

Методология

ПЕДАГОГИЧЕСКИЕ ИЗМЕРЕНИЯ В ШКАЛЕ ОТНОШЕНИЙ

Юрий Каргин, Ноябрьский колледж профессиональных
и информационных технологий
kargin04@yandex.ru

В данной работе обосновываются возможность и некоторые преимущества проведения педагогических измерений в шкале отношений, что может оказаться предпочтительней системы измерений Г. Раша (RM) в изложении Wright B.D., Stone M.H.¹

Ключевые слова: педагогические измерения, измерительная шкала отношений, система Раша.

Математические модели педагогических измерений

Идею построения системы совместных социальных измерений на основе аксиоматизации процесса взаимодействия (наложения) исследуемых свойств связывают чаще других с работой Luce R.D. & Tukey J.W.² Содержание этой идеи применительно к педагогическим измерениям состоит в математической аксиоматизации: а) правил взаимодействия абстрактных мер подготовленности испытуемых и трудности тестовых заданий; б) отклика на результат такого взаимодействия.

В системе Раша аксиома взаимодействия состоит в том, что латентные показатели уровней подготовленности испытуемого b и трудности тестового задания d входят в систему измерения аддитивно, в виде параметра $g = b - d$. Это условие предполагает введение единой интервальной шкалы с допустимыми правилами преобразования масштабов: сдвиг и деформацию. В частности, справедливо инвариантное преобразование — одновременное изменение значений b и d на произвольную величину не изменит значение параметра взаимодействия g , а значит, и отклик на результат этого взаимодействия. Отклик на результат взаимодействия искомых величин b и d определяется однопараме-

1

Wright B.D., Stone M.H.
Measurement Essentials.
WIDE RANGE, INC.
Wilmington, Delaware,
1999. 221 p.

2

Luce R.D., Tukey J.W.
Simultaneous conjoint
measurement: A new
type of fundamental
measurement. *Journal of
Mathematical Psychology*,
1, p. 1–27.

трической математической модели Раша, задающей вероятность правильного ответа испытуемым на тестовое задание в зависимости от параметра взаимодействия g (здесь и далее символом «R» обозначаем фрагменты, относящиеся к системе измерений Раша, символ «A» — к альтернативной системе измерений в шкале отношений):

$$P = \frac{\exp g}{1 + \exp g}. \quad (R1)$$

Эмпирическое обоснование приведённых аксиом должно подкрепляться и определёнными теоретическими принципами измерений. Основной из них состоит в условии независимости результатов измерений. В частности, расстояние между измеренными уровнями подготовленности испытуемых A и B не должно зависеть от уровня трудности тестового задания. Для модели Раша последнее требование приводит к равенству:

$$b_A - b_B = \ln \left(\frac{P_A}{(1 - P_A)} \times \frac{(1 - P_B)}{P_B} \right). \quad (R2)$$

Здесь выражение в скобках под знаком логарифма задаёт отношение уровней подготовленности испытуемых.

Из уравнения (R2) следует, что выражение в правой части не только не зависит от уровня трудности тестового задания d , но оно должно и адекватно отражать различия в уровнях подготовленности испытуемых. Только в этом случае и именно в этом смысле предложенная модель идентифицирует и располагает значения

оцениваемых показателей на измерительной шкале.

В альтернативной системе латентные показатели относительных уровней подготовленности испытуемого θ и трудности тестового задания β накладываются отношением $\gamma = \theta/\beta$, а отклик задаётся следующим математическим законом для вероятности правильного ответа:

$$P = 0,5^{1/\gamma}. \quad (A1)$$

Здесь, как и в модели Раша (R1), рост подготовленности испытуемых приводит к увеличению значений параметров взаимодействия g и γ и далее — к монотонному росту вероятности правильного ответа испытуемым. В этой измерительной системе предполагается введение единой измерительной шкалы отношений с допустимой деформацией масштаба и нулевой точкой начала отсчёта, а также допустимы следующие инвариантные преобразования — одновременное изменение значений показателей θ и β в определённое число раз не изменяет значения параметра взаимодействия γ .

