

КОМПЬЮТЕРНАЯ СИСТЕМА ТЕСТИРОВАНИЯ И ОЦЕНКА КАЧЕСТВА ТЕСТОВЫХ ЗАДАНИЙ

© Ф.Х.Зайнеев, С.В.Сушков

В статье описывается система компьютерного тестирования студентов. Система предназначена для автоматизации тестирования в учебном процессе, основанном на балльно-рейтинговой системе с модульной технологией обучения. Центральной частью системы является блок математико-статистической обработки результатов, основанный на современной теории тестирования – Item Response Theory (IRT).

Ключевые слова: компьютерное тестирование, балльно-рейтинговая система, статистическая обработка результатов тестирования, автоматизация учебного процесса в вузе.

Система компьютерного тестирования

Присоединение в 2003 году России к Болонскому процессу – единой европейской системе высшего образования – влечет за собой использование балльно-рейтинговой системы (БРС) оценки знаний студентов в учебном процессе. Татарский государственный гуманитарно-педагогический университет (ТГГПУ) приступил к внедрению БРС в учебном процессе в 2007 году на основе Положения о балльно-рейтинговой системе оценки качества подготовки высококвалифицированных кадров с использованием модульной технологии обучения, утвержденного решением Ученого совета ТГГПУ от 29 июня 2007 года.

В связи с увеличением контрольных точек в одном семестре, увеличением количества критериев, влияющих на итоговый результат, необходимостью постоянного учета текущей информации об успеваемости и выстраивании рейтинга студентов возникла острая потребность в единой информационной автоматизированной системе, позволяющей собирать и обрабатывать информацию об успеваемости студентов по каждому модулю. Для решения указанных задач нами была предложена схема компьютерной системы тестирования, в основе которой лежит модульная технология обучения (см. рис.1).

