

DOI: 10.32517/0234-0453-2023-38-6-75-85

# ИСПОЛЬЗОВАНИЕ МЕТОДОВ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ ОЦЕНКИ КОМПЕТЕНЦИЙ В ХОДЕ ТЕСТИРОВАНИЯ

В. Н. Гусятников<sup>1</sup> ✉, Т. Н. Соколова<sup>2</sup>, А. И. Безруков<sup>1</sup>, И. В. Каюкова<sup>1</sup><sup>1</sup> Саратовский государственный технический университет имени Гагарина Ю. А., г. Саратов, Россия<sup>2</sup> Саратовская государственная юридическая академия, г. Саратов, Россия

✉ victorgsar@rambler.ru

## Аннотация

Согласно современным стандартам высшего образования (ФГОС 3++) обучение направлено на формирование у выпускника набора универсальных, общепрофессиональных и профессиональных компетенций. В связи с этим возникает задача оценки уровня их сформированности. Поскольку компетенция является междисциплинарным понятием, для ее оценки требуется конструировать задание, моделирующее реальную ситуацию, при выполнении которого могут потребоваться комплексные знания и навыки, относящиеся к разным областям знаний. Одним из современных инструментов текущего и итогового контроля является компьютерное тестирование. В этом случае в ходе одного сеанса тестирования желательно оценить сразу несколько компетенций при ограниченном количестве заданий. Задача является актуальной, поскольку в используемых в настоящее время классических одномерных одно-, двух- и трехпараметрических моделях Раша—Бирнбаума предполагается, что все задания в тесте относятся к одной области знаний.

Цель статьи — предложить методику интеллектуальной оценки нескольких компетенций по итогам одного сеанса тестирования, основанную на модификации алгоритма Байеса, в которой выбор следующего задания проводится в результате анализа энтропии (неопределенности) отнесения испытуемого к типовым комбинациям (паттернам) уровней подготовленности по разным компетенциям.

Разработана оригинальная методика регуляризации алгоритма Байеса на основе оценки изменения энтропии распределения вероятностей по паттернам и дивергенции Кульбака—Лейблера, которая позволяет выявить момент, когда испытуемый дает ответ, не соответствующий тому, что от него ожидают, исходя из анализа предыдущих ответов. Предложенная методика не только повышает устойчивость байесовского алгоритма измерения вероятности принадлежности студента к заданному паттерну за счет некоторого увеличения необходимого количества заданий, но и позволяет смоделировать действия экзаменатора в процессе оценки компетенций. В целом разработанный подход дает возможность объективной оценки уровней сформированности нескольких компетенций по результатам одного тестирования.

**Ключевые слова:** компьютерное тестирование, модель Раша, компетентностный подход, адаптивное тестирование, алгоритм Байеса, дивергенция Кульбака—Лейблера.

## Для цитирования:

Гусятников В. Н., Соколова Т. Н., Безруков А. И., Каюкова И. В. Использование методов искусственного интеллекта для оценки компетенций в ходе тестирования. *Информатика и образование*. 2023;38(6):75–85. DOI: 10.32517/0234-0453-2023-38-6-75-85.

# USE OF ARTIFICIAL INTELLIGENCE METHODS TO ASSESS COMPETENCIES DURING TESTING

V. N. Gusyatnikov<sup>1</sup> ✉, T. N. Sokolova<sup>2</sup>, A. I. Bezrukov<sup>1</sup>, I. V. Kayukova<sup>1</sup><sup>1</sup> Yuri Gagarin State Technical University of Saratov, Saratov, Russia<sup>2</sup> Saratov State Law Academy, Saratov, Russia

✉ victorgsar@rambler.ru

## Abstract

According to modern standards of higher education (FSES 3++), teaching is aimed at developing a set of universal, general professional, and professional competencies in a graduate. In this regard, the problem of assessing the level of their formation arises. Since competence is a multidisciplinary concept, its assessment requires designing a task that simulates a real situation, which may require complex knowledge and skills related to different areas of knowledge. One of the modern tools for current and final control is computer testing. In this case, during one testing session, it is desirable to evaluate several competencies at once with a limited number of tasks. The task is relevant, because in the currently used classical one-dimensional one-, two- and three-parameter Rasch—Birnbbaum models, it is assumed that all tasks in the test belong to the same area of knowledge.

The purpose of the article is to propose a methodology for the intellectual assessment of several competencies based on the results of one testing session, based on the modification of the Bayes algorithm, in which the choice of the next task is carried out as a result of

analyzing the entropy (uncertainty) of the subject's attribution to typical combinations (patterns) of preparedness levels for different competencies.

An original technique for regularizing the Bayes algorithm based on the assessment of the change in the entropy of the probability distribution over patterns and the Kullback—Leibler divergence has been developed, which makes it possible to identify the moment when the student gives an answer that does not correspond to what is expected from them, based on the analysis of previous answers. The proposed technique not only increases the stability of the Bayes algorithm for measuring the probability of a student belonging to a given pattern by slightly increasing the required number of tasks but also allows to model the actions of an examiner in the process of competencies assessment. In general, the developed approach makes it possible to objectively assess the levels of formation of several competencies based on the results of single test.

**Keywords:** computer testing, Rasch model, competency-based approach, adaptive testing, Bayes algorithm, Kullback—Leibler divergence.

**For citation:**

Gusyatnikov V. N., Sokolova T. N., Bezrukov A. I., Kayukova I. V. Use of artificial intelligence methods to assess competencies during testing. *Informatics and Education*. 2023;38(6):75–85. (In Russian.) DOI: 10.32517/0234-0453-2023-38-6-75-85.

## 1. Введение

Проблема оценки компетенций, сформированных в ходе обучения, является актуальной для всех уровней современного образования. Компьютерное тестирование, являясь эффективным и широко применяемым инструментом для контроля уровня подготовленности обучающегося, активно используется для решения данной проблемы. Однако работ, посвященных научному обоснованию использования тестирования как средства оценки уровня сформированности компетенций, явно недостаточно.

Компетенция является, как известно, скрытой (латентной) характеристикой личности, которая не может быть измерена непосредственно [1]. Для ее оценки применяются различные модели, позволяющие связать наблюдаемые и латентные параметры. Наиболее известная на сегодня современная теория педагогических измерений (*англ.* Item Response Theory, IRT) стала основой создания десятков моделей, позволяющих оценивать уровень подготовленности обучающихся с учетом разнородных факторов:

- способность обучающегося решать трудные задачи;
- вероятность угадывания правильного ответа;
- возможность повторного выполнения заданий и т. д. [2–5].

В то же время применение теории IRT и лежащей в ее основе модели Раша для контроля компетенций вызывает ряд возражений. Главная проблема состоит в следующем: в классических одномерных одно-, двух- и трехпараметрических моделях Раша—Бирнбаума предполагается, что все задания в тесте относятся к одной области знаний. В *однопараметрической модели* (модели Раша) задания различаются только по уровню трудности. В *двухпараметрической одномерной модели* (модели Бирнбаума), кроме уровня трудности задания, учитывается его дифференцирующая способность, а в *трехпараметрической модели* принимается во внимание и вероятность угадывания правильного ответа.

