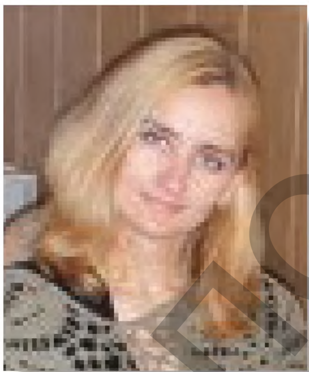
ПРОВЕРКА НЕКОТОРЫХ ПЕДАГОГИЧЕСКИХ ГИПОТЕЗ НА ОСНОВЕ СТАТИСТИЧЕСКОГО АНАЛИЗА РЕЗУЛЬТАТОВ ЦТ ПО МАТЕМАТИКЕ

Малиновский Василий Васильевич,
первый проректор
ВГУ имени П.М. Машерова,
кандидат педагогических наук

Чиркина Анна Александровна,
доцент кафедры информатики
и информационных технологий
ВГУ имени П.М. Машерова,
кандидат биологических наук



Булгакова Наталья Валентиновна,
старший преподаватель
кафедры информатики
и информационных технологий
ВГУ имени П.М. Машерова

ЗНАЧИМОСТЬ ПЕДАГОГИЧЕСКИХ ГИПОТЕЗ – АНАЛИЗ

В статье проведен анализ статистической значимости некоторых педагогических гипотез.

Введение. В повседневной жизни мы часто формулируем различные педагогические гипотезы. С математической точки зрения гипотеза – это высказывание, для которого нужно установить его истинность или ложность. Однако подход к проверке педагогических гипотез часто основан лишь на здравом смысле и в нем используются качественные характеристики: «логично», «очевидно», «понятно». В таком случае гипотеза фактически переводится в ранг аксиомы, то есть высказывания, истинность которого принимается без доказательства. Второй вариант проверки педагогических гипотез на достоверность состоит в применении некоторых количественных характеристик, позволяющих оценить интересующие нас результаты исследований, например: количество набранных за тест баллов; доля правильных ответов; время, затраченное на выполнение задания; количество ошибок в работе. Важно, какие параметры (свойства, характеристики, признаки) педагогического процесса выбраны для исследования. Некоторые привычные и кажущиеся логичными качественные оценки могут оказаться ошибочными, количественные характеристики – неприменимыми для того или иного конкретного случая.

Целью данной работы является анализ статистической значимости некоторых достаточно распространенных и «очевидных» педагогических гипотез на основе результатов централизованного тестирования по математике. Рассматривались такие количественные характеристики, как тестовый балл и количество выполненных заданий для каждого участника ЦТ.

Основная часть. Представленные гипотезы сформулированы на основе результатов ЦТ по математике абитуриентов, которые проходили тестирование на базе Витебского государственного университета имени П.М. Машерова с 2006 по 2014 год (8213 тестируемых). Отметим, что данные педагогические гипотезы не имеют математического содержания и могут быть сформулированы относительно абитуриентов и студентов других специальностей. Однако их истинность может не совпасть с нашими результатами. Укажем также, что любое заключение, полученное из статистического анализа, строится на конечном числе наблюдений, поэтому оно не полно и может быть не достоверно. Статистическая проверка достоверности любой гипотезы никогда не носит абсолютного характера, то есть истинность или ложность проверяемого утверждения имеет вероятностный характер – выполняется с некоторой вероятностью. В нашем исследовании при принятом уровне значимости α = 0,05 такая вероятность составляет 95%. Заметим также, что для работы с педагогическими гипотезами нужно точно определять категории, с помощью которых они формулируются. Иными словами, нужно точно описывать понятия «улучшается», «ухудшается», «меняется» и прочие.

1. Первая гипотеза: **«Сложность тестов по математике возрастает год от года».** Считается, что сложность теста как совокупности заданий возрастает, если достоверно установлено увеличение числа заданий более сложной категории за счет заданий более простой категории.

