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

DESIGNING INTEGRATED INTELLIGENT MONITORING ALGORITHMS FOR MANAGING DISTANCE LEARNING PROCESSES

Authors:
Sapaye Sapaboy Xikmatovich,
Yuldoshov Habibulla Alijon ugli,
Scientific Advisor: PhD.
Ashurov Jasur DJo'rayevich
Organization: Asia International University
Field: Information Technologies in Education

Abstract.: The widespread adoption of distance learning has created the need for real-time quality monitoring of the educational process. Traditional monitoring methods often rely on a single indicator (grades or the number of system logins) and detect negative changes in student learning activity with a delay. In this article, integrated intelligent monitoring (IIM) algorithms were designed for managing distance learning processes by combining various data streams (activity logs, test results, video and forum activity) into a unified adaptive system. The proposed model generates an overall risk score through weighted integration of an engagement index, a predictive model, and an anomaly detection module, and stabilizes it over time using exponential smoothing. Experimental evaluation demonstrated that the integrated approach achieved an accuracy of 0.90 and an AUC-ROC of 0.94 in identifying at-risk students, significantly outperforming the single-indicator baseline method and providing early warning by mid-semester.

Keywords: distance learning, intelligent monitoring, learning analytics, machine learning, predictive model, anomaly detection, LMS, risk score, early warning system, integrated algorithm.

Introduction

In recent years, driven by the rapid development of digital educational technologies and the impact of global circumstances, distance and blended learning formats have become an integral part of the higher education system. While the transition of the educational process to an electronic environment has expanded learning opportunities, it has also led to a weakening of direct communication between instructors and students. As a result, monitoring student learning activity, motivation, and knowledge acquisition has become more challenging.

In a traditional classroom setting, an instructor could assess a student's condition through direct observation. In a distance environment, however, this task must be performed based on digital traces collected by Learning Management Systems (LMS)—system logins, content access, assignment completion, video lecture viewing, and forum discussions. In practice, however, these data are often stored in a fragmented manner and analyzed separately, which hinders forming a holistic view of the student's overall condition.

The main shortcoming of existing monitoring systems is their frequent reliance on a single indicator. For example, a system based solely on midterm grades cannot detect a decline in student activity until grades are assigned; a system that only tracks login frequency cannot identify a situation where a student logs in but fails to learn anything. Such delayed and one-sided monitoring proves ineffective in preventing student dropout.

The aim of this article is to design integrated intelligent monitoring (IIM) algorithms that combine various data streams into a unified, adaptive, and interpretable system for managing distance learning processes. The scientific novelty and practical contributions of the research are as follows:

1) an integrated monitoring mathematical model combining activity, learning performance, and behavioral indicators through weighted integration was developed;

2) a modular (four-layer) system architecture covering the process from data collection to intervention was proposed;

3) an adaptive weight updating mechanism responsive to instructor decisions was implemented;

4) the proposed algorithm prototype was experimentally evaluated and compared with baseline methods.

Literature Review

Research on the analysis of e-learning data has developed primarily in two interrelated directions: learning analytics and educational data mining (EDM). Siemens defined learning analytics as the field of optimizing learning through measuring, collecting, and analyzing data about learners and the contexts in which they learn [1]. Romero and Ventura systematically summarized the developmental stages and main methods of EDM, demonstrating the application of classification, clustering, and association rules in education [2, 4].

Baker and Yacef outlined the prospects of EDM in student modeling and prediction [3]. Greller and Drachsler proposed a general conceptual framework for learning analytics, integrating data sources, objectives, and ethical constraints within a unified framework [7]. These works laid the theoretical foundation for monitoring systems, although most of them focused on analyzing a single data type.

A number of applied studies exist on predicting student outcomes based on LMS data. Macfadyen and Dawson demonstrated the feasibility of developing an “early warning system” for educators using activity indicators from the Moodle platform [6]. Conijn et al. analyzed models for predicting student success from LMS data across various courses, showing that prediction accuracy depends on course design [5]. Hussain et al. predicted student engagement levels using machine learning methods [8]. For large-scale research, Kuzilek et al. introduced the open OULAD dataset, which is widely used in this field [9].

Methodologically, monitoring algorithms rely on statistical and machine learning methods. Logistic regression, decision trees, and ensemble methods are widely used in classification tasks [10, 12]; the survey by Chandola et al. classified the main approaches to anomaly detection [11]; and deep learning methods show promise in sequential data modeling [13].

