

Построение альтернативной модели педагогических измерений по системе Г. Раша

Юрий Каргин

Ноябрьский колледж профессиональных
и информационных технологий

kargin04@yandex.ru

Опубликовано в ж. «Педагогические Измерения № 4, 2010 года.

В статье представлены исходные положения и процесс построения модели педагогических измерений, основанной на формальных определениях теории Георга Раша. Такой подход позволяет теоретически обосновать и проанализировать некоторые основные этой теории, глубже осмыслить используемые в ней понятия и привести более осмысленную интерпретацию её результатов.

Ключевые слова: теория педагогических измерений по модели Г. Раша, интерпретация свойств модели педагогических измерений.

Три этапа педагогических измерений по модели Г. Раша

Первый этап – обобщение эмпирических результатов. Графическое представление эмпирических результатов позволило выйти на более совершенный математический язык описания процессов тестирования. На рисунке 1 показан пример графического представления ответов испытуемых на некоторое тестовое задание. Ответы каждого испытуемого оцениваются по дихотомической шкале (1 – ответ правильный; 0 – ошибка). Каждой точке графика соответствует усреднённая по подгруппе испытуемых (или по группе заданий одного уровня трудности) доля правильных ответов. Эти точки аппроксимируются функцией.

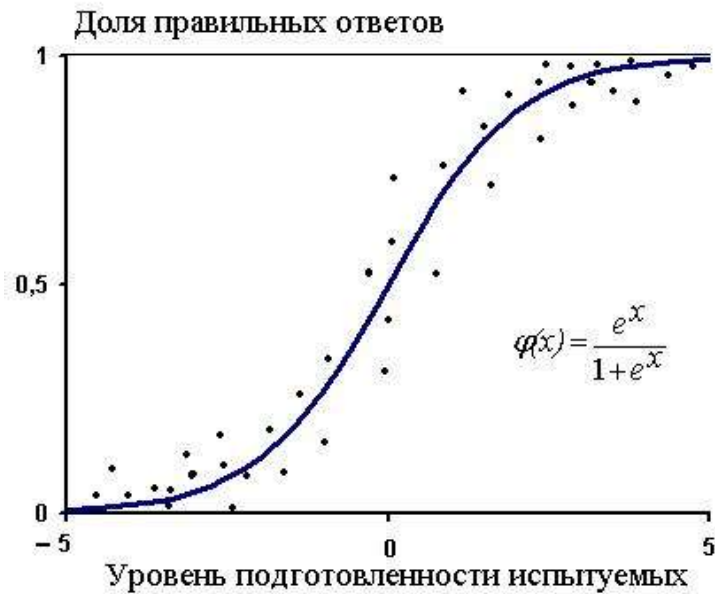


Рис. 1. Аналитическая аппроксимация эмпирических данных

Второй этап можно охарактеризовать как этап построения математической модели. На этом этапе, во-первых, определяется содержательный смысл латентных показателей уровня трудности тестового задания β и уровня подготовленности испытуемого θ . Далее, по виду функции $\varphi(x)$, записываются математические формулы для расчёта вероятности P правильного ответа усреднённых групп на тестовые задания в зависимости от значений латентных показателей:

$$P(\theta) = \frac{e^{\theta}}{1 + e^{\theta}} \quad , \quad P(\beta) = \frac{e^{-\beta}}{1 + e^{-\beta}} \quad . \quad (1)$$

И, наконец, сформулирована сама математическая модель педагогических измерений, как «объединение» формул (1) в единую взаимодействующую систему испытуемых и заданий теста:

$$P_{ij} = \frac{e^{(\theta_i - \beta_j)}}{1 + e^{(\theta_i - \beta_j)}} \quad , \quad (2)$$

где, P_{ij} – вероятность правильного ответа i – го испытуемого на j – ое задание теста. Отметим, что вероятность P является так же ненаблюдаемой величиной как и показатели β и θ , и в законе (2) имеет уже локальный (точечный) характер.

Третий этап – адаптационный. На этом этапе определяются условия и обосновываются правила применения модели в практике.

Именно на этом этапе разрешаются две достаточно важных и очевидных проблемы.

Первая. Прямой переход от формул (1) к обобщению (2) с математической точки зрения неправилен. Другими словами, математические закономерности для усреднённых (интегральных) вероятностей не обязательно выполняются для их локальных (дифференциальных) составляющих. Вторая проблема более глубокая. Показатели β и θ входят в модель (2) предполагая единую меру их измерения, т.е. в модели сопоставляются ранее несопоставимые свойства – трудность тестового задания и уровень подготовленности испытуемого.

