

Методология

МАТЕМАТИЧЕСКИЕ ОСНОВЫ RASCH MEASUREMENT

Олег Деменчёнок

Восточно-Сибирский институт МВД России
AskSystem@yandex.ru

Рассмотрены основы Rasch Measurement. Проанализированы достаточные статистики для педагогического измерения, выявлена зависимость существования этих статистик от метода подбора параметров модели Раша. Показана взаимосвязь модели Раша и пятым частным случаем закона Терстоуна. Сформулированы критерии выбора модели измерения.

Ключевые слова: математические модели педагогических измерений, модель Раша, Rasch Measurement, Item Response Theory, закон Терстоуна.

Введение

Rasch Measurement (RM) — направление современной теории педагогических и психологических измерений, основанное датским математиком и статистиком Георгом Рашем. Ключевым элементом RM является математическая модель. Эта модель и направление в целом находят применение не только в психологии и педагогике, но и в других областях, таких, как изучение общественного мнения, рыночной конъюнктуры и т.д. К сожалению, сам Раш не успел

конкретизировать положения своей революционной теории измерений до уровня, достаточного для решения многих прикладных задач.

Rasch Measurement явился исходной теоретической базой Item Response Theory (IRT) — теории измерений, использующей кроме модели Раша и другие математические модели. По мнению В.С.Аванесова, RM нельзя рассматривать как частный случай Item Response Theory. Научный статус RM много выше IRT, поскольку представляет собой открытие в области систем педагогических измерений. RM позволяет получить непересекающиеся графики заданий теста, что считается существенным преимуществом для построения интервальной шкалы в педагогических измерениях¹.

Математическая основа модели Раша

В работе², ставшей началом нового направления теории педагогических измерений, Георг Раш предпринял попытку вывода уравнения модели, определяющей вероятность правильного ответа испытуемого на тестовое задание.

По сути, модель Раша — это научная гипотеза, основанная на следующих предположениях:

1) мера уровня подготовленности любого испытуемого t_i (т.е. количественная характеристика уровня подготовленности испытуемого по определённому множеству заданий теста) не должна зависеть от уровня трудности тестовых заданий $t_i \in]0; \infty[$;

2) вероятность правильного ответа испытуемого P_i зависит только от уровня подготовленности испытуемого и от уровня трудности тестового задания $b \in]0; \infty[$ (т.е. количественной характеристики тестового задания, не зависящей от выборки испытуемых и отражённой на определённой шкале по конкретному разделу определённой области знания) или $P = f(t, b)$;

3) ответы испытуемых на задания теста статистически независимы, из чего следует взаимная независимость измеряемых параметров.

Предположим, что уровень трудности b_1 первого задания в k раз больше уровня трудности b_2 другого задания:

$$b_1 = k \cdot b_2. \quad (1)$$

С учётом принятых предположений первое задание должно оказаться в k раз труднее для каждого испытуемого.

Предположим также, что уровень подготовленности первого испытуемого t_1 в k раз

1
Аванесов В.С.
Применение тестовых форм в Rasch Measurement // Педагогические измерения. № 4. 2006.

2
Rasch G.
Probabilistic Models for Some Intelligence and Attainment Tests. Copenhagen, 1960, Danish Institute of Educational Research (Expanded edition, Chicago. 1980. Mesa Press. 199 p.).

больше уровня подготовленности другого испытуемого t_2

$$t_1 = k \cdot t_2. \quad (2)$$

Полагая уровни подготовленности и уровни трудности положительными величинами, разделим уравнение (1) на уравнение (2)

$$\frac{b_1}{t_1} = \frac{k \cdot b_2}{k \cdot t_2} = \frac{b_2}{t_2}. \quad (3)$$

Следовательно:

- вероятность правильного решения более подготовленным первым участником более трудного первого задания должна совпадать с вероятностью правильного решения вторым участником второго задания;
- вероятность правильного ответа определяется отношением уровня подготовленности испытуемого к уровню трудности тестового задания $P = f(t/b)$. Поэтому модель Раша называют однопараметрической, хотя она оперирует двумя переменными.

Конкретный вид аналитической зависимости $P = f(t/b)$ неизвестен, но эта функция должна отвечать ряду условий.

1. Значение вероятности правильного ответа должно быть в пределах от нуля до единицы, т.е. область значений $P \in [0; 1]$.
2. Функция $P = f(t/b)$ должна монотонно возрастать, так как при постоянном уровне трудности задания большему значению уровня подготовленности

должна соответствовать большая вероятность правильного ответа.

3. Логично предположить, что при $t/b > 1$ (уровень подготовленности тестируемого больше уровня трудности задания) вероятность правильного ответа P больше вероятности ошибки, т.е. $P > 0,5$. Чем больше отношение уровня подготовленности к уровню трудности задания, тем легче испытуемому решить задание. Значит, с увеличением параметра t/b вероятность правильного ответа должна возрастать. Но вероятность не может быть больше единицы, поэтому вероятность правильного ответа будет неограниченно приближаться к единице (асимптотически стремиться к единице), т.е.

$$\lim_{t/b \rightarrow \infty} P(t/b) = 1.$$

4. Аналогичный ход рассуждений приводит к выводу: если уровень подготовленности меньше уровня трудности задания, то вероятность правильного ответа $P < 0,5$; и с уменьшением параметра t/b вероятность правильного ответа P уменьшается, стремясь к нулю, т.е.

$$\lim_{t/b \rightarrow 0} P(t/b) = 0.$$

5. В случае равенства уровня подготовленности и уровня трудности задания вероятность правильного ответа $P = 0,5$. Характер взаимосвязи между параметром модели и вероятностью правильного ответа приведён на рис. 1.