Когда говорят о качестве и, в частности, сложности тестовых заданий, имеют в виду две характеристики: трудность и дифференцирующая способность (дискриминативность). Для расчета этих характеристик используются две теории: классическая теория тестирования (Classical Test Theory, CTT) и математическая теория измерений (Item Response Theory, IRT) [1; 2].

Трудность в классической теории тестов определяется как доля тестируемых, которые справились с заданием: $pj = Rj / N$, где N – общее количество испытуемых, Rj – число правильных ответов, полученных по заданию с номером j.

Математическая теория измерений IRT является психолого-педагогическим вариантом более общей методологии латентно-структурного анализа (Latent Structure Analyses, LSA), нацеленного на выявление латентных качеств личности посредством вероятностно-статистического моделирования. Латентными называются качества личности, недоступные для непосредственного измерения, например: «подготовленность обучаемых», «знание учебной дисциплины», «интеллектуальное развитие» и многое другое. Для количественной оценки характеристик тестовых заданий используются однопараметрическая и двухпараметрические модели IRT [3]. Однопараметрический вариант предложен Георгом Рашем (G. Rasch). Ключевая идея модели Раша может быть сформулирована следующим образом: вероятность правильного ответа на тестовое задание зависит от уровня подготовленности тестируемого и трудности тестового задания. Основная логистическая однопараметрическая модель Раша:

$$p_{ij} = \frac{e^{d(\theta_i - \beta_j)}}{1 + e^{d(\theta_i - \beta_j)}}, \qquad (1)$$

где p_{ij} – вероятность правильного ответа i-ого испытуемого на j-ое задание, θ_i – уровень подготовленности i-ого испытуемого, β_j – мера трудности j-ого задания, d – константа шкалирования, которая вводится в связи с использованием логистической функции для описания латентных параметров, распределение которых предполагается нормальным.

Было получено следующее распределение тестовых заданий по трудности (табл. 1).

Данные, представленные в табл. 1, показывают, что во все годы в тестах преобладали задания среднего уровня сложности. Чтобы выяснить, становились ли задания труднее с годами, был проведен ранговый анализ вариаций по Краскелу-Уоллису для характеристик «трудность тестового задания» и «доля правильных ответов». В обоих случаях выявлено, что существенные статистические различия между значениями выбранных показателей по годам отсутствуют.

Вторая характеристика тестовых заданий – это их дифференцирующая способность или дискриминативность, то есть способность задания разделять хорошо подготовленных и слабо подготовленных тестируемых. Двухпараметрическая модель Бирнбаума (A. Birnbaum):

$$p_{ij} = \frac{e^{a_j(\theta_i-\beta_j)}}{1+e^{a_j(\theta_i-\beta_j)}}, \qquad (2)$$

где параметр a_j характеризует дискриминативность задания теста.

По уровню дифференцирующей способности F. Baker делит тестовые задания на шесть уровней. В табл. 2 продемонстрировано распределение тестовых заданий ЦТ по математике по дискриминативности с использованием параметров IRT.

Сравнение групп по Краскелу-Уоллису для характеристики «дифференцирующая способность задания» показало, что значимые статистические различия по годам отсутствуют.

В целом полученные результаты показывают сбалансированность тестов по уровню трудности и дифференцирующей способности. Однако каждый тест по математике содержал ряд очень трудных заданий с низкой различающей способностью. В то же время большая часть заданий со средним уровнем трудности показала хорошую дискриминативность.

В результате проведения сравнительного анализа характеристик тестовых заданий в 2006–2014 годах существенных изменений в распределении заданий по трудности и дискриминативности не выявлено. Среди тестовых заданий преобладают задания среднего уровня сложности, что позволяет точнее определить средний уровень подготовки контингента участников тестирования. Большинство заданий обладает низкой и очень низкой дифференцирующей способностью. В то же время в тестах присутствуют очень трудные задания с низкой дискриминативностью, целесообразность включения этих заданий в тест вызывает сомнение. Гипотеза не получила своего подтверждения.