The analysis of existing literature reveals that most studies focus either solely on prediction or solely on activity measurement. Holistic solutions that combine various data streams into a unified, adaptive, and interpretable risk score and are self-adjusting based on instructor feedback have not been sufficiently developed. This article is aimed precisely at filling this gap.

System Architecture

The proposed integrated intelligent monitoring system was designed on a modular principle as a four-layer architecture (Figure 1). Each layer performs a specific function and transmits data to the upper layer through a standardized interface; this division enables system scalability, the addition of new data sources, and independent updating of individual modules.

1-rasm. Integrallashgan intellektual monitoring tizimining qatlamli arxitekturasi

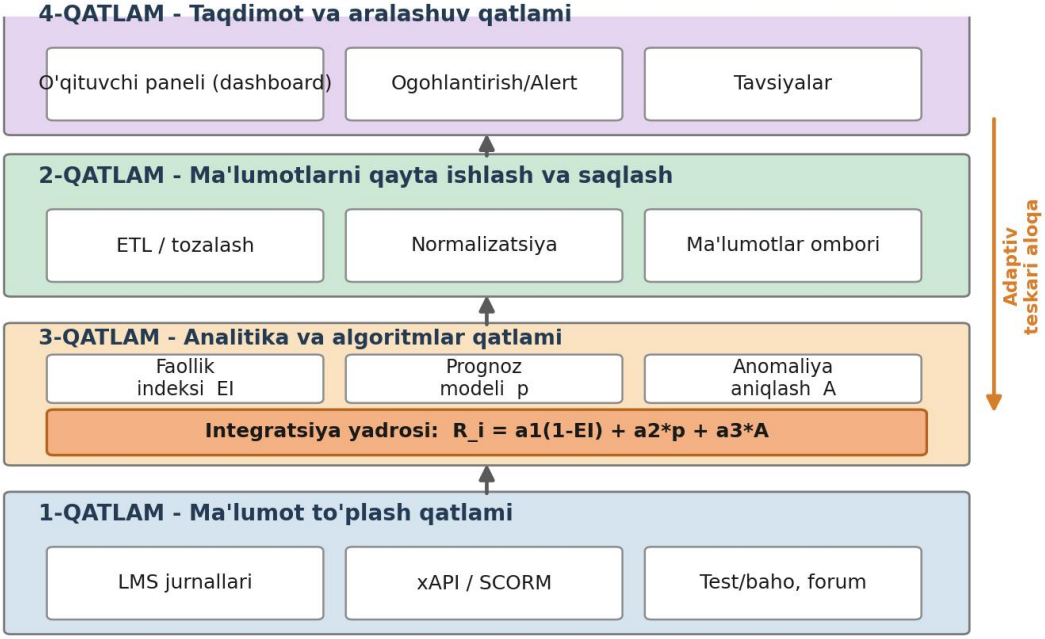


Figure 1. Layered architecture of the integrated intelligent monitoring system

The data collection layer, located at the lowest level, gathers digital traces of learning activity from various sources: LMS server logs (logins/logouts, page views), learning events in xAPI and SCORM standards, test and assignment scores, video lecture viewing statistics, as well as forum and chat interaction indicators are collected at set intervals—daily or weekly. These traces are transmitted to the upper layer, namely the data processing and storage tier. There, an ETL (Extract-Transform-Load) process is carried out: raw data are cleaned, duplicate and incomplete records are eliminated, indicators from various sources are merged by a unique student identifier, normalized using the min-max method, and stored in a data warehouse in time-series format.

The core of the system is the analytics and algorithms layer. It consists of three parallel analytical blocks: an engagement index calculation block, a predictive model for forecasting student outcomes, and an anomaly detection block that identifies abnormal changes in behavior. The outputs of these blocks are combined through weighted integration in the integration core and converted into an overall risk score—it is precisely this integration process that gives the system its “integrated” characteristic. The topmost presentation and intervention layer visualizes the analysis results through an interactive dashboard, sends automatic alerts when the risk level exceeds a threshold, and generates intervention recommendations. The instructor’s decision is fed back to the algorithms layer as feedback and is used for adaptive model tuning.