Безусловно, математик Г. Раш понимал эти проблемы. И именно ему приписывают способы их разрешения посредством введения дополнительных стандартизирующих методов шкалирования: пересчёт наблюдаемых значений к общей средней арифметической, равной нулю, и общему показателю вариации, равному единице. Такая корректировка исходных тестовых данных, существенно снижает влияние первой проблемы на тестовое измерение, и предлагает разрешение второй проблемы выбором общих единиц измерения – логитов, отражающих степень вариации наблюдаемых значений от среднего значения, принимаемого равным нулю.

Таким образом, процесс тестового измерения заключается:

- в подсчёте долей правильных ответов p_j всеми испытуемыми на j – ое задание теста и правильных ответов i – м испытуемым на все задания теста p_i ;

- в расчёте исходных потенциалов $\frac{1-\delta_j}{p_j}$, $\frac{p_i}{1-\delta_i}$, а также показателей

$\ln \frac{1-\delta_j}{p_j}$, $\ln \frac{p_i}{1-\delta_i}$ трудности тестовых заданий и подготовленности испытуемых;

- в пересчёте процедурами шкалирования исходных показателей в латентные β_j и θ_i , со стандартными единицами измерения – логитами.

Исходные положения и вывод альтернативного закона

Процесс построения модели педагогических измерений можно провести и в «обратной» логике, не от практики к теоретической модели, а от постулирования теоретических положений к подтверждению их выводов на практике. Именно такой подход демонстрируется далее.

Определение 1. Испытуемый A во столько раз лучше подготовлен к тестовому заданию чем испытуемый B , во сколько раз он с тем же успехом и в тех же условиях правильно выполнит больше заданий этого уровня, чем испытуемый B .

Определение 2. Тестовое задание A во столько раз легче чем задание B , во сколько раз больше заданий этого уровня правильно выполнит испытуемый с тем же успехом и в тех же условиях, чем задание B .

Фразу «с тем же успехом» на математический язык можно перефразировать как «с той же совместной вероятностью». Под фразой «в тех же условиях» можно понимать комплекс всех как внешних, так и внутренних условий, например, мотивацию. Однако в практике приходится учитывать только наблюдаемые факторы, среди которых наиболее важным, на наш взгляд, является промежуток времени выполнения задания. Тогда эту фразу можно воспринимать как «за один и тот же промежуток времени».

Дополнительно прокомментируем эти определения на двух примерах.

Допустим, испытуемый B за некоторое время правильно выполняет тестовое задание с вероятностью 0,25. Тогда, студент A в два раза лучше подготовлен к выполнению тестового задания, если он за это же время с совместной вероятностью 0,25 правильно выполнит два таких задания. Иными словами, испытуемый A одно задание правильно выполнит с вероятностью 0,5, за время, в два раза быстрее, чем испытуемый B .

Допустим, испытуемый за некоторое время правильно выполняет задание B с вероятностью 0,25. Тогда, задание A в два раза легче, чем задание B , если испытуемый правильно выполнит его с вероятностью 0,5, за время, в два раза быстрее. Иными словами, два таких задания он правильно выполнит за то же время, с той же совместной вероятностью $0,5^2 = 0,25$.

Этих определений достаточно, чтобы построить модель педагогических измерений¹.

Пусть дано некоторое j – ое задание теста, которое назовём опорным². Того испытуемого (пусть даже вымышленного), который с вероятностью 0,5 правильно выполняет опорное задание теста, так же назовём опорным. Выбор опорного испытуемого определяется выбором опорного задания.

Чем выше уровень подготовленности испытуемого, тем больше вероятность его правильного ответа на это задание. Вероятность правильного ответа на опорное задание i – м испытуемым обозначим через u_i . Этот показатель отражает уровень

¹ Каргин Ю.Н. Свойства мультипликативной модели педагогического теста – <http://www.uni-altai.ru/info/journal/vestnik/3365-nomer-1-2010.html>

² В теории педагогического тестирования выделяют нормативно-ориентированные и критериально-ориентированные подходы при интерпретации тестовых результатов. Если показатели альтернативной модели нормировать на опорное тестовое задание, т.е. на содержание теста, как это делается в данной статье, то в указанной классификации результаты теста можно интерпретировать как критериально-ориентированные. Если же результаты тестового измерения нормировать на выделенного опорного испытуемого, например, среднего в ранжированном списке, то их следует отнести к нормативно-ориентированному виду интерпретации.

