Learning Analytics in Massive Open Online Courses as a Tool for Predicting Learner Performance

Bystrova T., Larionova V., Sinitsyn E., Tolmachev A.

Tatiana Bystrova
Doctor of Sciences in Philosophy, Professor at Ural Institute for the Humanities, Ural Federal University named after the first President of Russia B. N. Yeltsin. Email: tatiana.bystrova@urfu.ru

Viola Larionova
Candidate of Sciences in Mathematical Physics, Associate Professor, Deputy Provost, Head of an academic department, Graduate School of Economics and Management, Ural Federal University named after the first President of Russia B. N. Yeltsin. Email: v.a.larionova@urfu.ru

Evgueny Sinitsyn
Doctor of Sciences in Mathematical Physics, Professor, Graduate School of Economics and Management, Ural Federal University named after the first President of Russia B. N. Yeltsin. Email: e.v.sinitcyn@urfu.ru.

Alexander Tolmachev
Senior Lecturer, Graduate School of Economics and Management, Ural Federal University named after the first President of Russia B. N. Yeltsin. Email: avtolmachev@urfu.ru

Address: 19 Mira St, 620002 Ekaterinburg, Russian Federation.

Abstract. Learning analytics in MOOCs can be used to predict learner performance, which is critical as higher education is moving towards adaptive learning. Interdisciplinary methods used in the article allow for interpreting empirical qualitative data on performance in specific types of course assignments to predict learner performance and improve the quality of MOOCs. Learning analytics results make it possible to take the most from the data regarding the ways learners engage with information and their level of skills at entry. The article presents the results of applying the proposed learning analytics algorithm to analyze learner performance in specific MOOCs developed by Ural Federal University and offered through the National Open Education Platform.

Keywords: massive open online courses, learning analytics, empirical evidence, online learning, assessment tools, checkpoint assignments, academic performance monitoring.

Due to the emergence of massive open online courses (MOOCs) that have swept the global education market [Semenova, Vilkova, Shcheglova 2018], online learning technologies have become widespread not only in informal education but in higher education and continuing professional development as well over the past decade [Eu-

ropean Association of Distance Teaching Universities 2018; Netology Group 2017]. Use of MOOCs in education programs [Roshchina, Roshchin, Rudakov 2018] has allowed universities and vocational schools to expand their educational choice options and create conditions for virtual mobility among students [Sancho, de Vries 2013], enhancing access to education and reducing college costs [Larionova, Tretyakov 2016]. In resorting to MOOCs, universities face the problem of selecting high-quality courses as well as the need to measure the effectiveness of online learning. The strategies for selecting online courses and the methods of assessing their effectiveness must be analyzed comprehensively in order to come up with well-defined decision-making criteria. Learning analytics in MOOCs is one of the key tools to improve education quality [O'Farrell 2017]. Not only does learning analytics data allow for monitoring learner performance and analyzing learner engagement but it also provides objective information on the effectiveness of online learning methods and techniques applied.

MOOC platforms offer diverse online courses [Hollands, Tirthali 2014]. The quality of MOOCs as a selection criterion is determined by how effective they are in achieving educational goals. In accordance with the experts' definition [Zagvyazinsky, Zakirova 2008; Samokhin et al. 2018], education effectiveness is understood as "the extent to which education outcomes are consistent to established goals", not just as an equivalent of economic efficiency defined as the ratio of actual education outcomes to the resources invested [Vishnyakova 1999]. The reliability of online learning effectiveness measurements depends on the adequacy of assessment tools and their consistence with the course performance requirements. Unlike with the conventional learning system, where the teacher provides a subjective face-to-face assessment of the student's knowledge and skills, MOOCs which imply exclusively distant interactions normally suggest that education outcomes are assessed using automated tests or peer reviews. Assessment objectivity requires fulfillment of the following conditions, which constitute the underlying principles of classical test theory and item response theory (IRT) [Crocker, Algina 2010].

- MOOC objectives must be formulated based on specific learning outcomes [Nekhaev 2016];
- Learning outcomes must be measurable;
- Assessment tools must be valid, reliable and sensitive to different levels of learner progress;
- Assessment results must be trustworthy and representative [Shmelev 2013].

The existing psychometric methods allow for assessing the quality of tests using the mathematical models and analytical procedures which are applied to analyze answers to specific test items [Mayorov 2002; Zvonnikov, Chelyshkova 2012]. The information theory-based algo-

rithm of assessing the informational value and quality of MOOC assessment tools proposed in this article expands the range of psychometric instruments and can be used to complement the conventional measures of test validity.

The social need for studying the effectiveness of digital technology in education has to do with the acute problem of organizing education in the information society with its high rates of technology turnover and lifelong accumulation of statistics on this type of learning. Reasons for low lifelong learning development rates may include, in particular, defects in the existing online courses and low motivation of students, who mostly belong to the so-called Generation Z, characterized by dependence on technology, impatience, drive for participation [Freitas, Morgan, Gibson 2015] and the habit of using the Internet to find information [Gryaznova, Mukovozov 2016; Guo, Kim, Rubin 2014; Tyler-Smith 2006]. Conventional teaching techniques prove to be low-effective for this cohort, so the need to modernize the learning process comes to the fore.

Apart from being socially relevant, research on the effectiveness of using online technology in education also has a pedagogical aspect. The content in online learning is still based on conservative mass education programs, and no allowance is made for the new educational paradigm [Jansen, Schuwer 2015; Kop, Fournier, Mak 2011]. Advocates of the traditional approach treat MOOC content as a series of video lectures and standard reading modules, although it has been about twenty years since education began to be understood not only as access to information but as the acquisition of specific practical skills as well [Lundvall, Borrás 1997; Nonaka, Takeuchi 2011]. As a result, MOOC statistics usually demonstrate a radical decrease in learning engagement and a gap between what learners expect and what MOOC providers have to offer [Brown, Lally 2017; Castano Muñoz et al. 2016]. A comparative study of the effectiveness of different online technologies will provide an opportunity to reduce that gap.

