



bezr_alex@mail.ru

Алексей Иосифович Безруков,

кандидат экономических наук,

доцент кафедры прикладной математики и информатики,
Саратовский социально-экономический институт (филиал)

РЭУ им. Г.В. Плеханова



graholskayalv@yandex.ru

Людмила Владимировна Грахольская,

кандидат экономических наук,

доцент кафедры прикладной математики и информатики,
Саратовский социально-экономический институт (филиал)

РЭУ им. Г.В. Плеханова



lv50@bk.ru

Лариса Вячеславовна Малышева,

кандидат педагогических наук,

доцент кафедры прикладной математики и информатики,
Саратовский социально-экономический институт (филиал)

РЭУ им. Г.В. Плеханова

УДК 340.4

МЕТОДЫ ОЦЕНКИ РЕЛЕВАНТНОСТИ РЕЗУЛЬТАТОВ ТЕСТИРОВАНИЯ В СИСТЕМЕ ПОДГОТОВКИ ЭКОНОМИСТОВ

Рассматривается проблема применения компьютерного тестирования при оценке уровня подготовленности экономистов. Для освоения экономической специальности студенты должны получить глубокие знания как по точным, так и по гуманитарным дисциплинам. Поэтому для обучения экономистов особенно актуально создание системы, обеспечивающей релевантную оценку знаний в различных областях. Перечислены типовые возражения против использования тестов для оценки знаний студентов. Рассмотрены требования, предъявляемые теорией педагогических измерений к содержанию, структуре тестов и процедуре тестирования. Показано, что большинство претензий может быть снято, если система тестирования соответствует модели Раша. Рассмотрены причины, вызывающие отклонение результатов тестов от требований модели. Исследовано воздействие каждой из них на результат. Предложен метод оценки времени, необходимого для выполнения теста. Показано, что правильная организация процесса тестирования позволяет существенно снизить влияние этих причин и повысить его релевантность. Сформулированы педагогические выводы, которые можно сделать при анализе невыполнения требований модели для всей исследуемой группы, ее части и отдельных студентов. Предложены мероприятия, позволяющие отслеживать соответствие результатов тестирования требованиям модели Раша.

Ключевые слова: тестирование, релевантность, статистический анализ результатов тестирования, IRT, модель Раша.

*A.I. Bezrukov,
L.V. Graholskaya,
L.V. Malysheva*

METHODS OF ASSESSING RELEVANCE OF TESTING RESULTS IN TRAINING OF FUTURE ECONOMISTS

The article studies the problem of applying computer testing in assessing the level of academic performance of economists. In order to get good economic education students must receive in-depth knowledge of both the exact and humanitarian disciplines. Therefore, for economic education it is especially important to create a system that provides tools for reliable assessment of knowledge in various fields. Typical objections to the use of tests for assessing students' knowledge are discussed. Requirements formulated by the theory of pedagogical assessment to the content, structure of tests and testing procedure are considered. It is stated that most objection can be withdrawn if the testing system corresponds to the Rasch model. Reasons of the deviation of test results from the requirements of the model are identified. The effect of each of them on the result is investigated. A method for estimating the time needed to complete a test is proposed. It is shown that proper organization of the testing process can significantly reduce the impact of these

reasons and increase its relevance. Educational conclusions are made after analyzing the non-fulfillment of the requirements of the model for the entire study group, its part and individual students. Measures are proposed to monitor the compliance of test results with the requirements of the Rasch model.

Keywords: testing, relevance, statistical analysis of test results, IRT, Rasch model.

Для освоения экономической специальности студент должен получить глубокие знания как по точным, так и по гуманитарным дисциплинам. Это делает особо актуальным создание единой системы контроля знаний, обеспечивающей сопоставимые оценки уровня его подготовленности в различных областях. Наиболее адекватным методом объективной оценки является компьютерное тестирование. При наличии специальных программ и добрых банков тестовых заданий этот метод позволяет достаточно быстро и с небольшими затратами получить оценку знаний, навыков и умений студентов в разных областях. Современные принципы организации процесса тестирования и использование автоматизированных систем препятствуют списыванию (каждый студент получает индивидуальный набор заданий), позволяют сохранить для последующего анализа множество сведений о процессе тестирования (например, о последовательности выполнения заданий, времени, затраченном студентом на выполнение теста и т.д.). Как правило, системы тестирования имеют несколько встроенных методов оценки уровня подготовленности студентов.

Однако в среде преподавателей существует критическое отношение к широкому применению тестов в учебном процессе. Перечислим наиболее аргументированные возражения скептиков.

Излишне механистическая процедура тестирования. Студент имеет дело с машиной, а не с преподавателем, который понимает смысл действий студента и может оценить их правильность, даже если результат был формально неверным.

