

ВОЗМОЖНОСТИ КОМПЬЮТЕРНЫХ ТЕХНОЛОГИЙ ПРИ ОЦЕНКЕ ТЕСТОВЫХ ЗАДАНИЙ В ПРАКТИКЕ ВЫСШЕЙ ПЕДАГОГИЧЕСКОЙ ШКОЛЫ

Е.Ю. Елизарова

Принципиальные изменения сложившейся практики оценивания в высшей школе состоят в переходе от субъективного оценивания к объективному измерению знаний обучаемых, к обработке результатов с помощью математических методов (системный анализ, теория больших чисел, статистические показатели) и компьютерных технологий (компьютерные программы: StatSoft Statistica, MS Excel). В статье рассматриваются некоторые аспекты эмпирической оценки тестовых заданий. Приводится описание статистических способов обработки и геометрические формы представления результатов тестирования, проводимых при обучении студентов-математиков. В статье рассматриваются аналитические характеристики тестового задания (индекс трудности и надежности и т.п.) и графические способы представления и описания (форма «решаемость задания», геометрический образ задания, гистограмма распределения результатов тестирования), примеры их изображения в StatSoft Statistica, MS Excel, анализ и интерпретация.

Ключевые слова: матрица ответов, индекс трудности, коэффициент корреляции, геометрический образ задания, гистограмма распределения результатов тестирования, MS Excel, Stat Soft Statistica.

Сложность и многогранность проблем оценивания в сфере образования, широкое использование тестов в школьной и вузовской практике, обусловили актуальность проблемы подготовки будущего учителя, владеющего теоретическими основами тестирования, имеющего представления о средствах оценивания результатов обучения, о способах педагогических измерения учебных достижений обучаемых с помощью математической обработки результатов тестирования.

Составление качественных тестов – очень трудоемкий процесс, поэтому важная задача вуза, обучающего студентов по направлению «Педагогическое образование»: подготовка специалистов в области педагогических измерений по вопросам теории и практики использования педагогических тестовых материалов в сфере образования.

Как известно, вычислительная техника с каждым годом оказывает все большее влияние на все этапы проведения тестов — от разработки и конструирования тестов до их проведения, от подсчета баллов до оперативного сообщения результатов и их интерпретации. Однако анализ целей использования компьютера при изучении тем курса, выбор программного обеспечения, отражающего содержание тем, которое будет доступно и интересно обучаемым, позволяет констатировать, что особенно эффективным и наглядным становится применение компьютера на стадии предъявления тестов, обработки и интерпретации результатов тестирования.

Рассмотрим одну из методик обработки и формы представления результатов тестирования, проводимых при обучении студентов-математиков с использованием компьютерной техники.

1. Построение матрицы ответов по эмпирическим данным тестирования. Получение статистических характеристик теста.

При вводе результатов тестирования для определенности номера или фамилии испытуемых вводим в столбец, а итоги тестирования по каждому испытуемому в соответствующую строку. Получаем матрицу ответов, пример которой представлен в приложении 1. Отметим, что результаты тестирования рассматриваются по дихотомической шкале: 1 балл ставится при правильном ответе на тестовое задание, 0 баллов – при неправильном ответе.

Анализ матрицы ответов может осуществляться в двух направлениях: получение статистических характеристик при анализе матрицы по вертикали (по столбцам) и по горизонтали (по строкам).

а) Анализ матрицы ответов по столбцам

Среди характеристик тестирования выберем следующие показатели:

- **Сумма (R_j)** – число правильных ответов на j задание.
- **Процент выполнения задания** вычисляется по формуле $\langle \frac{\text{Сумма}}{n} \cdot 100\% \rangle$, где n - число испытуемых.
- **Доля правильных ответов (P_j)** на j задание определяется по формуле $P_j = \frac{R_j}{n}$, где n - число испытуемых.
- **Доля учащихся, неверно выполнивших j задание (Q_j)** вычисляется по формуле $Q_j = 1 - P_j$.

Показатель Q_j является важной статистической характеристикой j задания, которая называется **индексом трудности**. Определение индекса трудности является обязательным требованием к тестовым заданиям. Если не известна эмпирическая мера трудности, то задание трудно назвать тестовым. Это связано с тем, что в педагогическом тесте задания должны быть упорядочены по степени возрастания трудности [1, с.165].