подготовленности испытуемого, который так и назовём – вероятностный уровень подготовленности испытуемого. Допустим i – й испытуемый в θ_i^0 раз лучше подготовлен опорного испытуемого, т.е. он, с учётом определения 1, выполнит с той же вероятностью 0,5 в θ_i^0 раз больше подобных заданий. Тогда можно записать $u_i \cdot \theta_i^0 = 0,5$ или $u_i = 0,5^{1/\theta_i^0}$

Показатель θ_i^0 назовём относительным уровнем подготовленности испытуемого.

Пусть дано j – ое задание теста, которое в β_j^0 раз труднее опорного задания, правильно выполняемого i – м испытуемым с вероятностью u_i , т.е. β_j^0 – относительный уровень трудности тестового задания. Тогда, из определения 2 следует, что j – ое задание теста он выполнит с вероятностью $u_i \cdot \beta_j^0$. Окончательно имеем искомый математический закон для расчёта вероятности правильного ответа на задание теста:

$$P_{ij} = 0,5^{\beta_j^0 / \theta_i^0}. \quad (3)$$

Если ввести вероятностный уровень трудности тестового задания $v_j = 0,5^{\beta_j^0}$, то закон (3) можно записать как в смешанных показателях $P_{ij} = u_i \cdot \beta_j^0 = v_j^{1/\theta_i^0}$, так и вероятностных:

$$\log_{0,5} P_{ij} = \log_{0,5} v_j \cdot \log_{0,5} u_i. \quad (4)$$

В законе (3) показатели β^0 и θ^0 измеряются в более совершенной шкале отношений с выбранной единицей для опорного уровня, в законе (4) показатели u и v имеют вероятностную трактовку. В IRT показатели модели (2) измеряются в интервальной шкале логитов. Переход к интервальным показателям можно осуществить и в нашей модели натуральным логарифмированием относительных показателей: $\beta = \ln \beta^0$, $\theta = \ln \theta^0$.

В качестве комментариев приведём некоторые следствия из закона (3).

1. Значение относительного показателя уровня трудности опорного задания равно единице, значение вероятностного показателя уровня трудности опорного задания равно 0,5.

2. Значение относительного показателя уровня подготовленности опорного испытуемого равно единице, значение вероятностного показателя уровня подготовленности опорного испытуемого равно 0,5.

3. Опорный испытуемый с вероятностью 0,5 правильно выполняет опорное задание теста.

4. Испытуемый с относительным уровнем подготовленности α правильно выполняет с вероятностью 0,5 тестовое задание уровня трудности α .

5. Относительные показатели принимают только положительные значения, которые изменяются в полуинтервале от нуля до бесконечности.

6. Значения относительных показателей показывают во сколько раз данное свойство ярче выражено (значительней проявляется) относительно опорного.

7. Вероятностные показатели могут принимать значения из интервала (0, 1).

8. Значения вероятностных показателей дают вероятность правильного ответа для опорного уровня.

9. Чем сложнее тестовое задание, тем выше значение относительного показателя и тем ниже значение вероятностного показателя.

10. Чем лучше подготовлен испытуемый, тем выше значения относительного и вероятностного показателей.

При построении математической модели педагогических измерений, закон (3, 4) представляется здесь как альтернативный по отношению к (2).

Эмпирическое обоснование альтернативного закона

В современном понимании педагогического теста³ есть требование относительно равномерно возрастающей трудности тестовых заданий. Как отмечает В.С. Аванесов, раньше этот уточняющий момент в определении теста не делался. В итоге задания некоторых т.н. «тестов» подбирались иногда с заметными «провалами» между заданиями, что сильно ухудшало метрические свойства метода; заметно снижалась точность измерений и дифференцирующая способность тестовых результатов. Можно с сожалением отметить, что ряд российских практиков и авторов этот критерий либо не признают, либо обходят стороной, как несущественный. Например, вместо понятия «система заданий» используют словосочетание «совокупность» или «множество заданий», как будто между ними нет разницы⁴.

Именно от этой мысли и оттолкнёмся в этом разделе.

Определение 3. Системой из m равномерно возрастающей трудности заданий будем называть такую совокупность заданий, уровень трудности которых определяется по закону: $v_k = \frac{2m-2k+1}{2m}$, где k – номер заданий возрастающей трудности, $k = 1, \dots, m$.