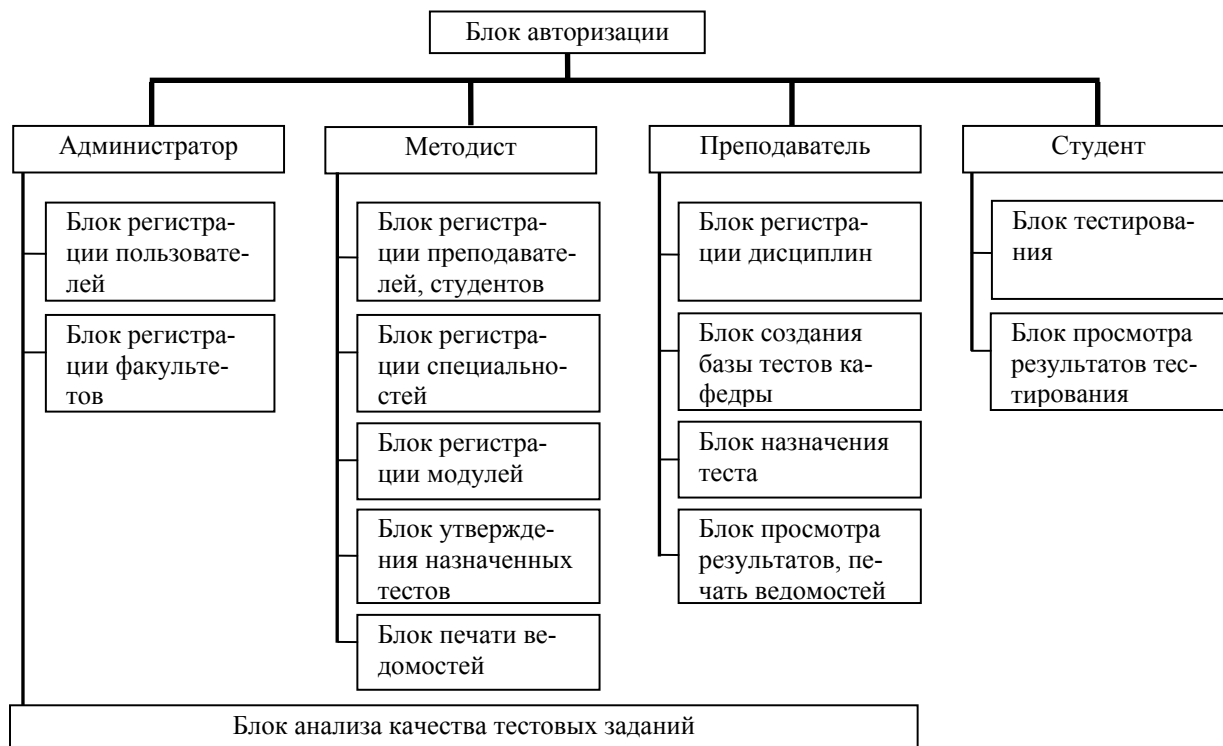


Рис.1. Структура системы компьютерного тестирования

В работе системы используются четыре роли: администратор системы, методист факультета, преподаватель кафедры, студент. Каждая роль имеет свою область ответственности и свои возможности. Вся цепочка взаимодействия от администратора системы до студента является полностью автоматизированной. В данный момент идет разработка системы на платформе Adobe Flex 4. Центральной частью системы является блок математико-статистической обработки результатов тестирования, основанный на современной теории тестирования – Item Response Theory (IRT) (см.: [1]).

Математико-статистическая обработка результатов тестирования

Часто использование в учебном процессе разработанных тестов вызывает закономерные вопросы о качестве и корректности отдельных вопросов и всего теста в целом. Анализ теста может быть сделан в рамках современной теории тестирования IRT на основе первичной апробации теста и статистической обработки ее результатов. Статистическая обработка делится на несколько этапов. Рассмотрим их подробнее.

1) Построение *бинарной матрицы* результатов тестирования. Как правило, на практике используется дихотомический случай. В этом случае ответы оцениваются двумя цифрами – 0 и 1. Бинарная матрица – это матрица результатов тестирования для дихотомического случая.

Таблица 1.

Бинарная матрица результатов тестирования

	Задания									
	1	2	3	4	5	6	7	8	9	10
Ученик 1	1	1	1	0	1	0	0	1	0	0
Ученик 2	1	1	0	1	1	0	1	0	0	0
Ученик 3	1	1	1	1	0	1	0	0	1	1
Ученик 4	1	0	0	1	0	0	0	0	0	0
Ученик 5	1	1	1	1	1	1	0	1	1	1
Ученик 6	1	1	0	1	0	0	0	0	0	0
Ученик 7	1	1	1	0	1	1	1	0	1	0
Ученик 8	1	1	1	1	1	0	1	0	0	0
Ученик 9	1	0	1	1	0	1	1	1	0	0
Ученик 10	1	1	1	1	1	1	1	1	1	1

2) На втором этапе строится *редуцированная бинарная матрица* результатов. Для этого из бинарной матрицы результатов удаляются вопросы, на которые были даны только верные или неверные ответы. Также из списка исключаются ученики, верно или неверно ответившие на все вопросы теста. Вычисляются индивидуальные баллы учащихся X_i и количество верных ответов на каждый вопрос R_j , после чего матрица упоря-

дочивается в порядке убывания вычисленных значений X_i и R_j .

Таблица 2.