Отметим, что задания, индекс трудности которых меньше 20% или больше 80 % либо исключаются из теста либо дорабатываются, поскольку в первом случае тестовые задания являются легкими для большинства учащихся, а во втором случае задания окажутся трудными (их решат меньше 20% учащихся). Однако в критериально-ориентированных тестах, направленных, например, на усвоение способов решения типовых задач, задания с высоким и низким индексом трудности могут быть оставлены в тесте. Они позволяют выявить типы задач, с которыми справляются большая (меньшая) часть учащихся, и внести коррективы в процесс обучения по результатам тестирования [3, с.32].

Таким образом, требование известной трудности задания является системообразующим признаком тестового задания.

б) Анализ матрицы по строкам

Для анализа рассмотрим ряд характеристик:

- **Сумма баллов** за тест для каждого испытуемого.
- **Средний балл (M)**, который равен частному **суммы баллов** за тест каждого испытуемого и n , где n - число испытуемых.
- **Мода (Mo)** - наиболее часто повторяющийся элемент. Подсчитывается, например, с помощью функции МОДА в MS Excel от суммы баллов по каждому испытуемому в формате $\langle \text{МОДА}(L2:L23) \rangle$.

- **Медиана (Me)** – значение признака у средней единицы ряда, записанного в возрастающем (убывающем порядке). Значение медианы может быть найдено в табличном редакторе MS Excel с помощью функции МЕДИАНА в формате «=МЕДИАНА(L2:L23)».
- **Ymax и Ymin** – это максимальное и минимальное значения суммы баллов по каждому испытуемому.
- **Размах баллов (RA)** равен разности Ymax и Ymin.
- **Процент (%)** набранной суммы баллов по отношению к максимально возможному количеству баллов.
- **Ранг** - ранг (место) каждого испытуемого. Функция Ранг в MS Excel может быть использована для подсчета данного признака.

На основе полученных данных определяем коэффициент трудности теста (**T**), как разность максимальной суммы баллов за тест и среднего арифметического значения суммы первичных баллов. Величина дисперсии (**Dj**) по результатам выполнения *j* задания определяется по формуле $D_j = \sum (q_j \cdot p_j)$ и может говорить о коэффициенте надежности гомогенного теста. Дисперсия распределения баллов за тест **Dy** (функция ДИСП в табличном редакторе MS Excel) и величина стандартного отклонения **Sy** (функция СТАНДОТКЛОН от столбца «сумма баллов» в MS Excel) определяются как мера рассеивания индивидуальных баллов за тест вокруг среднего арифметического баллов.

Одними из важных характеристик при анализе заданий теста являются коэффициент корреляции и критерий надежности теста.

Оценка надежности теста проводится по **коэффициенту Кьюдера—Ричардсона**, вычисляемого по формуле:

$$K = \frac{m}{m-1} \left(1 - \frac{\sum_{j=1}^m p_j q_j}{D_y} \right), \text{ где } m - \text{число заданий в тесте, } D_y - \text{дисперсия распределения баллов за тест [4, с.176].}$$

При проверке тестовых свойств заданий в качестве критерия также используется бисериальный **коэффициент корреляции**. Его вычисляют по формуле: $R_{jy} = \frac{M_1 - M_0}{S_y} \cdot \sqrt{\frac{n_0 n_1}{n(n-1)}}$, где M_1 – среднеарифметическое по всему тесту для испытуемых, получивших 1 балл за *j* задание, M_0 – среднеарифметическое по всему тесту для испытуемых, получивших 0 баллов за *j* задание, n_1 -число испытуемых, получивших 1 балл за *j* задание, n_0 – число испытуемых, неверно выполнивших *j* задание, $j=1,2...20$) [4, с.174].

Подсчитывая коэффициенты корреляции по каждому заданию, можно сделать выводы о качестве тестовых заданий (например, таблица 1).

Таблица 1

№ вопроса	1	2	3	4	5	6	7	8	9	10
R_{jy}	0,62	0,37	0,12	0	0,45	0,46	0,63	0,62	0	0,46
Вывод	+	±	-	-	±	±	+	+	-	±

Отметим, что нижняя граница включения задания в тест составляет **R_{jy} ≥ 0,2**, поэтому в таблице 1 в строке «Выводы» знак минус поставлен в тех столбцах, где условие не выполнено. Чем выше значение **R_{jy}**, тем больше вероятность превращения задания в тестовой форме в тестовое задание. Если коэффициент корреляции равен нулю или принимает отрицательные значения, то это свидетельствует об отсутствии у задания необходимых системных свойств, присущих для тестового задания. Если значение коэффициента находится в интервале от 0,2 до 0,5, то для решения вопроса о включении задания нужно проанализировать другие характеристики, например, меру трудности, меру корреляции с другими заданиями.