³ Аванесов В.С. Метрическая система Георга Раша — Rasch Measurement (RM). ПИ, №2, 2010 г. Стр. 59.

⁴ Там же. Стр. 72.

Несколько подробнее остановимся на этом определении. Оно даёт правило равномерного распределения тестовых заданий по вероятностному показателю трудности заданий. Например, для $m = 5$ тестовых заданий равномерно возрастающей трудности имеем значения v равные 0,9; 0,7; 0,5; 0,3; 0,1. Любопытно вычислить соответствующие им значения относительного показателя трудности тестовых заданий: β^0 примерно равно 0,152; 0,515; 1; 1,74; 3,32. Отсюда, в частности, самое сложное задание теста примерно в 22 раза труднее самого простого задания.

Если перейти к интервальным показателям, то получим значения β примерно равные -1,88; -0,66; 0; 0,55; 1,20. В IRT расстояние между соседними уровнями должно быть примерно одинаковым, желательно менее 0,5 или даже 0,25 логит. В данном подходе расстояние между соседними уровнями одинаковое, если оно измерено в вероятностных показателях. В этом случае, расстояния между соседними уровнями, измеренные в относительных и интервальных показателях, увеличиваются по мере удаления уровней от опорного.

Обращает на себя внимание и асимметрия значений показателей β и β^0 при симметрии показателей u для равномерного распределения тестовых заданий. Опора на вероятностные показатели не случайна. На наш взгляд, именно показатели u , v и их вероятностная трактовка, наиболее адекватно вписываются в вероятностную модель педагогических измерений. Далее акцентируем внимание ещё на одно отличие моделей. В IRT модель Раша (2) симметрична относительно параметров модели β и θ , а симметричность теста определяется симметричным распределением этих значений, и как следствие средним значением вероятности правильного ответа равным 0,5. Альтернативная модель симметрична относительно параметров u и v (4), асимметрична относительно параметров β^0 и θ^0 (3) со средним значением вероятности правильного ответа примерно равной 0,525. Т.е. в симметричном тесте доля правильных ответов несколько выше, чем неправильных. Выяснение причин такой асимметрии может служить проблемой отдельного исследования, разрешение которой лежит не в математической, а скорее педагогической или даже психологической области.

Рассмотрим тест с системой равномерно возрастающей трудности заданий и равномерным распределением испытуемых по уровню подготовленности u . В работе⁵ решаются задачи вычисления средней вероятности $P(\theta)$ правильного ответа испытуемого с уровнем подготовленности θ на все задания такого теста и вычисления

⁵ Каргин Ю. Модель педагогических измерений в шкале отношений//Педагогические измерения. 2010. №2. С.44-56.

средней вероятности $P(\beta)$ правильного ответа всех испытуемых на тестовое задание уровня трудности β :

$$\begin{aligned} D(\theta) &= \frac{1}{m} \sum_{j=1}^m P_{.j} \approx \int_0^1 v^{1/\theta^o} dv = \frac{\theta^o}{1+\theta^o} = \frac{e^\theta}{1+e^\theta}, \\ P(\beta) &= \frac{1}{n} \sum_{i=1}^n P_{ij} \approx \int_0^1 u^{\beta^o} du = \frac{1}{1+\beta^o} = \frac{e^{-\beta}}{1+e^{-\beta}}. \end{aligned} \quad (5)$$

Полученные результаты совпадают с эмпирическими закономерностями (1). На наш взгляд, это серьёзный аргумент в пользу закона (3). Формулы (5) дают и правила оценки значений мер трудности заданий и уровня подготовленности испытуемых по наблюдаемым долям правильных ответов для j – го задания p_j и i – го испытуемого p_i :

- для относительных показателей β^o и θ^o соответственно имеем оценки $(1-p_j)/p_j$ и $p_i/(1-p_i)$. В IRT эти показатели называют «неопределённым» потенциалом. Здесь они имеют однозначно определённый смысл (см. определения 1 и 2);

- для вероятностных показателей v и u соответственно имеем оценки, рассчитанные по формулам $v = 0,5^{\beta^{\hat{i}}}$ и $u = 0,5^{1/\theta^{\hat{i}}}$. В IRT иллюстративно обсуждаются вероятности правильных ответов в зависимости от значений параметра $(\theta - \beta)$. Здесь вероятностные показатели несут однозначно определённый смысл (см. следствие 8).