Редуцированная бинарная матрица

	Задания									
	4	2	3	5	7	6	9	8	10	X_i
Ученик 5	1	1	1	1	0	1	1	1	1	8
Ученик 3	1	1	1	0	0	1	1	0	1	6
Ученик 7	0	1	1	1	1	1	1	0	0	6
Ученик 8	1	1	1	1	1	0	0	0	0	5
Ученик 9	1	0	1	0	1	1	0	1	0	5
Ученик 2	1	1	0	1	1	0	0	0	0	4
Ученик 1	0	1	1	1	0	0	0	1	0	4
Ученик 6	1	1	0	0	0	0	0	0	0	2
Ученик 4	1	0	0	0	0	0	0	0	0	1
R_j	7	7	6	5	4	4	3	3	2	

Данная матрица имеет характерную особенность: почти все нули и единицы расположены по разные стороны диагонали, идущей из левого нижнего угла в правый верхний. Согласно Гуттману, это разграничение должно быть идеальным, т.е. если испытуемый верно ответил на трудное задание, то он тем более должен справиться с более легкими заданиями. Однако на практике так происходит не всегда. Если испытуемый верно ответил на трудные вопросы, но не справился с легкими, то говорят, что у него *инвертированный* профиль. Инвертированный профиль свидетельствует либо о неверной структуре знаний испытуемого, либо о нарушении процедуры тестирования (списывание, угадывание и т.д.), либо о недостатках тестовых заданий (по форме и/или по содержанию) [2].

3) После построения редуцированной бинарной матрицы строится таблица *частот тестовых баллов*.

Таблица 3.

Частоты тестовых баллов

X_i	1	2	4	5	6	8
Частота	1	1	2	2	2	1

На основе этой таблицы легко анализировать редуцированную бинарную матрицу. Для анализа можно использовать диаграммы в форме полигонов и гистограмм частот (см. рис.2, 3, 4).

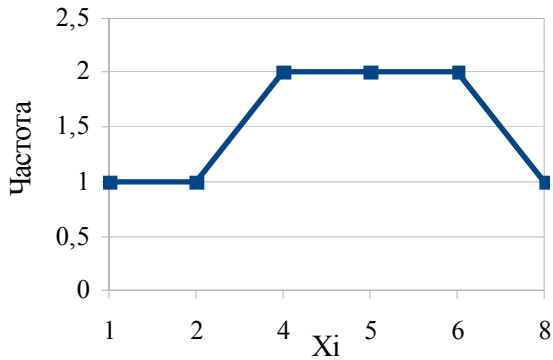


Рис.2. Полигон частот

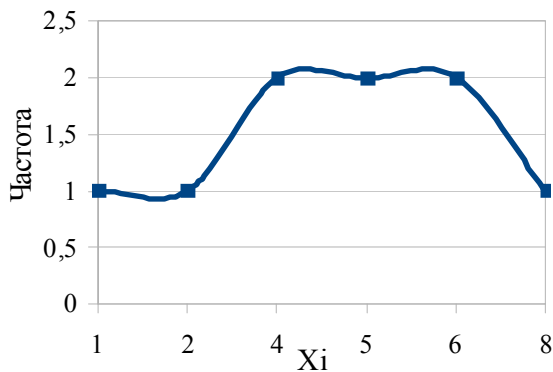


Рис.3. Сглаженная эмпирическая кривая

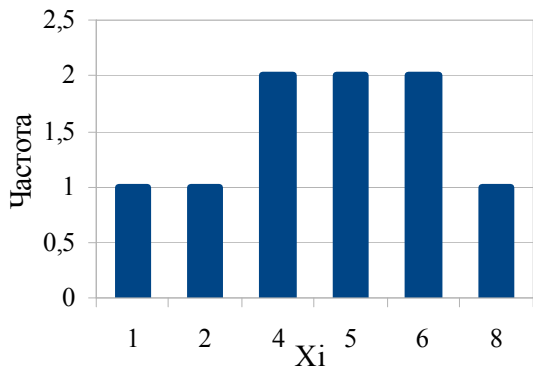


Рис.4. Гистограмма частот тестовых баллов

4) На четвертом этапе определяются *меры центральной тенденции* совокупности результатов, которые можно охарактеризовать тремя мерами центральной тенденции: мода, медиана, среднее. *Мода* – это такое значение во множестве наблюдений, которое встречается наиболее часто среди результатов выполнения теста. Если все значения в группе встречаются одинаково часто, то мода отсутствует. Например, в наших результатах (1, 2, 4, 4, 5, 5, 6, 6, 8) мода равна 4. Если два соседних значения имеют одинаковые частоты и они больше частоты любого другого значения, то мода есть среднее этих двух значений. Если же два несмежных значения в группе име-