Компетенция является мультидисциплинарным понятием, поэтому для ее оценки требуется конструировать задание, моделирующее реальную ситуацию. Для его выполнения могут потребоваться комплексные знания и навыки, относящиеся к раз-

ным областям знаний. При определении уровня подготовленности обучающегося желательно с помощью ограниченного количества заданий оценить сразу несколько компетенций в ходе одного сеанса тестирования. Однако это является сложной задачей, для решения которой требуется применение интеллектуальных методов формирования теста и анализа результатов тестирования, поскольку традиционные методы обработки данных тестирования не дают удовлетворительных результатов. Еще одно возражение против методов тестирования, которое часто звучит в профессиональном сообществе, состоит в том, что тест, в отличие от живого экзаменатора, не учитывает индивидуальные особенности обучающегося и тонкие нюансы его взаимодействия с преподавателем, которые раскрываются в ходе устного экзамена.

**Цель статьи** — предложить методику интеллектуальной оценки нескольких компетенций по итогам одного сеанса тестирования, основанную на модификации алгоритма Байеса, в которой выбор следующего задания проводится в результате анализа энтропии (неопределенности) отнесения испытуемого к типовым комбинациям (паттернам) уровней подготовленности по разным компетенциям. Методика позволяет смоделировать действия опытного экзаменатора в ходе оценочных процедур.

## 2. Материалы и методы

Для решения задачи оценки нескольких компонентов знаний с помощью тестов, включающих комплексные, многоэтапные задания, была предложена **многомерная теория педагогических измерений** (*англ.* Multidimensional Item Response Theory, MIRT) [6–8]. Данная теория рассматривает модели двух типов: **компенсаторные** и **некомпенсаторные** [9, 10]. В основе *компенсаторных моделей* лежит предположение, что недостаток какого-либо знания, навыка или компетенции при выполнении задания может компенсироваться достаточным уровнем других навыков, знаний и компетенций. Интегральный уровень развития компетенций или навыков, используемый в данной модели, определяется суммированием отдельных компонентов с их весовыми коэффициентами, а ве-

роятность правильного выполнения задания устанавливается следующим образом:

$$P_s = \frac{\exp\left(\sum_{i=1}^n \alpha_i \theta_i - \delta\right)}{1 + \exp\left(\sum_{i=1}^n \alpha_i \theta_i - \delta\right)}, \quad (1)$$

где:

$n$  — число исследуемых компетенций;

$\theta_i$  — уровень подготовленности испытуемого по компетенции  $i$ ;

$\alpha_i$  — чувствительность задания к компетенции  $i$  (значимость этой компетенции для выполнения задания);

$\delta$  — трудность выполнения задания.

В *некомпенсаторных моделях предполагается, что для выполнения задания необходимы все компетенции и отсутствие или недостаточный уровень одной компетенции не могут быть скомпенсированы другими компетенциями*. Вероятность выполнения задания в этом случае равна произведению вероятностей успешного проявления каждой из необходимых компетенций.

В одном из вариантов многомерной некомпенсаторной модели предполагается, что трудность задания может быть описана как линейная комбинация трудностей различных этапов выполнения этого задания, а вероятность его правильного выполнения устанавливается следующим образом:

$$P_s = \frac{\exp\left(\theta - \sum_{i=1}^n \alpha_i \delta_i\right)}{1 + \exp\left(\theta - \sum_{i=1}^n \alpha_i \delta_i\right)}. \quad (2)$$

Эта формула часто используется в педагогическом тестировании для оценки результатов выполнения заданий, решаемых в виде последовательности шагов [11]. К таким заданиям, в частности, относятся задания второй части ЕГЭ, или задания на установление соответствия. При этом данная формула не удовлетворяет в полной мере условию аддитивности трудностей, так как результирующая вероятность выполнения задания не равна произведению вероятностей выполнения отдельных шагов [12]. Наконец, существуют варианты комбинированных многомерных моделей, в которых вероятность правильного ответа записывается в виде [13]:

$$P_s = \frac{\exp\left(\sum_{i=1}^n \alpha_i (\theta_i - \delta_i)\right)}{1 + \exp\left(\sum_{i=1}^n \alpha_i (\theta_i - \delta_i)\right)}. \quad (3)$$

Главный недостаток всех моделей такого типа состоит в сложности их калибровки и правильной педагогической интерпретации результатов тестирования [14]. Именно поэтому многомерная теория

педагогических измерений до настоящего времени не нашла широкого применения в проводимых исследованиях ни в нашей стране, ни за рубежом.

В данной работе предлагается принципиально иной подход, обобщающий рассмотренные подходы к использованию модели Раша для оценки нескольких компетенций в процессе тестирования. При этом сделаны стандартные для современной теории педагогических измерений упрощающие предположения:

- 1) компетенции и навыки человека не меняются в ходе ответов на вопросы;
- 2) результаты выполнения каждого тестового задания не зависят от результатов выполнения других тестовых заданий;
- 3) связь между уровнем компетенции и характеристиками заданий линейна;
- 4) ответы являются бинарными.

Предположим для определенности, что при выполнении некоторого задания могут быть использованы три компетенции (обозначим их  $K_1, K_2, K_3$ ). Если для выполнения задания необходима каждая из трех компетенций (некомпенсаторная модель), то вероятность выполнения задания будет равна произведению вероятностей  $P_i$ , то есть **компетенции влияют на успешность выполнения задания по правилам логического умножения** (логическая операция конъюнкции, обозначается: «И», AND):

$$P_s = P_1 \cdot P_2 \cdot P_3. \quad (4)$$

Например, для выполнения задания кроме предметных знаний ( $K_1$ ) требуется знание русского языка, на котором оно изложено ( $K_2$ ), а также знание базовых терминов дисциплины, использованных в условиях задачи ( $K_3$ ). Проводя занятия с российскими студентами, которые учились в российских школах, мы не замечаем влияния последних двух компетенций, так как вероятности  $P_2$  и  $P_3$  для них близки к единице. Ситуация резко меняется, когда занятия проводятся с иностранными студентами-первокурсниками. Они часто не могут выполнить простые для них задания по предмету из-за незнания языка или базовых терминов дисциплины, хотя компетенция  $K_1$  у них развита в достаточной степени.

**Другой вариант проявления нескольких компетенций при выполнении одного задания (компенсаторная модель) может соответствовать логической функции «исключающего ИЛИ»** (логическая операция строгой дизъюнкции, обозначается также: «сложение по модулю 2», XOR). Например, испытуемому поставлено задание: переправиться через широкую реку. Для выполнения задания он может:

- использовать коммуникативную компетенцию и договориться с лодочником;
- при развитости должного уровня физической подготовленности переплыть реку;
- построить плот, проявив мастерство в его постройке.

