

Теория

АНАЛИТИЧЕСКИЙ МЕТОД РЕШЕНИЯ ОСНОВНОЙ ЗАДАЧИ ПЕДАГОГИЧЕСКИХ ИЗМЕРЕНИЙ

Юрий Каргин

Ноябрьский колледж
профессиональных и информационных технологий
kargin04@yandex.ru

В работе представлен аналитический метод решения основной задачи педагогических измерений — измерение уровня подготовленности испытуемых и уровня трудности тестовых заданий. Подробно представлен алгоритм и приведён пример решения поставленной задачи.

Ключевые слова: теория педагогических измерений, модель Раша, функция распределения, аналитический метод решения основной задачи педагогических измерений.

Введение

Основная задача теории педагогических измерений состоит в разработке методов измерения уровня подготовленности испытуемых и уровня трудности заданий. Классическим примером решения этой задачи является преобразование эмпирических тестовых дан-

ных в интервальную шкалу логитов на основе однопараметрической модели педагогических измерений Г. Раша.

В предыдущей работе автора¹ была предложена альтернативная однопараметрическая модель педагогических измерений, с идеей измерения уровня подготовленности испытуемых и уровня трудности заданий в шкале отношений. Сама технология проведения таких измерений только разрабатывается, и её возможности ещё подлежат исследованию. Поиску альтернативных путей решения основной задачи МТИ и посвящена данная работа.

Две логики построения математической модели педагогических измерений (МТИ)

МТИ разворачивается из об-суждения латентной природы и эмпирически-интуитивного толкования трудности тестового задания и подготовленности испытуемого, через понятие шанс правильного ответа на тестовое задание. Однако эти толкования могут быть различными. Различные толкования исходных понятий приводят и к различным моделям теории педагогических измерений, и к различным методикам их применения.

Сначала остановимся на модели Г. Раша, в рамках которой

приводится четыре меры трудности тестового задания и аналогичные четыре меры подготовленности испытуемого к заданиям теста².

Первая пара этих мер достаточно очевидна и легко интерпретируется: доля неправильных ответов испытуемых на j -е задание проектируемого теста q_j ; доля правильных ответов i -м испытуемым на тестовые задания p_i . Однако это не метрические показатели. Их значения есть результат счета, а не измерения.

Вторая пара мер — соответствующие долям ответов шансы³. Именно эти меры были предложены Г. Рашем для оценки результативности тестовых испытаний, которые и легли в основу его математико-статистической теории педагогических измерений. Зная доли ответов, шансы вычисляются по формулам: — шанс (потенциал) трудности тестового задания; — шанс (потенциал) подготовленности испытуемого. Следует отметить, что в левой части этих выражений стоят оцениваемые латентные показатели, а в правой — наблюдаемые значения. Остановимся подробнее на интерпретации этих показателей. Именно на этом этапе логика построения нашей альтернативной модели отклонилась от схемы Г. Раша.

Зададимся вопросом — как количественно выразить выска-

Теория

1

Каргин Ю.

Построение альтернативной модели педагогических измерений по системе Г. Раша // Педагогические измерения. 2010. № 4. С. 62–71.

2

Аванесов В.

Item Response Theory: основные понятия и положения // Педагогические измерения. 2007. № 2. С. 3–28.

3

В некоторых прикладных вопросах теории вероятностей вместо значения вероятности некоторого события используют величину «шанс». Шанс h есть отношение вероятности благоприятного события p к вероятности того, что это событие не произойдет: $h = p/(1 - p)$, или обратно $p = h/(1 + h)$. Иногда значение этой величины удобнее интерпретировать и вычислять прямым подсчетом количества равнозначных конфликтующих факторов. Например, при бросании игральной кости шанс выпадения благоприятной цифры «6» равен один к пяти, $h = 1/5$, т.е. один благоприятный фактор «6» против пяти противоположных факторов

ПЕД
измерения

«не 6». Отсюда вероятность благоприятного события равна $p = (1/5)/(1 + 1/5) = 1/6$, а вероятность противоположного — $q = 1/(1 + h) = 5/6$.

4

Убедимся в этом непосредственными расчётами. Пусть даны два испытуемых и два тестовых задания с известными количествами конфликтующих факторов: $Y_1 = 2, Y_2 = 1$ — для испытуемых, $Z_1 = 2, Z_2 = 1$ — для тестовых заданий. Тогда можно рассчитать шансы и вероятности правильных ответов первым испытуемым на отдельные задания теста и усреднённую вероятность правильных ответов на весь тест: $2/2$ и $2/1$ — соответствующие шансы; $1/2$ и $2/3$ — соответствующие вероятности; $7/12$ — усреднённую вероятность. Отсюда рассчитанный шанс (потенциал подготовленности) первого испытуемого для всего теста равен $7/5$. Аналогично для второго испытуемого получим: $1/2, 1/1$ и $5/7$ — соответствующие шансы для отдельных заданий и всего теста. Тогда отношение шансов подготовленности для отдельных заданий действительно равно 2, а для всего теста $49/25 \approx 2$.

зывание «первый испытуемый в два раза лучше второго подготовлен к тестовому заданию»? В логике Г. Раша ответ следующий — у первого испытуемого потенциал подготовленности (шанс выполнения задания) в два раза больше, т.е. $\Theta_1/\Theta_2 = 2$. Причём это отношение не зависит от тестового задания и справедливо не только для любого из них, но и (внимание, логическая ошибка⁴) для всего теста.

Далее, по схеме Раша, переходя от мультипликативных потенциалов B и Θ к их аддитивным эквивалентам $\beta = \ln B$ и $\theta = \ln \Theta$, имеем третью пару оценок латентных показателей по исходным тестовым баллам:

$$\beta_j = \ln \frac{q_j}{1 - q_j} \text{ — исходный уровень трудности тестового задания;}$$

$$\theta_i = \ln \frac{p_i}{1 - p_i} \text{ — исходный уровень подготовленности испытуемого.}$$

Эти рассуждения приводят к самой математической модели Г. Раша. Так как вероятность правильного ответа на тестовое задание выражается через отношение шансов по формуле

$$P = \frac{\Theta}{1 + \frac{\Theta}{B}}, \text{ то при } B = \exp \beta$$

$$\text{и } \Theta = \exp \theta, \text{ окончательно имеем}$$

$$P = \frac{\exp(\theta - \beta)}{1 + \exp(\theta - \beta)}.$$

Логическая ошибка, о которой говорилось выше, искажает исходные оценки. Если распределения близки к равномерным, то этот недостаток хорошо корректируются статистическим выравниванием — шкалированием. Такая процедура пересчёта исходных оценок в симметричные относительно нуля метрические значения β и θ и даёт окончательную, четвёртую пару мер трудности тестовых заданий и подготовленности испытуемых при педагогическом измерении. Правда, в этом случае теряется их исходная интерпретация, но относительное распределение участников теста в интервальной метрической шкале логитов вполне обосновано.

