

УДК 37

## Ранговая корреляция оценок тестирования

**Шайдуллина Наталья Константиновна**

Старший преподаватель,  
кафедра информатики и прикладной математики,  
Казанский национальный исследовательский технологический университет,  
420015, Российская Федерация, Казань, ул. Карла Маркса, 68;  
e-mail: nshaydullina@yandex.ru

### Аннотация

В статье рассматривается проблема повышения качества тестирования как части автоматизированной системы обучения. Электронная система обучения является метрической, т. е. все показатели качества обучения переводятся в числовой формат. Действует она и в рамках компетентностного подхода. В работе представлен способ повышения объективности оценки тестирования. Объективность рассматривается с нескольких сторон. Во-первых, необходимо нивелировать ответы «наугад». Во-вторых, следует максимально полно оценить знания. В-третьих, нужно согласовать между собой критерии, которые используются при выставлении оценки. Для получения итоговой оценки строятся два числовых многомерных вектора. Они отражают качество ответов на декларативные и процедурные вопросы. Их согласованность является основным показателем объективности. Критерием согласованности выступает ранговый коэффициент корреляции Спирмена, значение которого при необходимости корректируется динамически.

### Для цитирования в научных исследованиях

Шайдуллина Н.К. Ранговая корреляция оценок тестирования // Педагогический журнал. 2018. Т. 8. № 6А. С. 77-82.

### Ключевые слова

Обучение, автоматизированная система обучения, электронная система обучения, качество обучения, компетентностный подход.

## Введение

Современный этап развития образовательного процесса характеризуется наличием множества автоматизированных обучающих систем. Дидактические системы нового поколения должны обладать основными свойствами классических (традиционных) систем и в то же время еще дополнительными, которые в современных условиях позволяют вывести их на новый качественный уровень по эффективности [Нуриев, Старыгина, Гибадуллина, 2016]. Эти системы должны быть построены на принципиально новой основе с внедрением фундаментальных закономерностей педагогики, психологии, дидактики и инженерии. Многие функции преподавателя в них должны быть заменены элементами искусственного интеллекта [Нуриев, Старыгина, Дидактическая инженерия: подготовка..., 2016, 62].

Электронные образовательные системы могут являться как инструментом самообразования, так и средством контроля качества традиционного процесса обучения. Главным средством контроля качества в таких системах является тестирование [Нуриев, Старыгина, Крылов, 2015, 549]. Если автоматизированное тестирование является средством оценивания знаний, то система тестирования должна быть объективной и адаптивной. Самый простой способ достигнуть объективности – создать огромное количество вопросов, на которые должен будет ответить каждый обучающийся. Это чревато очень большими временными затратами с обеих сторон. Другой способ – усложнение самой системы. Сложность системы как программного продукта означает наличие большого количества программных модулей. Это, во-первых, повышает дороговизну разработки, во-вторых, усложняет бизнес-логику, в-третьих, увеличивает время разработки, отладки и тестирования информационной обучающей системы. Кроме того, сложная логика будет являться преградой для самих преподавателей в процессе создания курсов. В контексте вышесказанного мы хотим предложить свой критерий объективности оценки тестирования и способ ее повышения.

### Согласованность оценок тестирования

В работе «Автоматизированная система обучения с уровневой структурой» [Шайдуллина, 2016] нами была предложена модульная система обучения. Ее структура представлена в табл. 1.

**Таблица 1 - Структура модульной системы обучения**

	<b>Раздел 1</b>	<b>Раздел 2</b>	<b>...</b>	<b>Раздел n</b>
Уровень 1	Модуль (1,1)	Модуль (1,2)	...	Модуль (1,n)
Уровень 2	Модуль (2,1)	Модуль (2,2)	...	Модуль (2,n)
...	...	...	...	...
Уровень m	Модуль (m,1)	Модуль (m,2)	...	Модуль (m,n)

