable learning outcomes (engagement, academic performance, self-sufficiency).

Translation of signs into the SAR system. The key methodological contribution of this study is the development of a three-step procedure for “programming” a pedagogical feature for a SAR agent.

1. *Feature decomposition.* The annotated feature from the *action_description* field is decomposed into atomic multimodal components:

- Verbal: specific phrases and their sequence;
- Nonverbal: gestures, postures, position in space;
- Paraverbal: voice characteristics (tempo, volume, intonation contours).

2. *Mapping to SAR agent capabilities.* Each component is mapped to specific parameters for automatic recognition:

- Verbal → features for NLP models (tokens, embeddings, patterns);
- Nonverbal → features for computer vision (key point coordinates, motion trajectories);
- Paraverbal → features for audio analysis (low-frequency cepstral coefficients, pitch.

3. *Generation of training examples.* Atomic components are combined into labeled examples for training the SAR agent with temporal synchronization of modalities accurate to 100 ms.

4. *Experimental evaluation (SAR analysis).* The trained SAR agent undergoes a two-level evaluation:

- Technical evaluation – accuracy of feature recognition on a test sample (metrics: Precision, Recall, F1-score);
- Pedagogical evaluation – a controlled experiment to verify the causal relationship between the recognized feature and learning outcomes.

Statistical analysis of differences between conditions allows to draw a causal conclusion about the influence of the feature.

Results and discussion. To evaluate the effectiveness of the proposed approach, a comparative analysis was conducted with the traditional CLASS (Classroom Assessment Scoring System) observational methodology [4], which represents the standard for observational assessment of *features* (Table 2).

Table 2. **Comparative analysis of feature verification methodologies**

Criterion	CLASS (observational)	Proposed SAR approach
Type of knowledge obtained	Correlational (relationship between a feature and a result)	Causal (causal influence of a feature)
Objectivity of evaluation	Mean (depends on the observer's training, ICC 0.60-0.75)	High (based on statistics from a controlled experiment)
Ability to isolate the influence of a feature	Low (many confounders in a real classroom)	High (controlled conditions of the SAR experiment)
Scalability	Low (requires training of observers)	High (automatic reproduction of the experiment)
Ethical restrictions	High (impossibility of manipulation in a real classroom)	Minimal (experiments with a robot in controlled conditions)

Source: compiled by Timur M. Bosenko, Albina R. Sadykova.

The results of the comparison demonstrate the fundamental advantage of the SAR approach: the transition from establishing correlations to proving causality. Traditional CLASS methodology allows to assert: "Teachers who use feature X have students with higher engagement". SAR analysis allows to state: "Feature X causes an increase in student engagement".

Architectural differences between approaches. The methodological difference between correlational and causal verification is illustrated in Figures 2 and 3.

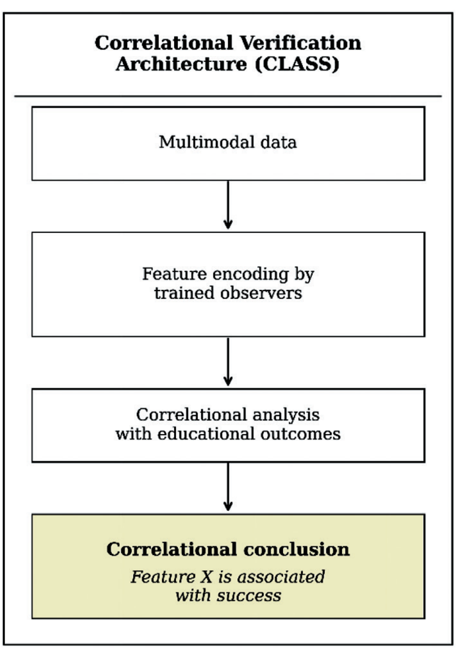

Figure 2. Verification without SAR analysis (correlation approach)

Source: created by Timur M. Bosenko, Albina R. Sadykova.

