

МОДЕЛЬ ПЕДАГОГИЧЕСКИХ ИЗМЕРЕНИЙ В ШКАЛЕ ОТНОШЕНИЙ

Юрий Каргин

Ноябрьский колледж профессиональных
и информационных технологий

kargin04@yandex.ru

Опубликовано в ж. «Педагогические Измерения№2, 2010

Работа посвящена решению задач теории педагогических измерений. В первой части, сопоставляя с моделью Раша, предлагается альтернативная модель, выделяются её свойства и достоинства в практике применения. Вторая часть носит одновременно гипотетический и прикладной характер. Здесь предпринимается попытка измерять основные показатели педагогического теста интервалами времени.

Ключевые слова: измерительная шкала, модель Раша, интерпретация показателей модели педагогических измерений

Одной из основных моделей педагогических измерений является математическая модель Раша (G. Rasch). Предлагаемую нами модель можно рассматривать как альтернативу модели Раша, сопоставлению этих моделей посвящена первая часть данной работы. Интерпретация значений основных показателей в альтернативной модели, позволяет предложить использовать в качестве меры их измерения интервалы времени. Эти вопросы обсуждаются во второй части работы.

Основная задача теории педагогического теста состоит в разработке такого нормированного по времени и содержанию набора тестовых заданий различного уровня трудности, который позволит максимально объективно дифференцировать, и оценить уровень подготовленности испытуемых. Отсюда и обязательное включение в математическую модель двух переменных величин, отражающих уровень трудности тестового задания и уровень подготовленности испытуемого к содержанию теста.

Исходным эмпирическим материалом для построения модели педагогического теста является таблица тестовых результатов X_{ij} (обычно, $X_{ij} = 1$ – задание выполнено правильно, $X_{ij} = 0$ – не правильно), m столбцов которой нумеруют тестовые задания, а n

строк – испытуемых. Тогда, подсчитанные доли правильных ответов $\tilde{o}_i = \frac{1}{m} \sum_{j=1}^m \tilde{O}_{ij}$ на

задания теста для отдельных испытуемых и доли правильных ответов $x_j = \frac{1}{n} \sum_{i=1}^n X_{ij}$

всех испытуемых на отдельное задание теста, являются исходными референтами как для определения параметров модели, так и для оценки её адекватности.

Математические модели

Математически наиболее простыми являются однопараметрические модели. Т.к. в педагогическом тесте присутствуют два основных множества его участников, набор тестовых заданий и группа испытуемых, то существует два простейших математических способа группировки числовых характеристик этих множеств в единый параметр: аддитивный (в виде суммы или разности) и мультипликативный (в виде произведения или частного). Если характеризовать анализируемые в работе модели с этих позиций, то модель Раша – аддитивная, а предлагаемая нами альтернативная модель – мультипликативная. Эти различия определяют и измерительные шкалы для численной оценки исследуемых качеств участников теста – в модели Раша уровни трудности тестовых заданий и подготовленности испытуемых измеряются в интервальной шкале, в нашей модели используется более совершенная шкала отношений.

Вероятность P_{ij} правильного ответа i – м испытуемым на j – ое задание теста задаётся параметром γ_{ij} , выраженным через две латентные переменные: θ_i – уровень подготовленности испытуемого и β_j – уровень трудности тестового задания. Вероятности правильных ответов отдельных испытуемых на все задания теста, и всех испытуемых на

отдельные задания теста рассчитываются по формулам: $p_i = \frac{1}{m} \sum_{j=1}^m P_{ij}$, $p_j = \frac{1}{n} \sum_{i=1}^n P_{ij}$.

Тогда, основным критерием адекватности модели может служить мера различий между теоретическими вероятностями и эмпирическими долями правильных ответов,

выраженная, например, в виде: $R = R_n + R_m = \sum_{i=1}^n |p_i - x_i| + \sum_{j=1}^m |p_j - x_j| \rightarrow \min$.

Представим сами модели. Для записи модели и далее, где нет необходимости, индексы, нумерующие тестовое задание и испытуемого, будем опускать.

Модель Раша (RM):

$$D = \frac{\partial \tilde{\delta} \partial \gamma}{1 + \partial \tilde{\delta} \partial \gamma}, \quad \gamma = \theta - \beta. \quad (1)$$

Переменные β и θ , после специальной процедуры шкалирования, измеряются в логитах, значения которых практически (более 99 % случаев) находятся в интервале от – 5 до + 5 с нейтральным нулём для средних уровней трудности задания и подготовленности испытуемого. Для понимания смысла значений этих единиц измерения приводятся вероятностные толкования¹: «Например, испытуемый с подготовкой $\theta = 5$ имеет вероятность правильного ответа на задание средней трудности 0,99».