Если найти (**R_{jy}**)² и умножить это число на 100%, то получим значение коэффициента детерминации [1, с.170-171].

Вычисление **коэффициента корреляции между двумя заданиями теста** осуществляется по формуле:

$$\varphi_{jk} = \frac{p_{jk} - p_j p_k}{\sqrt{p_j q_j p_k q_k}}, \text{ где } p_{jk} - \text{доля испытуемых, верно выполнивших оба задания, } p_j - \text{доля испытуемых верно выполнивших } j - \text{е задание, } p_k - \text{доля испытуемых верно выполнивших } k - \text{е задание, } q_j - \text{доля испытуемых неверно выполнивших } j - \text{е задание, } q_k - \text{доля испытуемых неверно выполнивших } k - \text{е задание.}$$

Отметим, что $\varphi_{jk} = \varphi_{kj}$, тогда таблица результатов вычислений φ_{jk} в MS Excel может быть представлена в виде таблицы 2.

Таблица 2

φ	1	2	3	4	5	6	7	8	9	10
1		0,21	0,12		0,22	0,16	0,29	0,26		-0,37
2	0,21		0,22		0,49	0,34	0,01	-0,13		-0,43
3	0,12	0,22			0,22	-0,37	-0,55	-0,48		-0,05
4										
5	0,22	0,49	0,22			0,34	0,01	0,12		-0,17
6	0,16	0,34	-0,37		0,34		0,66	0,50		-0,08
7	0,29	0,01	-0,55		0,01	0,66		0,12		0,66
8	0,26	-0,13	-0,48		0,12	0,50	0,12			0,24
9										
10	-0,37	-0,43	-0,05		-0,17	-0,08	0,66	0,24		

При анализе таблицы, можно сделать выводы на основе следующего замечания: если $\varphi_{jk} < 0$, то связи между

заданиями нет. В итоговом тесте стремятся к невысокому положительному коэффициенту φ_{jk} , т.е. $\varphi_{jk} \in (0; 0,3)$. Для тематического теста характерна высокая корреляция между заданиями [5, с.313].

Отметим, чтобы получить описательную статистику числовых переменных, можно воспользоваться программой по статистической обработке данных Stat Soft Statistica (Статистика) [2].

Для этого можно щелкнуть в диалоге Basic Statistic на кнопке Statistics... (Статистика) значок Descriptive Statistics. Откроется диалоговое окно Descriptive Statistics (Дескриптивные статистики).

В группе Advanced можно выбрать следующие характеристики:

- Mean (Среднее значение)
- Median (Медиана)
- Mode (Мода)
- Frequency (Частота)
- Sum (Сумма).
- Minimum (Минимум): наименьшее значение
- Maximum (Максимум): наибольшее значение
- Range (Размах)
- Variance (Дисперсия)
- Std. deviation (Стандартное отклонение)

2. Графическая интерпретация результатов тестирования

Выделим формы представления результатов, доступные в режиме анализа заданий.

1) Форма «Решаемость заданий»

Форма «Решаемость заданий» - это карта коэффициентов решаемости по заданиям (темам) теста. Коэффициент решаемости вычисляется как отношение суммы баллов по данному заданию, полученной всеми тестируемыми, к максимально возможному количеству баллов за тест. Например, максимально возможное количество баллов за тест составляет 20 баллов; число правильных ответов на второе задание равно 7, значит, коэффициент решаемости второго задания равен 0,35, т.е. $\frac{7}{20} = 0,35$.

Форму «Решаемость заданий» можно представить в программе StatSoft Statistica. Для этого выделяется в таблице исходных данных строка коэффициентов решаемости заданий K_j . С помощью команды Graphs (- Graphs of Block Data – Line Plot: Block Rows) на панели инструментов в программе StatSoft Statistica создаем поле с искомым графиком. Пример формы «Решаемость заданий» можно посмотреть на рисунке 1.