Иными словами, для выполнения задания необходимо иметь хотя бы одну достаточно развитую

компетенцию. В приведенном примере, если испытуемый решил построить плот, то не важно, насколько хорошо у него развиты другие компетенции. В основу представленной в работе имитационной модели, на которой будут сравниваться различные алгоритмы и методики тестирования, положен принцип использования наиболее развитой компетенции тестируемого. В этом случае вероятность выполнения задания будет равна максимальной вероятности из трех, если предполагать, что испытуемый сразу применит ту компетенцию, которая у него более развита и наиболее подходит для выполнения задания. Этот случай будет соответствовать ситуации, когда задания должны быть выполнены в ограниченное время и у испытуемого просто нет времени, чтобы делать попытки с помощью всех имеющихся у него компетенций:

$$P_s = \max(P_1, P_2, P_3). \quad (5)$$

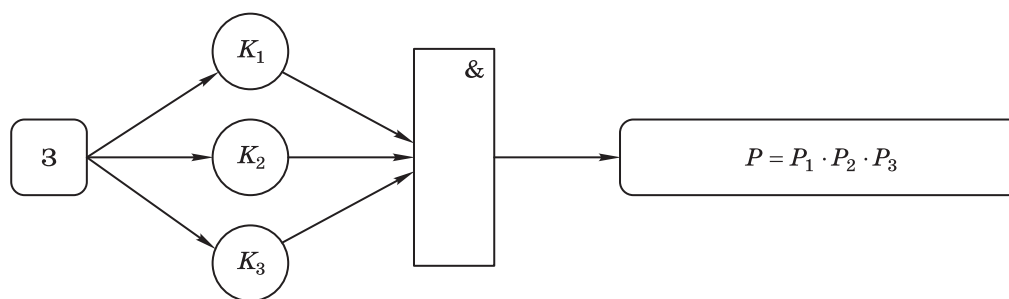


Схема AND

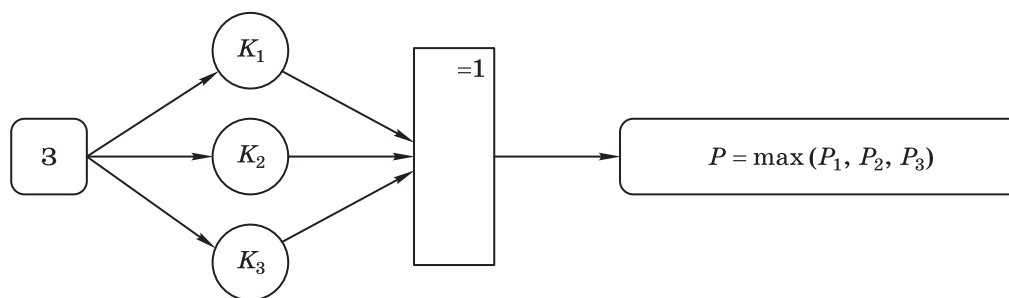


Схема XOR

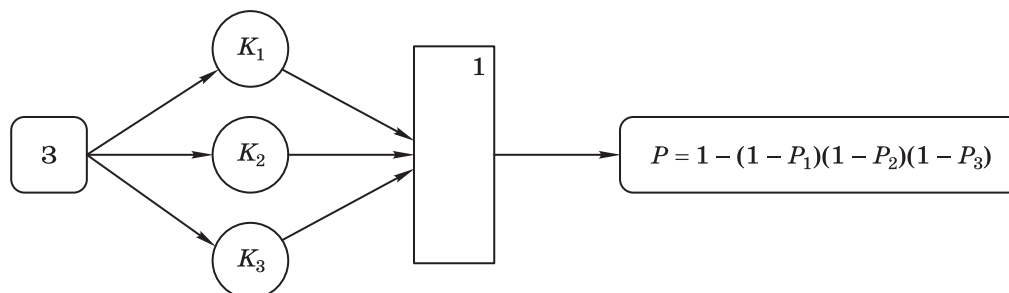


Схема OR

Еще один вариант выполнения задания с применением трех компетенций можно предложить, если рассмотреть ситуацию, когда время выполнения задания не ограничено и испытуемый может попытаться применить каждую из них. Тогда для вычисления вероятности выполнения задания можно использовать функцию логического сложения (логическая операция дизъюнкции, обозначается также: ИЛИ, OR). Вероятность невыполнения задания в этом случае будет равна произведению вероятностей его невыполнения с использованием каждой из компетенций. Следовательно, вероятность его выполнения можно описать следующей формулой:

$$P_s = 1 - (1 - P_1)(1 - P_2)(1 - P_3). \quad (6)$$

Рассмотренные варианты выполнения заданий с использованием нескольких компетенций иллюстрируются рисунком 1.

Рис. 1. Варианты определения вероятности выполнения задания, требующего нескольких компетенций  
Fig. 1. Options for determining the probability of completing a task requiring multiple competencies

Так как от трех логических переменных существует 256 логических функций, можно предложить много вариантов проявления компетенций при выполнении заданий. Эти варианты будут определяться характером самого задания, условиями проведения тестирования и индивидуальными особенностями испытуемого. При этом каждая компетенция проявляется и применяется независимо от других. Вероятность ее результативного применения описывается классической одномерной моделью Рапа, а их совместное применение описывается на более высоком уровне с помощью логических функций.

Вполне очевидно, что не всякое задание может служить для оценки нескольких компетенций и оценку не всяких компетенций можно объединить в одном задании. Однако есть области профессиональной деятельности, для которых в одном задании возможно использование нескольких компетенций. Например, в области информационных технологий одно и то же задание можно выполнить совершенно различными способами и инструментами.

В качестве поясняющего примера приведем следующую ситуацию. Предположим, что задание состоит в том, чтобы вычислить значение определенного интеграла с заданной точностью. Для решения этой задачи студент может выбрать различные методы. Хороший программист разработает небольшую программу, которая вычислит интеграл с заданной точностью методом Симпсона. Для укрупненной группы специальностей 09.03.00 «Информатика и вычислительная техника» использование данного метода будет соответствовать проявлению общепрофессиональной компетенции «Способен разрабатывать алгоритмы и программы, пригодные для практического применения»<sup>1</sup>, которая в ФГОС разных направлений подготовки имеет разные номера: ОПК-8 для направления подготовки 09.03.01 «Информатика и вычислительная техника»<sup>2</sup>, ОПК-6 для направлений подготовки 09.03.02 «Информационные системы и технологии»<sup>3</sup> и 09.03.04 «Про-

граммная инженерия»<sup>4</sup>, ОПК-7 для направления подготовки 09.03.03 «Прикладная информатика»<sup>5</sup>.

Знатор программы MatLab воспользуется для решения ее возможностями, проявив знание прикладных программ. Применение этого метода будет проявлением другой общепрофессиональной компетенции специалиста по направлению подготовки 09.03.03 «Прикладная информатика» — ОПК-2: «Способен понимать принципы работы современных информационных технологий и программных средств, в том числе отечественного производства, и использовать их при решении задач профессиональной деятельности»<sup>6</sup>. Наконец, хороший математик сумеет свести искомый интеграл к известным из справочников и по формуле Лейбница вычислит его значение, что будет соответствовать проявлению третьей общепрофессиональной компетенции из того же стандарта ОПК-1: «Способен применять естественнонаучные и общинженерные знания, методы математического анализа и моделирования, теоретического и экспериментального исследования в профессиональной деятельности»<sup>7</sup>.