Intelligent Monitoring Algorithms and Mathematical Model

In this section, the integrated monitoring process is presented in formal mathematical form; the overall processing pipeline of the model is shown in Figure 2. For each student i (i = 1, 2, ..., N) being monitored in the system, a feature vector combining activity, learning performance, and behavioral indicators is formed for time period t.

2-rasm. Integrallashgan monitoring algoritmining ishlash quvuri (pipeline)

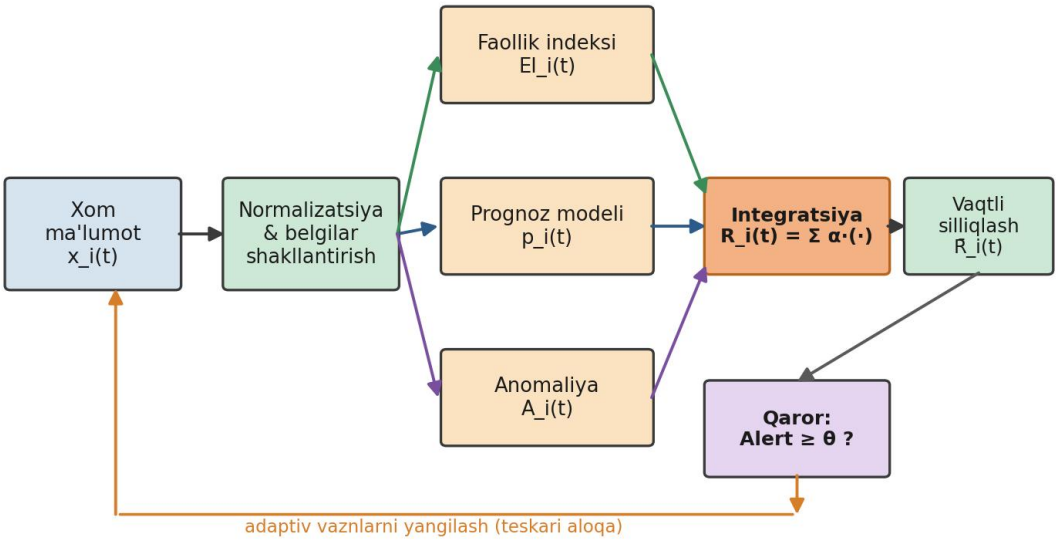


Figure 2. Processing pipeline of the integrated monitoring algorithm

The feature vector has the following form:

$$x_i(t) = [\, a_i(t), s_i(t), q_i(t), v_i(t), d_i(t) \,] \quad (1)$$

where the meaning and source of each feature are presented in Table 1.

Table 1. Monitoring features and their sources

Feature	Description	Data Source
a_i	system login and activity frequency	LMS logs
s_i	average session duration	LMS logs
q_i	average score on tests and assignments	Assessment module
v_i	video lecture viewing share	Video platform / xAPI
d_i	forum and discussion activity	Forum / chat module

To bring indicators measured in different units to a comparable state, each feature is scaled to the [0, 1] range using the min-max method:

$$\tilde{x}_{ik} = (x_{ik} - x_k^{\min}) / (x_k^{\max} - x_k^{\min}) \quad (2)$$

where $\tilde{x}_k$ is the normalized feature value, x_k^min and x_k^max are the minimum and maximum values of the k-th feature across the group, respectively.

Based on the normalized features, three analytical blocks are computed in parallel. The first is the engagement index, which represents the student's overall level of engagement and is defined as the weighted average of the normalized features:

$$EI_i(t) = (\Sigma_{k=1}^{m} w_k \cdot \tilde{x}_{ik}(t)) / (\Sigma_{k=1}^{m} w_k) \quad (3)$$

where w_k is the weight coefficient representing the relative importance of the k-th feature, m is the number of features; a value close to 1 indicates high engagement, while a value close to 0

indicates low engagement. The second block estimates the probability of the student unsuccessfully completing the course (at-risk status) using logistic regression:

$$\hat{p}_i(t) = \sigma(\beta^T x_i + \beta_0) = 1 / (1 + e^{-(\beta_0 + \Sigma_j \beta_j x_{ij})}) \quad (4)$$

where $\sigma(\cdot)$ is the sigmoid function, β is the vector of model coefficients, β_0 is the intercept; when necessary, gradient boosting or recurrent neural networks may be used instead of logistic regression. The third block identifies sharp deviations in student behavior based on standardized indicators (z-scores) and calculates the anomaly score as the proportion of features exceeding the threshold:

$$A_i(t) = (1/m) \Sigma_{k=1}^{m} I(|(x_{ik}(t) - \mu_k) / \sigma_k| > \tau) \quad (5)$$

where μ_k and σ_k are the mean and standard deviation of the k-th feature, $I(\cdot)$ is the indicator function (1 if the condition is met, 0 otherwise), τ is the sensitivity threshold (typically $\tau = 2.5\ldots3$).