ют равные частоты и они больше частот любого другого значения, то существуют две моды. В этом случае говорят, что группа оценок является бимодальной. *Медиана* – это значение, которое делит упорядоченное множество данных пополам, так что одна половина значений оказывается больше медианы, а другая — меньше. *Среднее арифметическое* определяется по формуле

$$\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i, \quad (1)$$

где N – количество учеников прошедших тестирование, X_i – индивидуальный бал i -го ученика [2]. В нашем случае мода равна 4, медиана равна 5, а среднее арифметическое равно 4,55.

5) Определение *характеристики рассеяния* отдельных значений вокруг среднего. Для определения характеристики рассеяния используют различные меры: размах, дисперсия, стандартное отклонение. *Размах* – это расстояние, в пределах которого изменяются все значения показателя в распределении. Вариационный размах используется редко, так как он является весьма приближенным показателем и вычисляется по крайним значениям, которые могут содержать в себе значительную ошибку измерения, поэтому чаще применяют *дисперсию*. Подсчет дисперсии основан на вычислении отклонений каждого значения показателя от среднеарифметического значения в распределении. Дисперсия обычно обозначается символом S_x^2 и вычисляется по формуле

$$S_x^2 = \frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N - 1}. \quad (2)$$

Низкий показатель дисперсии указывает на недостаточную дифференциацию испытуемых при тестировании. Высокая дисперсия же указывает на большой разброс по числу выполненных заданий, что является основанием говорить о возможных нарушениях процедуры тестирования, ошибках в формулировках заданий. Кроме дисперсии для характеристики меры изменчивости распределения используют показатель, называемый *стандартным отклонением*. Стандартное отклонение равно квадратному корню из дисперсии

$$S_x = \sqrt{S_x^2}. \quad (3)$$

В нашем случае размах равен 7, $S_x^2 \approx 4,52$, $S_x \approx 2,13$.

6) Следующим этапом исследования является вычисление показателей связи между результатами. Связь между результатами учеников по отдельным заданиям теста исследуют с помощью *коэффициента корреляции*. Коэффициент корреляции

ляции Пирсона для дихотомических данных называется коэффициентом "фи". Коэффициент φ_{mk} , описывающий связь между заданиями с номерами m и k записывается следующим образом

$$\varphi_{mk} = \frac{p_{mk} - p_m p_k}{\sqrt{p_m q_m p_k q_k}}, \quad (4)$$

где p_m – доля верных ответов для задания с номером m ; q_m – доля неверных ответов для задания с номером m ; p_k – доля верных ответов для задания k ; q_k – доля неверных ответов для задания с номером k ; p_{mk} – доля верных ответов для задания с номером m и k . Также вычисляется коэффициент *бисериальной корреляции* по следующей формуле:

$$r_{pb} = \frac{(\bar{X}_i)_j - (\bar{X}_0)_j}{S_x} \sqrt{\frac{(N_1)_j \cdot (N_0)_j}{N(N-1)}}, \quad (5)$$

где $(\bar{X}_i)_j$ – среднее значение индивидуальных баллов учеников, выполнивших верно j -е задание теста; $(\bar{X}_0)_j$ – среднее значение индивидуальных баллов учеников, выполнивших неверно j -е задание теста; S_x – стандартное отклонение по множеству значений индивидуальных баллов; $(N_1)_j$ – число учеников, выполнивших верно j -е задание теста; $(N_0)_j$ – число учеников, выполнивших неверно j -е задание теста; N – общее число учеников [1]. Ниже приводится сводная матрица коэффициентов корреляции.