2. Следующая гипотеза: **«С введением нижней границы ЦТ абитуриенты стали хуже справляться с тестовыми заданиями».** «Хуже» означает, что достоверно установлено увеличение числа абитуриентов с низкими баллами (ниже установленной границы) за счет абитуриентов с баллами выше установленной границы. В 2006 и 2007 годах тест по математике содержал 25 вопросов и задач, а с 2008 года – 30. Поэтому далее рассматриваем данные с 2008 года. В табл. 3 представлена информация о доле тестируемых, не набравших пороговый балл в ЦТ по математике.

Таблица 1 – **Распределение тестовых заданий ЦТ по трудности**

Градации трудности задания (β_j)	Год тестирования								
	2006	2007	2008	2009	2010	2011	2012*	2013	2014
Очень трудные (более 2,6)	2	5	4	7	4	7	4	5	9
Трудные (от 1,5 до 2,59)	7	2	3	4	5	4	3	3	3
Среднего уровня (от –1,49 до 1,49)	16	16	20	16	19	17	20	20	15
Легкие (от –2,59 до –1,5)	0	2	3	1	1	2	2	1	2
Очень легкие (менее –2,6)	0	0	0	2	1	0	0	1	1

* Из 30 заданий одно задание было исключено из рассмотрения, так как ни один из тестируемых его не выполнил.

Таблица 2 – **Распределение тестовых заданий ЦТ по дискриминативности**

Дискриминативность (aj)	Год тестирования								
	2006	2007	2008	2009	2010	2011	2012	2013	2014
Отсутствует (от 0 до 0,009)	0	0	0	0	0	1	0	0	0
Очень низкая (от 0,01 до 0,34)	6	1	10	9	11	8	5	12	9
Низкая (от 0,35 до 0, 64)	17	13	16	16	18	17	18	17	18
Средняя (от 0,65 до 1,34)	2	2	4	5	1	4	6	1	3
Высокая (от 1,35 до 1,69)	0	0	0	0	0	0	0	0	0

Таблица 3 – **Доля тестируемых, набравших менее 15 баллов по математике в ЦТ с 2008 по 2014 год**

Год тестирования	2008	2009	2010	2011	2012	2013	2014
Количество (%)	45%	47%	49%	56%	51%	54%	55%

Таким образом, пороговый балл, равный 15, не удалось преодолеть половине тестируемых, что значительно больше запланированных 30%. Имеется тенденция к увеличению этого показателя. Однако указанное увеличение не является следствием введения нижней границы, а обусловлено иными причинами, поскольку устойчивый его рост начался ранее. Гипотеза не подтвердилась.

На рис. 1 представлено распределение тестируемых по количеству выполненных тестовых заданий с 2008 по 2014 год. Наибольшая доля тестируемых выполнила 5 тестовых заданий (13,1%), только 8,2% справились с половиной и более заданий теста.

На следующем рисунке показано среднее количество баллов, соответствующих числу выполненных тестовых заданий по математике, и интервалы, характеризующие разброс баллов относительно среднего значения (±3s, где s – стандартное отклонение).

График, представленный на рис. 2, показывает, что при выполнении определенного количества заданий результаты 99,7% тестируемых попадут в указанный интервал баллов. Например, пятнадцати выполненным заданиям соответствует среднее значение количества набранных баллов, равное 41, при этом количество баллов в этом случае может варьироваться в интервале от 34 до 48. Полученный график приводит нас к обсуждению следующей гипотезы.