Когда распределения оценок носят явно нелинейный вид, то и стандартные процедуры шкалирования исходных тестовых баллов не позволяют получить качественные оценки латентных показателей на шкале логитов. В этом случае желаемый эффект может дать переход к известным модификациям модели Раша — двухпараметрической или трёхпараметрической модели Бирнбаума. Действительно, добавление в математическую модель корректирующих параметров обычно улучшает статистические показатели оценок, хотя и усложняет процесс их получения. Но даже и в этом случае эмпирические данные после обработки ме-

тодами IRT не всегда приводят к пригодным для практических целей результатам.

Вернёмся к вопросу — как можно количественно выразить высказывание «первый испытуемый в два раза лучше подготовлен к ответу на тестовое задание, чем второй»? Предложим альтернативный вариант — первый испытуемый в два раза лучше второго подготовлен к выполнению тестового задания, если он выполнит два таких задания с тем же шансом, что и второй выполнит только одно. Этот посыл приводит к следующей математической модели педагогических измерений:

$P = 0,5^{\frac{B}{\Theta}}$. Здесь применены те же обозначения B и Θ , но их содержательный смысл не эквивалентен строгому понятию шанс. В альтернативной модели они формулируются как относительные показатели уровня трудности тестового задания и уровня подготовленности испытуемого.

Существенным преимуществом этой модели является её относительная математическая простота. Это преимущество позволяет более строго оценивать искомые показатели. В частности, при равномерном распределении участников теста искомые показатели можно оценить по уже известным соотношениям:

$$B_j = \frac{q_j}{1 - q_j} \text{ и } \Theta_i = \frac{p_i}{1 - p_i}. \text{ Только}$$

в этом частном случае относительные показатели трудности тестовых заданий и подготовленности испытуемых можно интерпретировать как соответствующий шанс. В общем случае соотношения для оценки искомых показателей по наблюдаемым долям q_j и p_i такой интерпретации не имеют.

И, наконец, если в модели Раша основными показателями результатов анализа теста и их интерпретацией служат измеряемые в логитах уровни трудности тестовых заданий β и уровни подготовленности испытуемых θ , то в альтернативном подходе эти функции выполняют вероятностные показатели уровня трудности тестовых заданий $v = 0,5^B$ и уровня подготовленности испытуемых $u = 0,5^{1/\Theta}$. Введение вероятностных показателей позволяет записать и саму модель педагогических измерений в других видах: $P = u^B$ или $P = v^{1/\Theta}$. Методы измерения именно этих показателей (β и θ — в схеме Раша, v и u — в альтернативной схеме) анализируются и сопоставляются в следующем подразделе работы.

Обратная задача педагогического тестирования

Выше были приведены две мысли, которые попытаемся обосновать в этой части работы. Повторим их:

- не всегда обработка эмпирических данных методами IRT приводит к пригодным для практических целей результатам;

- относительная математическая простота альтернативной модели позволяет находить оценки искомых показателей аналитическими методами.

Если оценка уровня подготовленности испытуемых и уровня трудности заданий по эмпирическим данным формулируется как прямая и основная задача теории педагогических измерений, то обратную задачу педагогического тестирования можно сформулировать следующим образом: по известным значениям показателей уровня подготовленности испытуемых и уровня трудности заданий, в ходе тестирования получить, а затем и проанализировать совокупность эмпирических данных.

Конкретизируем и упростим эту задачу в следующем виде. Пусть дана группа испытуемых с известными уровнями подготовленности к тестовым заданиям — это входные значения. Проведём с этой группой тестовые испытания, и уже к наблюдаемым данным применим исследуемый метод их обработки. Получим выходные значения показателей уровня подготовленности испытуемых. Поэтому, насколько выходные значения соответствуют входным, можно не только оценить каче-

ство самих методов обработки эмпирических данных, но и области их применения. Именно эту задачу и попытаемся решить здесь, выделив в ней четыре этапа:

1. Определить входные значения характеристических показателей педагогического тестирования.

2. Провести имитационные тестирования и получить наборы эмпирических данных.

3. По наблюдаемым данным провести оценки характеризующих показателей.

4. Сопоставить полученные оценки с заданными на входе значениями.

1 этап. Количественный состав виртуальных участников теста определим следующим образом — 100 испытуемых и 100 тестовых заданий.

На входе в имитационное тестирование испытуемые всегда распределены в 10 равновеликих групп с определённым значением подготовленности к тестовым испытаниям (табл. 1). Для схемы Раша уровни подготовленности изменяются в диапазоне от -5 логит до $+5$ логит с шириной интервала для каждой группы 1 логит. Для альтернативной схемы вероятностные уровни подготовленности испытуемых изменяются во всей области определения от 0 до 1 с шириной интервала 0,1. Чем выше порядковый номер группы, тем выше подготовленность испытуемых.

Таблица 1

Равномерное распределение испытуемых

Номер группы	1	2	3	4	5	6	7	8	9	10
θ , уровень подготовленности, логит	-4,5	-3,5	-2,5	-1,5	-0,5	0,5	1,5	2,5	3,5	4,5
u , вероятностный уровень подготовленности	0,05	0,15	0,25	0,35	0,45	0,55	0,65	0,75	0,85	0,95
Количество испытуемых	10	10	10	10	10	10	10	10	10	10

Три набора тестовых заданий так же распределены по трудности в 10 групп. Количество заданий в каждой группе определяется в зависимости от выбора вида распределения (табл. 2, рис. 1): равномерное, линейное, нелинейное. Отметим, что в рассматриваемом примере, при линейном распределении тестовых заданий, количество заданий в каждой группе обратно пропорционально уровню трудности этих заданий, а при нелинейном распределении самые трудные задания вообще исключаются из теста.

2-й этап — имитационное тестирование. На этом этапе в подготовленную таблицу размером 100×100 «случайно вбрасываются» двоичные результаты ответов 0/1 на тестовые задания по формуле: $X = INT(P + Rnd)$, где P — рассчитанная по исследуемой модели вероятность правильного ответа, Rnd — случайное число из интервала от 0 до 1. Полученную таблицу наблюдаемых данных можно использовать для расчёта оценок показателей подготовленности испытуемого по соответствующей схеме — Раша или альтернативной.

Таблица 2

Распределения тестовых заданий

Номер группы	1	2	3	4	5	6	7	8	9	10
β , уровень трудности, логит	-4,5	-3,5	-2,5	-1,5	-0,5	0,5	1,5	2,5	3,5	4,5
v , вероятностный уровень трудности	0,95	0,85	0,75	0,65	0,55	0,45	0,35	0,25	0,15	0,05
Равномерное	10	10	10	10	10	10	10	10	10	10
Линейное	19	17	15	13	11	9	7	5	3	1
Нелинейное	22	34	24	13	5	2	0	0	0	0

