

М.В. Афолина, М.А. Ильина, Н.Т. Копылова

ИНТЕРПРЕТАЦИЯ РЕЗУЛЬТАТОВ ПЕДАГОГИЧЕСКИХ ИЗМЕРЕНИЙ

В статье рассматриваются вопросы, связанные с обработкой и интерпретацией результатов педагогических измерений.

Ключевые слова: педагогическое измерение, педагогический тест, нормативно-ориентированное тестирование, критериально-ориентированное тестирование.

V.M. Afonina, M.A. Ilyina, N.T. Kopylova

INTERPRETATION OF RESULTS OF PEDAGOGICAL MEASUREMENTS

The article deals with issues related to the processing and interpretation of the results of educational measurement, in particular – pedagogical testing.

Key words: pedagogical measurement, pedagogical test, standard oriented testing, criteria oriented testing.

Составной частью современных образовательных стандартов являются фонды оценочных средств (ФОС), на основе которых выполняются контроль и оценка достижений обучающихся. Функции педагогического контроля и оценки неразрывно связаны с разработкой и применением процедур и средств педагогических измерений.

Педагогические измерения трактуются как «конструирование числовой функции, осуществляющей изоморфное отображение некоторой эмпирической структуры в соответствующим образом подобранную числовую структуру» [1, с. 48]. Таким образом, получая изоморфную числовую структуру, педагог имеет возможность по ней судить о свойствах, признаках, особенностях, характеристиках измеряемых объектов, не обращаясь к ним напрямую. Очевидно, что осуществление педагогических измерений неразрывно связано с владением педагогом математико-статистическим аппаратом.

Процесс педагогических измерений включает следующие этапы:

- определение характеристик объекта измерения (латентных и наблюдаемых);
- выбор/разработка процедур измерения;
- разработка и применение инструментов измерения;
- определение шкалы измерения;
- отображение результатов измерения на шкалу;
- обработка, анализ и интерпретация результатов измерения.

Особенно важным этапом, собственно ради которого измерения и затеваются, является ана-

лиз и интерпретация результатов. Выполняя обработку и анализ результатов измерений, педагог отвечает на вопросы: позволяют ли выбранные инструменты измерения в достаточной степени определить уровни овладения обучающимися необходимыми знаниями, умениями и способами действий; каков процент учащихся, не освоивших (освоивших) программу на необходимом уровне; дают ли возможность применяемые инструменты дифференцировать достижения обучаемых; можно ли считать распределение результатов допустимым и как такое распределение можно охарактеризовать?

Одним из распространенных инструментов педагогических измерений является педагогический тест – это система взаимосвязанных заданий специфической формы (тестовые задания), определенного содержания, возрастающей сложности. Педагогический тест создается с целью объективно оценить структуру и качественно измерить уровень знаний.

Преимущества тестов очевидны:

- результаты оценки знаний и умений обучающихся не зависят от субъективного мнения преподавателя;
- результаты тестирования легко обрабатывать и визуализировать;
- стандартизированные тестовые задания легко формализуются и переводятся в компьютерную форму;
- высока скорость и надёжность выявления пробелов в знаниях и оценке уровня усвоения материала;
- возможен контроль большого количества обучающихся за относительно короткое время;

- тест может применяться для входного, текущего, промежуточного, итогового контроля и самоконтроля;

- тесты могут быть составлены с учетом индивидуального подхода к обучающимся.

Важнейшими требованиями к педагогическому тесту являются объективность, надежность и валидность.

Объективность требует, чтобы студенты были подвергнуты одному и тому же испытанию в аналогичных условиях, когда максимально исключены субъективные воздействия исследователей. Тестовые задания должны объективно показывать уровень знаний.

Под надежностью понимается характеристика точности, с какой может быть измерен тот или иной конкретный признак и устойчивость тестовых результатов к действию случайных факторов [2, с. 319].