Идея состоит в разбиении всего курса на учебные модули. Разделы соответствуют разделам учебной дисциплины. Уровни градируют материал по степени сложности. Продвижение по курсу осуществляется по спирали: сначала по горизонтали – все разделы изучаются на одном уровне сложности, потом по вертикали – при успешном прохождении определенного уровня осуществляется переход на следующий. Такой подход, по нашему мнению, способствует наиболее успешному обучению, так как позволяет обучаемому двигаться со своей скоростью, оставаясь в зоне ближайшего развития [Нуриев, Старыгина, Ахметшин, 2016, 561]. Успешность освоения уровня – это сложная комплексная оценка, порядок формирования которой подробно описан рядом исследователей [Нуриев и др., 2013, 421-422; Шайдуллина, 2016]. В данной работе

мы хотим остановиться на оценке результатов тестирования.

Все вопросы теста каждого модуля разбиваются на две категории: вопросы на полноту и вопросы на целостность [Нуриев, Старыгина, Ахметшин, 2015, 65-66]. Полноту знаний оценивают декларативные вопросы, обозначим их оценку через  $COM(i,j)$ , целостность – вопросы процедурные, их оценка обозначена как  $INT(i,j)$  [Нуриев, Старыгина, 2013, 53]. Если вопросы теста равнозначны по сложности, то оценка получается с помощью деления числа правильных ответов на число всех заданных вопросов. Если же в категории присутствуют вопросы разной сложности, то каждому вопросу ставится в соответствие степень его сложности из интервала  $[0,1]$ . Тогда оценка – это отношение суммы степеней сложности правильных ответов на сумму степеней сложности всех заданных вопросов. Имеются работы [Нуриев, Старыгина Гибадуллина, 2016, 401; Шайдуллина, Торкунова, 2016, www], в которых предлагается оценить глубину усвоенных знаний с помощью оценки  $DEP(i,j) = COM(i,j) * INT(i,j)$ .

Теперь мы хотим рассмотреть оценку глубины знаний несколько иначе. Мы попытаемся согласовать между собой оценки полноты и целостности. Согласованность оценок, во-первых, отразит системность знаний, во-вторых, нивелирует случайность выбранных ответов.

Формирование оценок полноты и целостности будем осуществлять следующим образом. Вопросы каждой категории разобьем на пять групп. Каждой группе поставим в соответствие степень сложности вопросов, т. е. очень легкие вопросы имеют степень 1, легкие – 2, средней трудности – 3, сложные – 4, очень сложные – 5. Количество вопросов в разных группах сложности одной категории и/или в одинаковых группах сложности разных категорий может не совпадать. После того как обучающийся ответит на определенное количество вопросов из обеих категорий, формируются два пятимерных вектора, в которых каждая компонента – отношение количества правильных ответов к количеству заданных вопросов из соответствующей группы, т. е. вместо точечных метрических оценок полноты и целостности мы получаем векторные метрические оценки:

$$COM(i,j) = (x_1(i,j), x_2(i,j), x_3(i,j), x_4(i,j), x_5(i,j)),$$

$$INT(i,j) = (y_1(i,j), y_2(i,j), y_3(i,j), y_4(i,j), y_5(i,j)).$$

Теперь мы можем ответить на вопрос о том, согласованы ли эти оценки между собой (если да, то насколько). Оценим меру тесноты связи между парой  $COM(i,j)$  и  $INT(i,j)$  с помощью коэффициента ранговой корреляции Спирмена. Полное совпадение рангов означает максимально тесную прямую связь, а полная противоположность рангов – максимально тесную обратную связь. В численном выражении значение коэффициента корреляции, равное 1, означает полную прямую связь, а значение, равное -1, – полную обратную. Это значение можно перевести в качественную характеристику по шкале Чеддока. В соответствии с ней коэффициент корреляции больше 0,7 говорит о высокой связи между нашими оценками. Если качественная связь оценок высока, то тестовые оценки полноты и целостности знаний можно принимать за достоверные. Если же нет, то целесообразно продолжить тестирование до достижения коэффициентом корреляции высокого уровня. В этом случае пересчет коэффициента корреляции Спирмена должен выполняться динамически. Заметим, что поднять значение коэффициента корреляции необходимо за минимальное количество дополнительных вопросов.