ПЕД
измерения

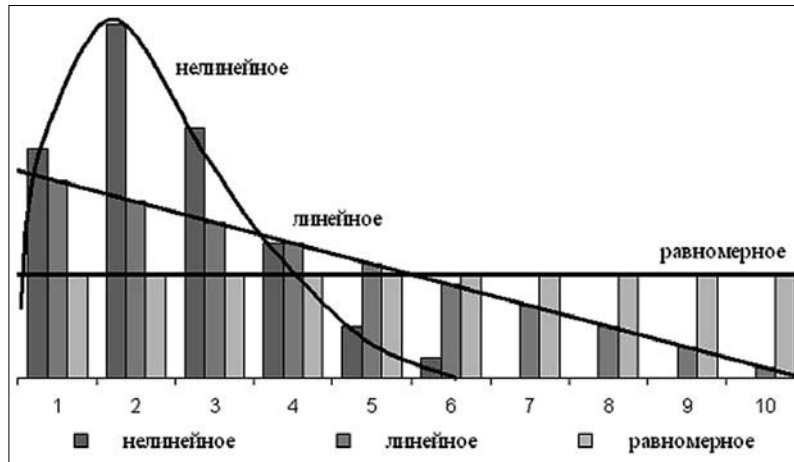


Рис. 1. Распределения тестовых заданий

3-й этап. Результаты таких расчётов представлены в табл. 3, 4. Они проиллюстрированы графически на рис. 2, 3. В первой строке таблиц указаны номера групп испытуемых, во второй — их истинные (входные) значения уровня подготовленности. На всех рисунках по оси абсцисс отложены входные (истинные) значения уровней подготовленности испытуемых, а по оси ординат — их вычисленные оценки. Прямая линия отражает «идеальное» расположение оценок. Чем ближе оценки к этой линии, тем адекватней процедура их получения.

Исходные значения уровня подготовленности по схеме Раша рассчитывались по наблюдаемым долям правильных ответов каждым испытуемым

$$\theta_i = \ln \frac{p_i}{1 - p_i} \text{ и последующем в ну-}$$

тригрупповом усреднении результатов. На рисунках эти значения обозначены серыми маркерами ■.

Корректировка исходных значений проводилась их линейным преобразованием к статистическим характеристикам истинных значений (среднему и дисперсии). На рисунках эти значения обозначены чёрными маркерами ■. Геометрически такая корректировка равносильна параллельному переносу исходных значений к нулевому среднему и одновременному повороту всех исходных значений в направлении линии расположения истинных показателей.

В альтернативной схеме исходные значения вероятностного уровня подготовленности рассчитывались по наблюдаемым долям правильных ответов каждым испытуемым по форму-

Таблица 3

Теория

Оценки уровня подготовленности испытуемых по схеме Раша

Номер группы		1	2	3	4	5	6	7	8	9	10
θ ист., логит		-4,5	-3,5	-2,5	-1,5	-0,5	0,5	1,5	2,5	3,5	4,5
равномерн.	θ исх.	-2,34	-1,61	-1,11	-0,58	-0,19	0,26	0,75	1,10	1,55	2,19
	θ кор.	-4,92	-3,38	-2,34	-1,22	-0,39	0,54	1,57	2,30	3,24	4,60
линейное	θ исх.	-1,75	-0,92	-0,27	0,21	0,67	1,23	1,64	2,13	3,01	3,83
	θ кор.	-4,73	-3,30	-2,17	-1,33	-0,54	0,43	1,16	2,00	3,52	4,96
нелинейн.	θ исх.	-1,06	-0,27	0,49	1,25	1,85	2,83	3,74	4,17	4,42	4,60
	θ кор.	-4,81	-3,64	-2,53	-1,40	-0,51	0,93	2,27	2,90	3,27	3,52

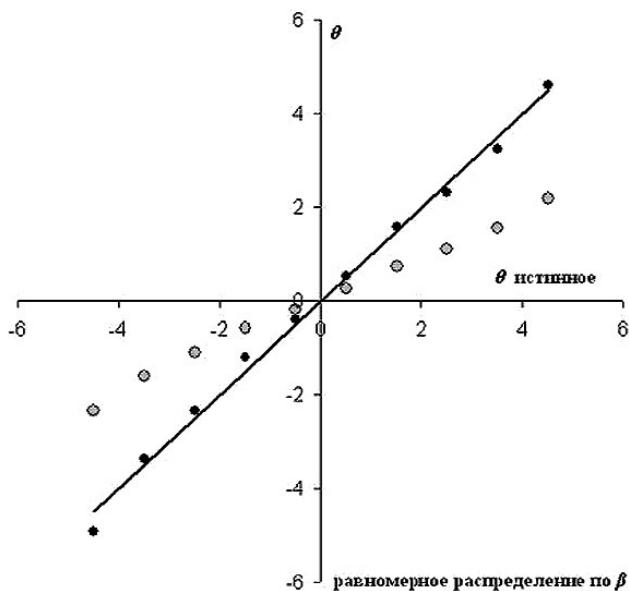


Рис. 2а. Шкалирование по схеме Раша: равномерное распределение тестовых заданий

ле из следующих преобразова-

ний: т.к. $p_i = \frac{\Theta_i}{1 + \Theta_i}$ отсюда

$u_i = 0,5^{\frac{1}{\Theta_i}} = 0,5^{\frac{(1-p_i)}{p_i}}$. Усреднённые внутри группы значения и

приведены в таблице. На рис. 3 эти значения обозначены серыми маркерами ■. В данном методе такие оценки справедливы лишь при равномерно распределённых тестовых заданиях (рис. 3а). В других случаях эти

ПЕД
измерения

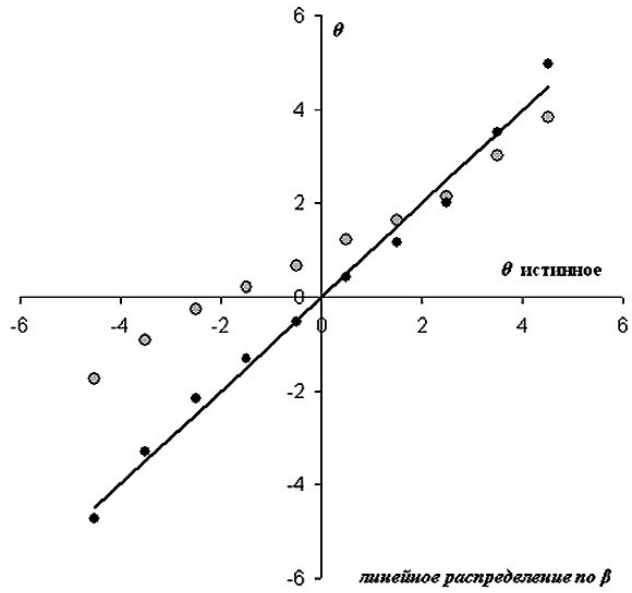


Рис. 2б. Шкалирование по схеме Раша: линейное распределение тестовых заданий

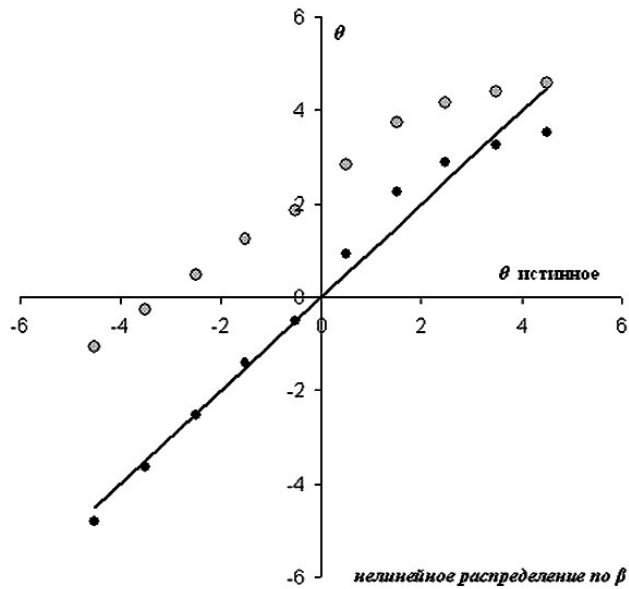


Рис. 2в. Шкалирование по схеме Раша: нелинейное распределение тестовых заданий

оценки неадекватны, что и отражено на рис. 3б,в.