Валидность теста – это характеристика качества теста, ориентированная на оценку адекватности теста поставленной цели его создания [2, с. 342]. Другими словами, тест валиден, если он позволяет измерить то, что и было целью измерить. Например, при отборе абитуриентов преследуется цель измерить уровень знаний, дифференцировать претендентов по уровню знаний и выявить лучших, соответственно тест должен полностью отвечать этой цели.

Интерпретация полученных баллов обучающихся зависит от видов тестов, от способа сопоставления баллов, от того, насколько тест апробирован.

В зависимости от способа сопоставления баллов обучающихся выделяют два вида тестов: нормативно-ориентированный и критериально-ориентированный.

При нормативно-ориентированном тестировании баллы тестируемых сопоставляются с нормами, как правило, это среднее значение тестовых баллов и показатель разброса (вариативности) вокруг среднего значения всех остальных баллов. Нормы получают эмпирически, на основе результатов выполнения теста определенной выборкой испытуемых – релевантной нормативной группой, репрезентативно представляющей генеральную совокупность тестируемых. Естественно, что прежде, чем применять нормативно-ориентированный тест, он должен быть тщательно, неоднократно проверен на выборках и скорректирован.

Нормативно-ориентированное тестирование применяют, если целью ставится ранжирование

испытуемых по уровню знаний и их дифференциация. Если результаты применения нормативно-ориентированного теста низкие, то при этом все равно можно ранжировать испытуемых и выделить «лучших» и «худших». При разработке такого теста, в результате апробации его на выборках испытуемых, важно исключить или доработать те задания, которые не дифференцируют обучающихся: задания, которые выполнили все испытуемые, и задания, с которыми не справился ни один человек.

При критериально-ориентированном подходе в педагогических измерениях результаты испытуемых интерпретируются по отношению к содержательной области или требованиям, установленным к учебным достижениям [1, с. 64]. Таким образом, критериально-ориентированный тест – это система заданий, позволяющая измерить уровень учебных достижений относительно полного объема знаний, умений и навыков, которые должны быть усвоены учащимися. Результаты такого тестирования показывают процент овладения обучающимся учебным материалом. Оценка результатов выполняется на основе критериев, установленных экспертами.

Критериально-ориентированное тестирование применяют, когда необходимо установить, знает ли испытуемый учебный материал (раздел, тему). В отличие от нормально-ориентированного тестирования при критериально-ориентированном возможно, что абсолютно все испытуемые успешно выполнили все задания, это будет означать лишь то, что группа хорошо освоила материал. Возможно и обратное: ни один испытуемый не справился с заданиями теста, это показывает, что учебный материал не усвоен. В любом случае тест работает и выполняет поставленную задачу, изменять его, дорабатывать в соответствии с полученными результатами нельзя.

Типичное распределение баллов по тестам для репрезентативной выборки обучающихся при нормативно-ориентированном и критериально-ориентированном подходах представлено на рис. 1.

Нормально-ориентированный педагогический тест может решать задачи и критериально-ориентированного теста. Если эти два подхода объединены, то тест является наиболее информативным для педагога. Примером разработки теста с расчетом на соотношение результатов с нормами и с содержанием являются контрольно-измерительные материалы ЕГЭ.