## Построение оценки

Приведем пример определения согласованности оценок полноты и целостности после получения ответов на вопросы некоторого модуля.

Допустим, для тестирования знаний по модулю (i,j) первоначально было предложено по четыре вопроса каждой группы сложности в каждой категории, всего 40 вопросов. В категории «полнота» было получено 1, 2, 2, 1, 3 правильных ответа в соответствующих группах сложности. В категории «целостность» – 4, 4, 3, 1, 2 ответа. По результатам тестирования обучающегося были построены два вектора:  $COM(i, j) = \left(\frac{1}{4}, \frac{2}{4}, \frac{2}{4}, \frac{1}{4}, \frac{3}{4}\right)$ ,  $INT(i, j) = \left(\frac{3}{4}, \frac{3}{3}, \frac{2}{4}, \frac{1}{4}, \frac{2}{4}\right)$ . Коэффициент корреляции Спирмена для данных числовых рядов  $\rho=0,108$ . Это означает, что связь между исследуемыми признаками прямая, так как он положителен, сила связи по шкале Чеддока слабая. Кроме того, критическая точка распределения Стьюдента по уровню значимости 0,05:  $T_{кр}=1,83 > \rho$ , что говорит об отсутствии статистической значимости ранговой корреляционной связи между двумя оценками.

Следовательно, для получения объективных результатов нужно продолжать тестирование. В табл. 2 приведены результаты расчетов коэффициента корреляции в зависимости от количества правильных ответов на дополнительные вопросы. В нашем примере было задано четыре дополнительных вопроса в категории «полнота»: три степени сложности 1 и один степени сложности 2, на которые тестируемый ответил верно. В результате был получен высокий коэффициент корреляции 0,942, который означает наличие тесной прямой связи между оценками, причем критическая точка показала статистическую значимость этого коэффициента.

Таблица 2 - Расчет ранговой корреляции Спирмена

Дополнительный вопрос	Всего вопросов в категории «полнота»	Оценка полноты $COM(i,j)$	Всего вопросов в категории «целостность»	Оценка целостности $INT(i,j)$	Коэффициент корреляции $\rho$	Критическая точка $T_{кр}$
0	20	$\left(\frac{1}{4}, \frac{2}{4}, \frac{2}{4}, \frac{1}{4}, \frac{3}{4}\right)$	20	$\left(\frac{3}{4}, \frac{3}{3}, \frac{2}{4}, \frac{1}{4}, \frac{2}{4}\right)$	0,108	1,83
1	21	$\left(\frac{2}{5}, \frac{2}{4}, \frac{2}{4}, \frac{1}{4}, \frac{3}{4}\right)$	20	$\left(\frac{3}{4}, \frac{3}{3}, \frac{2}{4}, \frac{1}{4}, \frac{2}{4}\right)$	0,463	1,63
2	22	$\left(\frac{3}{6}, \frac{2}{4}, \frac{2}{4}, \frac{1}{4}, \frac{3}{4}\right)$	20	$\left(\frac{3}{4}, \frac{3}{3}, \frac{2}{4}, \frac{1}{4}, \frac{2}{4}\right)$	0,663	1,38
3	23	$\left(\frac{4}{7}, \frac{2}{4}, \frac{2}{4}, \frac{1}{4}, \frac{3}{4}\right)$	20	$\left(\frac{3}{4}, \frac{3}{3}, \frac{2}{4}, \frac{1}{4}, \frac{2}{4}\right)$	0,663	1,38
4	24	$\left(\frac{4}{7}, \frac{3}{5}, \frac{2}{4}, \frac{1}{4}, \frac{3}{4}\right)$	20	$\left(\frac{3}{4}, \frac{3}{3}, \frac{2}{4}, \frac{1}{4}, \frac{2}{4}\right)$	0,942	0,62

## Заключение

В работе предложен способ повышения объективности системы тестирования. Предлагается оценить глубину знаний учащегося по его ответам на тестовые вопросы с помощью двух векторных оценок. Их согласованность оценивается с помощью коэффициента ранговой корреляции Спирмена. Если его значение достаточно высоко, то оценки признаются согласованными, а результаты тестирования достоверными. Если же нет, то предлагаются

дополнительные вопросы до тех пор, пока корреляция не станет удовлетворительной. К преимуществам способа можно отнести дешевизну разработки, небольшие временные затраты на формирование теста и его прохождение, простоту в использовании.