Совершенно очевидно, что данный подход полностью снимает и указанные выше две проблемы «Раша». Во-первых, сам переход из (3) к (1) строг и точен для предельного случая бесконечно большого числа равномерно распределённых по уровню трудности тестовых заданий и равномерно распределённых по уровню подготовленности испытуемых. Во-вторых, вероятностные и относительные единицы измерения показателей альтернативной модели, снимают проблему «сопоставления несопоставимых свойств». В конечном итоге дополнительные процедуры шкалирования (или искусственного статистического выравнивания) совершенно исключаются в данном подходе. На наш взгляд, и этот факт в пользу альтернативной модели.

В работе⁶ проводилась сравнительная оценка погрешностей аналитических расчётов по законам (2) и (3) с экспериментальным данным. Модель Раша уступает и по этому признаку.

Визуализация альтернативного закона

Визуализация аналитически выраженных законов является дополнительным способом лучшего их осмысления. Именно для этого приведём два рисунка,

относящихся к альтернативному закону (3, 4), и ограничимся кратким комментарием к ним.

На рисунке 2 представлена поверхность $P(u, v)$. Центральная, самая тёмная полоса, соответствует оптимальному режиму работы теста. В этой области изменения параметров модели u и v вероятность правильного ответа на тестовые задания изменяется в интервале (0,4; 0,6). Именно в этой области уровень трудности тестовых заданий лучше всего соответствует уровню подготовленности испытуемых, именно в этой области разрешения тестовых заданий самые высокие, что отражается в наибольшей «крутизне» поверхности. Области хорошего режима работы теста представлены двумя менее тёмными полосами поверхности. Нижняя полоса отражает изменения вероятности правильного ответа P в интервале (0,2; 0,4), а верхняя в интервале (0,6; 0,8). Самые светлые области поверхности P , особенно её крайние практически горизонтальные части, обладают низким разрешающим потенциалом. Нижнее плато поверхности отражает область «трудные задания – слабая подготовка испытуемых» где вероятность правильного ответа всех испытуемых близка к нулю, независимо от задания, и смысла в таких заданиях нет. Верхнее плато поверхности отражает противоположную область «лёгкие задания – сильная подготовка испытуемых». Здесь все испытуемые легко отвечают на любое задание теста и их дифференциация весьма затруднительна.