Для линейного распределения тестовых заданий $\rho(v) = 2 \cdot v$ расчёты дают:

$$p_i = \int_0^1 \rho(v) \cdot v^{\Theta_i} dv = \frac{2 \cdot \Theta_i}{1 + 2 \cdot \Theta_i}.$$

Отсюда выражаем Θ и находим

$u_i = 0,5^{\frac{1-p_i}{p_i}}$. Именно по этим формулам вычислялись скорректированные оценки вероятностных показателей u которые на рис. 3б обозначены чёрными маркерами ■. Аналогичные расчёты и иллюстрации проведены и при нелинейном распределении тестовых заданий.

Даже визуальный анализ графиков, отражающих истинные и оценочные значения подготовленности испытуемых, позволяет сделать следующие выводы:

- распределение тестовых заданий по уровню трудности учитывается как в методе получе-

ния искомым оценок в схеме Раша, так и альтернативном подходе;

- в схеме Раша линейные преобразования исходных показателей (шкалирование) достаточно эффективно при линейном распределении тестовых заданий по уровню трудности на всём множестве возможных значений.

Если задания распределены нелинейно, т.е. в наборе тестовых заданий есть количественно выраженные группы заданий близкого уровня трудности с «провалами» между ними, то расчётные процедуры схемы Раша перестают эффективно работать на всём множестве возможных значений. В частности, из данных на рис. 2в следует, что отсутствие в наборе тестовых заданий высокого уровня трудности существенно искажает оценки испытуемых высокого уровня подготовленности. Этот вывод хорошо известен в

Таблица 4

Оценки вероятностного уровня подготовленности испытуемых по альтернативной схеме

Номер группы		1	2	3	4	5	6	7	8	9	10
<i>u</i> истинное		0,05	0,15	0,25	0,35	0,45	0,55	0,65	0,75	0,85	0,95
Равномерн.	<i>u</i> исх.	0,04	0,12	0,26	0,36	0,44	0,53	0,63	0,73	0,85	0,96
	<i>u</i> кор.										
Линейное	<i>u</i> исх.	0,22	0,39	0,49	0,58	0,66	0,75	0,81	0,87	0,92	0,97
	<i>u</i> кор.	0,05	0,16	0,25	0,34	0,44	0,56	0,67	0,75	0,85	0,94
Нелинейн.	<i>u</i> исх.	0,42	0,60	0,68	0,75	0,80	0,86	0,90	0,92	0,96	0,99
	<i>u</i> кор.	0,06	0,16	0,25	0,33	0,43	0,55	0,67	0,73	0,86	0,95

ПЕД
измерения

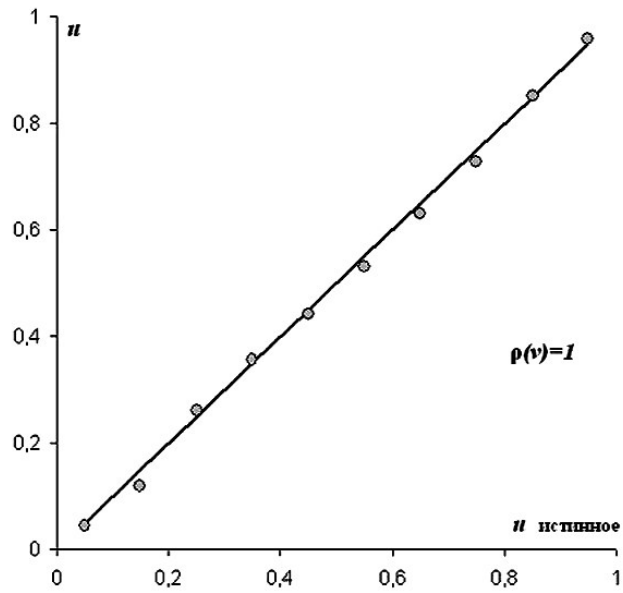


Рис. 3а. Расчёт по альтернативной схеме: равномерное распределение тестовых заданий

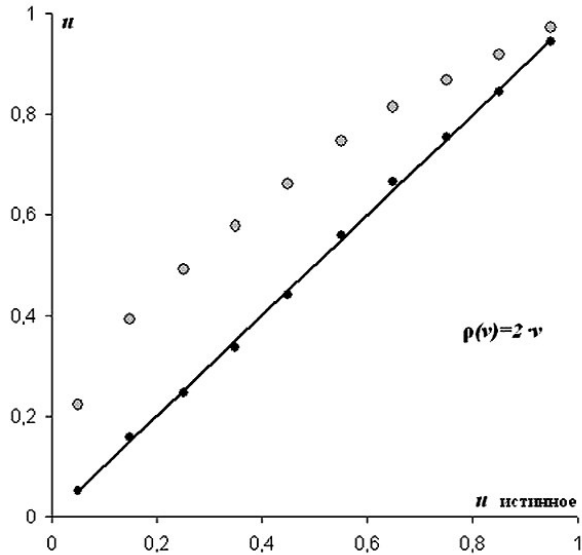


Рис. 3б. Расчёт по альтернативной схеме: линейное распределение тестовых заданий

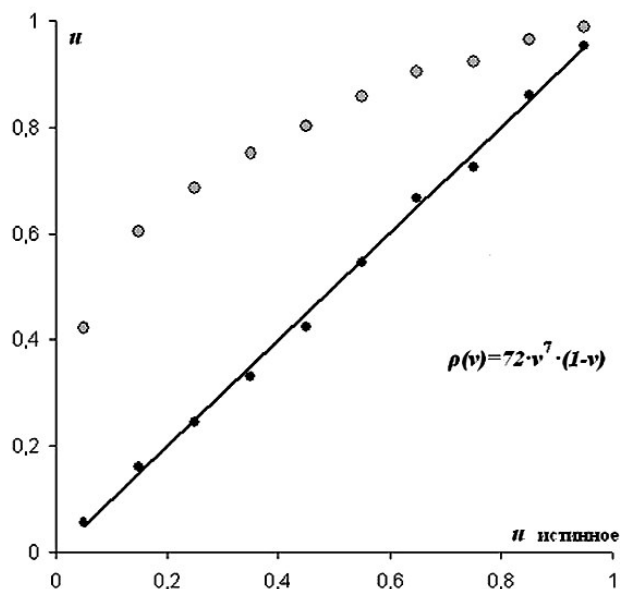


Рис. 3в. Расчёт по альтернативной схеме: нелинейное распределение тестовых заданий

IRT, как требование соответствия тестовых заданий группе испытуемых;

- аналитические методы оценивания показателей уровня подготовленности испытуемых в альтернативном подходе достаточно надёжны на всём множестве значений. Аналитический вид распределения тестовых заданий по уровню трудности позволяет получить адекватные этому распределению формулы для расчёта оцениваемых показателей.