### Библиография

1. Нуриев Н.К., Старыгина С.Д. Дидактическая инженерия: подготовка инженеров в техногенной образовательной среде // Образование и наука. 2016. № 9 (138). С. 61-79.
2. Нуриев Н.К., Старыгина С.Д. Дидактическая инженерия: технология подготовки // Образовательные технологии и общество. 2016. Т. 19. № 4. С. 412-420.
3. Нуриев Н.К., Старыгина С.Д. Эскизный проект дидактической системы природосообразно развивающего обучения // Alma mater. 2013. № 3. С. 51-55.
4. Нуриев Н.К., Старыгина С.Д., Ахметшин Д.А. Алгоритм оценки качества владения компетенцией на основе показателя глубины усвоенных знаний // Alma mater. 2015. № 11. С. 64-67.
5. Нуриев Н.К., Старыгина С.Д., Ахметшин Д.А. Дидактическая инженерия: проектирование электронного учебного курса с учетом «зоны ближайшего развития» студента // Образовательные технологии и общество. 2016. Т. 19. № 1. С. 558-566.
6. Нуриев Н.К., Старыгина С.Д., Гибадуллина Э.А. Дидактическая инженерия: проектирование систем обучения нового поколения // Интеграция образования. 2016. Т. 20. № 3. С. 393-406.
7. Нуриев Н.К., Старыгина С.Д., Ильмушкин Г.М., Шайдуллина Н.К. Проектирование дидактических систем нового поколения с использованием облачных технологий // Образовательные технологии и общество. 2013. Т. 16. № 4. С. 412-429.
8. Нуриев Н.К., Старыгина С.Д., Крылов Д.А. Дидактическая инженерия: метрическая оценка академической компетентности по технологии обучение – тест // Образовательные технологии и общество. 2015. Т. 18. № 3. С. 548-574.
9. Шайдуллина Н.К. Автоматизированная система обучения с уровневой структурой // Образовательные технологии и общество. 2016. Т. 19. № 4. С. 421-428.
10. Шайдуллина Н.К., Торкунова Ю.В. Дифференциация учебного материала при реализации модульного подхода к обучению // Современные проблемы науки и образования. 2016. № 6. URL: <http://www.science-education.ru/article/view?id=25711>

### The rank correlation of test scores

**Natal'ya K. Shaidullina**

Senior Lecturer,  
Department of information science and applied mathematics,  
Kazan National Research Technological University,  
420015, 68 Karla Marksa st., Kazan, Russian Federation;  
e-mail: [nshaidullina@yandex.ru](mailto:nshaidullina@yandex.ru)

### Abstract

The article deals with the problem of improving the quality of testing as part of an automated training system in the Russian Federation. Special attention is paid to increasing the reliability of the results of tests in the context of automation in the sphere of education. An e-learning system is metric, i. e. all the indicators, used for evaluating the quality of training, are converted to the number format. It also operates within the framework of the competency-based approach. The paper presents a way to improve the objectivity of evaluating test results. Objectivity is considered from several sides. Firstly, the author of the article points out that it is necessary to eliminate the influence of answers chosen at random. Secondly, it is very important to fully evaluate the knowledge of a

student. Thirdly, there is a need for using the same criteria in the process of evaluating test results. Two numerical multidimensional vectors are constructed to obtain the resulting estimate. They reflect the quality of responses to declarative and procedural questions. Their consistency is the main indicator of objectivity. Spearman's rank correlation coefficient is regarded as a criterion of consistency, its value can be dynamically adjusted if it is necessary.