Требование независимости результатов измерения уровней подготовленности двух испытуемых от трудности тестового задания имеет вид:

$$\frac{\theta_A}{\theta_B} = \frac{1/\log P_A}{1/\log P_B} = \frac{\log P_B}{\log P_A}. \quad (A2)$$

Здесь мы не указали основание логарифма, в данном уравнении оно может быть любым (из области определения).

В отличие от системы Раша, где разность (расстояние) между значениями показателей определяется числом логитов и сравнивать

испытуемых можно в терминах типа «испытуемый А подготовлен на два логита лучше испытуемого В», в альтернативной системе при сравнении значений показателей модели корректно выражаться в относительных терминах типа «испытуемый А в два раза лучше подготовлен к тесту, чем испытуемый В». Выражение в правой части уравнения (A2) также не только не зависит от уровня трудности тестового задания β , но и должно адекватно характеризовать различия в уровнях подготовленности испытуемых.

Сравнения в измерительных системах

Рассмотрим, как в рассматриваемых измерительных системах различаются испытуемые. Начнём с системы Раша. Используем пример¹ с прыгунами, комментирующий эти различия.

Пусть два прыгуна А и В преодолевают несколько раз определённую высоту. Моделируем количества удачных и неудачных прыжков в трёх ситуациях. Если в серии прыжков оба прыгуна показали одинаковый результат, т.е. одновременно взяли или не взяли высоту, то такие серии исключаются из рассмотрения. Такие числа «бесполезны для целей сравнения». Обозначим далее через N_{10} — количество серий, в которых прыгун А взял высоту, а прыгун В — нет, и N_{01} — наоборот. Количества таких чистых побед N_{10} и N_{01} являются «информативными для целей сравнения» в системе Раша, а их отношение отражает степень этого различия

N_{10}/N_{01} . Результаты такого сравнения представлены в таблице 1.

Таблица 1

Сравнение в системе Раша

	Статистика	Ситуация		
		I	II	III
Кол-во чистых побед прыгуна А	N_{10}	9	90	54
Кол-во чистых побед прыгуна В	N_{01}	1	10	46
Отношения	N_{10}/N_{01}	9	9	≈

Переход к вероятностям успехов по формуле

$$\frac{N_{10}}{N_{01}} \approx \frac{P_A}{(1-P_A)} \cdot \frac{(1-P_B)}{P_B}$$

позволяет рассматривать результаты последней строки таблицы как оценку показателей сравнения схемы Раша (R2).

Приведём краткую характеристику моделируемых ситуаций. В первых двух ситуациях отношение количества чистых побед спортсмена А к количеству чистых побед спортсмена В одинаково и равно 9. В первой и третьей ситуациях разность количества чистых побед спортсмена А и количества чистых побед спортсмена В одинакова и равна 8. В первых двух ситуациях спортсмен А существенно лучше спортсмена В подготовлен к соревнованиям, в третьей ситуации различия в уровнях подготовленности спортсменов не так значимы.

Такая схема сравнения имеет свою логику, но и вызывает вопрос. По этой схеме результат сравнения подготовленности прыгунов определяется только отношением чистых побед и не зави-

¹ Маслак А.А. Измерение латентных переменных в социально-экономических системах: теория и практика: Монография. Славянск-на-Кубани: Издательский центр СГПИ, 2007. 424 с.

сит от доли чистых побед в серии испытаний. Насколько это правомерно? Приведём контраргумент примером с игрой в шахматы. Рассмотрим две ситуации сравнения шахматистов в очном поединке. Если шахматисты будут играть с компьютером, результаты сравнения не изменятся.

В первой ситуации из трёх партий спортсмен А одержал победу 2 раза и потерпел поражение 1 раз. По схеме сравнения Раша спортсмен А в два раза лучше подготовлен спортсмена В. Такие заключения вполне можно принять.

Во второй ситуации из 100 партий количество чистых побед не изменилось, а 97 раз спортсмены разошлись ничьей. По схеме сравнения Раша и в этой ситуации спортсмен А подготовлен в два раза лучше спортсмена В. Но по шахматным правилам, да и, пожалуй, в соответствии со здравым смыслом, ситуация другая. Спортсмены получают соответственно 50,5 и 49,5 баллов, и даже если победа присуждается спортсмену А, уровни их подготовки

во второй ситуации скорее очень близки, нежели существенно различаются.