В заключение следует выделить одно важное замечание. Может показаться, что в альтернативном подходе распределение тестовых заданий влияет на

оценки уровня подготовленности испытуемых. Это не так. Правильней будет сказать, что учёт распределения тестовых заданий влияет на надёжность формул для расчёта этих оценок. Можно провести аналогию — рост человека не зависит от измерительной линейки, но измеренное значение роста зависит и от качества линейки и от умения правильно её применить.

Функция распределения

В IRT учёт распределений тестовых заданий и испытуемых

ПЕД
измерения

5

В математическом разделе специальных функций достаточно подробно изучена определённая в виде интеграла бета-функция (интеграл Эйлера первого рода):

$$B(x, y) = \int_0^1 u^{x-1} (1-u)^{y-1} du.$$

Свойства и методы вычисления этой функции достаточно хорошо изучены. Укажем лишь те из них, которые полезны в анализе тестовых данных альтернативными методами:

1. Бета-функция сходится при положительных x и y .
2. Бета-функция не изменяется при перестановке параметров $B(x, y) = B(y, x)$.
3. При натуральных значениях $y = n$ интеграл непосредственно вычисляется и равен:

$$B(x, n) = \frac{(n-1)!}{x(x+1)(x+2)\dots(x+n-1)}.$$

В частности:

$$B(x, 1) = \frac{1}{x};$$

$$B(x, 2) = \frac{2}{x(x+1)}.$$

носит качественный характер и проявляется в терминах соответствия эмпирических данных теории или соответствия тестовых заданий испытуемым. Разработаны статистические процедуры выявления этого соответствия, и как следствие, возможности применения самой технологии педагогического измерения. В альтернативном подходе учёт этих распределений уже носит количественный характер, а сама функция распределения имеет строгий математический вид и непосредственно участвует в обработке тестовых результатов.

Для задания функции распределения мы исходили из следующих положений: математический вид этой функции должен позволять проводить необходимые аналитические расчёты и преобразования; математический вид этой функции должен адекватно отображать эмпирические данные; параметры этой функции должны нести предметную интерпретацию. Учитывая эти обстоятельства, мы пришли к следующему функциональному виду (для группы испытуемых):

$$\rho(u) = k \cdot u^{m-1} (1-u)^{n-1}, \text{ где } k \text{ — нормирующий множитель, } m, n \text{ — параметры распределения. Множитель } k \text{ определяется из условия нормировки}$$

$$\int_0^1 \rho(u) du = k \cdot \int_0^1 u^{m-1} (1-u)^{n-1} du.$$

Интеграл вида

$$\int_0^1 u^{m-1} (1-u)^{n-1} du \text{ хорошо известен в математике, и в некоторых случаях непосредственно вычисляется}^5.$$

В частности, для натуральных m, n нормирующий множитель равен

$$k = \frac{(m+n-1)!}{(m-1)! \cdot (n-1)!}.$$

Тогда функция распределения принимает вид:

$$\rho(u) = \frac{(m+n-1)!}{(m-1)! \cdot (n-1)!} \cdot u^{m-1} (1-u)^{n-1}.$$

Явный вид функции распределения позволяет получить и некоторые основные числовые характеристики: математическое

ожидание $M(u) = \frac{m}{m+n};$ дис-

персия $D(u) = \frac{mn}{(m+n)^2 (m+n+1)};$

точка экстремума

$$u_{\text{экстр.}} = \frac{m-1}{m+n-2}.$$

Для иллюстрации приведём и проанализируем графики функции распределения $\rho(u)$ для некоторых наборов параметров $(m;n)$ (рис. 4). Даже беглый взгляд показывает, что варьированием параметров m, n можно получить достаточно разнообразный набор аппроксимирую-

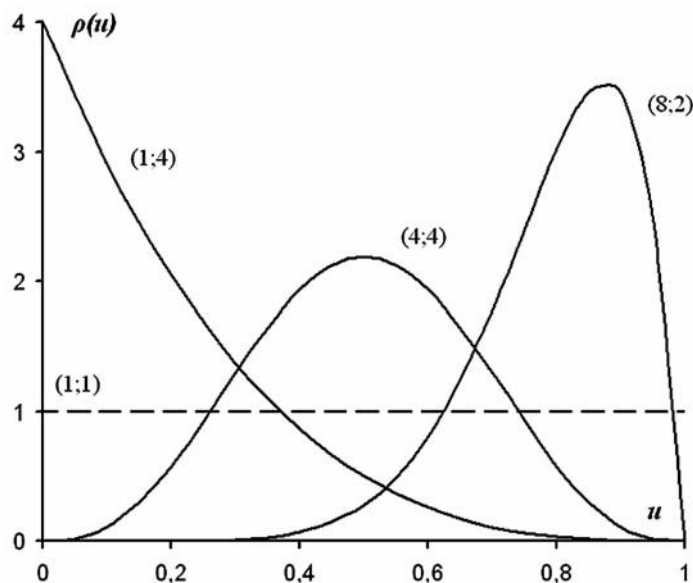


Рис. 4. Примеры функций распределения $\rho(u)$

щих функций для описания распределения эмпирических данных.

Интерпретацию введённой функции распределения можно провести, опираясь на описанное выше понятие шанса как отношения двух конфликтующих тенденций — правильно выполнить опорное (т.е. характерное для всего набора тестовых заданий) тестовое задание с вероятностью u , и неправильно выполнить это задание с вероятностью $(1 - u)$. Если параметры m , n интерпретировать как интенсивности проявления этих тенденций, то отношение m/n даёт шанс правильно ответить на опорное задание, отсюда среднее значение этой вероятности

для всей группы испытуемых

$$\text{совпадает } u_{cp.} = \frac{m}{m+n} \text{ с мате-}$$

матическим ожиданием $M(u)$. Прокомментируем приведённые примеры с этих позиций.

1. Распределение (1;1). Равномерное распределение испытуемых при $m = 1, n = 1$ принимает самый простой вид $\rho(u) = 1$. В этом случае интенсивности тенденций правильно — неправильно выполнить задание единичные или нейтральные, шанс правильно выполнить задание равен один к одному, среднее значение уровня подготовленности к тесту равно 0,5.

2. Распределение (1;4), т.е. интенсивность тенденции вы-

ПЕД	
	измерения

полнить правильно задание равна 1, а тенденция выполнить задание неправильно равна 4. Функция распределения имеет вид $\rho(u) = 4(1-u)^3$, шанс правильно выполнить тест группой испытуемых равен один к четырём, среднее значение уровня подготовленности к тесту равно 0,2. Эта группа слабо подготовлена к тестовым заданиям, в ней могут быть и испытуемые, которые ни на одно задание теста не отвечает правильно, т.е. с нулевым уровнем подготовленности $u = 0$.

3. Распределение (4;4) напоминает нормальный закон. Тенденции правильно и неправильно выполнить тест заметно выражены, и их интенсивности равны по 4. Функция распределения имеет вид $\rho(u) = 20u^3(1-u)^3$, шанс правильно выполнить тест группой испытуемых равен четыре к четырём, среднее значение уровня подготовленности к тесту равно 0,5.