В теории Раша оценку теоретических значений β и θ можно получить из таблицы

$$\text{тестовых результатов по формулам}^1: \quad \theta_i = \ln \frac{\tilde{o}_i}{1 - \tilde{o}_i}, \quad \beta_j = \ln \frac{1 - \tilde{o}_j}{\tilde{o}_j}, \quad \text{где } x_i -$$

эмпирические доли правильного ответа на тестовые задания i -м испытуемым, x_j – эмпирические доли правильного ответа группой испытуемых на j -ое тестовое задание.

Предлагаемую нами модель обозначим аббревиатурой ЕМ, отражая этим её «экспоненциальную» (показательную) основу¹:

$$D = 0,5^\gamma, \quad \gamma = \beta / \theta. \quad (2)$$

Если для исходных переменных β и θ ввести вероятностные переменные $v = 0,5^\beta$, $u = 0,5^{1/\theta}$, то закон (2) можно переписать в виде:

$$\ln P \cdot \ln 0,5 = \ln v \cdot \ln u \quad (2a)$$

Здесь, переменные β и θ изменяются в полуинтервале от нуля до бесконечности с нейтральной единицей для опорных уровней, а вероятностные переменные u и v – в интервале от нуля до единицы со значением 0,5 для опорных уровней. Вероятностные переменные u и v можно интерпретировать следующим образом: u – вероятность правильного ответа испытуемым на опорное задание теста; v – вероятность правильного ответа на задание теста опорным испытуемым. Введение новых переменных не только позволяет предложить разные формулировки самой модели, но и даёт дополнительные преимущества при её анализе и интерпретации свойств.

В частности, для ЕМ возможны достаточно внятные сравнения. Например, первый испытуемый подготовлен к тесту в два раза лучше второго, если он с той же совместной вероятностью, выполняет два опорных задания теста что и второй, выполняяший только одно опорное задание. Аналогичные рассуждения можно привести и для сравнения уровней трудности тестовых заданий.

Оценку теоретических значений β и θ по таблице тестовых результатов можно проводить различными способами, где эмпирические доли x_i и x_j определены выше:

а) по опорным показателям: $\theta_i = 1 / \log_{0,5} \tilde{o}_i$, $\beta_j = \log_{0,5} \tilde{o}_j$;

б) предполагая равномерное распределение участников теста относительно вероятностных переменных u и v :

¹ Каргин Ю.Н. Экспоненциальный спад в задачах дидактики// Научное творчество XXI века: Материалы II Всероссийской научной конференции, март 2010г. – Красноярск: Научно-информационный издательский центр, редакция журнала «В мире научных открытий», НЭБ, (www.nkras.ru)

$$\theta_i = \frac{\tilde{o}_i}{1-\tilde{o}_i}, \quad \beta_j = \frac{1-\tilde{o}_j}{\tilde{o}_j};$$

в) предполагая квадратичное распределение участников теста относительно вероятностных переменных u и v :

$$\theta_i = \frac{5\tilde{o}_i}{12(1-\tilde{o}_i)} \left(1 + \sqrt{1 + \frac{24(1-\tilde{o}_i)}{25\tilde{o}_i}} \right), \quad \beta_j = 2,5 \left(\sqrt{1 + \frac{24(1-\tilde{o}_i)}{25\tilde{o}_j}} - 1 \right).$$

Расчётные формулы, приведённые в пунктах б), в) следуют из цепочки предположений и математических преобразований, написанных, например, для j -го

тестового задания: $\tilde{o}_j \approx \delta_j = \frac{1}{n} \sum_{i=1}^n P_{ij} \approx \int_0^1 \rho(u) P du = \int_0^1 \rho(u) u^\beta du$. Вычисление

последнего интеграла и даёт расчётные формулы. Плотность распределения $\rho(u)$ испытуемых равна:

б) $\rho(u) = 1$, для равномерного распределение участников теста относительно вероятностной переменной u ;

в) $\rho(u) = 6 \cdot (u-u^2)$, для квадратичного распределения участников теста относительно вероятностной переменной u .

Аналогичные рассуждения и расчёты можно провести и для i -го испытуемого.

Приведённые распределения участников теста по уровню подготовленности u можно прокомментировать на примере. Пять участников теста равномерно распределены по уровню подготовленности к тесту, если: $u_1 = 0,1; u_2 = 0,3; u_3 = 0,5; u_4 = 0,7; u_5 = 0,9$. Пять участников теста по уровню подготовленности группируются ближе к опорному уровню, и распределены по квадратичному закону, если: $u_1 \approx 0,18; u_2 \approx 0,36; u_3 \approx 0,5; u_4 \approx 0,64; u_5 \approx 0,82$. Два рассмотренных простейших вида распределения проиллюстрированы на рисунке 1.