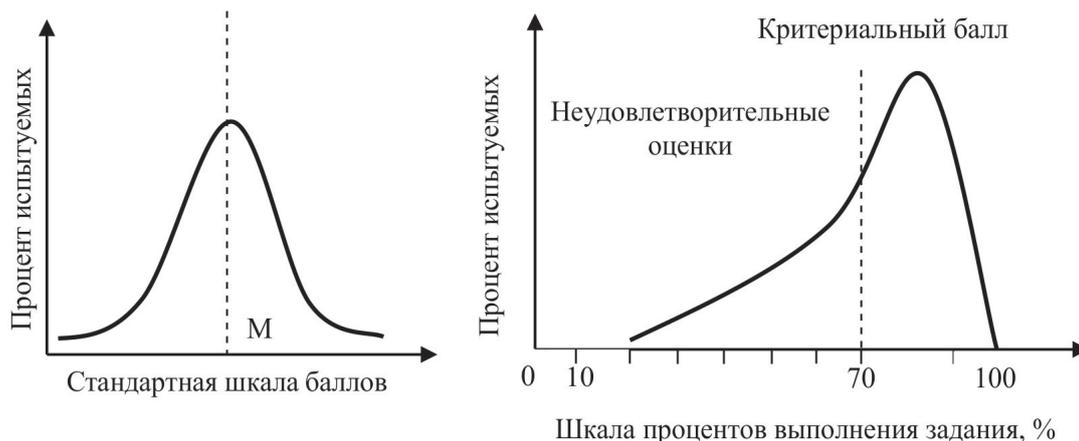


Рис. 1. Типичное распределение баллов по тестам для репрезентативной выборки

В статистической обработке результатов нормативно-ориентированного тестирования важную роль играет распределение Гаусса. Если неизвестны надежность и валидность нового теста, то для проверки его качества можно исследовать результаты тестирования на соответствие нормальному закону распределения, так как при качественных материалах и при больших выборках обычно получается распределение, близкое нормальному [3, с. 256]. Такой подход применяется, например, при визуализации результатов выпускных или вступительных экзаменов. Для демонстрации объективности теста вместе с баллами, набранными абитуриентами, приводится кривая распределения числа правильно решенных задач.

Анализу качества тестов и интерпретации результатов педагогических измерений уделяется большое внимание в современной научно-методической литературе [1–7].

В качестве примера рассмотрим обработку результатов тестирования группы из 65 студентов. В тесте предлагалось 16 заданий, каждое из которых оценивалось баллами, в сумме студент может получить максимально 100 баллов. Предварительно была проведена статистическая обработка результатов: построено 10 частичных интервалов $[c_i; c_{i+1}]$ (полученные баллы), найдены их середины x_i и подсчитаны наблюдаемые частоты n_i (число студентов, набравших эти баллы). Результаты группировки представлены в таблице.

№ интервала	c_i	c_{i+1}	x_i	n_i
1	0	10	5	1
2	10	20	15	1
3	20	30	25	3
4	30	40	35	7
5	40	50	45	11
6	50	60	55	16
7	60	70	65	15
8	70	80	75	6
9	80	90	85	3
10	90	100	95	2

Изучить случайную величину – это значит найти ее вид распределения, то есть подобрать функцию плотности распределения, наилучшим образом описывающую эксперимент. Гипотеза H_0 о виде распределения изучаемой случайной величины выдвигается на основе полигона частот (рис. 2).



Рис. 2. Полигон частот

Итак, выдвигаем гипотезу H_0 : случайная величина – количество набранных баллов – распределена по нормальному закону.

Параметры нормального закона a и σ^2 , являющиеся соответственно математическим ожиданием и дисперсией случайной величины X , неизвестны. Поэтому заменим их «наилучшими» точечными оценками, найденными по выборке: несмещёнными и состоятельными оценками математического ожидания a является выборочная средняя \bar{x}_e , а оценкой дисперсии σ^2 – «исправленная» выборочная дисперсия S^2 : $a \approx \bar{x}_e$, $\sigma^2 \approx S^2$.

Параметры вычислены с помощью Excel (рис. 3):

$$\bar{x}_e = \frac{\sum_{i=1}^m x_i n_i}{n} = 55,15,$$

По виду графика распределения баллов можно предположить нормальный закон распределения признака, так как кривая Гаусса хорошо сглаживает полученную линию. Обычно говорят: кривая аппроксимируется (приближается) к нормальной кривой.

$$S^2 = \frac{\sum_{i=1}^m (x_i - \bar{x}_e)^2 n_i}{n} = 17,76^2.$$

Для того, чтобы проверить соответствие предполагаемого закона эмпирическим данным, рассчитаем на его основе теоретические частоты $n'_i = n \cdot p_i$, где n – объём выборки, а p_i – вероятность попадания в интервал $[c_i, c_{i+1}]$.