Каждый из этих методов решения будет соответствовать проявлению «своей» компетенции. При этом сложность данного задания по отношению к разным компетенциям будет различной. По отношению к программированию это будет простой задачей, требующей умения организовать цикл и вычислять значения подынтегральной функции. Для математика это может быть значительно более сложным заданием, в зависимости от вида подынтегральной функции. Вероятность правильного выполнения этого задания будет вычисляться по формуле, соответствующей схеме XOR на рисунке 1.

Естественно, что по успешному результату выполнения одного такого задания мы не можем оценить, кто перед нами: хороший математик, или программист с навыками не ниже среднего уровня, или знатор программы MatLab не ниже начального уровня. Для получения полной картины мы должны предъявить испытуемому несколько заданий с различными уровнями трудности по каждой компетенции. Тогда по результатам их выполнения мы сможем оценить уровень развития каждой компетенции у испытуемого.

<sup>1</sup> Письмо Министерства науки и высшего образования Российской Федерации от 12 июля 2021 года № МН-5/4611 «О направлении информации». <https://www.garant.ru/products/ipo/prime/doc/402554726/>

<sup>2</sup> Приказ Министерства образования и науки Российской Федерации от 19 сентября 2017 года № 929 (ред. от 08.02.2021) «Об утверждении федерального государственного образовательного стандарта высшего образования — бакалавриат по направлению подготовки 09.03.01 «Информатика и вычислительная техника». [https://fgosvo.ru/uploadfiles/FGOS%20VO%203++/Bak/090301\\_B\\_3\\_15062021.pdf](https://fgosvo.ru/uploadfiles/FGOS%20VO%203++/Bak/090301_B_3_15062021.pdf)

<sup>3</sup> Приказ Министерства образования и науки Российской Федерации от 19 сентября 2017 года № 926 (ред. от 08.02.2021) «Об утверждении федерального государственного образовательного стандарта высшего образования — бакалавриат по направлению подготовки 09.03.02 «Информационные системы и технологии». [https://fgosvo.ru/uploadfiles/profstandart/090302\\_B\\_3\\_15062021.pdf](https://fgosvo.ru/uploadfiles/profstandart/090302_B_3_15062021.pdf)

<sup>4</sup> Приказ Министерства образования и науки Российской Федерации от 19 сентября 2017 года № 920 (ред. от 08.02.2021) «Об утверждении федерального государственного образовательного стандарта высшего образования — бакалавриат по направлению подготовки 09.03.04 «Программная инженерия». [https://fgosvo.ru/uploadfiles/FGOS%20VO%203++/Bak/090304\\_B\\_3\\_15062021.pdf](https://fgosvo.ru/uploadfiles/FGOS%20VO%203++/Bak/090304_B_3_15062021.pdf)

<sup>5</sup> Приказ Министерства образования и науки Российской Федерации от 19 сентября 2017 года № 922 (ред. от 08.02.2021) «Об утверждении федерального государственного образовательного стандарта высшего образования — бакалавриат по направлению подготовки 09.03.03 «Прикладная информатика». [https://fgosvo.ru/uploadfiles/FGOS%20VO%203++/Bak/090303\\_B\\_3\\_23062021.pdf](https://fgosvo.ru/uploadfiles/FGOS%20VO%203++/Bak/090303_B_3_23062021.pdf)

<sup>6</sup> Там же.

<sup>7</sup> Там же.

Очевидно, что алгоритм вычисления вероятностей успешного выполнения задания по схеме XOR не охватывает все множество возможных ситуаций. Например, некоторые задания могут потребовать проявления сразу нескольких компетенций (схема AND). В данной работе рассматривается только ситуация, соответствующая схеме XOR на рисунке 1.

Следует отметить, что в большинстве работ тестирование рассматривается как процедура оценки каких-либо характеристик испытуемого, то есть описывается как процесс измерения исследуемой характеристики. Такой подход вполне оправдан, когда тестирование применяется для проведения конкурсного отбора, например, при проведении ЕГЭ.

Однако большинство промежуточных и итоговых аттестаций обучающихся не требует точного ранжирования всех испытуемых. Типичный экзамен или зачет у большинства преподавателей проходит как процедура классификации обучающихся. На экзамене их надо распределить по четырем группам: «отличники», «хорошисты», «троечники» и «неуспевающие». На зачете — всего на две группы: «зачет» и «незачет».

Задачи классификации хорошо решаются методами искусственного интеллекта. Однако использование сложных методов анализа результатов тестирования с элементами искусственного интеллекта [15] порождает другую проблему, обусловленную недостаточным уровнем доверия к подобным системам оценки как со стороны обучающихся, так и со стороны преподавателей. Уровень доверия к интеллектуальным системам в значительной мере зависит от степени прозрачности такой системы, простоты и понятности используемых в ней алгоритмов. Например, если мы будем использовать для оценки искусственную нейронную сеть, то вряд ли удастся убедить пользователей этой системы оценивания в достоверности выдаваемых результатов, как бы хорошо мы ни проектировали и ни обучали ее, так как для пользователей она будет черным ящиком.

В данной работе решается задача оценки каждой из трех компетенций по четырехбалльной шкале. Таким образом, общее количество уникальных комбинаций уровней развития анализируемых компетенций у испытуемых составляет 64. Задачу оценки этих компетенций можно свести к задаче классификации, то есть отнесения испытуемого к одному из заранее определенных типов (паттернов) [16].

Будем считать, что:

- 1) уровень подготовленности испытуемого характеризуется тремя числами  $\theta_i$   $i = 1, 2, 3$ , соответствующими трем разным компетенциям и принимающими ряд дискретных значений в логитах  $(-3, -1, 1, 3)$ . Эти значения в логитах соответствуют традиционным оценкам: «неудовлетворительно», «удовлетворительно», «хорошо», «отлично»<sup>1</sup>;

- 2) трудность каждого задания также задается тремя числами  $\delta_i$   $i = 1, 2, 3$ , характеризующими его сложность по отношению к отдельной компетенции;

- 3) вероятность результативного проявления отдельной компетенции при выполнении задания описывается в рамках одномерной однопараметрической модели Раппа:

$$P_i = \frac{\exp(\theta_i - \delta_i)}{1 + \exp(\theta_i - \delta_i)}, \quad i = 1, 2, 3; \quad (7)$$

- 4) вероятность правильного выполнения задания вычисляется по правилу исключаящего ИЛИ:

$$P_s = \max(P_1, P_2, P_3). \quad (8)$$

Если предположить, что имеется банк тестовых заданий с известными параметрами трудности по отношению к каждой компетенции, то можно рассчитать вероятности выполнения любого задания из этого банка испытуемым, относящимся к любому паттерну [17]. Это дает возможность применить метод Байеса для вычисления вероятностей принадлежности испытуемого к тому или иному паттерну и уточнять полученные значения после каждого выполненного задания с учетом результата его выполнения.