Методология

ПЕД
измерения

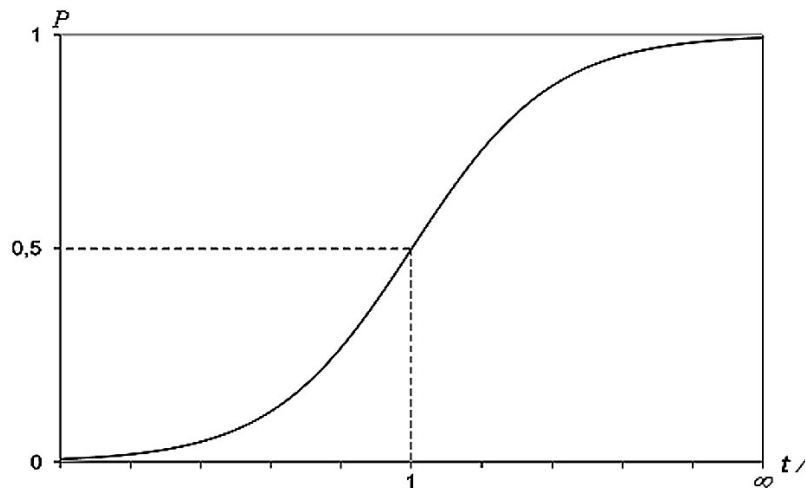


Рис. 1. Связь параметра модели Раша и вероятности правильного ответа

Георг Раш предложил первую форму записи своей модели в виде:

$$P = \frac{t}{t+b} = \frac{t/b}{1+t/b}. \quad (4)$$

Нетрудно убедиться, что уравнение (4) отвечает всем сформулированным выше требованиям к функции $P = f(t/b)$.

Модель Раша в виде логистической функции

Для построения шкалы измерений оказалось удобным выражать уровень подготовленности t и уровень трудности b в логарифмическом масштабе:

$$\theta = \ln(t),$$

$$\beta = \ln(b),$$

где θ и β — логарифмы уровней подготовленности и трудности или значения уровней подготовленности и трудности, измеряемые в логарифмическом масштабе (в соответствии с принятой терминологией, далее под уровнями подготовленности и трудности будем понимать θ и β).

Тогда модель Раша примет вид:

$$P = \frac{t/b}{1+t/b} = \frac{e^\theta/e^\beta}{1+e^\theta/e^\beta} = \frac{e^{\theta-\beta}}{1+e^{\theta-\beta}}. \quad (5)$$

Уравнение (5) — наиболее известная форма записи модели Раша. Заметим, что модель Раша относится к логистическим функциям³, общее уравнение которых имеет вид:

3
Логистической называется функция с S-образной формой графика, принимающая значения в интервале от нуля до единицы. (Электронный учебник по статистике. Москва, StatSoft; http://www.statsoft.ru/home/textbook/glossary/GLOSS_L.HTML).

$$f(x) = \frac{e^x}{1 + e^x}.$$

Логарифмическая шкала даёт возможность измерять уровни подготовленности θ и уровни трудности β в общем масштабе. За единицу шкалы принят логит (от англ. log odds unit — единица логарифмической разницы). Шкала характеризуется равными линейными интервалами (т.е. расстояние между 1 и 2 равно расстоянию между 2 и 3, между 10 и 11 и т.д.). Известны математические выкладки^{4,5}, подтверждающие корректность модели Раша в виде логистической функции.

Исходя из положенных в основу модели Раша предположений, сравним уровни трудности двух заданий. Задания i и j можно различить по уровню трудности только тогда, когда частоты ответов испытуемых на эти задания различаются. Для сравнения заданий i и j необходимо посчитать, как часто испытуемые отвечали правильно на задание i , в то время как на задание j отвечали неправильно, и соотнести полученное с тем, как часто происходили обратные события.

Количественное сравнение трудности заданий i и j на основе этой пары взаимных значений требует в качестве объективного инструмента вероятностную модель, т.е. сравнение вероятностей. Обозначим x_i — результат выполнения i -го зада-

ния ($x_i = 1$ при правильном ответе, $x_i = 0$ при неправильном), $P(x_i)$ — вероятность исхода x_i .

Учитывая независимость тестовых заданий, отношение уровней трудности заданий i и j можно записать в виде

$$\frac{P_{(x_i=1, x_j=0)}}{P_{(x_i=0, x_j=1)}} = \frac{P_{(x_i=1)} \cdot P_{(x_j=0)}}{P_{(x_i=0)} \cdot P_{(x_j=1)}} = \frac{P_i \cdot (1 - P_j)}{(1 - P_i) \cdot P_j}, \quad (6)$$

где P_i и P_j — вероятности правильного ответа на i -е и j -е задание.

В соответствии с предположением о зависимости вероятности правильного ответа от уровня подготовленности испытуемого и уровня трудности тестового задания, уравнение (6) не должно зависеть от испытуемых. Другими словами, отношение уровней трудности двух заданий должно быть равным для любой пары испытуемых n и m :

$$\frac{P_{ni} \cdot (1 - P_{nj})}{(1 - P_{ni}) \cdot P_{nj}} = \frac{P_{mi} \cdot (1 - P_{mj})}{(1 - P_{mi}) \cdot P_{mj}}. \quad (7)$$

Выполнение условия (7) означает независимость уровней трудности заданий от параметров испытуемых. Таким образом, соблюдается сформулированное Л.Л. Терстоуном и другими требование независимости (инвариантности) средств измерения от объекта измерения⁶.

Для упрощения оценки значений уравнения (7) мы можем выбрать $j = 0$ и $m = 0$. Тогда при

Методология

4

Wright B.D., Stone M.H. Measurement Essentials, 2nd Edition. Wilmington: Wide Range Inc, 1999.

5

Linacre J.M. The Rasch Model cannot be «Disproved»! // Rasch Measurement Transactions, 1996. 10:3. P. 512–514.

6

Thurstone L.L. Attitudes can be measured // American Journal of Sociology. Vol. 33. January. 1928. 529–544 pp.

шкалировании задание i будет сравниваться с заданием $j = 0$, а испытуемый n — с испытуемым $m = 0$. Можно также выбрать шкалу измерения так, чтобы испытуемый $m = 0$ с вероятностью 50% выполнял задание $j = 0$:

$$P_{mj} = P_{00} = 0,5$$

$$\frac{1 - P_{00}}{P_{00}} = 1.$$

Подставив $j = 0$ и $m = 0$ в выражение (7), получим:

$$\frac{P_{ni}(1 - P_{n0})}{(1 - P_{ni})P_{n0}} = \frac{P_{0i}(1 - P_{00})}{(1 - P_{0i})P_{00}},$$

$$\frac{P_{ni}}{1 - P_{ni}} = \frac{P_{n0}}{1 - P_{n0}} \cdot \frac{P_{0i} \cdot (1 - 0,5)}{(1 - P_{0i}) \cdot 0,5}, \quad (8)$$

$$\frac{P_{ni}}{1 - P_{ni}} = \frac{P_{n0}}{1 - P_{n0}} \cdot \frac{P_{0i}}{1 - P_{0i}} = t_n \cdot \frac{1}{b_i}.$$

Первый множитель правой части уравнения (8), зависящий только от свойств испытуемого n и изменяющийся от нуля до бесконечности, примем за уровень его подготовленности:

$$t_n = \frac{P_{n0}}{1 - P_{n0}}, \quad (9)$$

т.е. уровень подготовленности любого испытуемого определяется как отношение вероятности верного выполнения им стандартного задания, уровню трудности которого принят за единицу, к вероятности неверного выполнения этого задания.