4. Распределение (8;2) с явно выраженной тенденцией хорошо выполнить тест. Функция распределения имеет вид $\rho(u) = 72u^7(1-u)$, шанс правильно выполнить опорное задание группой испытуемых равен восемь к двум, среднее значение уровня подготовленности к тесту равно 0,8. Эта группа испытуемых очень хорошо подготовлена к тесту, но на все его задания никто правильно не отвечает. Нет испытуемых

с уровнем подготовленности $u = 1$.

Если в этой логике интерпретировать распределения тестовых заданий, то в качестве конфликтующих тенденций можно рассматривать способности выполнить тест опорным испытуемым. Например, распределение вида (8;2) характеризует набор достаточно лёгких тестовых заданий с шансом правильно выполнить его опорным испытуемым, равным 8 к 2, и средним вероятностным уровнем трудности теста $v_{cp} = 0,8$.

Таким образом, введение функции распределения для участников тестирования даёт не только возможность привести задачу педагогических измерений к аналитическим процедурам. Приведённая выше интерпретация значений её параметров m и n даёт новые групповые характеристики участников педагогического тестирования.

Аналитическое решение задач IRT

Введение функции распределения и математический закон альтернативной модели педагогических измерений дают возможность записать формулы для расчёта доли правильных ответов группой испытуемых с распределением $(m_u; n_u)$ на тестовое задание уровня трудности

$$B_j - p_i = k \int_0^1 u^{m_u - 1 + B_i} (1 - u)^{n_u - 1} du,$$

и для расчёта доли правильных ответов испытуемым с уровнем подготовленности Θ_i на распределённый с параметрами (m_u, n_u) набор тестовых заданий –

$$p_j = k \int_0^1 v^{m_v - 1 + \frac{1}{\Theta_j}} (1 - v)^{n_v - 1} dv.$$

Если удаётся непосредственно вычислить приведённые определённые интегралы, то задача педагогических измерений принимает аналитический вид. Итерационный алгоритм решения такой задачи можно представить следующим образом:

1 шаг. По наблюдаемым долям правильных ответов i -м испытуемым p_i и долям неправильных ответов на j -е задание теста q_j определяем исходные значения относительных показателей u .

2 шаг. Рассчитываем исходные значения вероятностных

показателей $\Theta_i = \frac{p_i}{1 - p_i}$ и $B_j = \frac{q_j}{1 - q_j}$.

3 шаг. По эмпирическим распределениям показателей u_i и v_j определяем соответствующие выборочные средние значения и дисперсии \bar{u} , S_u^2 , \bar{v} , S_v^2 .

4 шаг. Рассчитываем отдельные параметры m_u, n_u для функции $\rho(u)$ распределения испытуемых

$$m_u = \frac{\bar{u}}{S_u^2} (\bar{u}(1 - \bar{u}) - S_u^2),$$

$$n_u = \frac{1 - \bar{u}}{S_u^2} (\bar{u}(1 - \bar{u}) - S_u^2)$$

и m_v, n_v для функции $\rho(v)$ распределения тестовых заданий

$$m_v = \frac{\bar{v}}{S_v^2} (\bar{v}(1 - \bar{v}) - S_v^2),$$

$$n_v = \frac{1 - \bar{v}}{S_v^2} (\bar{v}(1 - \bar{v}) - S_v^2).$$

5 шаг. Уточняем параметры m, n . Для этого можно использовать графические иллюстрации и/или аналитические методы оптимизации, но удобными для дальнейших аналитических расчётов являются такие комбинации параметров, при которых параметр n является натуральным числом (1, 2, ...).

6 шаг. Записываем в явном виде уточнённые выражения для средней вероятности:

$$P(\Theta) = \frac{m_v (m_v + 1) \dots}{\left(\frac{1}{\Theta} + m_v\right) \left(\frac{1}{\Theta} + m_v + 1\right) \dots} \frac{\dots (m_v + n_v - 1)}{\dots \left(\frac{1}{\Theta} + m_v + n_v - 1\right)};$$

$$P(B) = \frac{m_u (m_u + 1) \dots}{(B + m_u) (B + m_u + 1) \dots} \frac{\dots (m_u + n_u - 1)}{\dots (B + m_u + n_u - 1)}.$$

Следует отметить, что в эти выражения для средней вероятности правильного ответа испы-

ПЕД
измерения

туемых входят параметры распределения тестовых заданий, а в выражения для тестовых заданий — параметры распределения испытуемых.

7 шаг. Из последних уравнений выражаем и вычисляем по наблюдаемым долям уточнённые относительные показатели уровня подготовленности испытуемых Θ_i и уровня трудности тестовых заданий B_j .

Если на шаге 5 удаётся подобрать функции распределения с параметром $n = 1$ или $n = 2$, то задача вычисления уточнённых оценок на шаге 6 сводится к элементарным расчётам. Для других натуральных значений n процесс вычисления искомых оценок сводится к численному решению достаточно простого алгебраического уравнения с единственным корнем.

Возможность аналитической обработки тестовых результатов позволяет получить и выражения для оценки погрешности наблюдаемых долей правильного ответа. Для дихотомического теста случайная величина «балл за правильный ответ на задание» распределена по закону Бернулли с дисперсией $p \cdot (1 - p)$. Используя этот результат, несложно получить расчётные формулы для дисперсии средней вероятности правильного ответа испытуемым на M заданий теста

$$D(\Theta) = \left[P(\Theta) - P\left(\frac{\Theta}{2}\right) \right] / M \text{ и для}$$

дисперсии средней вероятности правильного ответа N испытуемых на тестовое задание $D(B) = [P(B) - P(2B)] / N$.

В частности, при ответе на задание уровня трудности B группой равномерно распределённых испытуемых $\rho(u) = 1$ имеем выражения для средней вероятности правильного ответа

$$P = \frac{1}{(1+B)} \text{ и для дисперсии}$$

$$D(B) = \frac{B}{(1+B)(1+2B)N} = \frac{P(1-P)}{(2-P)N}.$$

Максимальное значение дис-

персии равно $D_{\max} = \frac{(\sqrt{2}-1)^2}{N} \approx$

$\approx \frac{0,17}{N}$, которое относится к те-

стовому заданию с уровнем

трудности $B = \frac{1}{\sqrt{2}} \approx 0,71$, со сред-

ней вероятностью правильного ответа на это задание $P = 2 - \sqrt{2} \approx 0,59$. Эти расчёты можно использовать для оценки верхней границы ошибки наблюдаемой доли правильного

ответа на задание $\Delta P \leq \frac{0,41}{\sqrt{N}}$.

Аналогичные оценки справедливы и для доли правильных ответов испытуемым на тестовые задания.

Оценки погрешности позволяют сформулировать требо-

вания по количеству участников теста. Например, если предполагается выделить среди испытуемых 5 групп, различимых по уровню подготовленности на основании доли правильных ответов, то погрешность таких измерений должна удовлетворять условию $\Delta P \leq 0,1$. Отсюда рекомендуемое число равномерно распределённых по уровню трудности тестовых заданий не менее 16.