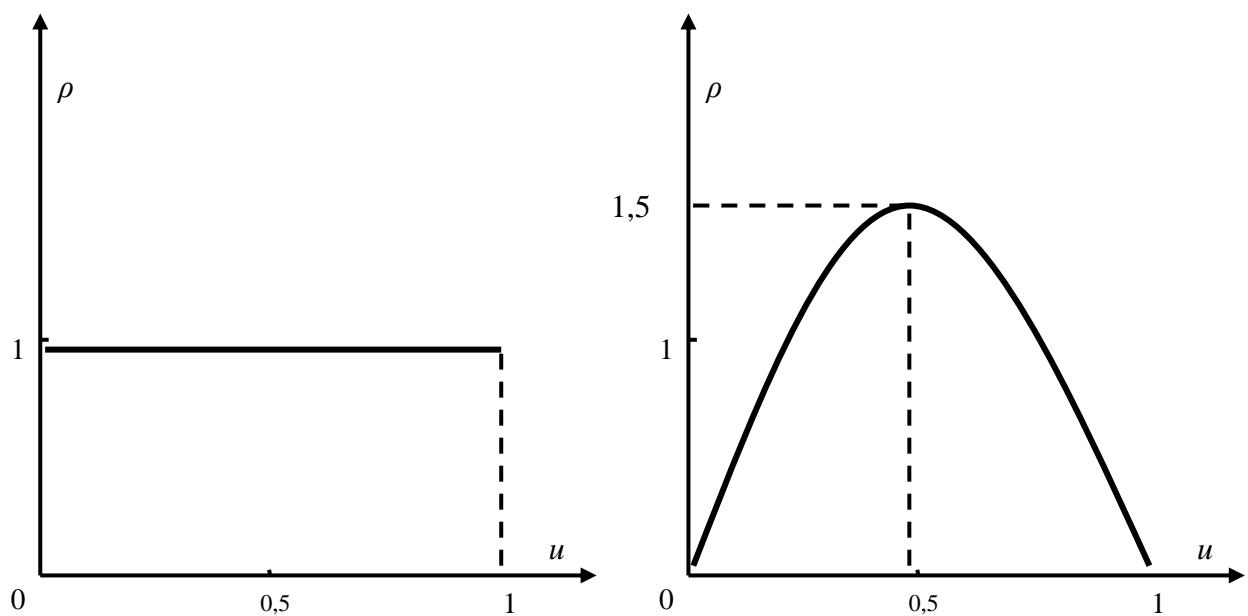


Рис. 1. Распределения участников теста

Сопоставим «работу» моделей на примере данных небольшой матрицы тестовых результатов, приведённых в работе В.С. Аванесова². Результаты расчётов выпишем в табл.1,2 отдельно для 13 испытуемых и для 10 тестовых заданий.

Таблица 1. Показатели 13 испытуемых

	i	1	2	3	4	5	6	7	8	9	10	11	12	13	Rn
	x_i	0,9	0,8	0,7	0,6	0,6	0,5	0,5	0,5	0,4	0,4	0,3	0,2	0,1	-
RM	θ_i	2,20	1,39	0,85	0,41	0,41	0	0	0	-0,4	-0,4	-0,8	-1,4	-2,2	-
	p_i	0,86	0,75	0,66	0,58	0,58	0,5	0,5	0,5	0,42	0,42	0,35	0,26	0,16	0,38
EM _A)	θ_i	6,58	3,11	1,94	1,36	1,36	1	1	1	0,76	0,76	0,58	0,48	0,3	-
	p_i	0,89	0,78	0,68	0,59	0,59	0,5	0,5	0,5	0,42	0,42	0,34	0,27	0,2	0,34
EM _B)	θ_i	9	4	2,33	1,5	1,5	1	1	1	0,67	0,67	0,43	0,25	0,11	-
	p_i	0,89	0,78	0,67	0,57	0,57	0,47	0,47	0,47	0,37	0,37	0,28	0,19	0,1	0,34
EM _B)	θ_i	7,69	3,52	2,13	1,43	1,43	1	1	1	0,71	0,71	0,5	0,33	0,19	-
	p_i	0,89	0,78	0,68	0,58	0,58	0,48	0,48	0,48	0,39	0,39	0,31	0,22	0,14	0,25

Таблица 2. Показатели 10 заданий теста

	j	1	2	3	4	5	6	7	8	9	10	Rm
	x_j	0,92	0,85	0,69	0,54	0,46	0,46	0,38	0,31	0,23	0,15	-
RM	β_j	-2,5	-1,7	-0,8	-0,2	0,15	0,15	0,47	0,81	1,2	1,7	-
	p_j	0,89	0,8	0,66	0,53	0,47	0,47	0,41	0,34	0,27	0,2	0,28

² Аванесов В. Проблема объективности педагогических измерений // Педагогические измерения. – 2008. - № 3. - С. 3 – 39.