Effectiveness of online learning is crucial for a modern learner, too. In the information age, people want their learning trajectories to be AI-personalized to suit their personal needs and abilities. MOOCs provide ample opportunity for customized education and lifelong learning [Deev, Glotova, Krevskiy 2015], in particular because they are adaptable to students' individual needs and characteristics.

The technological implications of this study are predetermined by the format of exclusively distance learning courses, which implies documentation of learning outcomes as a "learner footprint" in the digital learning environment. This allows for monitoring individual learning trajectories, identifying cause-effect relationships between learner engagement and learning outcomes, exploring possible reasons for failure, and predicting ultimate progress based on average student performance. In addition, learning analytics is one of the few objective indicators of MOOC quality and is actively used to improve it.

The central hypothesis of this study is that learning analytics can be used to obtain objective information on the effectiveness of online learning and predict the academic performance of different types of learners. The study aims at developing learning analytics algorithms in order to evaluate the quality of MOOC assessment tools, analyze patterns of learner performance, and predict the probability of success/failure using the statistics on MOOCs provided by Ural Federal University and available through the National Open Education Platform. Achieving this goal involves the following objectives: (i) analyze the quality of MOOC assessment tools based on empirical evidence; (ii) estimate and compare functions of learner performance distribution for all midterm and final tests; (iii) clusterize learners by their performance and analyze their progress in dynamics; (iv) construct a probability model of changes in performance among different types of learners throughout the course. The study also seeks to identify factors that have negative effects on student performance in MOOCs. Research findings will help develop recommendations for course developers, in order to enhance teaching methods in online learning and improve the quality of assessment tools, as well as for MOOC tutors and engineers.

1. Theoretical aspects of online learning effectiveness
1.1. Characteristics of learning with MOOCs

A massive open online course is understood here as an openly accessible, structured, theoretically substantiated, goal-oriented set of educational materials, assessment tools and other distance learning resources. An online course determines the teaching methods, progress checkpoints and tools for assessing learners' knowledge and skills. Student-teacher and student-student communication is provided using digital learning environment services. The well-elaborated pedagogical design of an online course ensures achievement of the learning outcomes, provided that entrants possess the required knowledge and skills and sufficient motivation for learning.

A MOOC can be taken by anyone regardless of age, location, educational background and financial opportunities. Most MOOCs are asynchronous, i. e. knowledge is transferred from teacher to student with a time lag. This allows MOOC learners to customize their learning schedules with due regard to their individual preferences and abilities and choose their own pace in accessing course materials and doing assignments. Self-paced courses are not bound to specific dates and are offered in the "on-demand" format, which means they can be accessed at any time which is convenient for the learner. To ensure a consistent pace and improve self-regulation among students, most courses set deadlines for application, webinars and tests, including final exams.

To obtain a certificate of completion, a MOOC learner must complete name verification and take an online proctored final exam. Certificates are issued to learners who meet the course passing threshold

(specified in course overview) and pass the final exam. Final exams with online proctoring are usually taken for a fee. University students may earn credits for MOOCs in their major or minor by submitting a certificate of MOOC completion. Credit transfer procedures are regulated locally by educational institutions.

1.2. Factors of online learning effectiveness

In contrast to digital teaching and learning packages as a series of syllabus-related teaching materials and assessment tools, important features of MOOCs include organization of the learning process and consistent monitoring of learner performance. In this regard, every MOOC is a set of unique teaching techniques. Their effectiveness is measured not so much by content quality as by teaching methods applied in the digital learning environment and by the quality of assessment tools allowing adequate measurement of learner progress.

Predictors of effective online learning include:

- Methodologically substantiated presentation of digital content in consistence with the learning cycle [Kolb 1985];
- Use of interactive learning technology;
- Monitoring of learning outcomes and detection of bugs and errors throughout the course;
- Organization of learners' interaction;
- Learner support and motivation strategies;
- Use of active online teaching methods;
- Collection and statistical analysis of learner feedback;
- Prompt changes and updates, when necessary [Jasnani 2013].

MOOC design is thus a complex pedagogical challenge that requires a high level of professional expertise, teaching experience, methodological and information technology skills. The key to designing an effective online course is the use of interactive technology based on active teaching strategies in the online format [Lisitsyna, Lyamin 2014].

As we can see, the use of a digital learning environment services allows for regulating the learning process distantly and running online courses without direct teacher-student interaction. Course maintenance is thus restricted to keeping the content up to date throughout and after the course as well as providing student counseling services. As maintenance is ensured with regard to original course content and teaching methods, it does not require the direct participation of the course designer just as it rarely requires in-depth knowledge of the subject matter from counselors. Therefore, the teacher's main function consists in creating an online course, while the learning process may be controlled by tutors who provide methodological and organizational support to students, advise them on the choice of MOOCs and credit transfer opportunities, and help them build personalized learning trajectories, creating the conditions for successful performance in midterm and final checkpoint assignments.

1.3. Use of learning analytics to support learners

As compared to traditional education, where teachers get feedback from students only in face-to-face interactions, online learning leaves a digital footprint, with all learner accomplishments and activities during the course being recorded in the digital learning environment. Analysis of such data—learning analytics—allows for monitoring learning consistency, student progress and assignment performance.

Learning analytics is based on analyzing big data on learning behaviors in MOOCs [Usha Keshavamurthy, Guruprasad 2014]. It can provide a lot of information on the causes of learner success and failure and allows for predicting future learning behaviors. Findings are used to fine-tune learning contexts, support students and adapt them to new environments [O'Farrell 2017]. The core objectives of learning analytics are as follows:

• Measure, collect and present data on user behavior;
• Analyze student performance throughout the course;
• Analyze behavioral patterns using big data;
• Establish cause-effect relationships between performance indicators and learning activities;
• Detect errors and methodological issues in MOOCs;
• Develop recommendations for course content revision;
• Predict student success or failure.