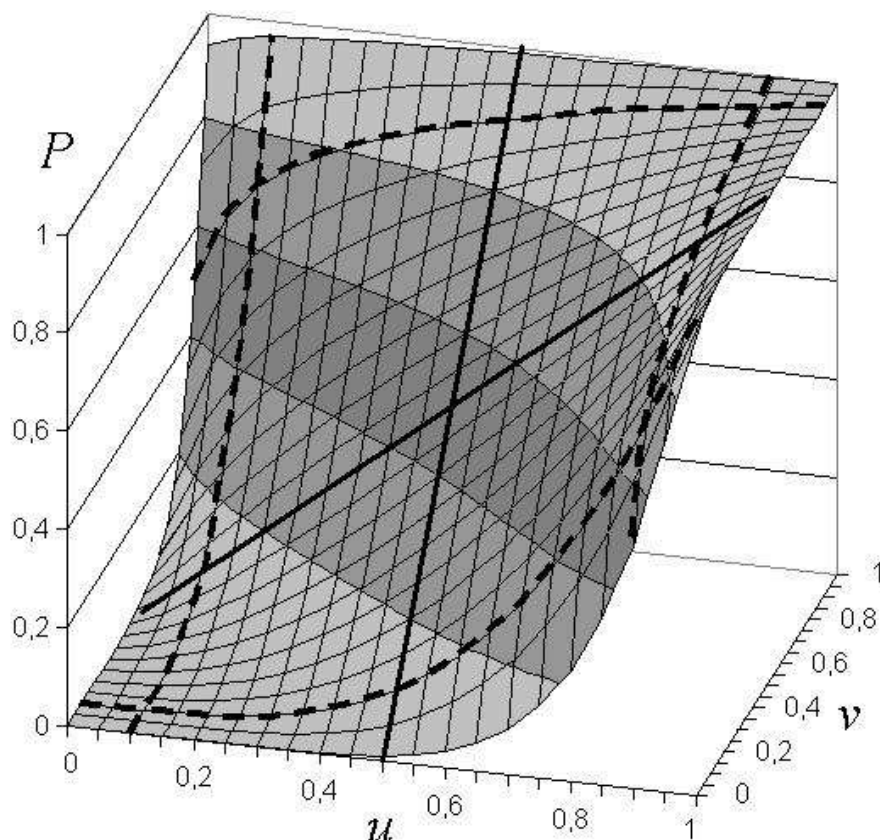


Рис. 2. Поверхность $P(u, v)$

Сплошными линиями мы выделили сечения поверхности $P(u, v)$ плоскостями опорного испытуемого $u = 0,5$ и опорного тестового задания $v = 0,5$. В этом случае вероятность правильного ответа определяется предельно простыми соотношениями $P(v) = v$ и $P(u) = u$, а графики этих функций – прямые линии. Пунктиром выделены сечения $u, v = 0,1$ и $u, v = 0,9$. Даже непосредственный визуальный анализ этих сечений даёт представление о характере изменения вероятности правильного ответа P для этих значений параметров модели.

На рисунке 3 представлен «вид сверху» на поверхность $P(u, v)$. Тон выделенных участков прежний. Узкие центральные полосы ещё раз отражают области хорошего разрешения тестовых заданий. Более подробно, с аналитическими расчётами, вопросы разрешения тестовых заданий обсуждались в работе⁶.

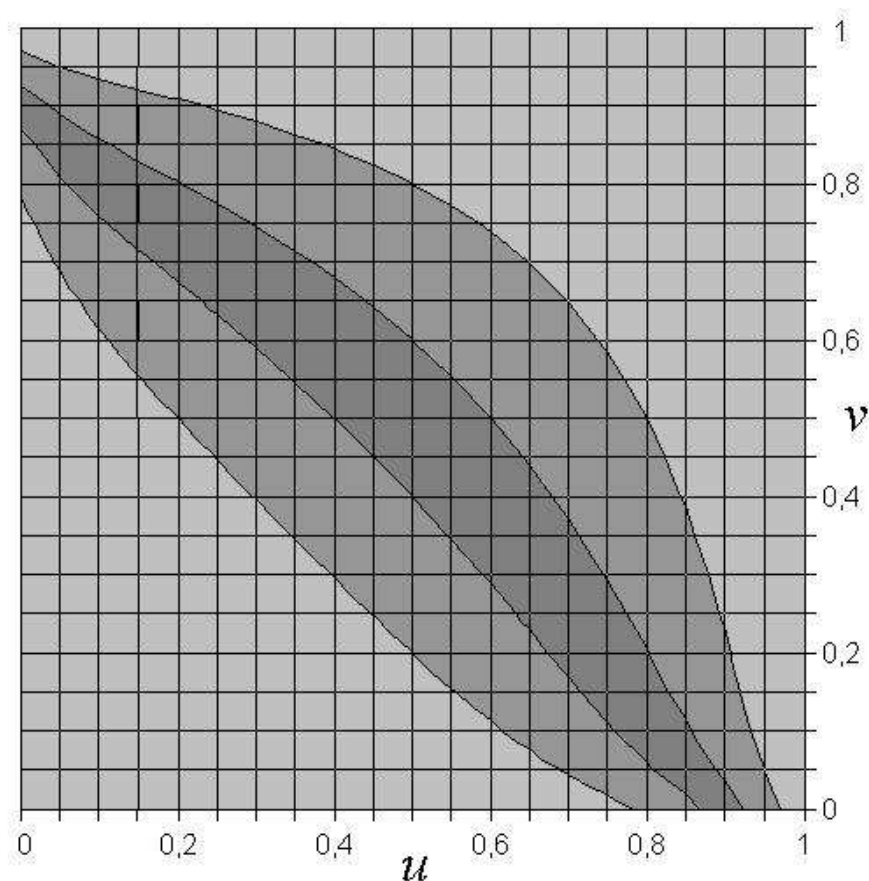


Рис. 3. Проекция поверхности $P(u, v)$ на плоскость изменения параметров модели

На наш взгляд, как в теории представленной модели, так и в практике её применения есть дискуссионные и даже проблемные моменты. На некоторые из них мы уже обращали внимание. Например, асимметрия теоретических положений и практика оценки латентных показателей для неравномерного теста. Однако сам переход от статистического нормирования эмпирических данных, реализованного в модели Раша, к аналитическому обобщению через постулирование ключевых понятий в данном подходе, может дать более адекватные решения проблемы педагогических измерений.