Аналитический анализ результатов педагогического теста

Отличительной особенностью альтернативного подхода к решению основной задачи МТИ является возможность её аналитического решения. Для проверки этих идей на практике нами разработаны и проведены серии тестовых испытаний с условным названием «Устный счёт», а затем по представленной выше методике обработаны результаты тестирования.

Общая характеристика теста. Вариативный тест содержит 10 заданий⁶ возрастающей трудности. Задания проверяют умения быстро и правильно выполнять арифметические действия с целыми числами в уме. Действия в каждом из четырёх вариантов задаются детерминированно, вариант и числа, с ко-

торыми производятся действия, задаются случайным образом. Для наглядности можно привести возможный вариант последовательности тестовых заданий: $6 + 8 =$; $7 - 4 \cdot 5 =$; $28 : 4 + 9 - 6 =$; и т.д. Последнее, десятое, задание проверяет умение выполнить в уме десять арифметических действий. Задания предлагаются на персональном компьютере, в автоматизированном режиме фиксируется вариант, результат и время выполнения каждого из них.

Относительная простота содержания экспериментального теста имеет ряд преимуществ. Во-первых, при этом не возникает проблема валидности измерений. Во-вторых, достаточно очевидна и контролируема независимая от применяемого метода мера трудности тестовых заданий, например, по числу арифметических действий в нём, что соответствует порядковому номеру заданий. И, наконец, с таким тестом можно достаточно быстро набрать большой массив тестовых результатов для последующей их обработки.

Испытуемые. Тестовые испытания проводились среди студентов колледжа в форме студенческой олимпиады. Сорок участников провели более чем 700 сеансов тестовых испытаний, после отбрасывания незавершённых сеансов испытаний осталось 616 строк по 10 элементов $1/0$ в каждой. Неко-

6
Педагогический тест обычно содержит 30 ± 10 заданий. Здесь используется схема экспериментального тестирования с меньшим числом заданий.

торые участники проводили десятки сеансов тестирования.

Приведём результаты анализа для трёх групп данных. В «смешанную» группу I мы включили результаты испытуемых, каждый из которых тестировался менее 12 раз. Таких испытуемых оказалось 29, общее число проведённых ими сеансов тестовых испытаний 194.

В группу II мы включили результаты тестирования «сильного» испытуемого А.Г., который не только показал самый высокий средний балл, но и провёл 105 сеансов тестирования.

В группу III включили данные относительно «слабого» испытуемого У.А., который провёл 24 сеанса тестирования.

Анализ результатов тестирования проводился по алгоритму, приведённому выше. Цель анализа — отдельно для каждой из выделенных трёх групп оценить показатели уровня трудности тестовых заданий.

По наблюдаемым долям правильных ответов i -м испытуемым p_i и долям неправильных ответов на j -е задание теста q_j определяем исходные значения

$$B_{\text{исх.}j} = \frac{q_j}{1 - q_j}$$

и вероятностных показателей $v_j = 0,5^{B_j}$. Результаты расчётов представлены в табл. 5. В первой строке каждой группы —

доли неправильных ответов; во второй строке — относительные показатели; в третьей строке — вероятностные показатели (рис. 5).

Для испытуемых, по наблюдаемым долям правильных ответов последовательно рассчитываем исходные относительные и вероятностные показатели уровня подготовленности u , по полученным рядам u для каждой группы рассчитываем выборочные средние значения и дисперсии. И, наконец, рассчитываем и уточняем (упрощаем для аналитических расчётов) параметры m_u, n_u функций распределения исходных показателей по группам. Результаты этих процедур представлены в табл. 6.

Следует отметить, что полученные распределения исходных значений показателя подготовленности испытуемых для всех групп далеки от линейных. В этом случае стандартные процедуры шкалирования исходных тестовых баллов могут дать заметные искажения. В аналитическом методе распределение исходных баллов испытуемых непосредственно учитывается в расчётных формулах.

С учётом исходных распределений для испытуемых корректируем формулы для расчёта относительных показателей трудности тестовых заданий по исходным показателям:

$$P = \frac{3}{B + 3} \Rightarrow B_{\text{кор.}j} = 3 \cdot B_{\text{исх.}j};$$

Таблица 5

Теория

Исходные значения показателя трудности тестовых заданий

Номер задания		1	2	3	4	5	6	7	8	9	10
I группа	<i>q</i>	0,06	0,11	0,18	0,19	0,23	0,31	0,35	0,36	0,40	0,46
	<i>B</i> исх.	0,07	0,13	0,22	0,24	0,30	0,45	0,54	0,56	0,66	0,85
	<i>v</i> исх.	0,96	0,92	0,86	0,85	0,81	0,73	0,69	0,68	0,63	0,56
II группа	<i>q</i>	0,08	0,10	0,10	0,10	0,17	0,21	0,28	0,29	0,29	0,35
	<i>B</i> исх.	0,08	0,11	0,11	0,12	0,21	0,27	0,38	0,40	0,40	0,544
	<i>v</i> исх.	0,94	0,93	0,93	0,92	0,87	0,83	0,77	0,76	0,76	0,69
III группа	<i>q</i>	0,13	0,25	0,33	0,46	0,58	0,63	0,75	0,83	0,88	0,92
	<i>B</i> исх.	0,14	0,33	0,50	0,85	1,40	1,67	3,00	5,00	7,00	11,00
	<i>v</i> исх.	0,91	0,79	0,71	0,56	0,38	0,31	0,13	0,03	0,008	0,0005

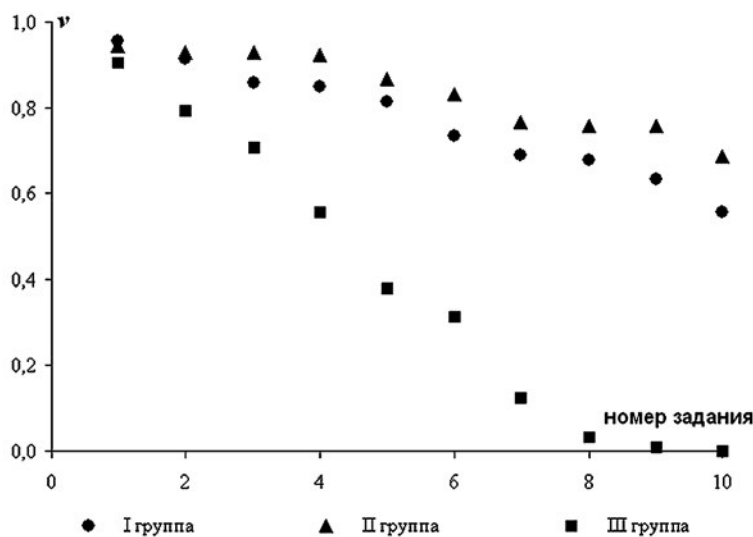


Рис. 5. Исходные значения показателя *v*

II группа —

$$P = \frac{4,9}{B + 4,9} \Rightarrow B_{кор. j} = 4,9 \cdot B_{исх. j};$$

$$= 1,7 \cdot \left(\sqrt{1 + \frac{10,56}{11,56} B_{исх. j}} - 1 \right).$$

III группа —

$$P = \frac{1,2 \cdot 2,2}{(B + 1,2) \cdot (B + 2,2)} \Rightarrow B_{кор. j} =$$

Получили таблицу корректированных оценок (табл. 7) и соответствующую геометрическую иллюстрацию (рис. 6).