Проведём сравнения тех же прыгунов по правилу, вытекающему из альтернативной системы (A2). Для этого зафиксируем количества чистых побед прыгунами и восстановим недостающие данные. Такое восстановление потребует информации об общем числе серий испытаний N . Проведём два сравнения: для минимально возможного количества серий $N = 200$ и для десятикратного увеличения этого количества, для $N = 2000$ серий. Обозначим через N_{11} количество серий с удачными прыжками спортсменов, через N_{00} — количество неудачных серий для обоих спортсменов. Корректное решение поставленной задачи о нахождении N_{11} и N_{00} приводит к двум вариантам решений. Используем вариант более сильной подготовки спортсменов. Именно он даёт возможность сопоставить рассматриваемые схемы. Результаты такого анализа представлены в таблицах 2 и 3.

Таблица 2

Сравнение в альтернативной системе ($N = 200$)

	Статистика	Ситуация		
		I	II	III
Кол-во побед прыгунов А и В	N_{11}	190	90	54
Кол-во неудач прыгунов А и В	N_{00}	0	10	46
Кол-во чистых побед прыгуна А	N_{10}	9	90	54
Кол-во чистых побед прыгуна В	N_{01}	1	10	46
Доля побед прыгуна А	$(N_{11} + N_{10})/N$	0,995	0,9	0,54
Доля побед прыгуна В	$(N_{11} + N_{01})/N$	0,955	0,5	0,5
Отношения	$\frac{\log[(N_{11} + N_{10})/N]}{\log[(N_{11} + N_{01})/N]}$	$\approx 9,2$	$\approx 6,6$	$\approx 1,12$

Таблица 3

Сравнение в альтернативной системе (N = 2000)

	Статистика	Ситуация		
		I	II	III
Кол-во побед прыгунов А и В	N_{11}	1990	1900	1899
Кол-во неудач прыгунов А и В	N_{00}	0	0	1
Кол-во чистых побед прыгуна А	N_{10}	9	90	54
Кол-во чистых побед прыгуна В	N_{01}	1	10	46
Доля побед прыгуна А	$(N_{11} + N_{10})/N$	0,9995	0,995	0,9765
Доля побед прыгуна В	$(N_{11} + N_{01})/N$	0,9955	0,955	0,9725
Отношения	$\frac{\log[(N_{11} + N_{01})/N]}{\log[(N_{11} + N_{10})/N]}$	≈ 9	≈ 9	$\approx 1,17$

Первое отличие альтернативной схемы от схемы Раша состоит в том, что здесь при сравнении спортивных достижений прыгунов учитываются доли всех успешных исходов (а не только чистых побед), или в пределе большого числа серий испытаний – вероятности успеха. Например, для спортсмена А:

$$\frac{N_{11} + N_{10}}{N} \approx P_A$$

Эти отличия сказываются и на результатах сравнения. Для достаточно большого числа серий (Таблица 3, N = 2000) результаты сравнения по альтернативной схеме совпадают с результатами схемы Раша. Отношения уровней подготовленности спортсменов не изменились. Для минимального числа серий (Таблица 2, N = 200) результаты несколько отличаются. Во второй ситуации спортсмен А сильнее спортсмена В примерно в 6,6 раз, в отличие от девятикратного преимущества по схеме Раша. Т.е., существуют как

области хорошего согласия схем сравнения, так и области, приводящие к различающимся оценкам.

Чтобы определить область согласия схем сравнения, достаточно сравнить характерные выражения:

$$C_R \approx \frac{P_A}{(1-P_A)} \cdot \frac{(1-P_B)}{P_B} \quad \text{и}$$

$$C_A = \frac{\ln P_B}{\ln P_A}$$

Совершенно очевидно, что при $P_A = P_B$ эти выражения совпадают. Другую область согласия можно определить разложением логарифмов. Для значений P_A и P_B , близких к единице, можно получить примерное равенство:

$$\frac{P_A}{(1-P_A)} \cdot \frac{(1-P_B)}{P_B} \approx \frac{\ln P_B}{\ln P_A}$$

Несложно убедиться в этих выводах простым сравнением выражений при $P_A \approx P_B$ и при значениях P_A и P_B , больших 0,8.