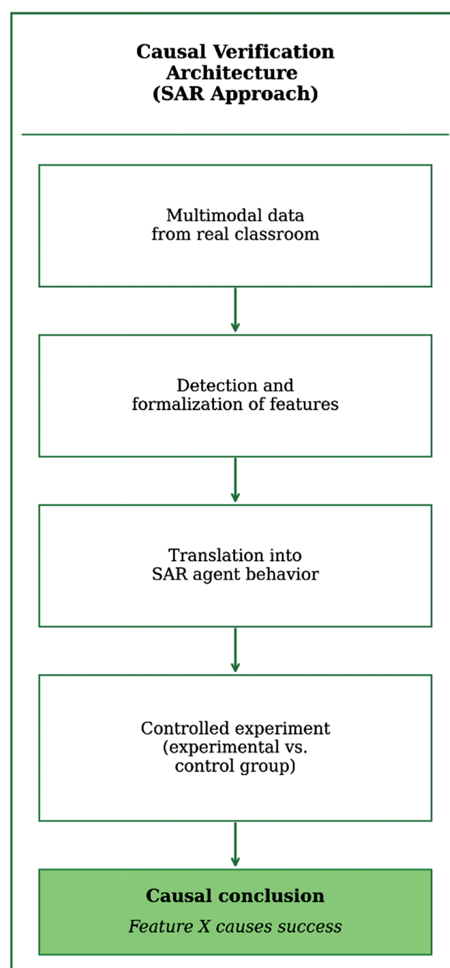

Figure 3. Verification using SAR analysis (causal approach)

Source: created by Timur M. Bosenko, Albina R. Sadykova.

The problem of subjectivity at the stage of forming a reference dataset. A critical analysis of the proposed methodology reveals a key problem of subjectivity in expert evaluation at the stage of forming a reference dataset (stages 2–5). Despite the use of a structured template, the fields category, subcategory, and especially context and comments are filled in by an expert methodologist, which introduces an element of interpretation.

Inter-expert agreement evaluation procedure. To quantitatively evaluate the reliability of the annotation, a pilot agreement study was conducted with 12 expert methodologists divided into two teams. A pool of 50 video recordings of lessons lasting from 15 to 30 minutes (total duration: ~18.5 hours, subject area: English language, grades 5–7) was analyzed by groups of three independent experts. Each trio labeled pedagogical events (an average of 8–10 events per video) according to two main categories with corresponding subcategories. Total amount: 450 pedagogical events. Agreement in the classification of events by category is presented in Table 3.

Table 3. **Inter-expert agreement on categories of pedagogical features**

Category	*k* (mean)	95% CI	Interpretation*
Pedagogical actions	0.71	0.65–0.77	Significant
Communicative modalities	0.65	0.58–0.72	Significant
Overall consistency	0.68	0.61–0.75	Significant

Note: interpretation according to Landis & Koch [16]: $k < 0.00$ poor agreement; 0.00–0.20 = weak; 0.21–0.40 = satisfactory; 0.41–0.60 = moderate; 0.61–0.80 = significant; 0.81–1.00 = almost complete agreement.

Source: compiled by Timur M. Bosenko.

The lowest agreement was observed in the subcategory of nonverbal features ($k = 0.58$), which can be explained by the subjectivity of interpreting gestures and facial expressions.

Cohen's kappa coefficient was calculated using the formula [16]:

$$k = (P_o - P_e) / (1 - P_e),$$

where P_o is the observed proportion of agreement between experts, P_e is the expected proportion of random agreement. The confidence interval (95% CI) was calculated using the bootstrap method (1000 iterations on a sample of 450 events).

Comparison with the performance of automated models. For context, inter-expert agreement was compared with the performance of modern automatic *feature* recognition models (Table 4).