### For citation

Shaidullina N.K. (2018) Rangovaya korrelyatsiya otsenok testirovaniya [The rank correlation of test scores]. *Pedagogicheskii zhurnal* [Pedagogical Journal], 8 (6A), pp. 77-82.

### Keywords

Training, automated training system, e-learning system, training quality, competency-based approach.

## References

1. Nuriev N.K., Starygina S.D. (2016) Didakticheskaya inzheneriya: podgotovka inzhenerov v tekhnogennoi obrazovatel'noi srede [Didactic engineering: training engineers in a technogenic educational environment]. *Obrazovanie i nauka* [Education and science], 9 (138), pp. 61-79.
2. Nuriev N.K., Starygina S.D. (2016) Didakticheskaya inzheneriya: tekhnologiya podgotovki [Didactic engineering: the technology of training]. *Obrazovatel'nye tekhnologii i obshchestvo* [Educational technology and society], 19 (4), pp. 412-420.
3. Nuriev N.K., Starygina S.D. (2013) Eskiznyi proekt didakticheskoi sistemy prirodosobrazno razvivayushchego obucheniya [The draft design of a didactic system for naturally developmental training]. *Alma mater*, 3, pp. 51-55.
4. Nuriev N.K., Starygina S.D., Akhmetshin D.A. (2015) Algoritm otsenki kachestva vladeniya kompetentsiei na osnove pokazatelya glubiny usvoennykh znaniy [An algorithm for assessing the quality of acquiring competency on the basis of the depth of acquired knowledge]. *Alma mater*, 11, pp. 64-67.
5. Nuriev N.K., Starygina S.D., Akhmetshin D.A. (2016) Didakticheskaya inzheneriya: proektirovanie elektronogo uchebnogo kursa s uchedom "zony blizhaishego razvitiya" studenta [Didactic engineering: designing an e-learning course with due regard to the "zone of proximal development" of a student]. *Obrazovatel'nye tekhnologii i obshchestvo* [Educational technology and society], 19 (1), pp. 558-566.
6. Nuriev N.K., Starygina S.D., Gibadullina E.A. (2016) Didakticheskaya inzheneriya: proektirovanie sistem obucheniya novogo pokoleniya [Didactic engineering: designing a new generation of training systems]. *Integratsiya obrazovaniya* [Integration of education], 20 (3), pp. 393-406.
7. Nuriev N.K., Starygina S.D., Il'mushkin G.M., Shaidullina N.K. (2013) Proektirovanie didakticheskikh sistem novogo pokoleniya s ispol'zovaniem oblachnykh tekhnologii [Designing a new generation of didactic systems by using cloud-based technology]. *Obrazovatel'nye tekhnologii i obshchestvo* [Educational technology and society], 16 (4), pp. 412-429.
8. Nuriev N.K., Starygina S.D., Krylov D.A. (2015) Didakticheskaya inzheneriya: metricheskaya otsenka akademicheskoi kompetentnosti po tekhnologii obuchenie – test [Didactic engineering: metric assessment of academic competency by using the training – test technology]. *Obrazovatel'nye tekhnologii i obshchestvo* [Educational technology and society], 18 (3), pp. 548-574.
9. Shaidullina N.K. (2016) Avtomatizirovannaya sistema obucheniya s urovnevoi strukturoi [An automated training system with a level structure]. *Obrazovatel'nye tekhnologii i obshchestvo* [Educational technology and society], 19 (4), pp. 421-428.
10. Shaidullina N.K., Torkunova Yu.V. (2016) Differentsiatsiya uchebnogo materiala pri realizatsii modul'nogo podkhoda k obucheniyu [Differentiation of educational material in the adoption of a modular approach to training]. *Sovremennye problemy nauki i obrazovaniya* [Modern problems of science and education], 6. Available at: <http://www.science-education.ru/article/view?id=25711> [Accessed 19/10/18].