Вероятность p_i попадания в интервал $[c_i, c_{i+1}]$ для непрерывной случайной величины может быть вычислена с помощью формулы:

$$p_i = P(c_i \leq X \leq c_{i+1}) = F(c_{i+1}) - F(c_i).$$

Для автоматизации расчетов применим функцию, связанную с нормальным распределением вероятностей:

НОРМРАСП (x ; среднее; стандартное_откл; интегральная).

	A	B	C	D	E	F	G	H
1	№ интервала	C_i	C_{i+1}	x_i	n_i	$P(C_i \leq X \leq C_{i+1})$	$n'_i = np_i$	$(x_i - \bar{x}_e)^2 n_i$
2	1	0	10	5	1	0,005	0,350	2515,408
3	2	10	20	15	1	0,018	1,194	1612,331
4	3	20	30	25	3	0,054	3,539	2727,763
5	4	30	40	35	7	0,118	7,697	2843,243
6	5	40	50	45	11	0,189	12,291	1134,107
7	6	50	60	55	16	0,222	14,411	0,379
8	7	60	70	65	15	0,191	12,408	1454,201
9	8	70	80	75	6	0,121	7,845	2363,219
10	9	80	90	85	3	0,056	3,641	2672,379
11	10	90	100	95	2	0,024	1,551	3175,432
12	Сумма				65	0,999	64,93	20498,462
13	$\bar{x}_e =$	55,15385						
14	$s =$	17,7584						

Рис. 3. Результаты расчётов

Выравнивающую теоретическую кривую Гаусса можно приближенно построить по точкам (x_i, n'_i) (рис. 4).

Если теоретическая кривая хорошо описывает эмпирическую линию, значит, вид и параметры распределения определены правильно и,

можно считать, что функция плотности распределения и функция распределения для изучаемой величины найдены. Однако визуально определить это невозможно, поэтому проверку правильности высказанной гипотезы проверяют с помощью критерия Пирсона χ^2 .

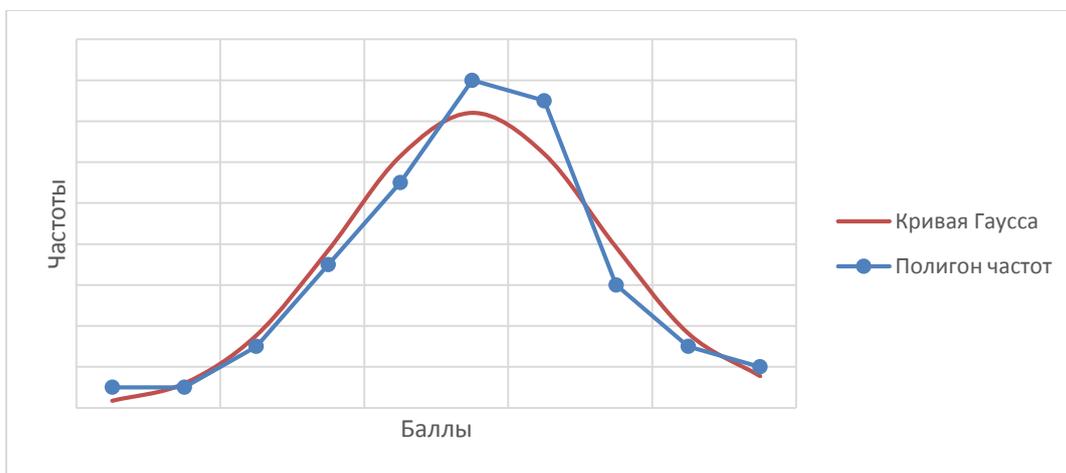


Рис. 4. Графики эмпирических и теоретических частот

Заметим, что статистика χ^2 имеет χ^2 -распределение лишь при $n \rightarrow \infty$, поэтому необходимо, чтобы в каждом интервале было достаточное количество наблюдений, по крайней мере 5 наблюдений. Имеет смысл объединить соседние интервалы № 1–3 и № 9, 10.

Определим меру расхождения эмпирических n_i и теоретических n'_i частот по наблюдаемому

значению критерия $\chi^2 = \sum_{i=1}^m \frac{(n_i - np_i)^2}{np_i} = 1,358$.