It is clear that automatic models based on verbal features (F1 = 0.52–0.67) are inferior to experts (k = 0.68–0.75), and the gap is even more significant for nonverbal features (F1 = 0.38–0.51). Full automation of labeling is not feasible at the current stage of technological development.

Solving the problem of subjectivity through an agent system based on SAR analysis. The SAR analysis methodology offers a solution to the problem of subjectivity through a mechanism of experimental feature filtering.

Experimental verification mechanism. The process functions as shown in Figure 4.

Table 4. **Comparison of the performance of experts and automated models**

Approach	Verbal features	Nonverbal features	Multimodal
Experts (MCU)	$k = 0.73^*$	$k = 0.58^*$	$k = 0.68$
Demszky et al. [19]	F1 = 0.52–0.67	–	–
Suresh et al. [20]	F1 = 0.61	–	–
Prieto et al. [8]	–	F1 = 0.38–0.51	–
Bosch et al. [21]	F1 = 0.71**	–	–

Notes:

* *k* for verbal = mean across the categories ‘Engagement’, ‘Cognitive Support’, ‘Feedback’; for nonverbal = ‘Recognition of Emotional Signals’.

** Only for classification of feedback types on a dataset >1000 hours.

Source: compiled by Timur M. Bosenko.

Subjective labeling. Experts form a dataset with inevitable subjectivity ($k \approx 0.68$). Some features receive high agreement (priority 1), while others receive low agreement (priority 2).

SAR verification. Each feature is tested in a controlled experiment. Approximately 60–70% of features show a statistically significant causal effect, while the remaining 30–40% are rejected regardless of the expert evaluation.

Critical result. Both false positives (features unanimously approved by experts but ineffective in the experiment) and false negatives (features with low expert agreement that showed a strong effect) are detected.

Knowledge base formation. Verified features form an objective basis for training neural network models for automatic recognition.

Thus, SAR analysis performs the function of an “objectivity filter”: objective knowledge about causally effective features is extracted from a subjectively labeled dataset through an objective experiment. Mathematical justification shows a 25–33% increase in the accuracy of identifying effective features compared to expert evaluation alone.

Economic efficiency evaluation. Labor costs were compared using video materials from a pilot study conducted as part of scientific cooperation between the Russian-Chinese Center for Artificial Intelligence: 50 video recordings of training sessions lasting 15–30 minutes each (total duration ~18.3 hours, subject area: English language, grades 5–7).

For the SAR approach, the entire chain was analyzed: *expert annotation (12 experts working in parallel in groups of three) → dataset formation → development of SAR scenarios for 12 features → conducting experiments.*

For the CLASS approach: *training of observers → video coding by two observers → agreement on evaluations.*

Results for the initial analysis of 50 videos: CLASS approach – 124 person-hours; SAR approach – 113 person-hours; Savings – 9% (costs are comparable due to the initial investment in SAR script development).

The key advantage of the SAR approach becomes apparent when scaling (Figure 5). After creating a library of SAR scripts (one-time costs), each new verification requires only the time to conduct experiments, which does not depend on the video length. The CLASS approach requires a linear increase in labor costs: each new video sample requires re-encoding.