Второй множитель правой части уравнения (8), зависящий только от свойств i -го задания и изменяющийся от нуля

до бесконечности, примем за величину, обратную его уровню трудности

$$\frac{1}{b_i} = \frac{P_{0i}}{1 - P_{0i}}, \quad b_i = \frac{1 - P_{0i}}{P_{0i}}, \quad (10)$$

т.е. уровень трудности любого задания отношением вероятности неверного выполнения этого задания стандартным испытуемым, уровню подготовленности которого принят за единицу, к вероятности правильного решения.

Шкала, определённая отношением $\frac{P_{ni}}{1 - P_{ni}}$, может быть пре-

образована в шкалу с равными линейными интервалами путём логарифмирования:

$$\ln\left(\frac{P_{ni}}{1 - P_{ni}}\right) = \ln\left(\frac{P_{n0}}{1 - P_{n0}}\right) +$$

$$+ \ln\left(\frac{P_{0i}}{1 - P_{0i}}\right) = \ln t_n - \ln b_i = \quad (11)$$

$$= \theta_n - \beta_i.$$

После очевидных преобразований получаем уже знакомое уравнение модели Раша (график этой функции приведён на рис. 2):

$$\frac{P_{ni}}{1 - P_{ni}} = e^{\theta_n - \beta_i},$$

$$P_{ni} = e^{\theta_n - \beta_i} - P_{ni} \cdot e^{\theta_n - \beta_i},$$

$$P_{ni} + P_{ni} \cdot e^{\theta_n - \beta_i} = e^{\theta_n - \beta_i},$$

$$P_{ni}(1 + e^{\theta_n - \beta_i}) = e^{\theta_n - \beta_i},$$

$$P_{ni} = \frac{e^{\theta_n - \beta_i}}{1 + e^{\theta_n - \beta_i}}.$$

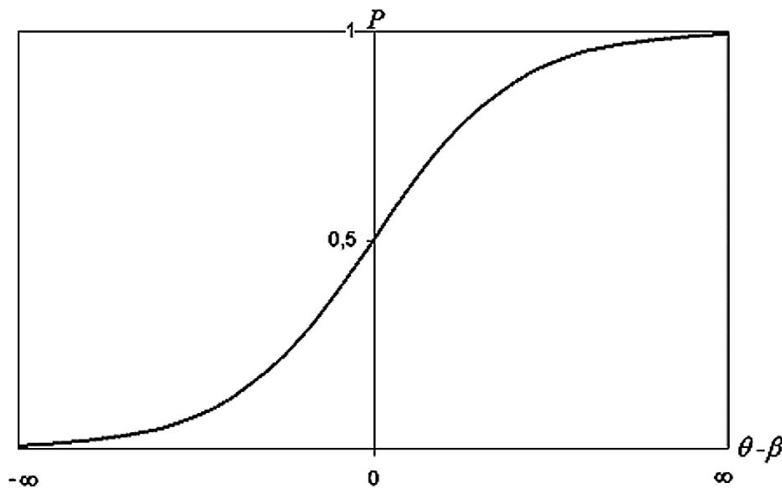


Рис. 2. График модели Раша

Достаточные статистики для модели Раша

Одно из преимуществ модели Раша — существование т.н. достаточных статистик⁷. Некоторая функция $\varphi(x)$, рассчитанная по результатам наблюдений $x_1, x_2 \dots x_n$, является достаточной статистикой тогда и только тогда, когда плотность распределения $f_w(x)$ может быть записана в виде⁸:

$$\prod_{i=1}^n f_w(x_i) = g_w(\varphi(x_i)) \cdot h(x_i). \quad (12)$$

Это означает, что условие существования достаточной статистики заключается в возможности представления плотности распределения в виде двух сомножителей, один из которых (h) не зависит от параметра w , а другой (g) зависит, но не

непосредственно от результатов наблюдений, а только через статистику $\varphi(x)$.

Достаточные статистики позволяют уменьшить объём статистических данных без потери информации об интересующем нас параметре.

Рассмотрим обоснование достаточных статистик для модели Раша.

Задача определения уровней подготовленности испытуемых и уровней трудности заданий по исходным данным тестирования заключается в подборе таких значений этих параметров, чтобы результаты тестирования и расчёт по модели наилучшим образом совпадали. Наиболее известный критерий оптимальности подбора параметров основан на методе максимального правдоподобия. Этот

Методология

7

Достаточная статистика — совокупность функций от результатов наблюдений, которые содержат ту же статистическую информацию о неизвестных величинах, что и сами результаты наблюдений. (Большая советская энциклопедия, электронная версия. М.: Большая Российская энциклопедия, 2002).

8

Андронов А.М.,
Копытов Е.А.,
Прийглаз Л.Я.
Теория вероятностей и математическая статистика: Учебник для вузов. СПб.: Питер, 2004. 461 с.

критерий требует выбора таких параметров, при которых вероятность получить в процессе измерения фактически наблюдаемые величины была бы максимальной. В формальной записи это соответствует максимуму произведения вероятностей всех наблюдаемых независимых событий⁹:

$$p_1 \cdot p_2 \cdot \dots \cdot p_N = \prod_{i=1}^N p_i \rightarrow \max. \quad (13)$$

По модели Раша можно найти вероятность результата x при решении i -м тестируемым j -го задания:

$$P_{ij}(x_{ij}) = \frac{e^{x_{ij}(\theta_i - \beta_j)}}{1 + e^{\theta_i - \beta_j}}, \quad (14)$$

где x_{ij} — результат выполнения задания, $x = 1$ при правильном ответе, $x = 0$ при неправильном ответе.