Критическое значение $\chi^2_{\alpha;k}$ при выбранном уровне значимости $\alpha = 0,05$ и числе степеней свободы $k = 4$ найдем в Excel с помощью встроенной функции ХИ2ОБР (рис. 5).

F28 fx =ХИ2ОБР(0,05;4)						
	A	B	C	D	E	F
	№	n_i		n'_i		$\frac{(n_i - n'_i)^2}{n'_i}$
16						
17	1	1	5	0,350	5,083	0,001
18	2	1		1,194		
19	3	3		3,539		
20	4	7		7,697		0,063
21	5	11		12,291		0,136
22	6	16		14,411		0,175
23	7	15		12,408		0,541
24	8	6		7,845		0,434
25	9	3	5	3,641	5,192	0,007
26	10	2		1,551		
27	Суммы	65		64,93	$\chi^2 =$	1,358
28					$\chi^2_{0,054} =$	9,488

Рис. 5. Вычисления критерия Пирсона

Фактически наблюдаемое значение χ^2 меньше критического, то есть $\chi^2 \leq \chi^2_{\alpha;k}$, следовательно, гипотеза H_0 не противоречит опытным данным.

Итак, получили, что случайная величина X – количество набранных баллов – распределена по нормальному закону. Следовательно, можно сделать вывод, что тест, предложенный студентам, отражает реальный уровень знаний. Таким образом, предложенная статистическая обработка результатов тестирования позволяет провести первичный анализ качества теста.

Полученные тестовые баллы не всегда распределены нормально. Часто встречаются случаи асимметричного расположения полученного вариационного ряда.

Необходимым и достаточным условием симметричного расположения вариационного ряда является равенство средней арифметической, моды и медианы: $\bar{X} = Me = Mo$.

Правосторонняя асимметрия имеет следующий вид распределения частот – результатов выполнения теста (рис. 6).

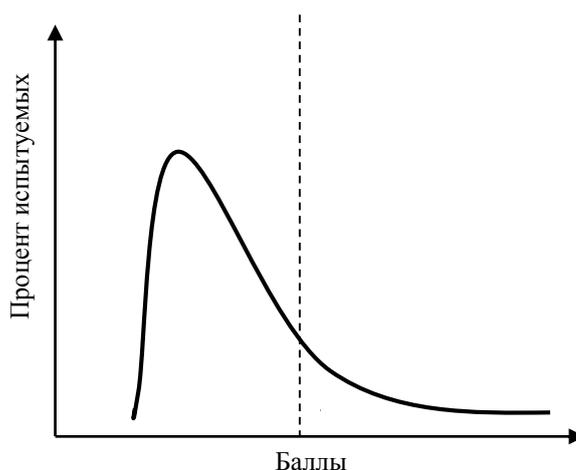


Рис. 6. Правосторонняя асимметрия

Три основные характеристики такого распределения соотносятся следующим образом:

$$\bar{x} > Me > Mo.$$

Если тест критериально-ориентирован, то по такому асимметричному расположению баллов испытуемых можно судить о не усвоении группой испытуемых учебного материала. Выполнены требуемые критерии лишь отдельными обу-

чаемыми. Если это нормативно-ориентированный тест, и он апробирован, скорректирован и выверен эмпирически, то можно сказать, что большинство испытуемых подготовлены на уровне, ниже среднего, лишь единицы опрошенных показали высокие результаты.

Левосторонняя асимметрия имеет следующий вид распределения (рис. 7).