Learning analytics includes diverse methods, from descriptive statistics to data mining. Additional sources of information, along with streaming data on user behavior fetched from MOOC platforms, may include administrative databases of educational institutions, surveys of learners and instructors, pre-test results, etc.

The global leaders in learning analytics include the National Forum for the Enhancement of Teaching and Learning in Higher Education, the National Research Center for Distance Education and Technological Advancements at the University of Wisconsin-Milwaukee and EdPlus at Arizona State University.

Research at Arizona State University is currently focused on finding efficient adaptive learning tools using big data on MOOC learner behaviors. By identifying behavioral patterns at the early stages of learning and classifying students based on their learning activities, researchers examine the factors that have a positive impact on student performance and use them to predict course completion (e. g. [Sharkey, Ansari 2014]).

2. Research methods

The algorithms described below are applied, among other things, to analyze the informational value and quality of MOOC assignments, which must differentiate between learners by level of performance as well as ensure and reflect their consistent progress. Another equally important objective consists in predicting checkpoint performance among students at different stages of their progress which is meas-

ured by average student performance. Such a prediction will allow for adapting learners with different performance levels to course requirements by additional counseling, personalized assignments, etc.

From the standpoint of the first objective, assignments that are either passed or failed by the great majority of learners should be recognized as equally ineffective, as they provide instructors with no information on course progress or the performance of individual students.

Informational value of assignments in terms of how well they are able to differentiate learners by their performance is assessed using standard information theory methods. If the distribution of checkpoint test grades (measured in scores) is labelled as (x), the fact that an individual learner has obtained a specific score will be loaded with the following number of information bits [Korn, Korn1973]:

(1) $$I = -\int_{0}^{100} \varphi(x) \cdot log_2\left(\varphi(x)\right) \cdot dx, \, bit$$

In practical calculations, the range of scores is divided into ten-point discrete intervals, and the integral is transformed into a sum of integral elements for such intervals. For convenience, this value will be compared to the maximum amount of information to which uniform distribution $\varphi_1(x) = 1/n$ corresponds, where n is the number of intervals:

$$I_{max} = log_2(n) = 3.22$$

In this case, the informational value of a checkpoint assignment will be described by measure

(2) $$inf = 100 \, \frac{I}{I_{max}}$$

rounded to the nearest whole number.

Statistical characteristics of individual learner performance in a series of checkpoint assignments must be analyzed to determine course progress and predict course completion. Our previous study [Larionova et al. 2018] examined changes throughout the course in the statistical distributions of scores among categories of learners identified based on their average performance in earlier periods (A students, B students, etc.).

To solve the problem of reflecting learner progress with the use of assessment tools, we will introduce three learner categories based on learner progress:

- Non-performers, who failed the assignment, i. e. scored under 40 ("Failure");
- Average performers, whose scores are ranged between 40 and 60 ("Pass"); and
- Constant performers, who scored 60 or higher ("Success").

There can be more categories, but three is enough to fully describe the level of learner progress and ensure that results are illustrative.

While taking a course and the checkpoint assignments within it, MOOC learners migrate from one category to another. If such transitions are traced for every student, the probability of cross-category transition for each checkpoint can be estimated. Accuracy of estimates depends on the number of learners in the sample: the larger the sample, the more accurate the probability of transition. Such estimates will allow for making inferences on how checkpoints reflect learner progress as well as predicting performance in checkpoint assignments among learners of different categories. Predictions like that require the accumulation of information on transition probabilities and the processing of large volumes of data on performance in the checkpoint assignments.

Let us label as $|i>$ and $|j>$ learner status before and after a checkpoint, respectively (status being understood as belonging to category i before the assignment and j after it; i, j = 1, 2, 3). Suppose each cross-category transition corresponds to operator T_{ij}, which is defined as follows:

(3) $T_{ij} \cdot |i> = |j>$

Operator T_{ij} is the operator of transition $i \rightarrow j$, transition probability being determined by the matrix

(4) $\hat{P} = \left\| \begin{array}{ccc} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{array} \right\|$

Matrix $\hat{P}$ is asymmetric, its entries satisfying the condition:

(5) $\sum_{j=1}^{3} P_{ij} = 1$

The number of learners in every category, at probabilities (4), can be estimated using the model proposed by Astratova et al. (2017), which allows for determining the probability that categories 1, 2, 3 will contain X_1, X_2, X_3 members, respectively, at the moment of time $t - P(X_1, X_2, X_3 \mid t)$. The equation for $P(X_1, X_2, X_3 \mid t)$ is written as follows:

(6) $\frac{\partial P(X_1, X_2, X_3 \mid t)}{\partial t} = P(X_1, X_2, X_3 \mid t) \cdot \left\{ (1-z) \cdot \sum_{j=1}^{3} P_{ii} - \sum_{j=1}^{3} X_i \right\} +$

$+ z \cdot \sum_{i=1}^{3} (X_i + 1) \cdot P(..., X_{i+1}, ... \mid t) + (1-z) \cdot \sum_{j=1, i \neq j}^{3} P_{ij} \cdot (X_i + 1) \times$

$\times P(..., X_{i+1}, ... \mid t \}$

In this equation, z is the probability of learner withdrawal per unit of time. Hereinafter, z will be considered equal to zero (for this purpose,

students who withdrew should be excluded from analysis at the pre-liminary stage).