Действительно, в случае правильного ответа (14) равноценно модели Раша:

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

Выражение (13) корректно и в случае неправильного ответа:

$$\begin{aligned} P_{ij}(x_{ij} = 0) &= \frac{e^0}{1 + e^{\theta_i - \beta_j}} = \\ &= \frac{1 + e^{\theta_i - \beta_j} - e^{\theta_i - \beta_j}}{1 + e^{\theta_i - \beta_j}} = \\ &= 1 - \frac{e^{\theta_i - \beta_j}}{1 + e^{\theta_i - \beta_j}} = 1 - P_{ij}(x_{ij} = 1). \end{aligned}$$

Тогда критерий оптимальности подбора параметров (13) примет вид:

$$\begin{aligned} \prod_{i=1}^n \prod_{j=1}^m P_{ij}(x) &= \\ &= \prod_{i=1}^n \prod_{j=1}^m \frac{e^{x_{ij}(\theta_i - \beta_j)}}{1 + e^{\theta_i - \beta_j}} \rightarrow \max, \quad (15) \end{aligned}$$

где n — количество испытуемых; m — число тестовых заданий.

Зная, что $e^a e^b = e^{a+b}$, преобразуем числитель:

$$\frac{e^{\sum_{i=1}^n \sum_{j=1}^m x_{ij}(\theta_i - \beta_j)}}{\prod_{i=1}^n \prod_{j=1}^m (1 + e^{\theta_i - \beta_j})} \rightarrow \max. \quad (16)$$

Введём обозначения: R_i — сумма результатов выполнения тестовых заданий i -м тестируемым (R_i также называют первичным или исходным баллом испытуемого):

$$R_i = \sum_{j=1}^m x_{ij},$$

s_j — сумма результатов выполнения j -го задания всеми тестируемыми (первичный или исходный балл задания):

$$s_j = \sum_{i=1}^n x_{ij}.$$

Используя R_i и s_j , заменим

$$\begin{aligned} \sum_{i=1}^n \sum_{j=1}^m x_{ij} \theta_i &= \sum_{i=1}^n R_i \theta_i, \\ \sum_{i=1}^n \sum_{j=1}^m x_{ij} \beta_j &= \sum_{j=1}^m s_j \beta_j. \end{aligned}$$

Таким образом, уравнение (16) принимает вид:

$$\frac{e^{\sum_{i=1}^n R_i \theta_i - \sum_{j=1}^m s_j \beta_j}}{\prod_{i=1}^n \prod_{j=1}^m (1 + e^{\theta_i - \beta_j})} = e^{\sum_{i=1}^n R_i \theta_i}.$$

9
Айвазян С.А.,
Мхитарян В.С.
Прикладная статистика
и основы эконометрики.
М.: ЮНИТИ, 1998.

$$\frac{1}{e^{\sum_{j=1}^m s_j \beta_j}} \cdot \frac{1}{\prod_{i=1}^n \prod_{j=1}^m (1 + e^{\theta_i - \beta_j})}. \quad (17)$$

Нетрудно заметить, что в выражение (17) не входят исходные данные тестирования x_{ij} , т.е. они не нужны для нахождения параметров модели. Данные x_{ij} преобразованы в исходный балл испытуемого (который вошёл в первый сомножитель правой части уравнения) и в исходный балл испытуемого (второй сомножитель правой части). Из уравнения (17) следует, что уровень подготовленности любого испытуемого полностью определяется суммой набранных им баллов, т.е. исходный балл — достаточная статистика для уровня подготовленности. Аналогично, для нахождения уровня трудности задания из всего объёма исходных данных требуется только сумма результатов выполнения этого задания всеми тестируемыми, т.е. сумма правильных ответов по заданию является достаточной статистикой для уровня трудности.

Заметим, что достаточные статистики модели Раша существуют только для решения по методу максимального правдоподобия. При выборе других критериев оптимальности параметров модели Раша достаточные статистики отсутствуют. Например, при подборе параметров модели Раша по методу на-

именных квадратов критерий подбора параметров принимает вид¹⁰:

$$\sum_{i=1}^n \sum_{j=1}^m \left(x_{ij} - \frac{e^{\theta_i - \beta_j}}{1 + e^{\theta_i - \beta_j}} \right)^2 \rightarrow \min. \quad (18)$$

В этом случае для подбора параметров нужны все значения исходных данных тестирования x_{ij} , которые нельзя заменить какими-либо достаточными статистиками. Аналогично можно доказать отсутствие достаточных статистик для уровней подготовленности испытуемых и уровней трудности заданий при педагогических измерениях по двух- и трёхпараметрическим моделям.

Преимущества достаточных статистик

Первое и самое очевидное — упрощение расчётов при нахождении уровней подготовленности испытуемых и уровней трудности заданий. Исходная матрица (таблица) данных тестирования x_{ij} размером $n \cdot m$ полноценно заменяется значениями сумм по каждой строке и каждому столбцу, количество которых равно $n + m$. Например, при обработке результатов выполнения теста из 50 заданий тысячей студентов 50000 ($n \cdot m = 50 \cdot 1000 = 50000$) результатов решения тестовых заданий можно заменить 1050 суммами

Деменчёнок О.Г.
Компьютерная программа для подбора параметров основных моделей Item Response Theory // Педагогические измерения. 2008. № 2. С. 28–42.

ПЕД
измерения

($n + m = 50 + 1000 = 1050$), т.е. объём данных может быть уменьшен в 47,6 раза. Однако вручную такую обработку выполнить нереально, а современная компьютерная техника способна обрабатывать данные без такого упрощения. Поэтому первое преимущество нельзя назвать весомым.

Второе преимущество — полное соответствие уровней подготовленности исходным баллам испытуемых, а уровней трудности — числу правильных ответов на задания. Это соответствие нелинейное, но сохраняющее упорядоченность, то есть большему исходному баллу будет соответствовать больший балл тестовый. Уровни подготовленности всех участников, набравших одинаковые исходные, будут совпадать, т.к. являются функциями равных достаточных статистик.

Другими словами, все испытуемые, решившие правильно одинаковое количество тестовых заданий (безразлично каких), получат одинаковую оценку подготовленности. Благодаря этому оценивание становится относительно прозрачным, поскольку испытуемый с большим исходным баллом может быть уверен в получении более высокого балла по уровню своей подготовленности на латентной шкале уровня подготовленности.

Модель Раша и закон Терстоуна

Одним из первых начал применять математические методы в психологии и социологии американский учёный Луис Леон Терстоун. Поиски «объективного в субъективном» привели Терстоуна к открытию в 1927 году т.н. закона (уравнения) сравнительного суждения, позволяющего при определённых допущениях сравнивать интенсивности не только количественных, но и качественных параметров. Это уравнение применяется в социальной психологии для оценки суждений при анализе общественного мнения, количественном изучении этноцентризма, психологии потребительских групп и т.д.