Таблица 3.

Матрица коэффициентов корреляции заданий

Задания	4	2	3	5	7	6	9	8	10	r_{pb}
4	1,000	-0,286	-0,378	-0,478	-0,060	-0,060	-0,189	-0,189	0,286	-0,118
2	-0,286	1,000	0,189	0,598	-0,060	-0,060	0,378	-0,189	0,286	0,414
3	-0,378	0,189	1,000	0,316	0,158	0,632	0,500	0,500	0,378	0,782
5	-0,478	0,598	0,316	1,000	0,350	-0,100	0,158	0,158	-0,060	0,470
7	-0,060	-0,060	0,158	0,350	1,000	0,100	-0,158	-0,158	-0,478	0,136
6	-0,060	-0,060	0,632	-0,100	0,100	1,000	0,791	0,316	0,598	0,755
9	-0,189	0,378	0,500	0,158	-0,158	0,791	1,000	0,000	0,756	0,546
8	-0,189	-0,189	0,500	0,158	-0,158	0,316	0,000	1,000	0,189	0,288
10	0,286	0,286	0,378	-0,060	-0,478	0,598	0,756	0,189	1,000	0,378
Сумма	-0,354	1,856	3,296	1,942	0,694	3,217	3,235	1,627	2,954	

Из данной матрицы можно сделать следующий вывод: в тесте есть 2 подозрительных вопроса с номерами 4 и 7. Об этом говорит отрицательная корреляция 4-го вопроса практически со всеми вопросами, сумма -0,354, а также отрицательный коэффициент бисериальной корреляции данного вопроса, равного -0,118. Несколько лучше ситуация с 7-м вопросом, но он также является проблемным. При создании теста необходимо придерживаться корреляции заданий друг с другом в пределах от 0 до 0,3. Если значение корреляции между двумя вопросами близко к единице, то одно из заданий является лишним. Отрицательное же значение корреляции между заданиями вызывает вопрос о корректности вопроса в тесте. Возможно, данный вопрос содержит ошибки в содержании, не указан верный ответ, либо этот вопрос принадлежит другой предметной области.

7) Важнейшей характеристикой теста является его *валидность*. Раньше валидность теста ассоциировали с понятием "качество теста". Теперь же валидность рассматривают как интерпретацию результатов тестирования к цели тестирования. От валидности заданий зависит валидность самого теста. Задание можно считать валидным, если коэффициент бисериальной корреляции $r_{pb} \approx 0,5$. Таким образом, задания, у которых значение $r_{pb} < 0$ или $r_{pb} \approx 1$, подлежат исключению из теста.

1. *Чельшкова М.Б.* Теория и практика конструирования педагогических тестов: учебн. пособ. – М: Логос, 2002. – 432 с.
2. *Ким В.С.* Тестирование учебных достижений: монография. – Уссурийск: УГПИ, 2007. – 214 с.

COMPUTER TESTING SYSTEM AND EVALUATION OF TESTS' QUALITY

F.H.Zayneev, S.V.Sushkov

The authors of the article speculate on the system of computer testing of students. The system is intended for automatizing the testing in the teaching process based on point-rating system with a modular technology of teaching. The central point of the system is a block of mathematical statistical processing of test results based on the modern testing theory – Item Response Theory (IRT).

Key words: computer testing, point-rating system, statistical processing of test results, automatizing the teaching process in higher education.

* * * * *

Зайнеев Фарид Хатибович – аспирант кафедры геометрии и математического моделирования Татарского государственного гуманитарно-педагогического университета.

E-mail: z-farid@mail.ru

Сушков Сергей Владимирович – доктор физико-математических наук, профессор кафедры геометрии и математического моделирования Татарского государственного гуманитарно-педагогического университета.

E-mail: Sergey_Sushkov@mail.ru

Поступила в редакцию 15.12.2010