Equation (6) can be solved in a general fashion, but for most types of problems, analysis of means and covariances will suffice:

(7) $\bar{X}_i = <X_i> = \int_0^{100} X_i \cdot P(X_1, X_2, X_3 \mid t) \cdot dX_i,$

$\sigma_{IJ} = \overline{(X_i - <X_i>) \cdot (X_j - <X_j>)} = \int_0^{100} X_i \cdot X_j \cdot P(X_1, X_2, X_3 \mid t) \cdot dX_i \cdot dX_j - <X_i> \cdot <X_j>.$

It can be shown that the following conditions are satisfied:

(8) $\bar{X}_i \sim N,$

$\sigma_{IJ} \sim \sqrt{N}$

N is the total number of learners in a MOOC. Therefore, where $N \rightarrow \infty$, variation coefficients tend to be zero:

$C_v(ij) \sim \frac{1}{\sqrt{N}} \rightarrow 0,$

which illustrates the law of large numbers. This way, if the number of learners N is high enough, their distribution among categories is hard-ly a coincidence and the size of category approaches $<X_i>$, where:

$<X_1> + <X_2> + <X_3> = N$

The equation for X_i is written as follows:

(9) $\frac{\partial X_l}{\partial t} = \sum_{k=1}^{3} \left[\tilde{P}_{kl} \cdot X_k - \tilde{P}_{lk} \cdot X_l \right],$

where

(10) $\tilde{P}_{kl} = 0$ for $k = l$

$\tilde{P}_{kl} = P_{kl}$ defined (4) for $k \neq l$

Transition matrix (4) can be linked to a problem of random walks on a directed graph whose vertices correspond to categories $i = 1, 2, 3$ and where the probabilities of cross-vertex transition are determined by (4) [Leskovec, Rajaraman, Ullman 2016].

Transition probabilities (4) determine unambiguously the influ-ence of checkpoints on the distribution of learners among perfor-mance categories and may be indirect indicators of assignment quality. However, using matrix (4) directly is inconvenient, first of all because of the abundance of parameters (9 transition probabilities) and their intricate, however unambiguous, relationship with the com-prehensible conventional characteristics of academic performance. For this reason, the role of an illustrative parameter will be assigned

to vector $\vec{a}_x = \{X_1, X_2, X_3\}$, which determines the steady-state distribution of learners among performance categories $j = 1, 2, 3$. This vector can be treated as a steady-state solution of equation (9), corresponding to continuous case $\left(\frac{\partial X_j}{\partial t} = 0\right)$, or as a limiting distribution that results after multiple transitions of the form $\vec{a}_x(n) = \hat{P} \cdot \vec{a}_x(n-1)$ on a graph relative to matrix (4) [Astratova et al. 2017], provided that $n \rightarrow \infty$. This limiting case corresponds to a hypothetical situation where the checkpoint assignment is taken a number of times by categories of learners with statistically equivalent characteristics of academic performance. It is easy to show that $\vec{a}_x(n \rightarrow \infty) = \hat{P} \cdot \vec{a}_x$ satisfies the equation [Ibid.]:

(11) $\vec{a}_x = \hat{P} \cdot \vec{a}_x$

Hence, $\vec{a}_x$ is an eigenvector of $\hat{P}$ (4) with eigenvalue 1. Using (5), (10), it can be shown that $\vec{a}_x$ in (11) corresponds to steady-state solution (9) for $\frac{\partial X_j}{\partial t} = 0$.

A formula analogous to (11) can also be used with known matrix $\hat{P}$ (4) to predict checkpoint performance. Suppose that $\vec{a}_x(0)$ is a vector describing the distribution of learners among performance categories before the checkpoint and $\vec{a}_x(1)$ after the checkpoint; then, in compliance with the theory of Markov processes [Maksimov 2001], these two vectors are related by the following formula:

(12) $\vec{a}_x(1) = \hat{P} \cdot \vec{a}_x(0)$

where $\hat{P}$ is a matrix of the form (4) corresponding to the checkpoint analyzed.

3. Application and discussion

A case study illustrating how the algorithm described above can be applied involves analysis of data on the online course *Engineering Mechanics* offered by Ural Federal University and available through the National Open Education Platform[1]. The course includes the following assessment tools (checkpoint assignments):

- theory tests (T);
- home assignments (HA);
- project assignments (PA);
- the final test (FT).

In the source database, each checkpoint assignment was assessed on a 100-point scale, and each of them was assigned weight coefficient

---

[1] https://openedu.ru/course/urfu/ENGM/

k_p, $p = 1, ..., 4$. Weight coefficients $0 \leqslant k_p \leqslant 1$ and scores $0 \leqslant B_j(C) \leqslant 100$ obtained by learners in each checkpoint, where $C = T, HA, PA, FT$, were used to calculate the following indicators:

• Average student current performance

(13) $Avg = k_1 \frac{1}{16} \sum_{j=1}^{16} B(T)_j + k_1 \frac{1}{18} \sum_{j=1}^{18} B(HA)_j + k_1 \frac{1}{5} \sum_{j=1}^{5} B(PA)_j$.

• Final course grade

(14) $Grade = Avg + k_4 B(FT)$.

• In accordance with course design, coefficients k_p took on values: $k_1 = 0.16$; $k_2 = 0.34$; $k_3 = 0.1$; $k_4 = 0.4$. Therefore, maximum Avg value is 60. To facilitate comparison of results in different checkpoints to this value for every learner with identifier i, the maximum Avg value was translated to a 100-point scale using the formula

(15) $Avg_j(100) = 100 \cdot \dfrac{Avg_i}{Max\{Avg_i | i = 1, ..., N\}}$.

where N is the total number of learners in the MOOC.

Each checkpoint can be assigned the following characteristics:

– Average checkpoint grade;
– Task solvability coefficient: $k_i = \dfrac{c_i}{N}$,
   where c_i is the number of learners who solved the task and N is the total number of learners in the MOOC;

• Checkpoint assignment grade probability density function (a more complex characteristic).

The latter can be used to assess the informational value of checkpoint (2).

Let us analyze Test 1 as an example. The grade distribution function is displayed in Figure 1. Normal distribution is shown in the same figure for comparison. Even when no special criteria are used, it can be seen that deviations in actual distribution of scores for Test 1 from normal distribution are significant and cannot be explained by random processes. The load of information contained in the fact "learner was awarded a specific number of scores for Test 1", calculated using formula (1), is I=1.47 bits.