Закон сравнительного суждения Терстоуна на русском языке можно сформулировать следующим образом: искомые шкальные значения каких-либо двух объектов (т.е. их субъективные веса) тем далее отстоят друг от друга, чем чаще респондент предпочитает один объект другому (отмечает, что один объект «тяжелее» другого) при многократном предъявлении ему соответствующей пары объектов¹¹.

Аналитически закон Терстоуна представлен в виде зависимости¹²:

11
Толстова Ю.Н.
Измерение в социологии: Учебное пособие.
М.: Инфра-М, 2009.
288 с.

12
Thurstone L.L.
A law of comparative judgment
// Psychological Review.
Vol 34(4). Jul 1927.
273–286 pp.

$$S_i - S_j = z_{ij} \sqrt{\sigma_i^2 + \sigma_j^2 - 2r_{ij}\sigma_i\sigma_j}, \quad (19)$$

где S_i и S_j — значения i -го и j -го параметра на шкале измерений; σ_i^2 и σ_j^2 — дисперсии i -го и j -го параметра; z_{ij} — частота, с которой зарегистрированные значения S_i оказались больше значений S_j , выраженная единицами нормированного стандартного отклонения (см. формулу 20); r — коэффициент корреляции между i -м и j -м параметрами.

Замена реальных значений единицами нормированного стандартного отклонения z используется для приведения эмпирического распределения к стандартному нормальному распределению, для которого среднее значение случайной величины равно нулю, а дисперсия и стандартное отклонение — единице¹³. Замена производится по формуле:

$$z = \frac{x - \bar{x}}{\sigma}, \quad (20)$$

где x — значение случайной величины; \bar{x} — среднее значение; σ — стандартное отклонение.

С точки зрения психометрической теории, закон сравнительного суждения Терстоуна определяет модель измерения. Ввиду высокой сложности практических измерений на основе исходной формы закона (19) Терстоун выделил несколько частных случаев (вариантов) закона сравнительного суждения.

Первый случай — это сама исходная общая форма закона (19). Второй случай предусматривает изменение экспериментальной методики, обращаясь от оценок, производимых одним испытуемым, к групповым оценкам. Каждый испытуемый в этом случае производит только одно сравнение. Третий, четвертый и пятый частные случаи основаны на дополнительных допущениях¹⁴, которые изменяют аналитическое выражение закона Терстоуна.

Так, III случай закона Терстоуна справедлив при отсутствии корреляции ($r_i = 0$), что приводит общую формулу (19) к следующему виду:

$$S_i - S_j = z_{ij} \sqrt{\sigma_i^2 + \sigma_j^2}.$$

IV случай закона Терстоуна основывается на допущении, что $r_i = 0$ и что дисперсии мало отличаются друг от друга, т.е. $\sigma_i = \sigma_j + d$, где d мало по сравнению с σ_i . Тогда выражение (19) преобразуется:

$$S_i - S_j = z_{ij} \sqrt{(\sigma_j + d)^2 + \sigma_j^2}.$$

Раскрывая скобки и делая ряд преобразований и упрощений, получаем окончательное выражение IV случая закона:

$$S_i - S_j = z_{ij} k (\sigma_i + \sigma_j),$$

где k — постоянный множитель.

Наиболее известен V случай закона Терстоуна, в котором значения параметров считаются независимыми (т.е. коэффициент

Методология

13

Кибзун А.И. и др.
Теория вероятностей и математическая статистика. Базовый курс с примерами и задачами. М.: Физматлит, 2002. 224 с.

14

*Гусев А.Н.,
Измайлов Ч.А.,
Михалевская М.Б.*
Измерение в психологии: общий психологический практикум. 2-е изд. М.: Смысл, 1998. 286 с.

ПЕД
измерения

15
Maydeu-Olivares A.
Thurstone's Case V
Model: a Structural
Equations Modeling
Perspective. //Recent
Developments on
Structural Equation
Models. Series:
Mathematical Modelling:
Theory and Applications,
Vol. 19, 2004. 360 p.

16
Bradley R.A., Terry M.E.
Rank analysis of incom-
plete block designs. I. the
method of paired compar-
isons. //Biometrika,
1952. № 39. 324–345 p.

17
Luce R.D.
Individual Choice
Behaviours:
A Theoretical Analysis.
New York: J. Wiley, 1959.

18
Andrich D.
Relationships Between the
Thurstone and Rasch
Approaches to Item
Scaling // Applied
Psychological
Measurement, 1978. № 2.
451–462 p.

19
Birnbaum A.
Some Latent Trait Models
and Their Use in Inferring
an Examinee's Ability / In:
F.M. Lord and M.R.
Novick. Statistical Theories
of Mental Test Scores.
Reading, Mass: Addison-

корреляции равен нулю), а дисперсии — равными:

$$z_{ij} = \frac{S_i - S_j}{\sqrt{\sigma_i^2 + \sigma_j^2}} = \frac{S_i - S_j}{\sigma},$$

где

$$\sigma = \sqrt{\sigma_i^2 + \sigma_j^2} = \sqrt{2\sigma_i^2} = \sqrt{2} \cdot \sigma_i. \quad (21)$$

При этом вероятность того, что S_i больше S_j , может быть получена непосредственно из частоты случаев, в которых j -й параметр оценен больше, чем i -й, в предположении, что значения разности $S_i - S_j$ распределены согласно функции нормального распределения¹⁵:

$$P(S_i > S_j) = \frac{1}{2\pi} \int_{-\infty}^z e^{-\frac{z^2}{2}} dz. \quad (22)$$

Как известно, значения интегральной функции нормированного нормального распределения с точностью до сотых совпадают со значениями логистической функции при изменении масштаба аргумента с коэффициентом 1,7:

$$\left| \frac{e^{1,7z}}{1 + e^{1,7z}} - \frac{1}{2\pi} \int_{-\infty}^z e^{-\frac{z^2}{2}} dz \right| \leq 0,01. \quad (23)$$

Тогда уравнение (22) можно переписать следующим образом:

$$P(S_i > S_j) \approx \frac{e^{1,7z}}{1 + e^{1,7z}} = \frac{e^{1,7(S_i - S_j)/\sigma}}{1 + e^{1,7(S_i - S_j)/\sigma}} = \frac{e^{c(S_i - S_j)}}{1 + e^{c(S_i - S_j)}}, \quad (24)$$

где $c = 1,7/\sigma$ — константа, не зависящая ни от объектов, ни от субъектов измерения.