3. Гипотеза: **«Набрать количество баллов, не меньшее нижней границы, можно простым случайным проставлением меток в первом разделе заданий ЦТ».**

Интервал варьирования важно знать для определения количества заданий, которые нужно решить, чтобы гарантированно набрать пороговый балл, позволяющий поступать в вуз. Для того чтобы набрать 15 баллов, нужно решить в среднем 6,5 заданий, а чтобы гарантированно преодолеть пороговый балл, должно быть выполнено, как минимум, 9 заданий. В то же время, по данным тестирования, семь заданий и более решает 47% тестируемых, 9 заданий и более – 29,5% тестируемых. То есть всего треть тестируемых гарантированно набирает проходной балл. При этом 9 заданий – это половина раздела А теста по математике. Просто угадать – теоретически можно, но вероятность этого события чрезвычайно мала. Гипотеза не подтвердилась.

На следующем рисунке отображена средняя доля правильных ответов на тестовые задания по математике. Задания раздела В (с 19 по 30) выполняются хуже, чем задания раздела А (с 1 по 18). Интересно отметить, что почти во все рассматриваемые годы седьмое задание выполнялось плохо, вне зависимости от его содержания и представленного раздела математики.

В результате получен инструмент, позволяющий прогнозировать количество набранных баллов в зависимости от числа выполненных заданий. Если сложность тестов не будет меняться, то для гарантированного преодоления порогового значения баллов необходимо правильно выполнить не менее 9 тестовых заданий.

Рисунок 1 – **Распределение тестируемых по количеству выполненных тестовых заданий.**

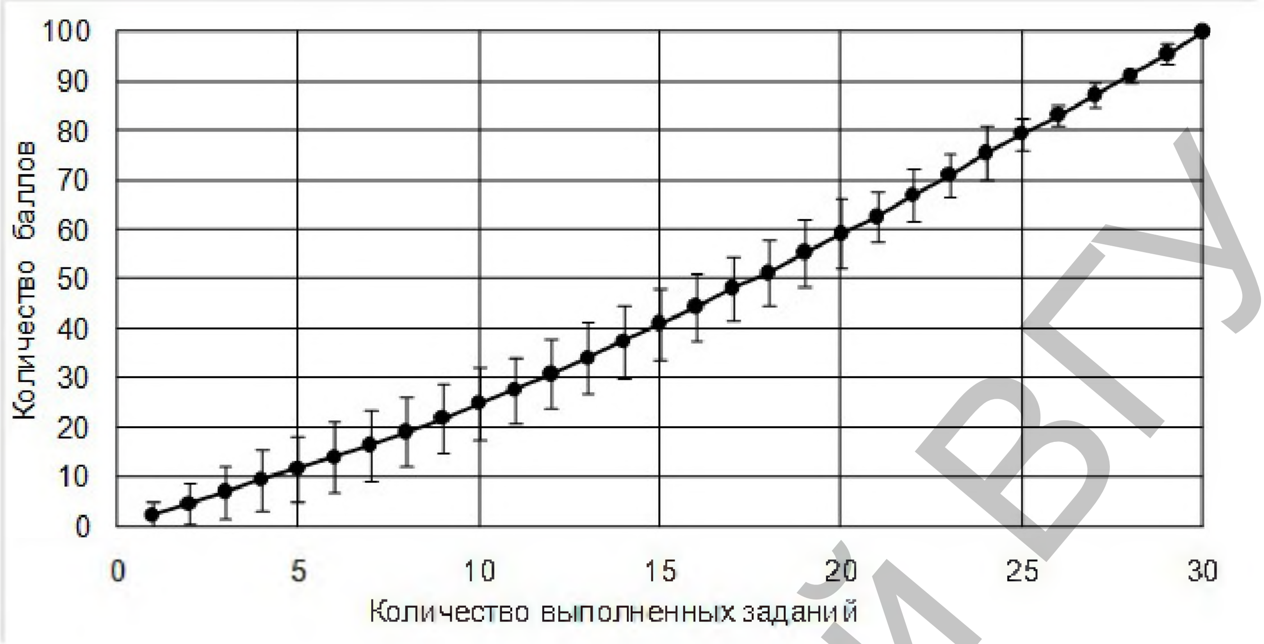


Рисунок 2 – **Количество баллов, соответствующих числу выполненных тестовых заданий.**