EM _a)	β_j	0,12	0,24	0,53	0,89	1,12	1,12	1,38	1,7	2,12	2,7	-
	p_j	0,91	0,83	0,67	0,53	0,47	0,47	0,4	0,34	0,28	0,22	0,24
EM _b)	β_j	0,08	0,18	0,44	0,86	1,17	1,17	1,6	2,25	3,33	5,5	-
	p_j	0,91	0,82	0,67	0,52	0,44	0,44	0,36	0,28	0,19	0,11	0,26
EM _b)	β_j	0,1	0,21	0,49	0,88	1,14	1,14	1,48	1,94	2,62	3,76	-
	p_j	0,91	0,83	0,67	0,52	0,45	0,45	0,38	0,31	0,23	0,16	0,1

В приведённых таблицах следует обратить особое внимание на последний столбец. Критерий R, отражающий адекватность модели экспериментальным данным, для правильного применения EM примерно в два раза ниже, чем в RM, т.е. аналитические решения EM адекватней описывают рассматриваемое явление.

Свойства моделей.

Сопоставим два наиболее важных свойства моделей: графики и дифференцирующие характеристики.

Графики вероятности правильного ответа в зависимости от уровня подготовленности испытуемых для рассматриваемых моделей представлены на рисунке 2.

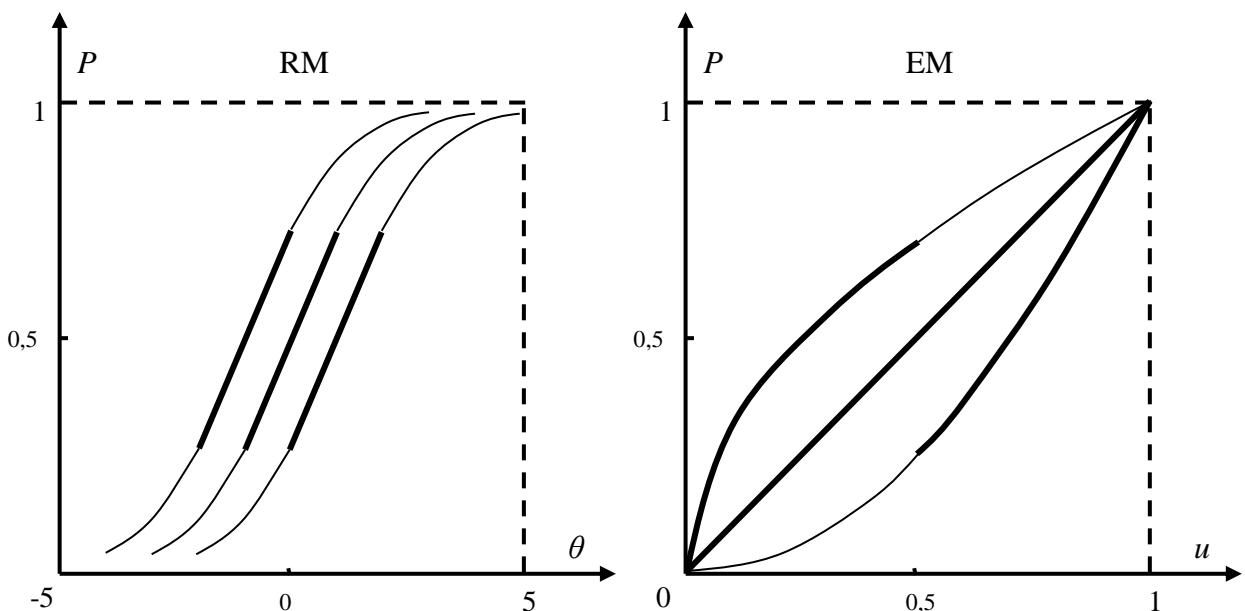


Рис. 2. Графики $P(\theta)$ и $P(u)$ для трёх заданий различного уровня трудности

Внешние различия графиков RM и EM связаны с различием осей ординат. В обоих случаях, средняя линия отражает вероятность правильного ответа испытуемым уровня подготовки θ или u на среднее (опорное) задание теста. Левее расположена линия

вероятности правильного ответа на задание, которое легче опорного, а правее – на более трудное задание. В терминологии RM лёгкое задание соответствует уровню трудности минус один логит, $\beta = -1$, а трудное – плюс один логит, $\beta = 1$. В терминологии EM лёгкое задание в два раза легче опорного, т.е. $\beta = 1/2$, а трудное задание в два раза труднее опорного, т.е. $\beta = 2$. На наш взгляд, второе толкование не просто более внятное, но и несёт в себе конкретный смысл (см. выше).