На рисунке 1 представлены графики выражений $C_R(P_A)$ и $C_A(P_A)$ при двух значениях параметра P_B . Эти рисунки наглядно демонстрируют результаты

расчёта отношений подготовленности двух испытуемых по схеме сравнения Раша (чёрная линия) и по альтернативной схеме сравнения (серая линия).

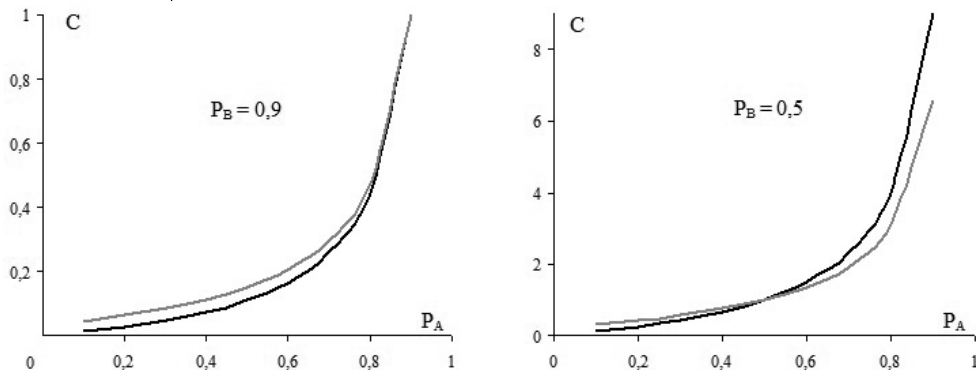


Рис. 1. Графики двух испытуемых

Анализ схем сравнения испытуемых в различных измерительных системах не только указывает на их качественное подобие, но и даёт некоторое преимущество альтернативной схемы, связанное с привычной логикой «здравого смысла» учёта общей доли удач.

Аксиомы измерительной системы Раша

Иногда заслуги Г. Раша представляют как удачный выбор вида функции (R1). Если учесть, что этот выбор следует из кропотливого перебора большого числа других статистических функций и массовым эмпирическим обоснованием именно этой закономерности, из анализа возможностей, достоинств и недостатков её применения, а под удачей понимать высокую продуктивность построенной на её основе

измерительной системы, то с этим мнением можно согласиться. Действительно, выбор логистической функции в качестве «преобразователя» наблюдаемых данных в значения латентных показателей, отложенных на интервальной шкале логитов, можно трактовать как аксиоматику измерительной системы Раша. И уже исходя из принятых положений, их следствий и результатов интерпретировать значения измеряемых показателей. Примерно в этой логике мы в предыдущем разделе выводили, что различия способностей прыгунов определяются отношением чистых побед.

Но можно идти по этой логике и в обратном порядке — от аксиоматизации различия способностей испытуемых к правилам преобразования наблюдаемых данных в измеряемые значения. Исходным для такого подхода яв-

ляется понятие шанса (odds) некоторого события как отношение числа факторов (интенсивности факторов), благоприятных этому событию, к числу неблагоприятных факторов. Если в качестве благоприятного события рассматривать правильный ответ на тестовое задание, то шанс правильного ответа h можно определить как отношение подготовленности испытуемого B к трудности тестового задания D : $h = \frac{B}{D}$.

$$h = \frac{B}{D}.$$

Пока совершенно непонятно содержание понятий, значение величин, обозначенных символами B и D , но именно это и следует определить и использовать как исходные положения для измерительной системы. Однако предварительно следует сделать два важных замечания. Первое. Математическая теория измерений, а именно в этом направлении проводятся данные исследования, не связана с психолого-педагогическими аспектами педагогического тестирования. И в определениях «подготовленности испытуемого» и «трудности тестового задания» не следует ждать психолого-педагогического содержания. И второе. Развёртывание этих понятий проводится не в абсолютных категориях, а в относительных единицах. Их величины определяются относительно себе подобных величин внутри самого теста.