Formula (2) is used to calculate the informational value of all "test"-type checkpoints. The results are presented in Figure 2.

In particular, Figure 2 makes it clear that Tests 2, 14 and 15 have the highest informational value, which means that they are effective in differentiating learners by the level of progress. Meanwhile, Tests 3, 6 and 16 are the least informative: they are probably too easy, as the great majority of learners perform them successfully. Table 1 compares the highest and lowest informational values of the tests with other checkpoint characteristics.

Figure 1. **Probability Distribution Function for Grades Obtained for Test 1 $\varphi_1(x)$.**
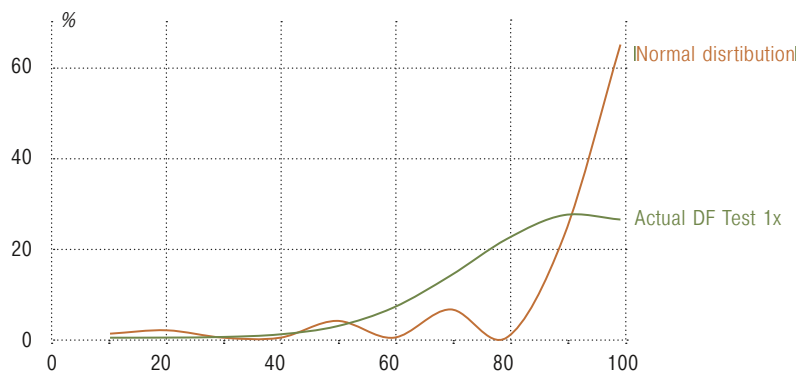

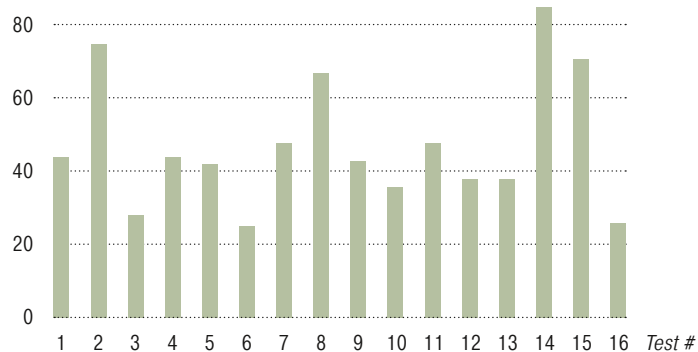Figure 2. **Informational Value of Tests Calculated Using Formula (2).**

Table 1. **Characteristics of the Most and the Least Informative Tests**

Test #	Informative Value *inf* (2)	Average Grade	Solvability Coefficient
Test 6	25	96.3	0.993
Test 16	26	95.8	0.985
Test 2	75	80.2	0.898
Test 14	85	68.5	0.797
Test 15	71	72.4	0.869
Relative difference between the highest value and the lowest one (Max – Min)/Min, %	240	40.5	24.6

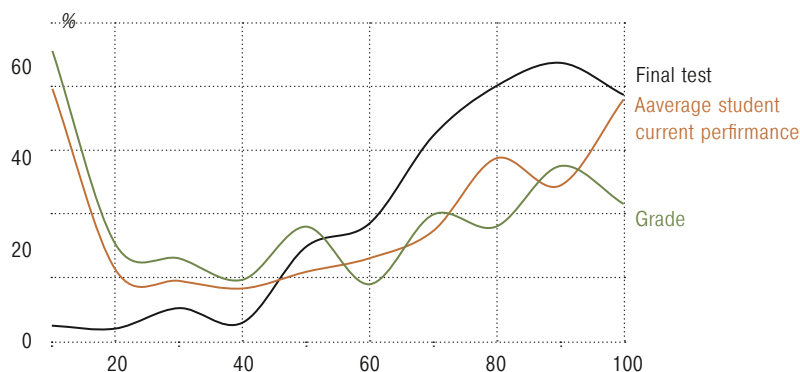Figure 3. **Grade Probability Density for Average Student Current Performance, Final Test and Grade.**

Table 2. **Informational Value of Checkpoints**

#	Checkpoint	*Inf,* formula (2)
1	Average student performance (*Avg*)	93
2	Final test	84
3	Grade	94

The difference in informational value between the most and the least informative tests, calculated using formula (2), is essentially higher than the relevant differences in such characteristics as average grade and solvability coefficient. Therefore, informational value is the most convenient tool for comparing checkpoint assignments and their quality.

Of all the types of checkpoints, the following is of the most interest:

- Average student current performance (*Avg*);
- the Final test (FT);
- Grade, i. e. integral estimate of course completion which includes *Avg* and FT.

Grade probability densities for these types of checkpoints are shown in Figure 3.

All the three checkpoints in Figure 3 have a rather broad range of grades, i. e. each of them is a good differentiator of learners. Data on the informational value *inf* of relevant checkpoints, calculated using formula (2), is given in Table 2.

As we can see, such integral checkpoints as *Avg* and Grade, which reflect learner progress throughout the course, have a high informa-

Table 3. **Steady-State Distribution of Learners among Performance Categories and its Informational Value (i) for Different Checkpoints**

Proportion of Category in the Sample	Test	HA	PA	FT
x_1 (non-performers)	0.336	0.436	0.658	0.296
x_2 (average performers)	0.002	0.087	0	0.129
x_3 (constant performers)	0.662	0.477	0.342	0.575
inf (informational value)	59	84	58	86

tional value, which is not always true for individual checkpoint assignments (see Table 1). The informational value of the final test is somewhat lower but still pretty high.