Иллюстрация моделей графиками ещё раз подчёркивает их аддитивный и мультипликативный характер. Изменение уровня трудности задания «параллельно переносит» график вероятности правильного ответа в RM, и «деформирует» его в EM.

Графическое сопоставление самих моделей проиллюстрировано на рис. 3 в виде зависимостей $p_i(x_i)$, построенных для трёх испытуемых (по данным ¹):

- а) $i = 1$ – самый высокий уровень подготовленности испытуемого, который и на самые трудные задания правильно отвечает с вероятностью более чем 0,5;
- б) $i = 7$ – средний уровень подготовленности;
- в) $i = 13$ – самый слабый испытуемый, который с вероятностью около 0,5 отвечает лишь на самые легкие задания теста.

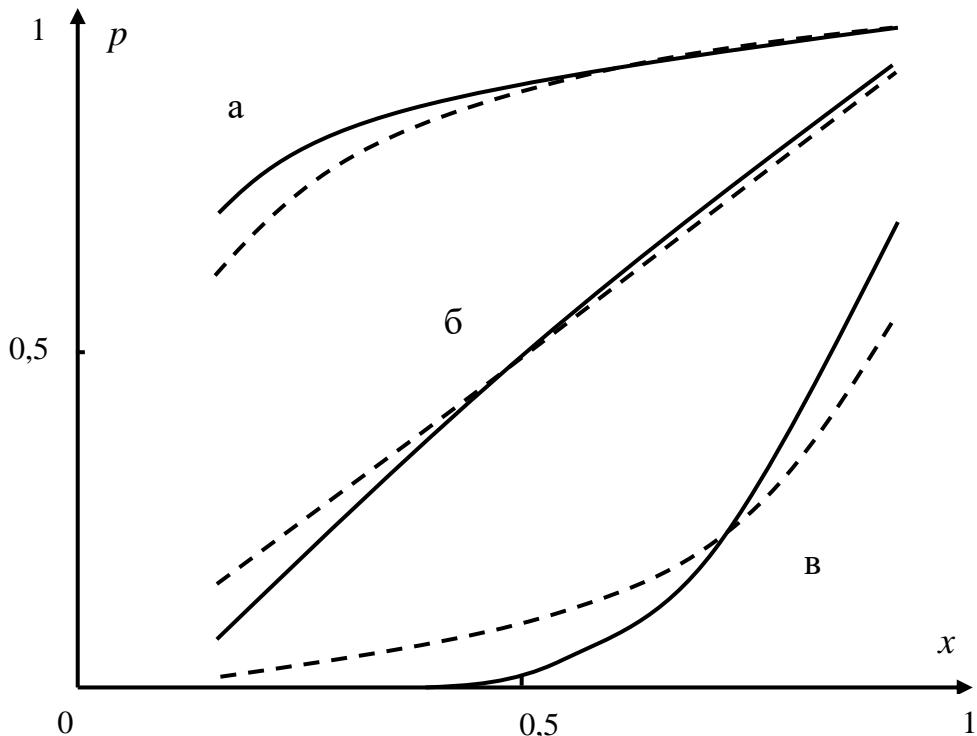


Рис. 3. Графическое сопоставление моделей (--- RM; — EM)

Существенные различия графиков в области параметров теста «низкий уровень подготовленности испытуемого – высокий уровень трудности тестового задания», не столько ставят под сомнения какую либо модель, сколько указывает на существование

области их эффективного применения. В теории педагогических измерений показано, что как трудные задания бессмысленно предлагать слабо подготовленным испытуемым, так и очень лёгкие задания не оценивают знания у хорошо подготовленных учащихся. Те области параметров теста, в которых тестовое задание соответствует уровню подготовленности испытуемого, называют областями высокого разрешения или высокой дифференцирующей способности. Существование и выявление таких областей является второй задачей теории педагогических измерений, наряду с задачей оценки значений основных параметров теста.