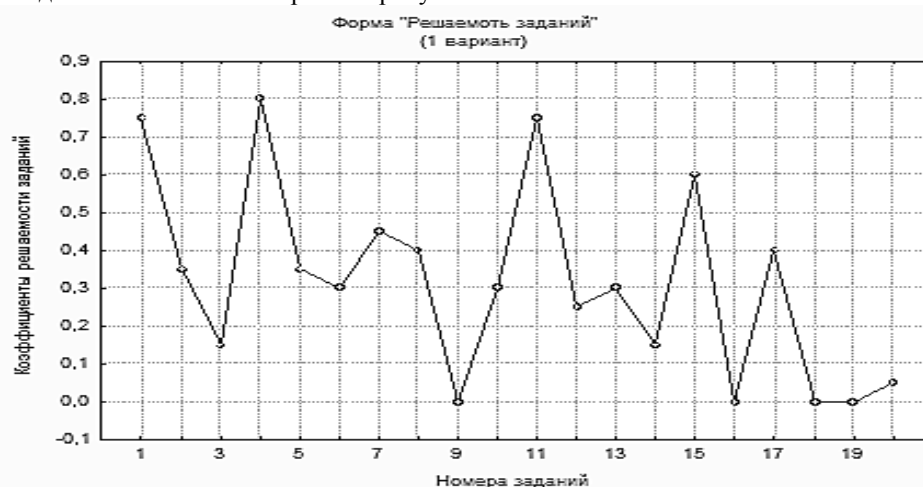


Рисунок 1. Форма «Решаемость заданий»

Отметим, что для большей наглядности форма «Решаемость заданий» для нескольких вариантов может быть сведена в одно графическое поле.

Анализируя форму «Решаемость заданий» можно определить номера заданий, входящих в зону «легких» и «трудных» заданий, доля заданий, составляющих зону «средних заданий». Например, по первому варианту самыми трудными оказались 9, 16, 18, 19 задания (100% не выполнения) и легким – 4 задание (0% не выполнения).

2) Геометрический образ задания

Для построения геометрического образа задания нужно определить вероятность выполнения каждого j задания, как отношение числа учащихся, верно выполнивших j задание к числу учащихся, получивших Y баллов за это задание.

Приведем пример геометрического образа пятого задания, построенного с помощью команд меню Graphs - Graphs of Block Data – Line Plot: Block Rows в программе StatSoft Statistica (Рис. 2).

Специалисты, анализируя геометрический образ задания, дают педагогическую, тестологическую и статистическую интерпретацию. Например, по рисунку 2 можно сформулировать ряд выводов:

- задание плохо сформулировано, поэтому о правильном ответе можно только догадываться. При его выполнении ошибаются и слабые и сильные учащиеся;
- задание требует доработки для внесения большей ясности в его смысл;
- график имеет высокую крутизну, что означает высокую дифференцирующую способность задания;
- задание средней сложности, т.к. с ним справляются учащиеся, получившие наименьшее и наибольшее количество

баллов за тест.