Изменив масштаб шкалы измерения с коэффициентом пропорциональности c , можно привести (24) к виду:

$$P(S_i > S_j) = \frac{e^{S_i - S_j}}{1 + e^{S_i - S_j}}. \quad (25)$$

Выражение (25) определяет вероятность того, что параметр S_i больше параметра S_j . Это уравнение известной в психологии и социологии модели Бредли–Терри–Люса (Bradley–Terry–Luce)^{16,17}, которая применяется для выявления предпочтений путём попарного сравнения.

Применительно к педагогическим измерениям, считая параметрами S_i и S_j уровень подготовленности испытуемого θ и уровень трудности задания β , получим уравнение модели Раша¹⁸:

$$P = \frac{e^{\theta - \beta}}{1 + e^{\theta - \beta}}.$$

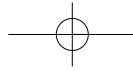
Таким образом, модель Раша соответствует V случаю закона Терстоуна. Следует отметить, что этому случаю соответствует целый класс моделей вида:

$$P = \frac{e^{c(\theta - \beta)}}{1 + e^{c(\theta - \beta)}}, \quad (26)$$

где c — константа, не зависящая от тестовых заданий и испытуемых (для модели Раша $c = 1$).

Уравнение (26) очень напоминает двухпараметрическую модель, предложенную А. Бирнбаумом¹⁹:

$$P = \frac{e^{a(\theta - \beta)}}{1 + e^{a(\theta - \beta)}}, \quad (27)$$



где a — второй параметр модели, называемый различающей способностью тестового задания (чем больше значение различающей способности a , тем ближе к вертикали центральная часть графика задания).

Однако между уравнением (26) и двухпараметрической моделью есть принципиальное различие: хотя константа c постоянна, каждое задание характеризуется собственным значением различающей способности (т.е. $a_j \neq \text{const}$). Двухпараметрическая модель совпадёт с уравнением (26) и соответствует V случаю закона Терстоуна толь-

ко при одинаковой разрешающей способности всех заданий ($a_j = \text{const}$).

Графически это можно проиллюстрировать графиками, представленными на рис. 3.

На графиках показана зависимость вероятности правильного ответа от уровня подготовленности испытуемого для пяти тестовых заданий с равномерно возрастающим уровнем трудности $b_1 = -2, b_2 = -1, b_3 = 0, b_4 = 1$ и $b_5 = 2$. Три первые диаграммы (рис. 3, а–в) соответствуют уравнению (26) при различных значениях константы: $c = 1$ (т.е. модель Раша, рис. 3, а), $c = 2$

Методология

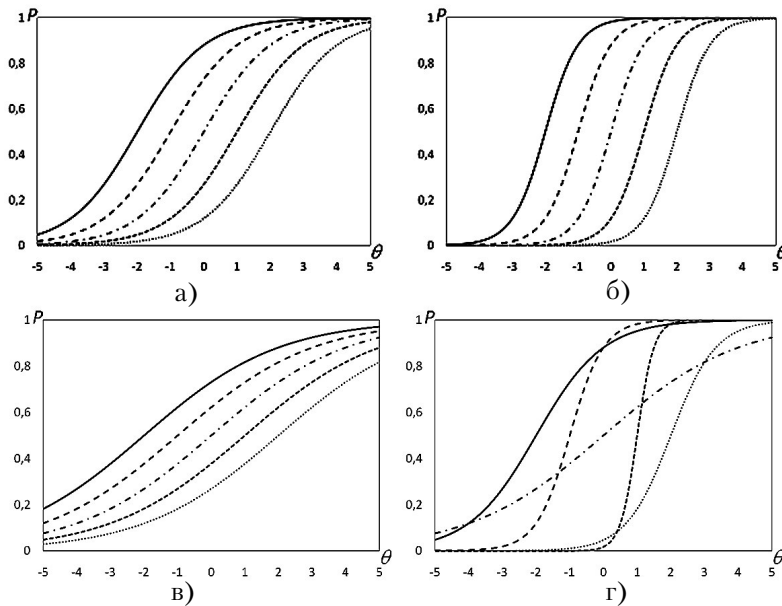
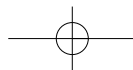


Рис. 3. Зависимость вероятности правильного ответа от уровня подготовленности испытуемого (— задание с уровнем трудности $\beta = -2$; ---- задание с $\beta = -1$; - · - · - задание с $\beta = 0$; - - - - - задание с $\beta = 1$; ····· задание с $\beta = 2$)



(рис. 3, б) и $c = 0,5$ (рис. 3, в). Легко заметить, что графики заданий на этих рисунках не пересекаются.

На рис. 4 представлены результаты расчёта по двухпараметрической модели при значениях различающей способности $a_1 = 1$, $a_2 = 2$, $a_3 = 0,5$, $a_4 = 4$ и $a_5 = 1,5$ (рис. 3, в). На этом рисунке графики заданий пересекаются, что является наглядным свидетельством несоответствия двухпараметрической модели V частному случаю закона Терстоуна. Нетрудно показать, что соответствие недостижимо и для трёхпараметрической модели. Указанное обстоятельство выделяет модель Раша из ряда моделей педагогических измерений, что многие теоретики считают важным преимуществом^{20,21}.

Однако, на взгляд автора, это преимущество не стоит переоценивать. Вспомним условия, при которых достигается соответствие модели Раша и закона Терстоуна:

- независимость измеряемых параметров;
- равенство дисперсий измеряемых параметров;
- распределение разности параметров по нормальному закону распределения.

Первое условие — независимость измеряемых параметров — стандартное для педагогических измерений предположение, и его справедливость ос-

паривать не будем. Выполнение третьего условия — распределение разностей измеряемых параметров по нормальному закону распределения — также не очевидно.