С геометрической точки зрения вероятность правильного ответа $P(\beta; \theta)$ можно рассматривать как поверхность, заданную на множестве значений переменных β и θ . Некоторые сечения этой поверхности представлены на рисунке 2. В общих чертах свойства этой поверхности достаточно понятны. Максимальные значения точек поверхности расположены в области «очень легкие задания теста – очень высокий уровень подготовленности испытуемых». В этой области вероятность правильного ответа P близка к единице и результат решения тестового задания не сильно зависит от параметров теста.

По мере усложнения заданий и снижения уровня подготовленности испытуемых вероятность правильного ответа на задание теста всё стремительней убывает, поверхность P становится круче, и уже в области «низкий уровень подготовленности испытуемого – высокий уровень трудности тестового задания» вероятность правильного ответа асимптотически стремиться к нулю. В самой нижней части, как и самой её верхней части, поверхность P практически горизонтальная и результат решения тестового задания опять слабо зависит от значений переменных β и θ . Таким образом, в поверхности P можно выделить как «почти горизонтальные» области низкой дифференциации участников теста (низкого разрешения), так и «крутые» области с высоким разрешением. Математической величиной, отражающей скорость изменения уровня поверхности P , является градиент этой функции.

Обозначим градиент функции $P(\beta; \theta)$ через $G(\beta; \theta)$ и в качестве опорного примем его значение в нейтральной точке, где разрешение теста оптимальное. Для модели Раша такой нейтральной точкой являются значения переменных $\beta = 0, \theta = 0$. В этом случае $G_0 = G(0;0) = \sqrt{2} / 4$. В современной теории теста близкой по смыслу к функции G является информационная функция I , значение которой в нейтральной точке равно $I(0) = 1 / 4$, а сама информационная функция отражает скорость изменения уровня поверхности P , но только в направлении изменения переменной трудности тестового задания β .

Для ЕМ нейтральной точкой являются значения $\beta = 1$, $\theta = 1$, или для вероятностных переменных $v = 0,5$, $u = 0,5$. В этой точке $G_0 = G(0,5;0,5) = \sqrt{2}$.

Область изменения переменных теста будем называть хорошо разрешённой, если выполняется условие: $G(\beta;\theta) > G_0 / \sqrt{2}$. Т.е., в хорошо разрешённой области дифференцирующая способность теста пусть и не оптимальная, но меньше её не более чем в $\sqrt{2}$ раз. Математически, в RM разрешённая область определяется приближённым неравенством $|\beta - \theta| \leq 1,21$, когда значения переменных отличаются на более чем на 1,21 логит. В ЕМ область хорошего разрешения определяется двумя приближёнными неравенствами, первое из которых исключает область «трудное задание – слабая подготовка», второе неравенство исключает область «легкое задание – сильная подготовка»: $v^{1,35} + u^{1,35} \leq 0,5^{1,35}$, $(1-v)^{1,79} + (1-u)^{1,79} \leq 0,5^{1,79}$. Наглядно эти области представлены на рисунке 4.

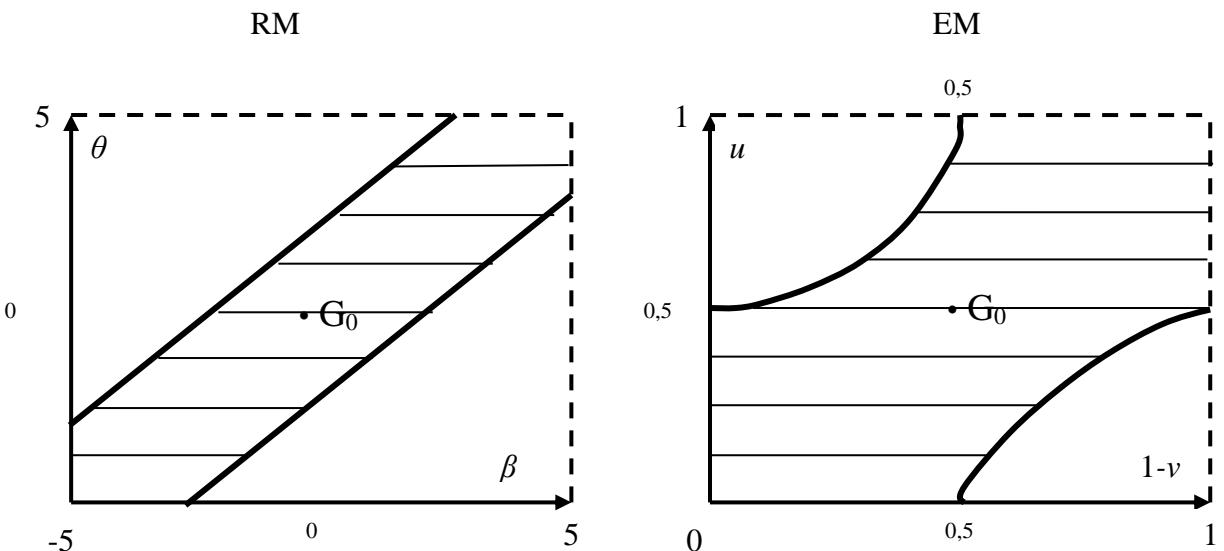


Рис. 4. Область хорошего разрешения теста

Теперь можно отметить, что на рисунке 2 жирной линией выделены те участки графика, на которых тестовые задания обладают хорошим разрешением в том понимании, как указано выше.