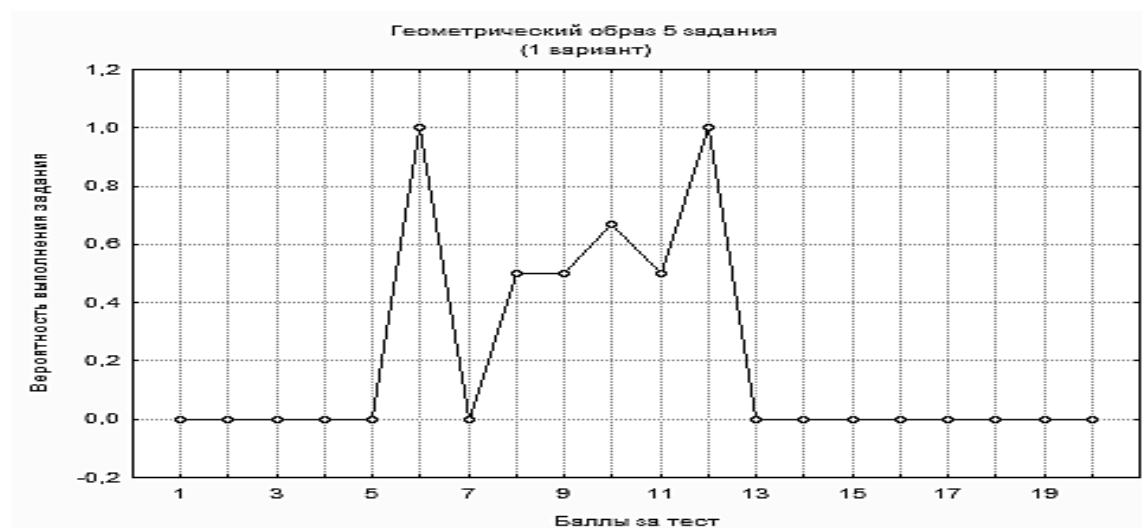


Рисунок 2. Геометрический образ задания

3) Гистограмма распределения результатов тестирования

Для построения гистограммы в матрице ответов выделяем столбец **процент «%»** набранной суммы баллов по отношению к максимально возможному количеству баллов. Выбираем команды Graphs – Histograms. После этого в разделе Variables (Переменные) устанавливаем имя столбца таблицы «%». Нажимаем кнопку ОК. После внесения соответствующих заголовков получаем следующую диаграмму (рис. 3)

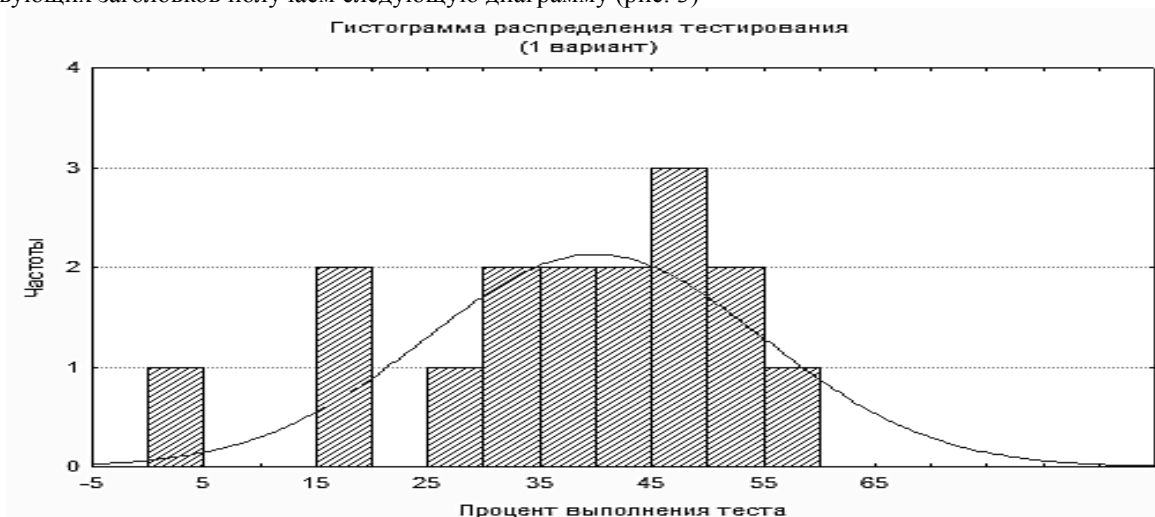


Рисунок 3. Гистограмма распределения тестирования

Таким образом, применение описанных методик анализа результатов тестирования при обучении студентов по направлению «Педагогическое образование» позволит сформировать у будущего учителя теоретико-методологические основы тестирования: способы обработки результатов с помощью математических методов с последующим определением характеристик точности и надежности полученных результатов. Применение компьютерных программ (например, StatSoft Statistica, MS Excel) при анализе результатов тестирования позволяют не только быстро получить статистические показатели, наглядные геометрические образы, но и соединяют теоретические знания и практические умения и навыки студентов в едином процессе учебной деятельности.

Приложение

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Сум балл		%	Ранг	T
Вариант 1																									
1	1	0	0	1	1	1	1	1	0	0	1	0	1	0	0	0	0	0	0	0	8	M	40	9	8
2	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	7,94	5	16	3
3	1	1	0	1	0	1	1	0	0	0	1	0	0	0	1	0	1	0	0	0	8		40	10	8
4	1	0	0	1	0	0	1	1	0	1	1	0	1	1	1	0	1	0	0	1	11	Me	55	3	10
5	1	1	1	1	1	0	0	0	0	0	1	1	0	1	1	0	0	0	0	0	9	8,5	45	7	9
6	1	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	4		20	11	5
7	1	0	0	1	1	1	1	1	0	1	1	1	1	0	1	0	0	0	0	0	11	Ymax	55	3	10
8	1	0	0	1	0	0	1	1	0	1	1	1	1	0	1	0	1	0	0	0	10	12	50	3	9
9	1	1	1	1	1	0	0	0	0	0	1	1	0	1	1	0	1	0	0	0	10		50	3	9
10	1	1	0	1	1	1	1	1	0	1	1	1	1	0	1	0	0	0	0	0	12	Ymin	60	2	11
11	1	1	0	1	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	6	1	30	6	7