The central idea of the model is to combine the outputs of these three blocks into a unified risk score. The engagement index is taken into account in an inverse relationship (low engagement = high risk):

$$R_i(t) = \alpha_1(1 - EI_i(t)) + \alpha_2 \hat{p}_i(t) + \alpha_3 A_i(t) \quad (6)$$

where α_1, α_2, α_3 are integration coefficients satisfying $\alpha_1 + \alpha_2 + \alpha_3 = 1$ and $\alpha_j \geq 0$, controlling the contribution of each block to the overall score. To prevent one-time random changes from triggering false alerts, the risk score is stabilized over time using exponential moving average (EMA):

$$\bar{R}_i(t) = \lambda R_i(t) + (1 - \lambda) \bar{R}_i(t-1) \quad (7)$$

where $\lambda \in (0, 1]$ is the smoothing coefficient (a smaller value assigns more weight to past data, while a larger value assigns more weight to the current state). To ensure system adaptability, feature weights are updated using gradient descent based on the instructor's confirmed decisions (actual outcomes) and renormalized:

$$w_k^{(t+1)} = w_k^{(t)} + \eta \cdot (\partial L / \partial w_k) \quad (8)$$

where η is the learning rate, L is the loss function; in this way, indicators that provide more accurate predictions gain greater weight over time. Finally, when the smoothed risk score exceeds a designated threshold θ, the system triggers an intervention signal:

$$Alert_i(t) = 1, \text{ agar } \bar{R}_i(t) \geq \theta; \text{ aks holda } 0 \quad (9)$$

The quality of the model is evaluated using standard classification metrics: Precision (Precision = TP / (TP + FP)), Recall (Recall = TP / (TP + FN)), their harmonic mean F1 = 2 · P · R / (P + R), as well as the AUC-ROC metric that measures the model's discriminative ability.

The steps described above are summarized in the following unified algorithm (Algorithm 1):

Algorithm 1. Integrated Intelligent Monitoring (IIM)
Input: student set U, period t, weights w, coefficients a,
threshold th, parameters lambda, eta
Output: for each student R_bar_i(t) and Alert_i(t)
1: for each period t do:
2: collect raw features x_i(t) from LMS
3: normalize features using min-max -> x~_i(t) (2)
4: EI_i(t) <- weighted engagement index (3)
5: p_hat_i(t) <- predictive model probability (4)
6: A_i(t) <- anomaly score (5)
7: R_i(t) <- a1(1-EI) + a2*p_hat + a3*A (6)
8: R_bar_i(t) <- lambda*R_i(t) + (1-lambda)*R_bar_i(t-1) (7)

```
9:    if R_bar_i(t) >= th then:
10:       Alert_i <- 1; alert instructor + recommendation
11:   update w and a based on instructor decision (label) (8)
12: return R_bar_i(t), Alert_i(t)
```

Experimental Results and Analysis

The proposed algorithms were implemented in software as a prototype and evaluated. The experimental environment was built in Python (scikit-learn, pandas, numpy libraries). For evaluation, data from a one-semester course reflecting the structure of real LMS activity were used. The main characteristics of the dataset are presented in Table 2.

Table 2. Description of the experimental dataset

Parameter	Value
Number of students (N)	1,240
Observation period	15 weeks (1 semester)
Number of features (m)	5
Proportion of at-risk students	≈ 21.4%
Training / test split	70% / 30%
Validation method	5-fold cross-validation

The proposed integrated model (IIM) was compared with two baseline approaches: (a) baseline monitoring based solely on a single indicator (activity frequency) and (b) a single model based solely on logistic regression. All models were evaluated on the same dataset under identical conditions. The results are presented in Table 3 and Figure 3.