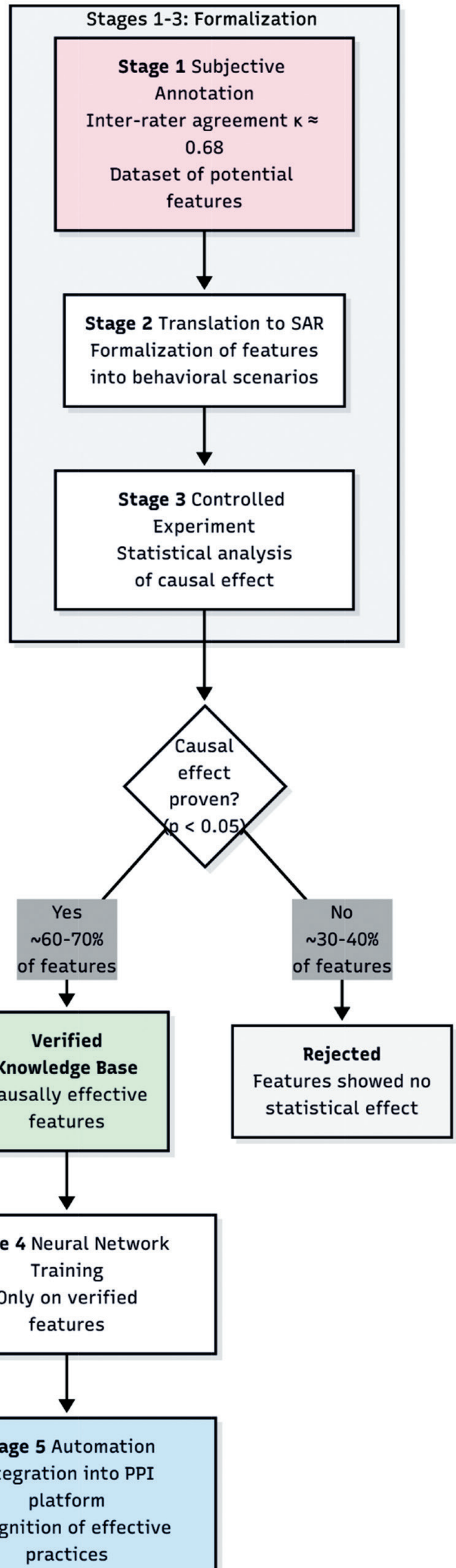

Figure 4. Mechanism for overcoming subjectivity through SAR verification

Source: created by Timur M. Bosenko.

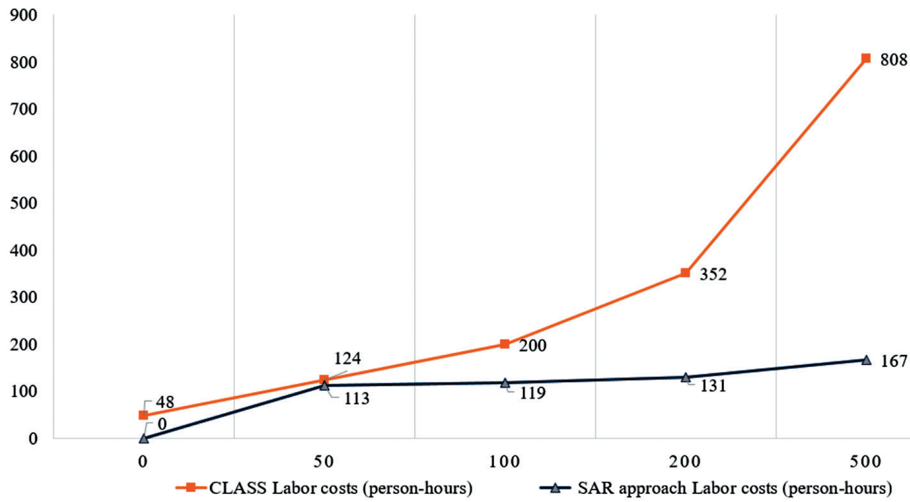

Figure 5. Dependence of labor costs on analyzed material amount

Source: created by Timur M. Bosenko.

Limitations and prospects for further development. The current version of the methodology has a number of limitations that determine the directions for further research. The first limitation is related to the need to form sufficiently large reference datasets for training SAR agents: to achieve high accuracy in *feature* recognition, at least 500 hours of labeled video material is required for each feature category, which requires significant expert resources at the initial stage. The second limitation is due to the technical capabilities of modern computer vision and natural language processing systems: automatic recognition of subtle nonverbal features (micro-expressions, complex spatial configurations of interactions) and context-dependent verbal patterns currently achieves an accuracy of F1 = 0.58–0.71, which requires additional expert validation of the results of SAR agents. The third limitation concerns the need to adapt trained SAR agents to different educational contexts: models trained on data from a single subject area or age group [22] may demonstrate reduced performance when transferred to other contexts without fine-tuning [23].