Представляется весьма сомнительным выполнение второго условия. Истинные значения дисперсии уровней подготовленности D_{ui} и уровней трудности заданий D_{bj} неизвестны, но можно оценить предельные значения этих величин²²:

$$D_{\theta_i} = \sigma_{\theta_i}^2 = \frac{1}{\left(\sqrt{\sum_{j=1}^m P_{ij}(1-P_{ij})}\right)^2} =$$

$$= \frac{1}{\sum_{j=1}^m \left[\frac{e^{\theta_i - \beta_j}}{1 + e^{\theta_i - \beta_j}} \left(1 - \frac{e^{\theta_i - \beta_j}}{1 + e^{\theta_i - \beta_j}}\right) \right]} =$$

$$= \frac{1}{\sum_{j=1}^m \frac{e^{\theta_i - \beta_j}}{(1 + e^{\theta_i - \beta_j})^2}},$$

$$D_{\beta_j} = \sigma_{\beta_j}^2 = \frac{1}{\sum_{i=1}^n P_{ij}(1-P_{ij})} =$$

$$= \frac{1}{\sum_{i=1}^n \frac{e^{\theta_i - \beta_j}}{(1 + e^{\theta_i - \beta_j})^2}}.$$

Для проверки предположения о равенстве дисперсий проведены пробные расчёты. Использовалась компьютерная программа Estimate3PL (сайт www.asksystem.narod.ru), исход-

20

Partchev I.

A visual guide to item response theory. — Jena: Friedrich-Schiller-Universität, 2004. 61 p.

21

Аванесов В.С.

Применение тестовых форм в Rasch Measurement // Педагогические измерения. № 4. 2006.

22

Демещёнок О.Г.

Проблема анализа погрешности педагогических измерений. // Педагогические измерения. № 1, 2009.

ные данные — матрица результатов тестирования, опубликованная в известной работе Б.Д. Райта и М.Х. Стоуна²³ (эта матрица заложена в программу Estimate3PL в качестве примера). Результаты расчётов приведены в табл. 1. Оценки дисперсии уровней подготовленности D_{θ_i} составили от 0,68 до 1,33; дисперсий уровней трудности заданий D_{β_j} — от 0,21 до 1,19. Максимальная оценка дисперсии превышает минимальную более чем в шесть раз, ввиду чего равенство истинных значений дисперсий всех измеряемых параметров представляется маловероятным.

Таким образом, в практике педагогических измерений проблематично обеспечить соблюдение обязательных условий соответствия модели Раша V частному случаю закона Терстоуна, что ставит соответствие под сомнение.

По мнению автора этой статьи, взаимосвязь модели Раша и закона Терстоуна, интересная с теоретической точки зрения, особого практического значения не имеет. Модель Раша самодостаточна и не нуждается в дополнительном теоретическом обосновании в виде закона Терстоуна. На технологию педагогических измерений наличие или отсутствие соответствия модели Раша V частному случаю закона Терстоуна какого-либо влияния не оказывает. По-

этому целесообразно сосредоточиться на более важных вопросах, таких, как изучение причин недостаточной точности и надёжности результатов педагогических измерений.

Rasch Measurement and Item Response Theory

Модель Раша — это самая простая математическая модель педагогических измерений. Наряду с многочисленными успехами Rasch Measurement, к сожалению, известны случаи, когда результаты педагогического измерения оказывались непригодны для практического использования ввиду недостаточной точности и надёжности результатов. Такие случаи заставляют задуматься над извечными вопросами: кто виноват и что делать.

Ж.М. Линакре в своей статье «Модель Раша не может быть опровергнута!»²⁴ всю вину возлагает на неподходящие для модели Раша исходные данные. Если модель Раша не работает для реальных данных, то причина в тех искажениях, которые содержатся в этих данных. Действительно, искажение исходных баллов тестирования может быть обусловлено следующими факторами:

- несанкционированный доступ испытуемых к правильным

Методология

²³ Wright B.D., Stone M.H. Best Test Design. Chicago: Mesa Press. 1979.

²⁴ Linacre J.M. The Rasch Model cannot be «Disproved»! // Rasch Measurement Transactions, 1996. 10:3. pp. 512–514.

ПЕД
измерения

Таблица 1

Результаты выполнения теста

	Тестовые задания														D_{θ}
	4	5	7	6	9	8	10	11	13	12	14	15	16	17	
25	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0,78
4	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0,68
33	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0,68
1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0,69
27	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0,69
11	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0,79
12	1	1	1	0	1	0	1	0	0	0	0	0	0	0	0,79
17	1	0	1	1	1	1	1	0	0	0	0	0	0	0	1,02
19	1	1	1	1	1	1	0	0	0	0	0	0	0	0	1,02
30	1	1	1	1	1	1	0	0	0	0	0	0	0	0	1,02
2	1	1	1	1	1	1	1	0	0	0	0	0	0	0	1,33
3	1	1	1	1	1	1	1	0	0	1	0	0	0	0	1,33
5	1	1	1	1	1	1	1	0	0	0	0	0	0	0	1,33
6	1	1	1	1	1	1	1	0	0	0	0	0	0	0	1,33
8	1	1	1	1	1	1	1	0	0	0	0	0	0	0	1,33
9	1	1	1	1	1	1	1	0	0	0	0	0	0	0	1,33
13	1	1	0	0	1	1	1	1	0	1	0	0	0	0	1,33
16	1	1	1	1	1	1	0	1	0	0	0	0	0	0	1,33
26	1	1	1	1	1	1	1	0	0	0	0	0	0	0	1,33
28	1	1	1	1	1	1	0	1	0	0	0	0	0	0	1,33
29	1	1	0	1	1	0	1	1	0	0	1	0	0	0	1,33
31	1	1	1	1	1	1	1	0	0	0	0	0	0	0	1,33
10	1	1	1	1	1	1	1	1	0	0	0	0	0	0	1,22
18	1	1	1	1	1	1	1	0	1	0	0	0	0	0	1,22
14	1	1	1	1	1	1	1	1	0	0	0	0	0	0	1,22
32	1	1	1	1	1	1	1	1	0	0	0	0	0	0	1,22
20	1	1	1	1	1	1	1	0	1	0	0	0	0	0	1,22
21	1	1	1	1	1	1	1	1	1	0	0	0	0	0	1,04
22	1	1	1	1	1	1	1	1	0	1	0	0	0	0	1,04
23	1	1	1	1	1	1	1	0	1	0	1	0	0	0	1,04
34	1	1	1	1	1	1	1	0	0	1	1	0	0	0	0,95
15	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0,95
7	1	1	1	1	1	1	1	1	1	1	0	1	0	0	0,92
24	1	1	1	1	1	1	1	1	1	1	0	0	0	1	0,92
D_{β}	0,68	0,51	0,43	0,51	0,30	0,43	0,24	0,21	0,33	0,30	0,52	1,19	1,19	1,19	

ответам (списывание, использование запрещённых справочных материалов, подкуп должностных лиц и т.д.);

- случайное угадывание ответа;
- ошибки при вводе ответа;
- некорректность тестовых заданий;
- фрагментарность знаний испытуемого (модель предполагает одинаковый уровень знаний по всему кругу проверяемых тестом знаний) и др.