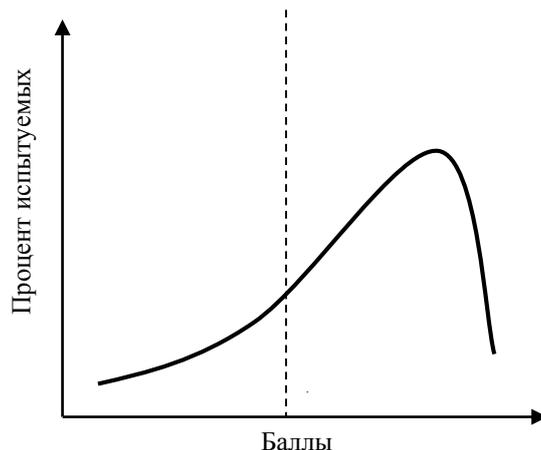


Рис. 7. Левосторонняя асимметрия

Три основные характеристики такого распределения соотносятся следующим образом:

$$\bar{x} < Me < Mo.$$

Если тест критериально-ориентирован, то по такому расположению баллов испытуемых можно судить о высоком уровне усвоения ими учебного материала. Если это апробированный нормативно-ориентированный тест, то можно сказать, что группа испытуемых «сильная», но в данной группе по применяемому тесту трудно дифференцируются испытуемые, показавшие высокие результаты. Для более четкой дифференциации стоит усложнить тест.

Для выявления асимметрии вовсе не обязательно строить диаграммы и графики, достаточно рассчитать показатель асимметрии – коэффициент асимметрии Пирсона:

$$K_a = \frac{\bar{x} - Mo}{\sigma}.$$

Если $K_a > 0$, асимметрия правосторонняя, если $K_a < 0$ – левосторонняя. Равенство $K_a = 0$ говорит о симметричности ряда.

Примером постоянной доработки теста являются контрольно-измерительные материалы ЕГЭ. В частности, контрольно-измерительные материалы ЕГЭ по информатике и ИКТ за 2014–2016 годы претерпели изменения в сторону усложнения тестовых заданий, так как данный экзамен не обязателен, является профильным и рассчитан на испытуемых, изучавших информатику углубленно.

На рисунках 8 и 9 представлены гистограммы распределения тестовых баллов участников ЕГЭ по информатике и ИКТ в Алтайском крае в 2014 и 2015 годах.

По распределению баллов видно, что и в 2014 и в 2015 годах группы испытуемых условно делятся на 2 подгруппы: первая – участники, не подготовленные к ЕГЭ по информатике и ИКТ, освоившие в лучшем случае дисциплину на базовом уровне; вторая – участники, с хорошим уровнем подготовки по углубленному курсу информатики и ИКТ, знакомые с контрольно-измерительными материалами ЕГЭ и методами решения задач. Как видно, первая группа испы-

туемых и в 2014 г. и в 2015 г. практически не набрали даже минимально необходимого коли-

чества баллов (40), учитываемого при поступлении в вузы.

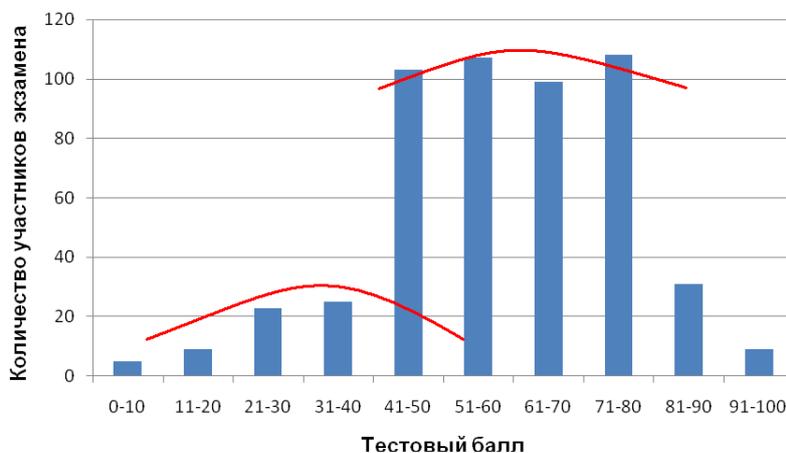


Рис. 8. Распределение тестовых баллов участников ЕГЭ по информатике и ИКТ в Алтайском крае в 2014 г.