Next, a series of checkpoints corresponding to different types of assignments (T, HA, PA, FT) are analyzed instead of individual checkpoints. The state after the first checkpoint in a set is taken as the input state here. It thus becomes possible to analyze all the sets of checkpoints independently; besides, it solves the problem of no entrance testing in most MOOCs (information on entrants' skills is usually unavailable). The results are shown in Table 3. Analysis results can be presented even more concisely if factor *inf* (2) is used. In this case, it reflects the informational value of post-checkpoint learner distribution.

Assessment tools of the types "test" and "project assignment" in fact split learners into constant performers and non-performers, the intermediate category of average performers being virtually indistinguishable. This data indicates, in particular, the low informational value of the respective types of checkpoints, which is illustrated by the last row in Table 2. Indeed, learners either fail or obtain high grades in these checkpoint assignments. Perhaps, the assignments are too easy or results are assessed as pass/fail, which is especially typical of project assignments. Of course, there can be other reasons for the stratification observed. Anyway, the analysis performed obviously provides course designers with useful information to measure the quality of assessment tools.

Data on average student performance (*Avg*) can be used when taking the final test as input state $|i>$. In this case, transitions among performance categories as a result of the final test will be calculated: $Avg(i) \rightarrow FT(j)$ (*i* and *j* are performance categories here). The resulting pairs *{ij}* for post-FT transitions among performance categories yield the following matrix:

$$(16) \quad \hat{P} = \begin{Vmatrix} 0.320 & 0.586 & 0.218 \\ 0.200 & 0.106 & 0.098 \\ 0.480 & 0.308 & 0.684 \end{Vmatrix}$$

Figure 4. **Oriented Graph for Transitions among Performance Categories Generated by the "Final Test" Checkpoint. Numbers correspond to transition probabilities (16).**
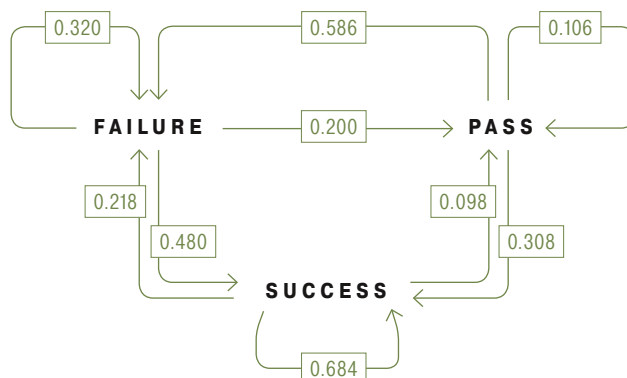
Transition probabilities can be presented as a directed graph, as shown in Figure 4.

As can be seen from Figure 4, transitions from "Success" to "Success" and from "Pass" to "Failure" are the most probable ones. The probability of transition from "Failure" to "Success" is also surprisingly high. However, researchers at Arizona State University have also observed this personality type in students, referring to them as "kangaroos" ([Johnson 2018]).

Let's suppose that learners are distributed uniformly among performance categories just before the final test:

$$x_1 = x_2 = x_3 = \frac{1}{3},$$

According to estimated transition probabilities (16), the predicted distribution of learners after the final test in compliance with (12) will be the following: $x_1 = 0.375$; $x_2 = 0.135$; $x_3 = 0.491$. If the predicted distribution is unacceptable for instructors (e. g. an increase of the non-performer category as compared to the current state is predicted in the case analyzed), they can take some provisional measures to support students and increase overall performance.

Let us now compare efficiency of this learning analytics algorithm for different online courses. Since every analyzed MOOC has its own structure of checkpoints, it makes sense to compare transitions $Avg(i) \rightarrow FT(j)$ (*i, j* are categories "Failure", "Pass", "Success"), as data on average student performance and the final test is available in any course. The findings are presented in Table 4.

The predicted proportion of constant performers in Descriptive Geometry and Technical Drawing is the lowest, while that of the "Failure" category is, vice versa, the highest among the courses analyz-

Table 4. **Predicted Steady-State Distribution of Learners among Performance Categories for Transitions _Avg → FT_ in Different MOOCs.**

	x_1 (Failure)	x_2 (Pass)	x_3 (Success)	_inf_ (informatio-nal value)
Engineering Mechanics	0.296	0.129	0.575	86
Construction Materials Engineering	0.197	0.105	0.698	73
Descriptive Geometry and Technical Drawing	0.48	0.149	0.371	91

Table **5. Probabilities of Transition among Performance Categories _Avg → FT_ in the Descriptive Geometry and Technical Drawing MOOC.**

Descriptive Geometry and Technical Drawing	Failure	Pass	Success
Failure	0.40	0.98	0.38
Pass	0.25	0.00	0.08
Success	0.35	0.02	0.54

ed. The final test will be the most informative assessment tool in this course.

Probabilities of transition $Avg(i) → FT(j)$ among performance categories in Descriptive Geometry and Technical Drawing are given in Table 5[2].

The probability of transition from "Pass" to any other category is extremely low, while that of transition from "Failure" to "Success" (the "kangaroo" personality type) is rather high (0.35). The "Success" category tends towards stratification at the FT checkpoint: students classified under this category based on their average performance either pass into the "Failure" category (with a probability of 0.38) or, more likely (0.54), retain their positions among constant performers.

The "kangaroo" personality type manifests itself more in Construction Materials Engineering (probability of relevant transitions being equal to 0.47), whereas the probability of transition from "Success" to "Failure" after the final test is low here (0.15). Most students in the "Success" category remain high performers with a probability of 0.83. The probability of transition from "Success" to "Pass" is the lowest for this course (Table 6).

---

[2] The matrix presents probabilities of transition from categories corresponding to columns to those corresponding to rows (the sum of elements in each column thus being 1).

Table **6. Probabilities of Transition among Performance Categories**
Avg → FT **in the Construction Materials Engineering MOOC**

Construction Materials Engineering	Failure	Pass	Success
Failure	0.13	0.64	0.15
Pass	0.40	0.07	0.03
Success	0.47	0.29	0.83

The Engineering Mechanics MOOC was analyzed earlier in this article (see Figure 4). It differs significantly from the other two MOOCs in transition probabilities $Avg(i) → FT(j)$ and provides the most adequate distribution of final course grades, which indicates sufficient reliability of the assessment system in this online course, a high level of instructor support, and theoretically substantiated course content that contributes to learner progress.