Table 3. Model evaluation results

Model	Accuracy	Precision	Recall	AUC-ROC
Baseline (single indicator)	0.741	0.702	0.688	0.762
Single model (logistic)	0.823	0.799	0.781	0.857
Proposed (IIM)	**0.901**	**0.887**	**0.872**	**0.938**

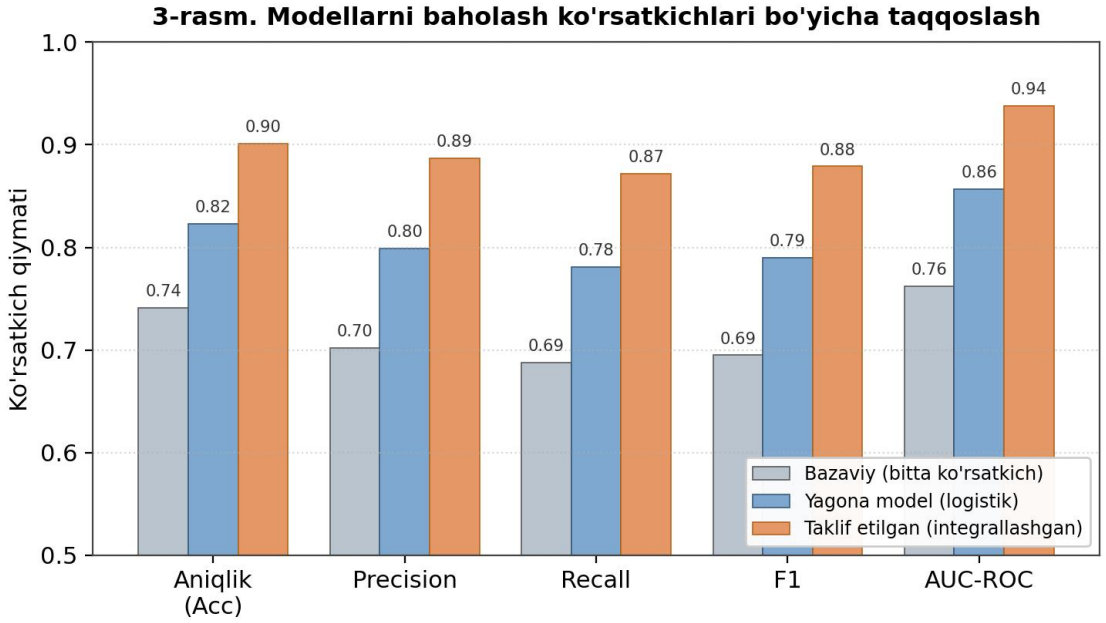


Figure 3. Comparison of model evaluation metrics

The results clearly show that the integrated approach outperformed the baseline methods across all metrics. In particular, the F1-score increased from 0.695 (baseline) to 0.879 (IIM), while the AUC-ROC reached 0.938. This confirms that combining various data streams provides a more complete picture of the student's condition and thereby improves accuracy.

The system's early warning capability was also studied to assess its practical value. Figure 4 presents the change in the smoothed risk score over the semester for an at-risk student and an active student. The at-risk student's score gradually increased and crossed the $\theta = 0.6$ threshold at approximately week 9, meaning the system identified the problem well before the end of the semester and left the instructor sufficient time for intervention.