Завершая сопоставление моделей, отметим, математически важное свойство модели Раша – её аддитивную симметрию относительно значения $\gamma = 0$. Это свойство проявляется на рисунках 2 и 4, это свойство отражается и в математическом законе (1), и его следствии для двух взаимно обратных событий: $D(\gamma) + D(-\gamma) = 1$. Это следствие можно толковать примерно следующим высказыванием: «С какой вероятностью хорошо подготовленный учащийся может совершить ошибку в очень лёгком тестовом задании, с той же вероятностью слабо подготовленный учащийся вдруг выполнит очёнь трудное тестовое задание».

ЕМ – несимметричная модель относительно γ , это свойство также отражено на рисунках 2 и 4. Для двух взаимно обратных событий справедливо неравенство $D(\gamma) + D(1/\gamma) \leq 1$, и знак равенства выполняется только при $\gamma = 1$. Для ЕМ справедливо другое высказывание, которое нам представляется более правдоподобным: «Вероятность того, что хорошо подготовленный учащийся совершил ошибку в очень легком тестовом задании, больше вероятности правильного решения очень трудного задания слабо подготовленным учеником».

Переходя ко второй части работы, отметим, сама модель Раша не предполагает какого-либо учёта времени и оценки количества тестовых заданий. Эти вопросы обсуждаются лишь с позиций статистической надёжности результатов тестирования и здоровьесберегающих технологий в образовании в виде рекомендаций: «Допустимая надёжность результатов тестирования может быть обеспечена при количестве тестовых заданий не менее 30, время выполнения теста не должно превышать 40 мин для исключения эффекта утомления испытуемых». ЕМ позволяет рассматривать и вопросы, связанные с нормировкой времени.

Е – Т модель педагогического теста.

В основе Е-Т модели лежит закон экспоненциального спада интенсивности мыслительных действий с течением времени²: $P \sim \exp(-T)$. Если этот закон приложить к педагогическому тесту, то формулу (2) можно переписать в виде:

$$D = 0,5 \frac{\dot{O}/\dot{O}_0}{\tau}, \quad \dot{O} = \tau \cdot \beta, \quad \dot{O}_0 = \tau \cdot \theta . \quad (3)$$

Прокомментируем новые обозначения. Масштабный множитель τ имеет размерность интервалов времени (удобно минуты), а его значение равно времени, которое затратит опорный испытуемый, чтобы правильно выполнить опорное задание с вероятностью 0,5. Переменные β и θ определяются в ЕМ. Переменные T и T_0 можно интерпретировать следующим образом: задание уровня трудности T означает такое задание, которое за это время правильно с вероятностью 0,5 решит опорный испытуемый; испытуемый уровня подготовки T_0 – это такой испытуемый, который за это время с совместной вероятностью 0,5 правильно выполнит T_0/τ опорных заданий.

Оценка трудности тестового задания или уровня подготовленности испытуемого интервалами времени достаточно широко применяется и в работе практикующего учителя. Например, характеризуя контрольную работу, учитель оценивает меру её трудности фразой «примерно на 30 мин», или, пытаясь подчеркнуть хорошую подготовку к ней одного из учеников, учитель считает, что он способен выполнить её за

10 мин. Эти, взятые из опыта, субъективно-интуитивные оценки принимают вполне конкретный и измеряемый смысл, если использовать нашу трактовку. Тридцать минут на контрольную работу означает, что за это время опорным учащимся будут выполнены все её задания с вероятностью 0,5 каждое, а хорошо подготовленный ученик (в три раза лучше, чем опорный) за отведённое время выполнит каждое задание с вероятностью $0,5^{1/3} \approx 0,8$.