4. Conclusion

Online learning is a new educational paradigm generated by recent sociocultural processes, communicational ones in the first place. It implies better feedback for learners, which shapes personalized learning trajectories and ultimately promotes lifelong learning. Education has moved from monologue to dialogue, making the student an active participant in learning. The method of predicting MOOC performance proposed in this article will allow for providing learners with better feedback and more personalized learning trajectories; it could become an integral part of online learning over time.

The results of learning analytics research show that:

— Analysis of the informational value of assessment tools based on the method described herein may provide course developers with useful information on the quality of checkpoint assignments in addition to traditional psychometric analysis;

— Monitoring of learners' checkpoint performance trajectories and the probabilities of learner transition among performance categories estimated based on the monitoring data can be used to assess post-checkpoint redistribution of learners, which provides additional information to assess the quality of assessment tools;

— Knowing the probabilities of learner transition among performance categories, instructors can predict the final distribution and take necessary measures to enhance their teaching efforts.

References

Astratova G., Sinicin E., Toporkova E., Frishberg L., Karabanova I. (2017) Mechanism of Information Model Development for Company Brand Assessment within Marketing Strategy. Proceedings of the *International Conference on*

Trends of Technologies and Innovations in Economic and Social Studies (Tomsk, 28–30 June 2017), Tomsk: Atlantis, pp. 20–25.

Brown K., Lally V. (2017) It Isn't Adding Up: The Gap between the Perceptions of Engineering Mathematics Students and Those Held by Lecturers in the First Year of Study of Engineering. Proceedings of the *10th Annual International Conference of Education, Research and Innovation—ICERI2017 (Seville, Spain, 16–18 November 2017)*, Valencia: IATED, pp. 317–321.

Castaño Muñoz J., Punie Y., Inamorato dos Santos A., Mitic M., Morais R. (2016) *How Are Higher Education Institutions Dealing with Openness? A Survey of Practices, Beliefs and Strategies in Five European Countries. Institute for Prospective Technological Studies. JRC Science for Policy Report, EUR27750EN.* Available at: http://publications.jrc.ec.europa.eu/repository/bitstream/JRC99959/lfna27750enn.pdf (accessed 2 October 2018).

Crocker L., Algina J. (2010) *Vvedenie v klassicheskuyu i sovremennuyu teoriyu testov* [Introduction to Classical and Modern Test Theory], Moscow: Logos.

Deev M. V., Glotova T.V, Krevskiy I. G. (2015) Individualized Learning Trajectories Using Distance Education Technologies. *Creativity in Intelligent, Technologies and Data Science. Ser.: Communications in Computer and Information Science*, part XI, vol. 535, pp. 778–792.

European Association of Distance Teaching Universities (2018) *The 2018 OpenupEd Trend Report on MOOCs.* Available at: https://www.openuped.eu/images/Publications/The_2018_OpenupEd_trend_report_on_MOOCs.pdf (accessed 2 October 2018).

Freitas S. I., Morgan J., Gibson D. (2015) Will MOOCs Transform Learning and Teaching in Higher Education? Engagement and Course Retention in Online Learning Provision. *British Journal of Educational Technology*, vol. 46, no 3, pp. 455–471.

Gryaznova Y., Mukovozov O. (2016) *Pilotnoe issledovanie RASO "Kak pokolenie Z vosprinimaet informatsiyu"* [Pilot Study by the Russian Public Relations Association: How Generation Z Perceives Information]. Paper presented at 4th International Applied Research Conference on Communication in Sociology, Humanities, Economics and Education (Minsk, Belarus, 7–9 April, 2016), Minsk: Belarusian State University.

Guo P.J, Kim J., Rubin R. (2014) How Video Production Affects Student Engagement: An Empirical Study of MOOC Videos. Proceedings of the *First ACM Conference on Learning and Scale (Atlanta, GA, 4–5 March 2014)*, New York: ACM, pp. 41–50.

Hollands F. M., Tirthali D. (2014) *MOOCs: Expectations and Reality. Full Report*. Available at: http://cbcse.org/wordpress/wp-content/uploads/2014/05/MOOCs_Expectations_and_Reality.pdf (accessed 2 October 2018).

Jansen D., Schuwer R. (2015) *Institutional MOOC Strategies in Europe. Status Report Based on a Mapping Survey Conducted in October-December 2014*. Available at: http://www.eadtu.eu/documents/Publications/OEenM/Institutional_MOOC_strategies_in_Europe.pdf (accessed 2 October 2018).

Jasnani P. (2013) *Designing MOOCs. A White Paper on Instructional Design for MOOCs*. Available at: http://www.tatainteractive.com/pdf/Designing_MOOCs-A_White_Paper_on_ID_for_MOOCs.pdf (accessed 2 October 2018).

Johnson D. (2018) *Driving Performance & Persistence: How ASU is Improving Learner Outcomes with an Active Adaptive Approach*. Available at: https://youtu.be/ASekB3jElBs (accessed 2 October 2018).

Kolb D. A. (1985) *Learning Style Inventory. Technical Manual*. Boston: McBer and Company.

Kop R., Fournier H., Mak S. F.J (2011) A Pedagogy of Abundance or a Pedagogy to Support Human Beings? Participant Support on Massive Open Online

Courses. *The International Review of Research in Open and Distance Learning*, vol. 12, no 7, pp. 74–93.

Korn G., Korn T. (1973) *Spravochnik po matematike dlya nauchnykh rabotnikov i inzhenerov* [Mathematical Handbook for Scientists and Engineers], Moscow: Chief Editorial Board for Physics and Mathematics Literature.