### 3. Результаты исследования

В предыдущих исследованиях [18, 19] авторы разработали имитационную модель процесса тестирования, использующую предложенный подход к решению задачи оценки нескольких компетенций в ходе одного сеанса. В этой модели перед выполнением первого задания предполагается, что вероятность принадлежности испытуемого к каждому из паттернов равна. После выполнения очередного задания значения вероятности принадлежности к каждому из паттернов пересчитываются по формуле Байеса. По мере выполнения заданий теста значения вероятности принадлежности к каким-то паттернам уменьшаются, а к каким-то растут. При этом каждое следующее задание выбирается таким образом, чтобы его информационная функция (количество информации об испытуемом, которое можно получить с помощью данного задания) была максимальна для паттерна, текущая вероятность принадлежности испытуемого к которому максимальна. Такой выбор следующего задания реализует алгоритм адаптивного тестирования и хорошо согласуется с используемым методом Байеса.

Исследования, проведенные на разработанной имитационной модели [19] с использованием ком-

в реальных тестах. Перевод в стандартные оценки четырехбалльной шкалы выполняется с помощью линейного преобразования (результат в логитах умножается на 0,5 плюс константа 3,5) (например, см. [5]).

<sup>1</sup> Диапазон уровней подготовленности от  $-3$  до  $3$  логитов определяется диапазоном трудностей заданий в тесте и соответствует типичным значениям, наблюдаемым

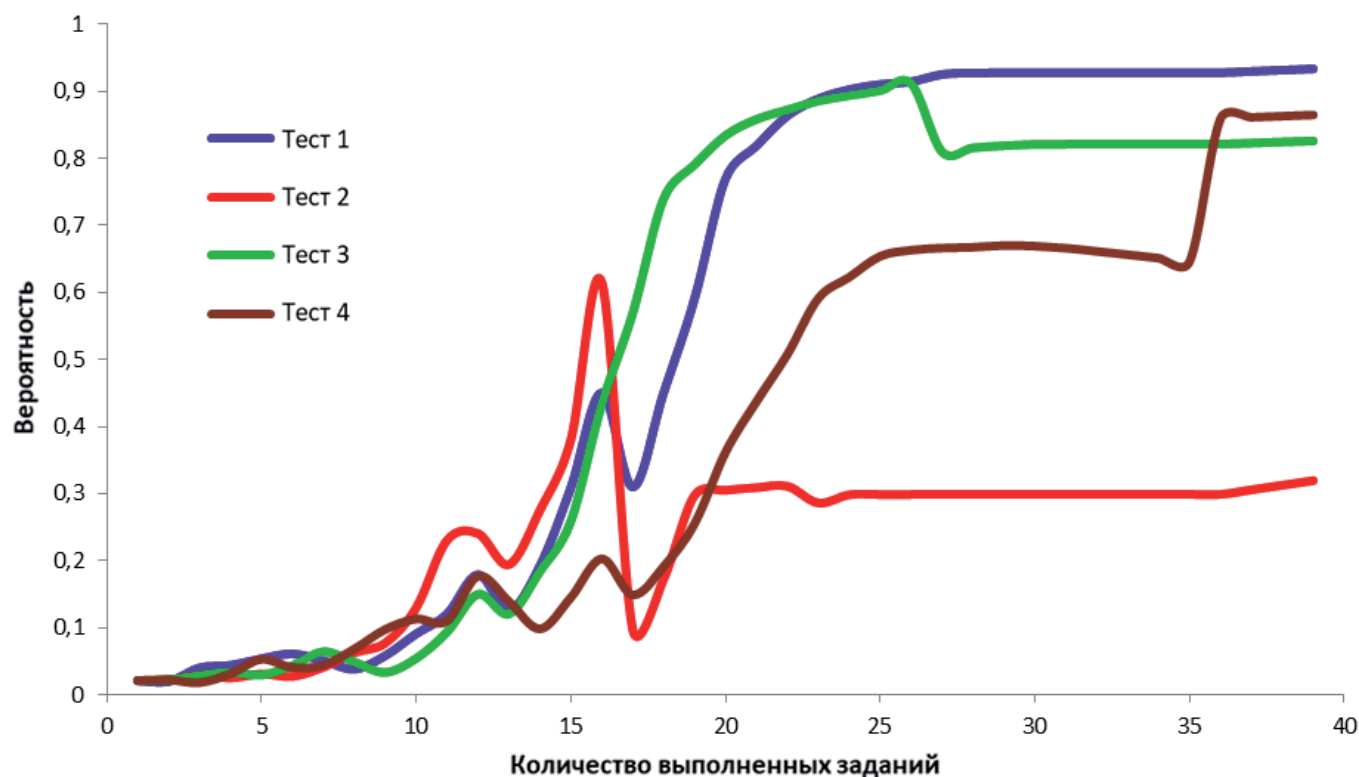


Рис. 2. Зависимость вероятности принадлежности студента заданному для него паттерну от числа выполненных заданий при нескольких имитациях процесса тестирования

Fig. 2. Dependence of the student's probability of belonging to the pattern set for him on the number of completed tasks with several simulations of the testing process

пенсаторной модели успешного выполнения задания по схеме XOR (см. рис. 1), выявили низкую устойчивость алгоритма Байеса к случайным вариантам ответов испытуемого.

На рисунке 2 представлены результаты четырех реализаций численного эксперимента для студента, имеющего хороший уровень подготовки по двум компетенциям и удовлетворительный уровень по третьей компетенции.

Кривые, отображающие изменение вероятности принадлежности испытуемого к заданному для него паттерну, показывают, что по мере выполнения заданий теста байесовская вероятность изменяется по разным сценариям. Причиной такой неустойчивости является особенность самого алгоритма Байеса, который становится излишне чувствительным к случайным факторам, влияющим на результат выполнения задания. При этом существенное влияние на траекторию изменения вероятности оказывают ответы на первые 10–15 вопросов, а последующие ответы слабо влияют на результат. В системах с искусственным интеллектом, использующих алгоритм Байеса, для контроля процесса классификации используют анализ поведения энтропии распределения вероятностей и дивергенции Кульбака—Лейблера, характеризующей изменение количества информации об испытуемом после получения от него очередного ответа [20]. На рисунке 3 показано изменение

энтропии распределения вероятностей по паттернам, дивергенции Кульбака—Лейблера и вероятности принадлежности студента заданному для него паттерну (см. кривая «Тест 2» на рис. 2) в ходе процесса тестирования.

Хорошо видно, что резкие изменения вероятности принадлежности студента к заданному паттерну сопровождаются «нетипичными» изменениями энтропии и резкими всплесками дивергенции. Каждый следующий вопрос, выбранный адаптивным алгоритмом, должен плавно снижать энтропию (степень неопределенности оценки испытуемого). Однако выполнение или невыполнение задания — случайное событие, и его вероятность тем ближе к  $1/2$ , чем ближе уровень трудности задания к уровню подготовленности испытуемого. В этом случае адаптивный алгоритм становится неустойчивым, его «мнение об испытуемом» может существенно поменяться по результатам выполнения одного задания.

Резкое уменьшение энтропии распределения вероятностей по паттернам или ее увеличение, а также резкий всплеск дивергенции Кульбака—Лейблера после очередного выполненного задания могут свидетельствовать о том, что с полученным ответом не все в порядке. Например, студент, показывающий средний уровень знаний, неожиданно ответил на сложный вопрос. Возможно, он угадал правильный ответ или воспользовался подсказкой.