Прямая связь показателя уровня трудности задания и времени его выполнения не только предполагается из соображений здравого смысла, но и проявляется в экспериментальных исследованиях. В работе³ приведён пример анализа данных демоверсии ЕГЭ по математике 2010 на основе ЕМ: доли правильных ответов на задания В1 – В12 взяты из статистики сайта (<http://live.mephist.ru>), по этим данным не сложно оценить уровень трудности заданий β , ориентиры времени T (мин) выполнения каждого из заданий приведены в сопровождающих экзаменационных материалах. Коэффициент корреляции между значениями β и T примерно равен 0,87, а уравнение линейной регрессии имеет вид: $T \approx 2,66 \cdot \beta$ ($R^2 \approx 0,74$).

В этом примере предполагаемая линейная связь между показателями β и T , статистически достаточно обоснована. В общем случае подобные выводы делать пока преждевременно, но наличие такой связи и возможности непосредственной оценки латентных показателей мы рассматриваем в качестве исследовательской гипотезы.

Любая теория оценивается широкой практикой использования её следствий. Попытаемся и мы предложить учителю вариант формирования нормированного по времени педагогического теста с хорошей разрешающей способностью.

1 шаг. Предположить некоторого опорного испытуемого, принять для него уровень подготовленности $\theta_0 = 1$. Это может быть как реальный испытуемый так и вымышленный, как удобно учителю. Важно только то, что он делит группу испытуемых, ранжированных по уровню их подготовленности к материалу теста, на две примерно равные части.

2 шаг состоит в предварительном определении времени τ . В практике это время изменяется от 0,5 до 5 минут. Значение τ зависит от количества и уровня трудности

³ Каргин Ю.Н., Каргина Н.Ю. Мультиплективная модель анализа педагогического текста [Текст] / Ю.Н.Каргин, Н.Ю. Каргина // Фундаментальные науки и образование : материалы III Всероссийской науч.-практ. конф. (Бийск, 31 января - 3 февраля 2010 г.) / Бийский гос. пед. ун-т им. В.М. Шукшина. - Бийск : БПГУ им. В.М. Шукшина, 2010. - с.219-225.

тестовых заданий. После выполнения последнего шага, возможно, Вам придётся подкорректировать это значение.

3 шаг. Составить предварительно опорное тестовое задание с уровнем трудности $\beta_0 = 1$, которое опорный испытуемый должен правильно выполнить за время τ с вероятностью 0,5.

4 шаг. Выделить самого подготовленного испытуемого (реального или вымышленного) и оценить уровень его подготовленности θ_{max} из следующих соображений: θ_{max} равно количеству опорных заданий, которые выполнит этот испытуемый за время τ с вероятность 0,5. Т.е. он в θ_{max} раз лучше подготовлен к тесту, чем опорный испытуемый.

5 шаг. Составляем самое трудное тестовое задание. Его самый подготовленный испытуемый с вероятностью 0,5 должен выполнить за время τ . Уровень трудности этого задания $\beta_{max} = \theta_{max}$. Условие для проверки – это задание опорный испытуемый должен выполнить с вероятностью 0,5 за время $T_{max} = \theta_{max} \cdot \tau$.

Повторяем 4 и 5 шаги для самого слабо подготовленного испытуемого, оцениваем уровень его подготовленности θ_{min} (он меньше единицы, отражает долю опорного задания, выполненного с вероятностью 0,5 за время τ), составляем самое легкое задание $\beta_{min} = \theta_{min}$ и рассчитываем время его выполнения опорным испытуемым $T_{min} = \theta_{min} \cdot \tau$.

Составляем группу тестовых заданий так, чтобы они располагались в диапазоне от самого легкого до самого трудного, и само их распределение по уровню трудности было подобно распределению испытуемых по уровню подготовленности (например, равномерно или с большим количеством среднего уровня). Рассчитываем время правильного выполнения с вероятностью 0,5 каждого задания опорным испытуемым.

Последний шаг. Суммируем время выполнения всех заданий опорным испытуемым, это и есть расчётное время выполнения всего теста. Если это время Вас устраивает, то тест готов для применения. Если нет, то требуется или изменять количество тестовых заданий, или изменять масштабный множитель τ и трудность опорного задания, и опять переходить к шагу 2.

После апробирования теста на группе испытуемых, можно оценить его качество из следующих соображений: чем ближе предварительные оценки уровней трудности тестовых заданий к уровням, рассчитанным по тестовым результатам, тем лучше разработан тест, тем лучше ранжированы и оценены участники тестирования.

В рамках этой работы не обсуждаются вопросы надёжности и валидности тестовых испытаний. Эти вопросы разработаны в рамках классической теории педагогического теста, и вполне применимы для нашего теста.

Громоздкий, на первый взгляд, алгоритм составления теста, после двух-трёхкратного применения становится, тем не менее, естественным и эффективным педагогическим инструментом разработки и использования в практике специфически объективного измерителя качества знаний учащихся.