Сама процедура тестирования, когда важно нажать правильную кнопку, не понимая смысла изученного материала, ориентирует студентов, и преподавателей на натаскивание на тест вместо глубокого изучения предмета.

Часть заданий, включенных в тесты, сформулированы некорректно. Это допускает множественность их интерпретации, назначение «правильного ответа» отражает субъективное мнение составителя теста.

Зачастую и студентами, и преподавателями тест рассматривается как лотерея, результат которой зависит от удачи.

Отдельным испытуемым, вполне подготовленным, не хватает времени на выполнение всех заданий теста.

В работе анализируются перечисленные возражения с целью сформулировать методы оценки relevance системы тестирования предъявляемым требованиям.

Чтобы оценить обоснованность возражений рассмотрим основные понятия и предположения теории педагогических измерений (*Item response theory – IRT*) [1]. Согласно этой теории тестирование рассматривается как случайный процесс, результат которого зависит от латентных (непрямую наблюдаемых) пара-

метров испытуемого, трудности выполняемых заданий и множества других случайных факторов, искажающих результат. Вероятность P_{ij} правильного выполнения j -м студентом i -го задания теста зависит от разности уровня подготовленности θ_j студента и трудности задания β_i :

$$P_{ij} = F(\theta_j - \beta_i). \quad (1)$$

Чтобы уровень подготовленности и трудность заданий измерялись в одинаковых единицах, в IRT используется следующий прием. Уровень подготовленности оценивается как логарифм отношения шансов студента выполнить задание к шансам не выполнить, а уровень трудности как логарифм отношения числа студентов, не выполнивших задание к числу выполнивших.

$$\theta_j = \log\left(\frac{M_S}{M_N}\right); \beta_i = \log\left(\frac{N_N}{N_S}\right), \quad (2)$$

где M_S – число заданий, выполненных j -м студентом; M_N – число невыполненных им заданий; N_S – число студентов, выполнивших i -е задание; N_N – число студентов, не выполнивших данное задание.

Такие единицы измерения получили название «логиты» и широко используются в статистических исследованиях.

Г. Раш [9] предложил в качестве функции F использовать логистическую функцию. Для учета различий разброса величины удачных выполнений каждого задания, а также вероятности угадывания правильного ответа (c_i) и чувствительности тестовых заданий к уровню подготовленности студентов (q_i) разработана модифицированная модель Раша – Бирнбаума [7]:

$$P(\theta_j - \beta_i) = c_i + (1 - c_i) \cdot \frac{\exp(\alpha_i \cdot (\theta_j - \beta_i))}{1 + \exp(\alpha_i \cdot (\theta_j - \beta_i))}. \quad (3)$$

Главное преимущество модели Раша заключается в том, что она дает аналитический вид вероятности успешного выполнения заданий. Это позволяет использовать для оценки уровня подготовленности студентов метод максимального правдоподобия. Многочисленные публикации, посвященные методам обработки результатов тестирования, подтверждают, что этот метод дает наиболее надежные и точные оценки [8].

Покажем, что, если тест соответствует требованиям модели Раша, большинство типичных возражений теряет актуальность.

Оценка студента является статистической характеристикой, полученной в результате тестирования, поэтому механическая ошибка при выполнении одного задания скорее всего окажется несущественной при достаточно большом количестве заданий теста.

Большое количество разнообразных заданий, накопленных в банке тестовых заданий (БТЗ), возмож-

ность составления множества комбинаций в каждом тесте делает бессмысленным натаскивание на конкретные шаблоны.

Одной из самых обсуждаемых причин неадекватности теста считается наличие в нем некорректных заданий. Для выявления таких причин в работах [2] предложен метод статистического анализа результатов тестирования. По тому, как выполнен весь тест, оценивается уровень подготовки студентов. Полученные оценки используются для построения зависимости частоты успешности выполнения каждого задания от уровня подготовленности студентов. Если зависимость сильно отличается от предсказанной моделью Раша, это, как правило, свидетельствует о некорректности задания. Методика была апробирована на результатах, хранящихся в системе компьютерного тестирования АСТ Саратовского социально-экономического института [4]. Алгоритм анализа реализован в виде компьютерной программы [5], что позволяет проводить подобный анализ во всех вузах, использующих систему АСТ.

Расчеты по модели Раша показывают, что существует четкая корреляционная связь между уровнем подготовленности студентов и их оценками [3]. Следовательно, если тест соответствует модели Раша, то рассматривать тест как лотерею неправомерно.

Однако возникает вопрос: для всех ли результатов тестирования можно использовать модель Раша?

При реальном тестировании часть требований модели Раша может не выполняться, поэтому, прежде чем использовать эту модель, нужно проверить статистическую гипотезу о ее применимости к полученным результатам. На рисунке перечислены причины, по которым модель Раша может быть неадекватна результатам тестирования.