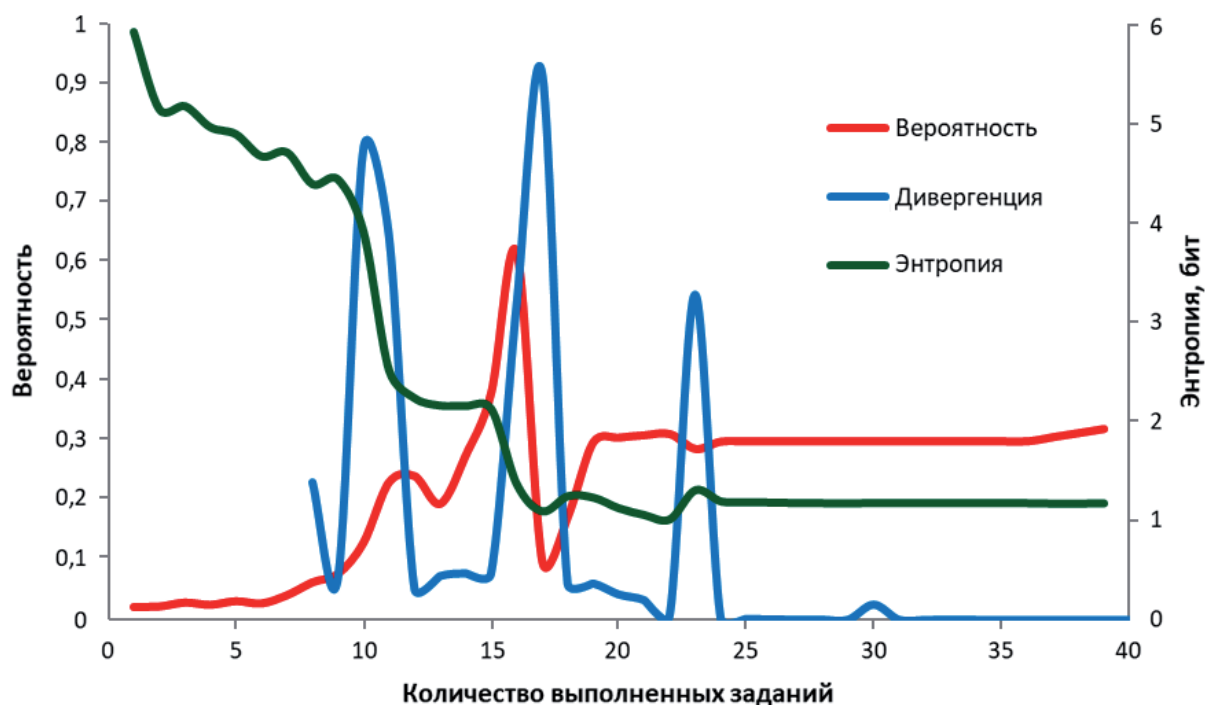


Рис. 3. Зависимость энтропии распределения вероятностей по паттернам для кривой «Тест 2» (см. рис. 2) и дивергенции Кульбака—Лейблера для этой же кривой от числа выполненных заданий

Fig. 3. The dependence of the entropy of the probability distribution over patterns for the Test 2 curve (in Fig. 2) and the Kullback—Leibler divergence for the same curve on the number of completed tasks

Опытный экзаменатор при получении ответа, не соответствующего сложившемуся мнению об испытуемом, задает дополнительный вопрос того же уровня трудности и только после этого формирует свое мнение. Анализ изменения энтропии и дивергенции Кульбака—Лейблера позволяет распознать такие моменты резкой смены «мнения об испытуемом» в ходе тестирования.

Для регуляризации получаемых решений предложена следующая процедура. Если после выполнения очередного задания дивергенция Кульбака—Лейблера испытывает резкий всплеск или энтропия распределения вероятностей по паттернам увеличивается, то результат выполнения данного задания не учитывается при расчете байесовской вероятности. Следующее задание, которое предъявляется студенту, имеет такие же параметры сложности, как и только что выполненное. Данная процедура не применяется во время выполнения первых восьми заданий, когда формируется «первичное» мнение об испытуемом.

На рисунке 4 представлены результаты тех же четырех реализаций тестирования, что и на рисунке 2, но с применением предложенного метода регуляризации байесовского алгоритма.

Рисунок 4 демонстрирует, что применение регуляризации на основе анализа изменения энтропии распределения вероятностей по паттернам и дивергенции Кульбака—Лейблера позволяет повысить устойчивость байесовского алгоритма измерения вероятности принадлежности студента к заданному

паттерну за счет некоторого увеличения необходимого количества заданий. При этом количество вопросов, необходимых для оценки уровня сформированности сразу трех компетенций в ходе одного сеанса тестирования с использованием четырехбалльной шкалы оценивания, не превышает двух-трех десятков.

Одним из недостатков существующих систем тестирования и анализа результатов теста является часто критикуемый формальный, механистический подход, не учитывающий индивидуальные особенности обучаемого. Поэтому часто звучат призывы вернуться в российском образовании к старой, «советской» системе оценивания. **Представленный подход позволяет имитировать действия опытного экзаменатора в процессе оценки результатов обучения с помощью тестирования.** На основе анализа поведения энтропии и дивергенции Кульбака—Лейблера для распределения вероятностей принадлежности паттернам разработан алгоритм выявления «нетипичных» ответов испытуемого, не соответствующих тому, что от него ожидают. Предложено в этом случае задавать дополнительный вопрос с теми же параметрами, имитируя поведение преподавателя на очном экзамене.

#### 4. Выводы

Разработанная методика установления паттерна, к которому принадлежит испытуемый, легко встраивается в системы адаптивного тестирования, так