Larionova, V., Brown K., Bystrova T., Sinitsyn E. (2018) Russian Perspectives of Online Learning Technologies in Higher Education: An Empirical Study of a MOOC. *Research in Comparative and International Education*, vol. 13, no 1, pp. 70–91.

Larionova V., Tretyakov V. (2016) Otkrytye onlayn-kursy kak instrument modernizatsii obrazovatelnoy deyatelnosti v vuse [Open Online Courses as a Tool for Modernization of Educational Process in Universities. *Higher Education in Russia*, no 7, pp. 55–66.

Leskovec J., Rajaraman A., Ullman J. (2016) *Analiz bolshikh naborov dannykh* [Mining of Massive Datasets], Moscow: DMK Press.

Lisitsyna L., Lyamin A. (2014) Approach to Development of Effective E-Learning Courses. *Smart Digital Futures*, vol. 262, pp. 732–738.

Lundvall B. A., Borrás S. (1997) *The Globalising Learning Economy: Implications for Innovation Policy. Report Based on Contributions from Seven Projects under the TSER Programme DG XII.* Available at: www.globelicsacademy.org/2011_pdf/Lundvall%20Borras%201997.pdf (accessed 2 October 2018).

Maksimov Y. (ed.) (2001) *Veroyatnostnye razdely matematiki* [Probability-Related Areas of Mathematics], Saint Petersburg: Ivan Fedorov.

Mayorov A. (2002) *Teoriya i praktika sozdaniya testov dlya sistemy obrazovaniya* [Theory and Practice of Developing Tests for the Education System], Moscow: Intellekt-Tsentr.

Nekhaev I. (2016) Analiz kachestva protsessa obucheniya s ispolzovaniem onlayn-kursov [Analyzing the Quality of Learning with MOOCs]. Proceedings of the *Best Practices in Digital Learning: 2nd Methodological Conference (Tomsk, 26–27 May 2016),* Tomsk: Tomsk State University, pp. 8–14.

Netology Group (2017) *Issledovanie rossiyskogo rynka onlayn-obrazovaniya i obrazovatelnykh tekhnologiy* [Research in Russia's Online Learning Market and Education Technology]. Available at: http://edumarket.digital (accessed 2 October 2018).

Nonaka I., Takeuchi H. (2011) *The Company Is the Creator of Knowledge. Origin and Development of Innovations in Japanese Firms*. Moscow: Olimp-Business.

O'Farrell L. (2017) *Using Learning Analytics to Support the Enhancement of Teaching and Learning in Higher Education*. Paper presented at National Forum for the Enhancement of Teaching and Learning in Higher Education. Available at: https://www.teachingandlearning.ie/wp-content/uploads/2018/01/Final_LA-Briefing-Paper_Web-with-doi.pdf (accessed 2 October 2018).

Roshhina J., Roshchin S., Rudakov V. (2018) Spros na massovye otkrytye onlayn-kursy (MOOC): opyt rossiyskogo obrazovaniya [The Demand for Massive Open Online Courses (MOOC): Evidence from Russian Education]. *Voprosy obrazovaniya / Educational Studies Moscow*, no 1, pp. 174–199. DOI: 10.17323/1814–9545–2018–1–174–199

Samokhin I., Sergeyeva M., Sokolova N., Marchenko E. (2018) Soderzhanie ponyatiya "effectivnost obrazovaniya" v kontekste inklyuzivnykh tendentsiy sovremennjy shkoly [Concept of "Educational Efficiency" in Context of Inclusive Trends of Modern School]. *Nauchnyy dialog*, no 1, pp. 278–288.

Sancho T., de Vries F. (2013) Virtual Learning Environments, Social Media and MOOCs: Key Elements in the Conceptualisation of New Scenarios in Higher Education. *Open Learning*, vol. 28, no 3, pp. 166–170.

Semenova T., Vilkova K., Shcheglova I. (2018) Rynok MOOK: perspektivy dlya Rossii [The MOOC Market: Prospects for Russia]. *Voprosy obrazovaniya / Educational Studies Moscow*, no 2, pp. 173–197. DOI: 10.17323/1814–9545–2018–2–173–197

Sharkey M., Ansari M. (2014). Deconstruct and Reconstruct: Using Topic Modeling on an Analytics Corpus. *LAK Data Challenge*. Available at: http://ceur-ws.org/Vol-1137/lakdatachallenge2014_submission_1.pdf (accessed 2 October 2018).

Shmelev A. (2013) *Prakticheskaya testologiya: testirovanie v obrazovanii, prikladnoy psikhologii i upravlenii personalom* [Testing in Education, Applied Psychology and Human Resource Management], Moscow: Maska.

Tyler-Smith K. (2006) Early Attrition among First Time E-Learners: A Review of Factors that Contribute to Drop-Out, Withdrawal and Non-Completion Rates of Adult Learners Undertaking E-Learning Programmes. *Journal of Online Learning Teaching*, vol. 2, no 2, pp. 73–85.

Usha Keshavamurthy, Guruprasad H. S. (2014) Learning Analytics: A Survey. *International Journal of Computer Trends and Technology (IJCTT)*, vol. 18, no 6, pp. 260–264.

Vishnyakova S. (1999) *Professionalnoe obrazovanie: Slovar. Klyuchevye ponyatiya, terminy, aktualnaya leksika* [Vocational Education: Vocabulary. Keywords, Terminology, Language], Moscow: Research and Methodology Center for Secondary Vocational Education.

Zagvyazinsky V., Zakirova A., Strokova T. et al. (2008) *Pedagogicheskiy slovar* [Pedagogical Vocabulary], Moscow: Akademiya.

Zvonnikov A., Chelyshkova M. (2012) *Kontrol kachestva rezultatov obucheniya pri attestatsii (kompetentnostny podkhod). Ucheb. Posobie* [Monitoring Educational Quality in Teacher Appraisal (Competency-Based Approach). Study Guide], Moscow: Logos.