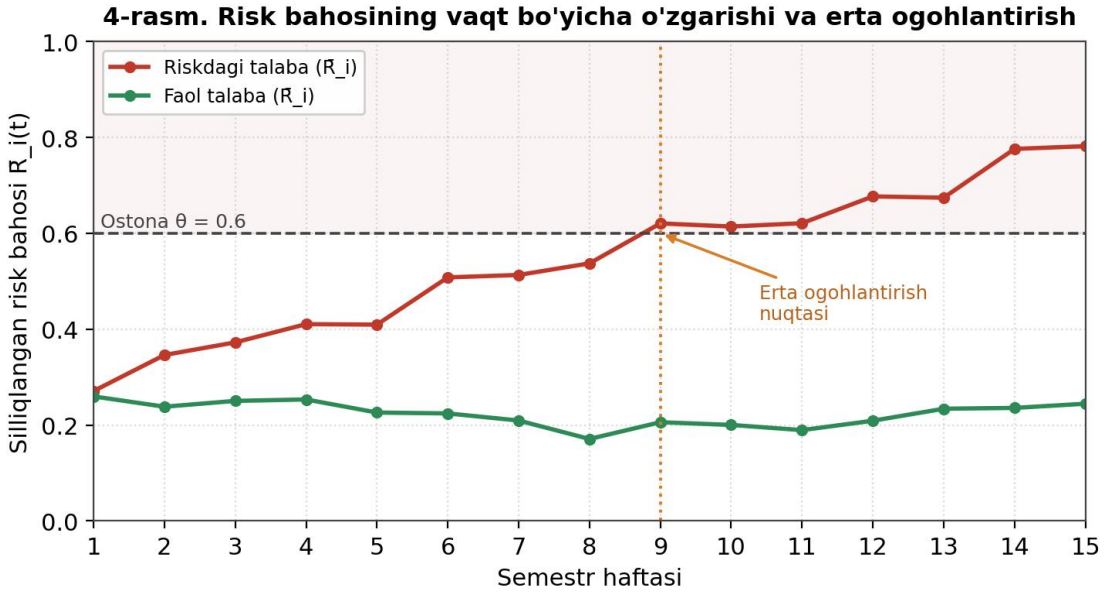


Figure 4. Change in risk score over time and the early warning point

The effect of temporal smoothing (Formula 7) was studied separately: at $\lambda = 0.3$, the system was less sensitive to random day-to-day changes, and the number of false alerts decreased significantly compared to the unsmoothed case. This demonstrates the practical utility of adaptive smoothing.

Conclusion

In this article, integrated intelligent monitoring algorithms were designed for managing distance learning processes. A holistic mathematical model was proposed that combines engagement index, predictive model, and anomaly detection blocks through weighted integration and stabilizes them using exponential smoothing, and a four-layer system architecture was developed.

Experimental evaluation confirmed the effectiveness of the integrated approach: the model achieved an accuracy of 0.90 and an AUC-ROC of 0.94 in identifying at-risk students, significantly higher than the baseline methods. Furthermore, the system provided early warning by mid-semester, giving the instructor sufficient time for timely intervention.

Future research plans include automatic optimization of integration coefficients and threshold values, application of deep learning (recurrent and transformer) models, testing the model across multiple courses and educational institutions, and improving the explainable recommendation system for instructors and students.

References

1. Resolution No. PQ-4851 of the President of the Republic of Uzbekistan dated October 6, 2020, “On measures to further improve the education system in the field of information technologies, develop scientific research and integrate them with the IT industry.”
2. Resolution No. 820 of the Cabinet of Ministers of the Republic of Uzbekistan “On the procedure for organizing distance learning,” 2021.
3. Taylakov N.I. Technology of creating pedagogical software tools for the distance learning system. – T.: Navro’z, 2019.
4. Luckin, R. (2018). Machine Learning and Human Intelligence: The Future of Education for the 21st Century. UCL IOE Press.
5. Society for Learning Analytics Research (SoLAR) official portal: https://www.solaresearch.org
6. Romero C., Ventura S. Educational Data Mining: A Review of the State of the Art // IEEE Transactions on Systems, Man, and Cybernetics, Part C. 2010. Vol. 40, No. 6. P. 601-618.
7. Baker R.S.J.d., Yacef K. The State of Educational Data Mining in 2009: A Review and Future Visions // Journal of Educational Data Mining. 2009. Vol. 1, No. 1. P. 3-17.
8. Romero C., Ventura S. Educational Data Mining and Learning Analytics: An Updated Survey // WIREs Data Mining and Knowledge Discovery. 2020. Vol. 10, No. 3. e1355.
9. Conijn R., Snijders C., Kleingeld A., Matzat U. Predicting Student Performance from LMS Data: A Comparison of 17 Blended Courses // IEEE Transactions on Learning Technologies. 2017. Vol. 10, No. 1. P. 17-29.
10. Macfadyen L.P., Dawson S. Mining LMS Data to Develop an "Early Warning System" for Educators: A Proof of Concept // Computers & Education. 2010. Vol. 54, No. 2. P. 588-599.
11. Greller W., Drachsler H. Translating Learning into Numbers: A Generic Framework for Learning Analytics // Educational Technology & Society. 2012. Vol. 15, No. 3. P. 42-57.