The prospects for further development of the methodology are related to the automation of the verification process and the scaling of SAR agent applications. The first direction involves the creation of self-learning SAR agents capable of automatically replenishing reference datasets through active learning: the agent identifies cases with low prediction confidence and requests expert annotation only for them, which can reduce the need for expert resources by 60–70%. The second direction involves the development of multitasking SAR agents capable of simultaneously recognizing several categories of *features* and their interrelationships, which will allow analyzing the causal impact on learning outcomes. The third direction is related to the integration of SAR agents into real-time systems to provide teachers with immediate feedback during lessons, which requires the development of effective neural network architectures capable of processing video streams with a delay of no more than 200 ms, and can be used in both intelligent learning systems and adaptive robotic platforms for education.

Conclusion. The proposed scientific and methodological approach based on SAR analysis represents a paradigm shift in the verification of pedagogical *features*: from passive observation of correlations to active experimental proof of causality. The methodology solves a fundamental problem of traditional approaches – the inability to isolate the influence of individual features in the complex environment of a real classroom.

The methodological contribution of the work is the development of a three-step procedure for translating multimodal pedagogical features into SAR agent training parameters, enabling automated recognition and verification of these *features*. The empirical contribution consists in demonstrating that comparative analysis showed a reduction in verification labor costs during scaling by more than 79% (from 808 to 167 person-hours per 500 videos) when switching from observational to experimental methods and increasing the objectivity of the assessment from ICC 0.60–0.75 to a statistically controlled experiment. The practical contribution is expressed in the creation of a context-enriched reference dataset template that allows not only observable features to be recorded, but also the critically important pedagogical context for training SAR agents. The conceptual contribution lies in the justification of a mechanism for overcoming the subjectivity of expert annotation through multiple SAR verification, which forms an objective knowledge base for training automated *feature* recognition systems.

It is important to emphasize the systemic nature of the proposed approach. SAR analysis is not an end in itself, but a critical stage in a larger architecture for building evidence-based pedagogy. The knowledge base created through experimental verification is the foundation for training SAR agents capable of automatically recognizing effective pedagogical practices. The ultimate goal of the research program is to integrate trained SAR agents into the PPI platform for automatic analysis of pedagogical activities in real time. Such a system will be able to instantly recognize proven effective techniques in teacher behavior, automating the work of expert methodologists and providing teachers with timely, objective, and scientifically based feedback for professional growth. Moreover, trained SAR agents can be integrated into robotic educational platforms, expanding the possibilities of adaptive learning.

The proposed methodology paves the way for the digital transformation of teaching theory and methodology, creating a technological basis for the transition of pedagogy from art to evidence-based science and the formation of a new generation of intelligent learning systems based on experimentally verified principles of effective learning and automated recognition of pedagogical *features* through SAR agents.

References / Список литературы

[1] Zagvyazinsky VI, Atakhanov R. *Methodology and Methods of Psychological and Pedagogical Research: Textbook for Students of Higher Professional Educational Institutions*. 7th ed., stereotyped. Moscow: Publishing Center Akademiya; 2012. 208 p. (In Russ.)
Загвязинский В.И., Атаханов Р. Методология и методы психолого-педагогического исследования : учеб. пособие для студ. учреждений высш. проф. образования. 7-е изд., М. : Издательский центр «Академия», 2012. 208 с.

[2] Pearl J, Mackenzie D. *The Book of Why: The New Science of Cause and Effect.* New York: Basic Books; 2018. 432 p.