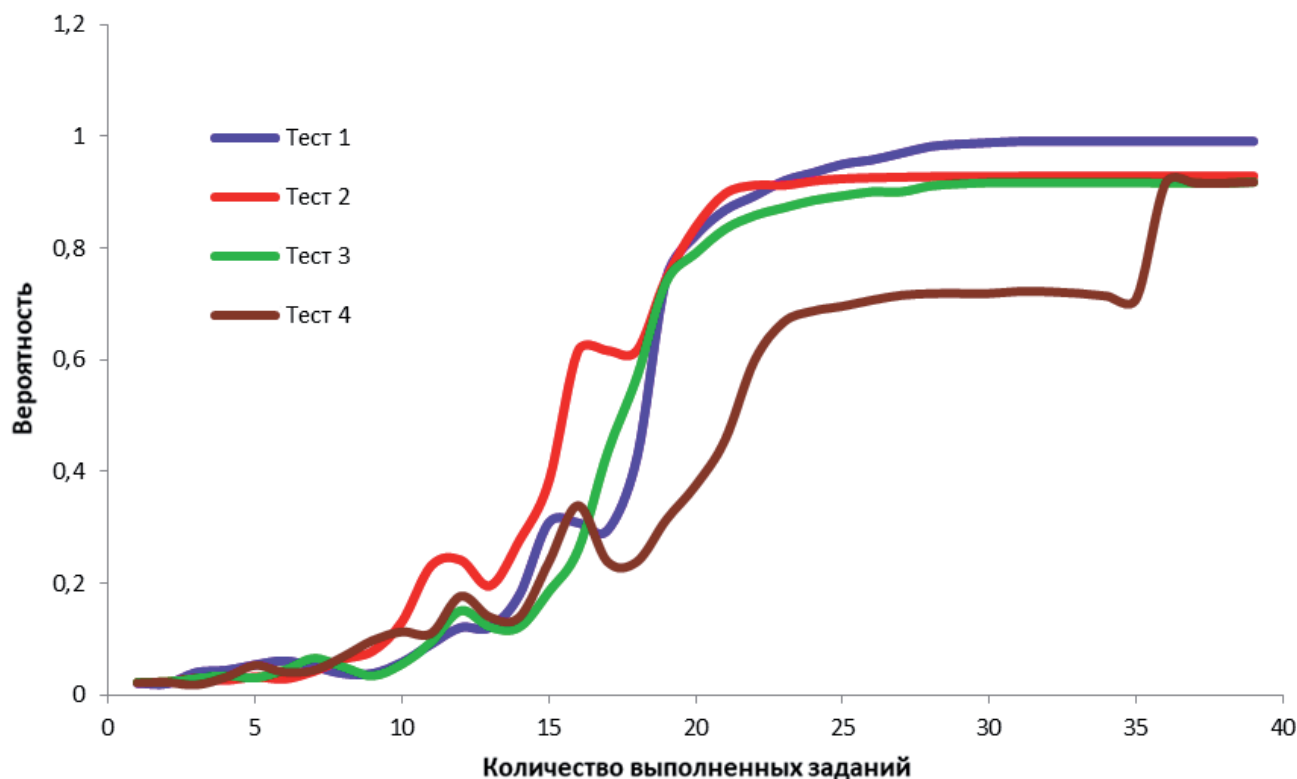


Рис. 4. Зависимость вероятности принадлежности студента заданному для него паттерну от числа выполненных заданий при нескольких имитациях процесса тестирования с регуляризацией на основе анализа энтропии и дивергенции Кульбака—Лейблера

Fig. 4. Dependence of the student's probability of belonging to the pattern set for him on the number of completed tasks with several simulations of the testing process with regularization based on entropy analysis and Kullback—Leibler divergence

как позволяет организовать итерационный процесс уточнения информации об испытуемом, рассчитывая новые значения условных вероятностей принадлежности испытуемого паттернам после каждого выполненного задания. При этом имеется возможность выбирать следующее задание, исходя из максимума информационной функции заданий, рассчитанного на основе уже имеющихся данных об испытуемом и о каждом тестовом задании, оставшемся в банке заданий.

Разработанная оригинальная методика регуляризации алгоритма Байеса на основе оценки изменения энтропии распределения вероятностей по паттернам и дивергенции Кульбака—Лейблера позволяет выявить момент, когда испытуемый дает ответ, не соответствующий тому, что от него ожидают, причем опорой служит анализ предыдущих ответов. Как поступает опытный экзаменатор, когда испытуемый дает неожиданный ответ, не соответствующий уже сложившемуся мнению о нем? Вместо того, чтобы резко изменить свое мнение об испытуемом, экзаменатор постарается задать ему дополнительный вопрос того же уровня трудности. И только после этого сформирует свое мнение. Таким образом, разработанная методика не только повышает устойчивость байесовского алгоритма измерения вероятности принадлежности студента к заданному паттерну за счет

некоторого увеличения необходимого количества заданий, но и позволяет смоделировать действия экзаменатора в процессе оценки компетенций.

В целом разработанный подход дает возможность объективной оценки уровней сформированности нескольких компетенций по результатам одного тестирования.

#### Список источников / References

1. Маслак А. А. Теория и практика измерения латентных переменных в образовании: монография. М.: Юрайт; 2022. 255 с. EDN: RFKEIY.  
[Maslak A. A. Theory and practice of measuring latent variables in education: Monograph. Moscow, Yurait; 2022. 255 p. (In Russian.) EDN: RFKEIY.]
2. Нейман Ю. М., Хлебников В. А. Введение в теорию моделирования и параметризации педагогических тестов. М.: Прометей; 2000. 168 с.  
[Neiman Yu. M., Khlebnikov V. A. Introduction to the theory of modeling and parameterization of pedagogical tests. Moscow, Promethey; 2000. 168 p. (In Russian.)]
3. Ким В. С. Тестирование учебных достижений: монография. Уссурийск: УГПИ; 2007. 215 с. EDN: QWGVHH.  
[Kim V. S. Testing of learning achievements: Monograph. Ussuriysk, UGPI; 2007. 215 p. (In Russian.) EDN: QWGVHH.]
4. Аванесов В. С. Item response theory: основные понятия и положения. Педагогические измерения. 2007;(2):3–28. Режим доступа: <https://testolog.narod.ru/Theory59.html>

[Avanesov V. S. Item response theory: Basic concepts and provisions. *Pedagogicheskiye Izmereniya*. 2007;(2):3–28. (In Russian.) Available at: <https://testolog.narod.ru/Theory59.html>]

5. Чельшкова М. Б. Теория и практика конструирования педагогических тестов: учебное пособие. М.: Логос; 2002. 432 с.

[Chelyshkova M. B. Theory and practice of designing pedagogical tests: Study guide. Moscow, Logos; 2002. 432 p. (In Russian.)]

6. McDonald R. P. A basis for multidimensional item response theory. *Applied Psychological Measurement*. 2000;24(2):99–114. DOI: 10.1177/01466210022031552.

7. Ackerman T. A., Gierl M. J., Walker C. M. Using multidimensional item response theory to evaluate educational and psychological tests. *Educational Measurement: Issues and Practice*. 2003;22(3):37–53. DOI: 10.1111/j.1745-3992.2003.tb00136.x.

8. Bartolucci F. A class of multidimensional IRT models for testing unidimensionality and clustering items. *Psychometrika*. 2007;72(2):141–157. DOI: 10.1007/s11336-005-1376-9.

9. Freeman L. Assessing model-data fit for compensatory and non-compensatory multidimensional item response models using Vuong and Clarke statistics. *Theses and Dissertations*. University of Wisconsin, Milwaukee, US, 2016:1366. Available at: <https://dc.uwm.edu/etd/1366>

10. Fox J. P., Entink R. K., Avetisyan M. Compensatory and non-compensatory multidimensional randomized item response models. *British Journal of Mathematical and Statistical Psychology*. 2014;67(1):133–152. DOI: 10.1111/bmsp.12012.

11. Карданова Е. Ю. Доказательство применимости полиномической модели Г. Раша. *Вестник Новгородского государственного университета*. 2006;(39):13–15. EDN: MNJEKL.

[Kardanova E. Yu. Proof of the applicability of G. Rasch's polytomous model. *Vestnik NovSU*. 2006;(39):13–15. (In Russian.) EDN: MNJEKL.]