Очевидно, что влияние различных причин может проявиться одинаковым образом. Поэтому статистический анализ результатов тестирования может выявить только группу проблем, что само по себе не является конечным результатом, но увеличивает эффективность последующего содержательного анализа. Рассмотрим проявления некоторых из них.

Если результаты тестирования не соответствуют модели Раша для всей исследуемой группы, следует проанализировать время, отведенное на тест и саму процедуру тестирования. Недостаток времени проявляется в том, что большинство студентов даже не приступали к части заданий теста (в системе компьютерного тестирования АСТ это задания, на которые было истрачено 0 единиц времени).

Анализ имеющейся в АСТ статистики позволяет обосновать оптимальное время, отводимое на выполнение каждого теста, а также проанализировать зависимость времени, затраченного на выполнение теста от полученной оценки. Типичная зависимость представляет собой выпуклую вверх кривую. Как правило, наименьшее время затрачивают студенты, получившие неудовлетворительные оценки, а наибольшее – получившие средние оценки. Студенты, имеющие твердые знания, выполняют тест быстрее троекников и хорошистов.

Причиной снижения уровня выполнения тестов могут быть также помехи, психологические и физиологические проблемы, возникающие в ходе тестирования (например, шум и духота в помещении, неисправная техника, запугивание студентов и т. д.).

При аномальных результатах части исследуемой группы следует проверить, не было ли нарушений в процессе тестирования. Например, повышенный уровень выполненных заданий может говорить о наличии



Причины отклонений результатов тестирования от модели Раша

подсказок. Особенно ярким свидетельством подсказок является неожиданное выполнение слабыми студентами трудных заданий.

Значительные отклонения индивидуальных результатов можно интерпретировать с помощью профиля Гуттмана [6]. Например, в результате анализа может быть выявлено, что студент не знает некоторые темы, включенные в тест.

Неумение студента распределить свои силы и время на выполнение всех заданий проявляется как наличие заданий, к выполнению которых студент даже не приступал. Кроме того, об этом могут свидетельствовать досадные ошибки при выполнении простых заданий. Пренебрежение выполнением простых заданий может свидетельствовать о том, что студент недостаточно мотивирован. Не понимая, что тест – часть процедуры аттестации, студент получает оценки значительно ниже уровня его подготовленности.

Можно сформулировать следующие рекомендации по организации процесса тестирования и ведению банков тестовых заданий для оценки знаний студентов экономических специальностей.

В процессе тестирования должна накапливаться статистика, позволяющая добиться соответствия процесса тестирования требованиям, предъявляемым моделью Раша.

Необходимо исследовать все случаи отклонения, выявлять и устранять породившие их причины.

При выполнении требований, предъявляемых моделью Раша к содержанию и структуре тестов, а также к процессу тестирования, мы получим мощный, на-

дежный и эффективный инструмент проверки уровня подготовленности студентов в различных областях знаний.

1. Аванесов В.С. Проблема объективности педагогических измерений // Педагогические измерения. 2008. № 3. С. 3–40.
2. Безруков А.И., Погожильская Г.Г. Алгоритм визуализации статистики результатов выполнения тестовых заданий, накопленных в системе ACT // Известия Российского экономического университета им. Г.В. Плеханова. 2015. № 4 (22). С. 178–188.
3. Гусятников В.Н., Безруков А.И., Соколова Т.Н., Каюкова И.В., Погожильская Г.Г. Методы и модели оценки качества обучения в вузе на основе компетентностного подхода. Саратов, 2016.
4. Малышева Л.В., Погожильская Г.Г. Применение методики статистического анализа банков тестовых заданий // Наука и общество. 2016. № 2 (25). С. 32–36.
5. Свидетельство о регистрации программы ЭВМ и базы данных № 2016612887. Программа анализа качества банков тестовых заданий на основе модели Раша – Бирнбаума / В.Н.Гусятников, А.И. Безруков, Т.Н. Соколова, И.В. Каюкова, Г.Г. Погожильская. 25.05.2016 г. М.: Роспатент, 2016.
6. Челышкова М.Б. Теория и практика конструирования педагогических тестов: учеб. пособие. М.: Логос, 2002.
7. Birnbaum A. Some Latent Trait Models and Their Use in Inferring and Examinee's Ability // Lord F.M., Novick M. Statistical Theories of Mental Test Scores. Addison-Wesley Publ. Co. Reading, Mass. 1968. P. 397–479.
8. Dartolucci F., Bacci S., Gnaldi M. Statistical Analysis of Questionnaires. A Unified Approach Based on R and Stata. CRC Press Taylor & Francis Group Boca Raton, London New York, 2016.
9. Rasch G. On general laws and the meaning of measurement in psychology / In proceedings of the fourth Berkley symposium on mathematical statistics and probability. Berkley: Univ. of California Press, 1961.