[3] Korzhuev AV, Babaskin VS, Sadykova AR. Reflexion as a strategy in the process of lifelong pedagogical education. *Vysshee Obrazovanie v Rossii = Higher Education in Russia*. 2013;(7):77–80. (In Russ.) EDN: QZDWVR

Коржуев А.В., Бабаскин В.С., Садыкова А.Р. Педагогическая рефлексия как компонент непрерывного образования преподавателя высшей школы // Высшее образование в России. 2013. № 7. С. 77–80. EDN: QZDWVR

[4] Pianta RC, La Paro KM, Hamre BK. Classroom Assessment Scoring System™: Manual K-3. Baltimore: Brookes; 2008. 116 p.

[5] Slavin RE. Evidence-based education policies: transforming educational practice and research. *Educational Researcher*. 2002;31(7):15–21. https://doi.org/10.3102/0013189X031007015

[6] Blikstein P, Worsley M. Multimodal learning analytics and education data mining: using computational technologies to measure complex learning tasks. *Journal of Learning Analytics*. 2016;3(2):220–238. https://doi.org/10.18608/jla.2016.32.11

[7] Ochoa X. Handbook of learning analytics. In: Lang C, Siemens G, Wise A, Gašević D. (eds.). *The Handbook of Learning Analytics.* Beaumont, Alberta; 2017. p. 129–141. https://doi.org/10.18608/hla17.011

[8] Prieto LP, Sharma K, Kidzinski L, Rodríguez-Triana MJ, Dillenbourg P. Multimodal teaching analytics: automated extraction of orchestration graphs from wearable sensor data. *Journal of Computer Assisted Learning*. 2018;34(2):193–203. https://doi.org/10.1111/jcal.12232

[9] Worsley M, Martinez-Maldonado R, D'Angelo C. A new era in multimodal learning analytics: twelve core commitments to ground and grow MMLA. *Journal of Learning Analytics*. 2021;8(3):10–27. https://doi.org/10.18608/jla.2021.7361

[10] Schlotterbeck D, Uribe P, Araya R, Jiménez A, Caballero D. What classroom audio tells about teaching: a cost-effective approach for detection of teaching practices using spectral audio features. In: *LAK21: 11th International Learning Analytics and Knowledge Conference, 12–16 April 2021, Irvine, CA, USA.* New York: Association for Computing Machinery; 2021. p. 132–140. https://doi.org/10.1145/3448139.3448152

[11] Martinez-Maldonado R, Yan L, Deppeler J, Phillips M, Gašević D. Classroom analytics: telling stories about learning spaces using sensor data. In: Gil E, Mor Y, Dimitriadis Y, Köppe C. (eds.). *Hybrid Learning Spaces.* Cham: Springer; 2022. p. 185–203.

[12] Belpaeme T, Kennedy J, Ramachandran A, Scassellati B, Tanaka F. Social robots for education: a review. *Science Robotics*. 2018;3(21):eaat5954. https://doi.org/10.1126/scirobotics.aat5954

[13] Matarić M.J., Scassellati B. Socially assistive robotics. In: Siciliano B., Khatib O. (eds.). *Springer Handbook of Robotics*. 2nd ed. Cham: Springer; 2016. p. 1973–1994. https://doi.org/10.1007/978-3-319-32552-1_73

[14] Bosenko TM, Frolov YuV. Development of an adaptive learning platform architecture based on the institute of digital education of MCU. In: Barannikov KA, Bosenko TM, Voronkov AA, et al. *Modern Digital Didactics*. Moscow: Intellect-Center Publishing House; 2024. p. 97–102. (In Russ.) EDN: ASGYYE

Босенко Т.М., Фролов Ю.В. Разработка архитектуры платформы адаптивного обучения на базе института цифрового образования МГПУ // Современная цифровая дидактика : монография / К.А. Баранников, Т.М. Босенко, А.А. Воронков [и др.]. М. : Издательство Интеллект-Центр, 2024. С. 97–102. EDN: ASGYYE