12. Попов А. П. Новое направление в компьютерном тестировании. *Наука и мир*. 2013;(1(1)):38–46. EDN: RQCKTZ.

[Popov A. P. The recent trends in computer-based testing. *Science and World*. 2013;(1(1)):38–46. (In Russian.) EDN: RQCKTZ.]

13. Wu M., Davis R. L., Domingue B. W., Piech C., Goodman N. D. Variational item response theory: Fast, accurate, and expressive. *Proc. of the 13th Int. Conf. on Educational Data Mining*. Montreal, Canada, 2020:257–268. DOI: 10.48550/arXiv.2002.00276.

14. Reckase M. D. Multidimensional item response theory. Springer New York, US; 2009. 349 p. DOI: 10.1007/978-0-387-89976-3.

15. Дворяткина С. Н. Интеграция фрактальных и нейросетевых технологий в педагогическом контроле и оценке знаний обучаемых. *Вестник Российского университета дружбы народов. Серия: Психология и педагогика*. 2017;14(4):451–465. EDN: ZRKOQH. DOI: 10.22363/2313-1683-2017-14-4-451-465.

[Dvoryatkina S. N. Integration of fractal and neural network technologies in pedagogical monitoring and assessment of knowledge of trainees. *RUDN Journal of Psychology and Pedagogics*. 2017;14(4):451–465. (In Russian.) EDN: ZRKOQH. DOI: 10.22363/2313-1683-2017-14-4-451-465.]

16. Куравский Л. С., Юрьев Г. А., Ушаков Д. В., Юрьева Н. Е., Валуева Е. А., Лаптева Е. М. Диагностика по тестовым траекториям: метод паттернов. *Экспериментальная психология*. 2018;11(2):77–94. EDN: TKLBMZ. DOI: 10.17759/exppsy.2018110206.

[Kuravsky L. S., Yuryev G. A., Ushakov D. V., Yuryeva N. E., Valueva E. A., Lapteva E. M. Diagnostics basing on testing paths: The method of patterns. *Experimental Psychology*. 2018;11(2):77–94. (In Russian.) EDN: TKLBMZ. DOI: 10.17759/exppsy.2018110206.]

logy. 2018;11(2):77–94. (In Russian.) EDN: TKLBMZ. DOI: 10.17759/exppsy.2018110206.]

17. Liu Y., Hu G., Cao L., Wang X., Chen M.-H. A comparison of Monte Carlo methods for computing marginal likelihoods of item response theory models. *Journal of the Korean Statistical Society*. 2019;48(4):503–512. DOI: 10.1016/j.jkss.2019.04.001.

18. Соколова Т. Н., Гусятников В. Н., Безруков А. И., Каюкова И. В. Методика оценки набора компетенций на основе результатов тестирования. *Фундаментальные исследования*. 2020;(12):209–215. EDN: IHYEON. DOI: 10.17513/fr.42935.

[Sokolova T. N., Gusyatinikov V. N., Bezrukov A. I., Kayukova I. V. Methodology for assessing a set of competencies based on testing results. *Fundamental Research*. 2020;(12):209–215. (In Russian.) EDN: IHYEON. DOI: 10.17513/fr.42935.]

19. Гусятников В. Н., Соколова Т. Н., Безруков А. И., Каюкова И. В. Адаптивная модель тестирования нескольких компетенций на основе алгоритма Байеса. *Современные наукоемкие технологии*. 2022;(1):40–46. EDN: KPDACI. DOI: 10.17513/snt.39007.

[Gusyatinikov V. N., Sokolova T. N., Bezrukov A. I., Kayukova I. V. Adaptive model for testing several competencies based on the Bayes algorithm. *Modern High Technologies*. 2022;(1):40–46. (In Russian.) EDN: KPDACI. DOI: 10.17513/snt.39007.]

20. Кульбак С. Теория информации и статистики. М.: Наука; 1967. 408 с.

[Kulbak S. Information theory and statistics. Moscow, Nauka; 1967. 408 s. (In Russian.)]

#### Информация об авторах

Гусятников Виктор Николаевич, доктор физ.-мат. наук, профессор, профессор кафедры «Информационно-коммуникационные системы и программная инженерия», Институт прикладных информационных технологий и коммуникаций, Саратовский государственный технический университет имени Гагарина Ю. А., г. Саратов, Россия; ORCID: <https://orcid.org/0000-0002-8560-4588>; e-mail: victorgsar@rambler.ru

Соколова Татьяна Николаевна, канд. экон. наук, доцент, доцент кафедры информационного права и цифровых технологий, Саратовская государственная юридическая академия, г. Саратов, Россия; ORCID: <https://orcid.org/0000-0001-6175-5325>; e-mail: tnsokol@yandex.ru

Безруков Алексей Иосифович, канд. экон. наук, доцент, доцент кафедры «Информационно-коммуникационные системы и программная инженерия», Институт прикладных информационных технологий и коммуникаций, Саратовский государственный технический университет имени Гагарина Ю. А., г. Саратов, Россия; ORCID: <https://orcid.org/0000-0002-8726-6404>; e-mail: bezr\_Alex@mail.ru

Каюкова Инна Викторовна, аспирант кафедры «Информационно-коммуникационные системы и программная инженерия», Институт прикладных информационных технологий и коммуникаций, Саратовский государственный технический университет имени Гагарина Ю. А., г. Саратов, Россия; ORCID: <https://orcid.org/0000-0002-2691-1756>; e-mail: i.v.kayukova@mail.ru

#### Information about the authors

Victor N. Gusyatinikov, Doctor of Sciences (Physics and Mathematics), Professor, Professor at the Department “Information and Communication Systems and Software Engineering”, Institute of Applied Information Technologies and Communication, Yuri Gagarin State Technical University of Saratov, Saratov, Russia; ORCID: <https://orcid.org/0000-0002-8560-4588>; e-mail: victorgsar@rambler.ru

Tatyana N. Sokolova, Candidate of Sciences (Economics), Docent, Associated Professor at the Department of Information Law and Digital Technologies, Saratov State Law Academy, Saratov, Russia; ORCID: <https://orcid.org/0000-0001-6175-5325>; e-mail: tnsokol@yandex.ru

**Aleksey I. Bezrukov**, Candidate of Sciences (Economics), Docent, Associated Professor at the Department “Information and Communication Systems and Software Engineering”, Institute of Applied Information Technologies and Communication, Yuri Gagarin State Technical University of Saratov, Saratov, Russia; *ORCID*: <https://orcid.org/0000-0002-8726-6404>; *e-mail*: bezr\_Alex@mail.ru

**Inna V. Kayukova**, a postgraduate student at the Department “Information and Communication Systems and Software

Engineering”, Institute of Applied Information Technologies and Communication, Yuri Gagarin State Technical University of Saratov, Saratov, Russia; *ORCID*: <https://orcid.org/0000-0002-2691-1756>; *e-mail*: i.v.kayukova@mail.ru

*Поступила в редакцию / Received*: 29.05.23.

*Поступила после рецензирования / Revised*: 17.08.23.

*Принята к печати / Accepted*: 22.08.23.