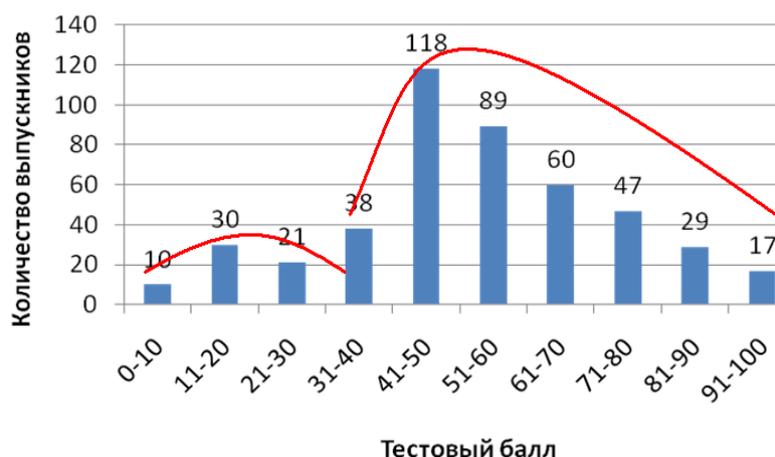


Рис. 9. Распределение тестовых баллов участников ЕГЭ по информатике и ИКТ в Алтайском крае в 2015 г.

Во второй группе испытуемых результаты экзамена в 2014 г. и в 2015 г. заметно разнятся. В 2014 году основная масса учащихся, подготовленных к ЕГЭ по информатике и ИКТ, равномерно распределена по баллам от 41 до 80, и небольшая часть получили баллы больше 80. Результаты на рисунке 8 не дают больших возможностей для дифференциации хорошо подготовленных испытуемых и выделения лучших из них. В 2015 году, после изменения тестовых заданий, направленного на усиление профилирования теста, группа хорошо подготовленных испытуемых демонстрирует результаты, сходящиеся к кривой с пра-

восторонней асимметрией, что подтверждает увеличение сложности заданий. Однако в таком варианте наблюдается хорошая возможность четкой дифференциации участников экзамена с высоким уровнем подготовки, так как правая часть кривой, описывающей гистограмму, сходится к нормальной кривой (рис. 9).

Таким образом, применение статистических методов обработки данных позволяет выявить характеристические особенности подготовки групп обучаемых, а так же принять решения о характере необходимых качественных изменений тестовых заданий.

Библиографический список

1. Звонников, В. И. Современные средства оценивания результатов обучения : учебное пособие для студентов высш. учеб. заведений / В. И. Звонников, М. Б. Челышкова. — Москва : Академия, 2009. — 224 с.
2. Челышкова, М. Б. Теория и практика конструирования педагогических тестов : учебное пособие / М. Б. Челышкова. — Москва : Логос, 2002. — 432 с.
3. Родкевич, О. Б. Объективная диагностика качества знаний / О. Б. Родкевич, В. В. Мамонов, Н. Ф. Стась // Механизмы гарантии качества образования: системы, технологии, инновации : сборник тезисов докладов международной научно-практической конференции. — Барнаул : Изд-во АлтГТУ, 2009. — 282 с.
4. Афонина, М. В. ЕГЭ как показатель результатов обучения информатике и ИКТ [Электронный ресурс] / М. В. Афонина, Е. И. Апольских // Педагогическое образование на Алтае. — 2015. — № 1. — Режим доступа: <http://journals.altspu.ru/pedagogical-education/article/view/219/214>, свободный. — Загл. с экрана.
5. Ильина, М. А. Анализ качества тестов с помощью статистических методов обработки данных / М. А. Ильина, Н. Т. Копылова // Математические методы и информационные технологии в социально-экономической сфере : сборник статей по материалам IV всероссийской научно-практической конференции. — Уфа, 2015. — С. 171–178.
6. Ильина, М. А. Новый подход к тестированию в рамках заочного образования / М. А. Ильина, Н. Т. Копылова // Экономика и социум. — 2014. — № 4–6 (13). — С. 227–230.
7. Щербинина, М. В. Педагогическое тестирование как средство измерения уровня знаний школьников по информатике / М. В. Щербинина // Вестник Алтайской государственной педагогической академии. — 2001. — № 1. — С. 31–39.