12		1	1	0	1	1	1	1	0	0	1	0	0	0	1	0	1	0	0	0	10		50	3	9
13		1	0	1	1	0	0	0	0	0	1	1	0	1	0	1	0	0	0	0	7	RA	35	5	7
14		1	0	0	1	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	4	11	20	7	5
15		1	0	0	1	0	0	1	1	0	1	1	0	0	0	1	0	0	0	0	7	T	35	7	7
16		1	1	0	1	0	1	1	1	0	0	1	0	0	0	1	0	1	0	0	9	12,06	45	5	9
	Сумма	15	7	3	16	7	6	9	8	0	6	15	5	6	3	12	0	8	0	0	1	127			
	% выполнения	93,75	43,8	18,75	100	43,75	37,5	56,25	50	0	37,5	93,8	31,3	37,5	18,8	75	0	50	0	0	6,25	Mo			
	Pj	0,94	0,44	0,19	1	0,44	0,38	0,56	0,5	0	0,38	0,94	0,31	0,38	0,19	0,75	0	0,5	0	0	0,06	10			
	Qj	0,06	0,56	0,81	0	0,56	0,63	0,4	0,5	1	0,63	0,06	0,69	0,63	0,81	0,25	1	0,5	1	1	0,94				
	Dj	0,06	0,25	0,15	0	0,25	0,23	0,25	0,25	0	0,23	0,06	0,21	0,23	0,15	0,19	0	0,25	0	0	0,06				
	Dy	8,996																							
	Sy	2,999																							
	ΣDj	2,824																							
	M _i	8,4	9,14	8,67	7,94	9,43	9,67	9,56	9,75	0	9,67	8,4	10,4	9,83	10	9	0	8,5	0	0	11				
	M ₀	1	7	7,77	0	6,78	6,9	5,86	6,13	7,94	6,9	1	6,82	5,67	7,46	4,75	7,94	7,38	7,94	7,94	7,73				
	R _{jy}	0,62	0,37	0,12	0	0,453	0,46	0,63	0,62	0	0,46	0,62	0,57	0,69	0,34	0,63	0	0,19	0	0	0,27				
	R _{jy} ² (%)	38,05	13,4	1,46	0	20,5	21,3	39,91	39	0	21,3	38	32,7	48,2	11,6	40,2	0	3,75	0	0	7,41				

A fundamental change in the established practice of assessment in higher education consist in the transition from subjective assessment to objective measurement of students knowledge, to the processing of the results by means of mathematical methods of system analysis, the theory of large numbers, statistics and computer technology (computer programs: StatSoft Statistica. MS Excel). The article discusses some aspects of the empirical evaluation of test tasks. The description of the statistical methods of processing and geometric presentation of the results of test carried out in teaching mathematics students. The article discusses the analytical characteristics of the test task (index of difficulty, the correlation coefficient, etc.) and graphic methods of representation and description (solutions form, geometrical image of test item, histogram of the distribution of test results), examples of their image in StatSoft Statistica, MS Excel, analysis and interpretation.

Keywords: *matrix of responses, the difficulty index, the correlation coefficient, geometric image of test item, the histogram of the distribution of test results, MS Excel, Stat Soft Statistica.*

Список литературы

1. Аванесов В.С. Композиция тестовых заданий. Учебная книга. 3 изд. доп. - М.: Центр тестирования.- 2002г, 240 с.
2. Боровиков В. Statistica: искусство анализа данных на компьютере. – СПб.: Питер. – 2003, 344 с.
3. Майоров А.Н. Теория и практика создания тестов для системы образования. – М: Интеллект-центр.-2001, 296 с.
4. Сидоренко Е.В. Методы математической обработки в психологии. – СПб: ООО «Речь». -2003, 350 с.
5. Челышкова М.Б. Теория и практика конструирования педагогических тестов. – М.: Логос.- 2002, 432 с.

Об авторе

Елизарова Е.Ю. – старший преподаватель кафедры математики и математического образования ФГБОУ ВПО «Нижегородский государственный педагогический университет им. Козьмы Минина», *Elizarova-EU@yandex.ru*