ПЕД
измерения

Таблица 6

Распределения исходных значений вероятностного показателя подготовленности испытуемых

Характеристики распределения	Среднее значение	Дисперсия	Рассчитанные		Уточнённые	
			<i>m</i>	<i>n</i>	<i>m</i>	<i>n</i>
I группа	0,75	0,043	2,47	0,81	3,0	1
II группа	0,83	0,022	4,40	0,90	4,9	1
III группа	0,38	0,064	1,00	1,65	1,2	2

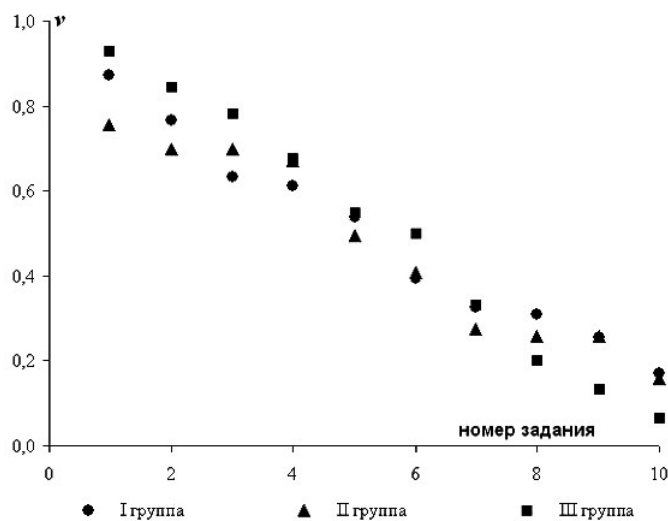


Рис. 6. Корректированные значения показателя *v*

Можно провести уточняющую корректировку искомым оценкам. Для этого достаточно пересчитать (корректировать) исходные показатели подготовленности испытуемых по корректированным значениям показателя трудности тестовых заданий, определить функции распределения этих показателей, и уже по более точному распределению провести более уточняющую корректировку

показателей трудности тестовых заданий. Однако наш опыт обработки результатов аналитическими методами показывает, что уже первая корректировка даёт оценки в пределах погрешности измерений и вторая корректировка может оказаться излишней.

В нашем эксперименте вторая корректировка привела к незначительным изменениям в показателях трудности тестово-

Таблица 7

**Корректированные значения показателя трудности
тестовых заданий**

Номер задания		1	2	3	4	5	6	7	8	9	10
I группа	<i>B</i> кор.	0,20	0,38	0,66	0,71	0,90	1,35	1,62	1,69	1,97	2,54
	<i>v</i> кор.	0,87	0,77	0,63	0,61	0,54	0,39	0,33	0,31	0,25	0,17
II группа	<i>B</i> кор.	0,40	0,52	0,52	0,57	1,01	1,30	1,87	1,96	1,96	2,67
	<i>v</i> кор.	0,756	0,699	0,699	0,672	0,495	0,406	0,274	0,257	0,257	0,158
	<i>v</i> кор.,	0,751	0,694	0,694	0,667	0,488	0,399	0,266	0,250	0,250	0,152
III группа	<i>B</i> кор.	0,11	0,24	0,35	0,56	0,87	1,00	1,59	2,31	2,92	3,95
	<i>v</i> кор.	0,93	0,85	0,78	0,68	0,55	0,50	0,33	0,20	0,13	0,06

го задания только для группы II. Её влияние отражено в дополнительной строке табл. 7. Для других групп эти изменения ещё менее значимы.

Любые измерения желательно дополнить оценкой погрешности результатов. Для групп I и II (для этих групп $n = 1$) дисперсия средней вероятности неправильного ответа Q на тестовое задание определя-

ется по формуле
$$D = \frac{Q(1-Q)}{(1+Q)N}$$
.

Для группы III расчётная формула для дисперсии не приводится. Для $n > 1$ она несколько усложняется, но её можно получить из общей формулы для дисперсии, приведённой выше.

Если эти формулы применить для вычисления среднеквадратичного отклонения наблюдаемой доли неправильного ответа и пересчитать эти отклонения для измеряемых значений показателя трудности тестовых заданий, то получим следующие интервалы (табл. 8).

Таблица 8

**Интервальные оценки значений показателя трудности
тестовых заданий**

Номер задания		1	2	3	4	5	6	7	8	9	10
I группа	<i>v</i>	0,84	0,72	0,58	0,56	0,49	0,34	0,28	0,26	0,21	0,14
		0,91	0,81	0,68	0,66	0,59	0,44	0,37	0,36	0,30	0,21
II группа	<i>v</i>	0,68	0,62	0,62	0,59	0,41	0,32	0,20	0,19	0,19	0,11
		0,83	0,78	0,78	0,75	0,58	0,48	0,34	0,32	0,32	0,21
III группа	<i>v</i>	0,89	0,78	0,71	0,59	0,44	0,39	0,21	0,09	0,04	0,01
		0,97	0,90	0,85	0,76	0,64	0,60	0,44	0,30	0,23	0,14

ПЕД	
	измерения

Графически эти результаты отличаются от рис. 6 изображением представленных на рис. 6а, который планок погрешности измерений.

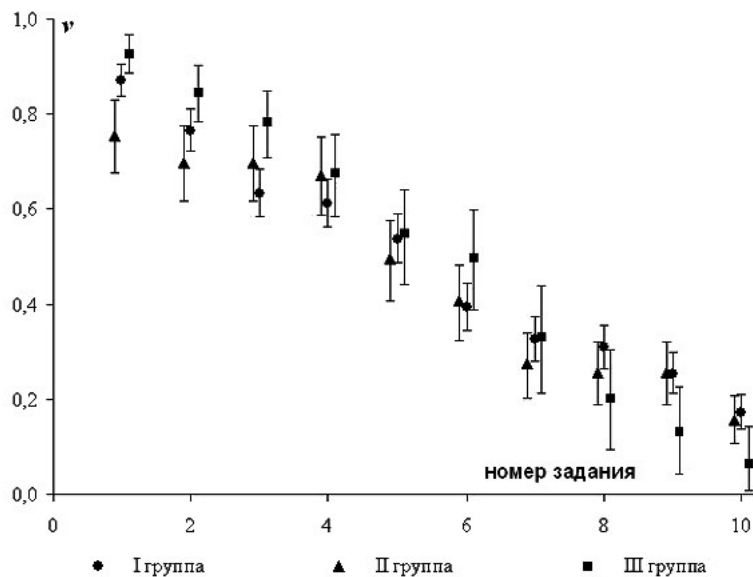


Рис. 6а. Корректированные значения показателя v с погрешностью измерений

Достаточно хорошее совпадение результатов измерения по трём независимым исследуемым группам позволяет говорить о надёжности и объективности самой процедуры их измерения.

Заключение. Аналитический метод решения основной задачи IRT можно рассматривать как альтернативный подход к обработке результатов педагогического тестирования,

который даёт возможность:

- получать независимые от классического метода оценки, что усиливает надёжность педагогических измерений;
- давать дополнительные интерпретации измеряемым показателям;
- получать дополнительные характеристики распределений участников теста.