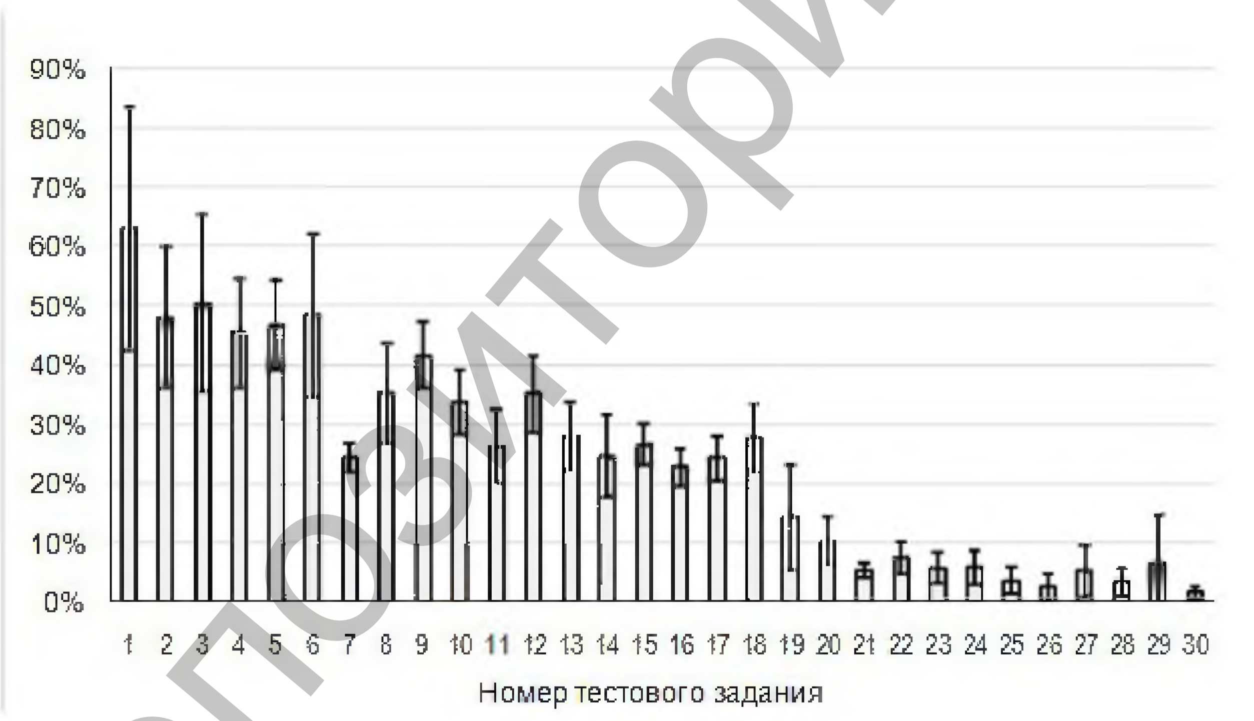


Рисунок 3 – **Доля правильных ответов на тестовые задания.**

Заключение. Таким образом, формулируя те или иные педагогические гипотезы для принятия решений, следует понимать, что истинность таких высказываний, даже если она кажется очевидной на основе имеющегося опыта, не всегда оказывается верной.

ЛИТЕРАТУРА

1. Аванесов, В.С. Композиция тестовых заданий / В.С. Аванесов. – М.: Изд-во Центра тестирования Минобразования РФ, 2002.
2. Челышкова, М.Б. Теория и практика конструирования педагогических тестов: учеб. пособие / М.Б. Челышкова. – М.: Логос, 2002.
3. Феськов, Н.С. Математические методы интерпретации результатов нормативно-ориентированного тестирования / Н.С. Феськов, А.П. Якобчук // Адукацыя і выхаванне. – 2007. – № 3. – С. 76–94.