[15] Bosenko TM. Integration of multimodal technologies into the educational process on the example of developing machine learning model on the Apache Spark platform. *MCU Journal of Informatics and Informatization of Education*. 2025;(1):80–89. (In Russ.) https://doi.org/10.24412/2072-9014-2025-171-80-89

Босенко Т.М. Интеграция мультимодальных технологий в образовательный процесс на примере разработки модели машинного обучения на платформе Apache Spark // Вестник МГПУ. Серия: Информатика и информатизация образования. 2025. № 1(71). С. 80–89. https://doi.org/10.24412/2072-9014-2025-171-80-89

[16] Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159–174. https://doi.org/10.2307/2529310

[17] Dweck CS. *Mindset: The New Psychology of Success.* New York: Random House; 2006. 276 p.

[18] Hattie J, Timperley H. The power of feedback. *Review of Educational Research*. 2007;77(1):81–112. https://doi.org/10.3102/003465430298487

[19] Demszky D, Liu J, Mancenido Z, et al. Measuring conversational uptake: a case study on student-teacher interactions. In: Zong C, Xia F, Li W, Navigli R. *The 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing.* Stroudsburg: Association for Computational Linguistics (ACL); 2021. p. 1638–1653. https://doi.org/10.18653/v1/2021.acl-long.130

[20] Suresh A, Jacobs J, Lai V, Tan C, Ward W, Martin JH, Sumner T. Using transformers to provide teachers with personalized feedback on their classroom discourse: the talk moves application. *ArXiv Preprint*; 2021. https://doi.org/10.48550/arXiv.2105.07949

[21] Bosch N, D'Mello S, Mills C. What Emotions Do novices experience during their first computer programming learning session? In: Lane CH, Yacef K, Mostow J, Pavlik P. (eds.). *Artificial Intelligence in Education, 16th International Conference, AIED 2013, 9–13 July 2013, Memphis, TN, USA.* Berlin: Springer; 2013. p. 11–20. https://doi.org/10.1007/978-3-642-39112-5_2

[22] Kory-Westlund JM, Breazeal C. Assessing children's perceptions and acceptance of a social robot. In: *IDC'19: Interaction Design and Children,* 12–15 June 2019, Boise, ID, USA. Boise; 2019. p. 38–50. https://dl.acm.org/doi/pdf/10.1145/3311927.3323143

[23] Tanaka F, Matsuzoe S. Children teach a care-receiving robot to promote their learning: field experiments in a classroom for vocabulary learning. *Journal of Human-Robot Interaction*. 2012;1(1):78–95. https://doi.org/10.5898/JHRI.1.1.Tanaka

Bio notes:

Timur M. Bosenko, Associate Professor at the Department of IT, Management and Technology, Institute of Digital Education, Moscow City University, 28 Sheremetyevskaya St, Moscow, 129594, Russian Federation. ORCID: 0000-0002-5375-096X; SPIN-code: 7117-2458. E-mail: bosenkotm@mgpu.ru

Albina R. Sadykova, Professor at the Department of Informatics, Management and Technology, Institute of Digital Education, Moscow City University, 28 Sheremetyevskaya St, Moscow, 129594, Russian Federation. ORCID: 0000-0002-1413-200X; SPIN-code: 7107-7576. E-mail: sadykovaar@mgpu.ru

Сведения об авторах:

Босенко Тимур Муртазович, доцент департамента информатики, управления и технологий, Институт цифрового образования, Московский городской педагогический университет, Российская Федерация, 129594, Москва, ул. Шереметьевская, д. 28. ORCID: 0000-0002-5375-096X; SPIN-код: 7117-2458. E-mail: bosenkotm@mgpu.ru

Садыкова Альбина Рифовна, профессор департамента информатики, управления и технологий, Институт цифрового образования, Московский городской педагогический университет, Российская Федерация, 129594, Москва, ул. Шереметьевская, д. 28. ORCID: 0000-0002-1413-200X; SPIN-код: 7107-7576. E-mail: sadykovaar@mgpu.ru