Влияние указанных факторов приводит к нарушению заложенных в модель теоретических предположений. Вследствие этого модель не может удовлетворительно объяснить исходные данные, и результаты педагогического измерения становятся непригодными для практического использования. По мнению Linacre J.M., исправлять нужно не модель Раша, а исходные данные. Высокоточные косвенные измерения латентных параметров по подвергшихся неизвестным искажениям данным — это задача, непосильная для любой математической модели. Поэтому целесообразность устранения (минимизации) искажений исходных данных тестирования не подлежит сомнению.

С другой стороны, теория математического моделирования в случае неудачи рекомендует улучшение и уточнение модели²⁵. Отказ от некоторых допущений и упрощений, вклю-

чение в модель ранее неучитываемых факторов, способны повысить адекватность модели, сблизить расчётные и экспериментальные данные. В свете этого, предпринятые в IRT шаги по улучшению модели Раша путём введения дополнительных параметров не противоречат общей теории и потенциально могут привести к положительным результатам. Так, например, кроме известных двух- и трёхпараметрической моделей существует также пятипараметрическая модель тестирования²⁶:

$$P_{ij} = c_i c_j + (1 - c_i c_j) \cdot \left(1 + \exp \left(\frac{\beta_j - \theta_i}{\sqrt{\sigma_i^2 + \sigma_j^2}} \right) \right)^{-1}, \quad (28)$$

где c_i — вероятность того, что на i -й испытуемый прибегнет к угадыванию правильного ответа; c_j — вероятность угадывания правильного ответа на j -е задание; σ_i и σ_j — стандартные отклонения θ_i и β_j .

Теоретический предел количества параметров моделей определяется числом степеней свободы, которое обычно равно уменьшенному на единицу числу экспериментальных данных.

На взгляд автора, между двумя этими подходами, несмотря на внешнюю противоречивость, нет антагонизма. Никакая модель не в состоянии корректно устранить все искаже-

Методология

25

Самарский А.А.,
Михайлов А.П.
Математическое моделирование: Идеи. Методы. Примеры. 2-е изд.
М.: Физматлит, 2001.
320 с.

26

Кроммер В.В.
О многопараметрической оценке уровней подготовленности испытуемых и трудности заданий // Педагогические измерения. № 3. 2005. С. 65–72.

ния исходных данных. Поэтому актуальность задачи получения пригодных исходных данных высока и не зависит от выбранной математической модели.

Теперь рассмотрим целесообразность улучшения модели Раша. Усложнённая модель более гибко адаптируется к исходным данным, расчётные данные становятся ближе к экспериментальным.

Однако не будем забывать о косвенном характере измерений, погрешность которого равна квадратичной сумме произведений частных производных и погрешностей измерения каждой переменной²⁷:

$$\Delta y = \sqrt{\left(\frac{\partial y}{\partial x_1} \Delta x_1\right)^2 + \left(\frac{\partial y}{\partial x_2} \Delta x_2\right)^2 + \dots + \left(\frac{\partial y}{\partial x_n} \Delta x_n\right)^2}, \quad (29)$$

где Δy — погрешность измерения величины y , рассчитываемой по значениям переменных величин $x_1, x_2, \dots, x_i, \dots, x_n$;

$\frac{\partial y}{\partial x_i}$ — частная производная функции

$y(x_1, x_2, \dots, x_i, \dots, x_n)$ по переменной x_i ;

Δx_i — погрешность измерения переменной x_i .

Это означает, что каждая дополнительная переменная вводит дополнительное слагаемое в уравнение погрешности (28). Конечно, изменятся и

значения других слагаемых. Поэтому усложнение модели может привести как к уменьшению, так и к увеличению погрешности измерений. Следовательно, усложнение модели оправдано только тогда, когда дополнительная переменная учитывает статистически значимый фактор. Такое изменение позволяет существенно сблизить расчётные и исходные данные, снизить погрешность измерения.

Напротив, усложнение математической модели путём введения малозначимого фактора при несущественном сближении модели и экспериментальных данных может привести к значительному росту погрешности измерения. Такие модели фактически бесполезны. Зачастую более простая модель позволяет лучше и надёжнее исследовать реальную систему, чем более сложная (и, формально, «более правильная»).

Таким образом, из всех моделей, обеспечивающих достаточную точность и надёжность измерения, следует выбрать наиболее простую. Введение в модель дополнительных параметров оправдано только тогда, когда адекватность простой модели неудовлетворительна.

Проведённый анализ особенностей Rasch Measurement приводит к следующим выводам.

1. Отличительная черта модели Раша — математически доказанная независимость (инвариантность) средств и субъектов измерения (тестовых заданий и испытуемых).

2. Существование достаточных статистик — ещё одна отличительная черта модели Раша. Это свойство проявляется только при решении задачи педагогического измерения методом максимального правдоподобия. Основным достоинством достаточных статистик полное соответствие уровней подготовленности первичным баллам испытуемых, а уровней трудности — первичным баллам заданий. Вследствие этого все правильно решившие одинаковое количество тестовых заданий получают одинаковую оценку подготовленности, а испытуемый с большим первичным баллом может быть уверен в более высокой оценке подготовленности.

3. Модель Раша полностью совпадает с V частным случаем закона Терстоуна при выполнении следующих условий:

- независимость измеряемых параметров — уровней подготовленности испытуемых и

уровней трудности тестовых заданий;

- равенство дисперсий измеряемых параметров;

- распределение разности параметров по нормальному закону распределения.

4. В частном случае закона Терстоуна также соответствует двухпараметрическая модель при равных значениях различающей способности всех тестовых заданий. В этом случае графики заданий теста не пересекаются.

5. Необходимым условием корректности педагогического измерения служит устранение (минимизация) искажений исходных данных, вызываемых несанкционированным доступом испытуемых к правильным ответам, случайным угадыванием ответа, некорректностью тестовых заданий, ошибками при вводе ответа и другими факторами.

6. Из всех моделей, обеспечивающих достаточную точность и надёжность измерения, следует выбрать наиболее простую. Выбор более сложной модели оправдан только тогда, когда адекватность простой модели не удовлетворительна.

Методология

Методология