На первый взгляд может показаться, что такой подход теряет объективность, независимость и является только внутренним свойством тестовой системы. Относительно внутреннего свойства

системы это действительно так. Но на этом этапе мы и не ставим цели определить подготовленность испытуемого или трудность задания в универсальных единицах. Мы пытаемся определить только взаимные отношения этих латентных показателей внутри системы¹.

Определение R1. Испытуемый A в k раз лучше испытуемого B подготовлен к тестовому заданию, т.е. $\frac{B_A}{B_B} = k$, если отношение шансов испытуемых правильно выполнить задание равно k :

$$\frac{h_A}{h_B} = k.$$

Из определения следует, что отношение подготовленности испытуемых не зависит от трудности тестового задания, действительно,

$$\frac{h_A}{h_B} = \frac{B_A / D}{B_B / D} = \frac{B_A}{B_B} = k.$$

Определение R2. Тестовое задание A в k раз труднее задания B , т.е. $\frac{D_A}{D_B} = k$, если отношение шансов правильно выполнить эти задания испытуемым равно $1/k$:

$$\frac{h_A}{h_B} = \frac{1}{k}.$$

Из определения следует, что отношение трудности заданий не зависит от подготовленности испытуемого, действительно,

$$\frac{h_A}{h_B} = \frac{B / D_A}{B / D_B} = \frac{D_B}{D_A} = \frac{1}{k} \quad \text{или}$$

$$\frac{D_A}{D_B} = k.$$

1
Эту мысль можно проиллюстрировать ситуацией из хорошо известного мультфильма, где размеры удава в «попугаях» заметно больше чем в «мартышках». Но даже в этих «эталонах», в этой измерительной системе, точнее, независимо от выбранного эталона, удав всегда и в определённое число раз длиннее слонёнка.

Следствие R1. Одновременное изменение трудности тестового задания и подготовленности испытуемого в k раз не изменит шанс (и вероятность) правильного выполнения этого задания. Действительно, обозначим $B_2 = k \cdot B_1$, $D_2 = k \cdot D_1$ тогда

$$h_2 = \frac{B_2}{D_2} = \frac{k \cdot B_1}{k \cdot D_1} = h_1.$$

Следствие R2. Увеличение трудности тестового задания в k раз эквивалентно уменьшению подготовленности испытуемого тоже в k раз. Доказательство аналогично.

Шанс правильного ответа h связан с вероятностью p отноше-

$$\text{нием: } h = \frac{p}{1-p} \text{ или } p = \frac{h}{1+h}.$$

Отсюда, вероятность правильного ответа испытуемым с подготовленностью B на тестовое задание труд-

$$\text{ности } D \text{ равно: } p = \frac{B/D}{1+B/D}.$$

Переходя к аддитивным переменным, $b = \ln B$ — уровень подготовленности испытуемого и $d = \ln D$ — уровень трудности тестового задания, переходим к известной математической модели Раша R1¹.

Аксиомы альтернативной измерительной системы

Прежде чем привести аналогичные определения для альтернативной системы измерений, приведём рассуждения, которые к ней привели.

Представим, что учебное задание состоит из некоторых элементарных учебных единиц, усвоение которых и проверяется тестовым

заданием. Определение этих единиц и степень их дискретизации в данном рассуждении не имеют значения. Важно предположить только то, что все они равнозначны с точки зрения вероятности правильного выполнения испытуемым. Пусть испытуемый правильно выполняет одну элементарную единицу с вероятностью u . Тогда задание, содержащее n таких единиц, испытуемый правильно выполнит с вероятностью u^n , а сам показатель степени можно рассматривать как характеристику трудности задания.

Поясним на примере. Требуется вычислить: $2 + 2 \times 2$. Для правильного решения задания испытуемый должен правильно выполнить три учебных действия: определить порядок арифметических действий, выполнить два арифметических действия. Ошибка хотя бы в одном из трёх действий приведёт и к ошибочному результату. Если вероятность правильного выполнения одного элемента равна u , то вероятность правильного выполнения задания из трёх элементов равна $u \times u \times u = u^3$. Т.е. трудность задания можно связать с числом включённых в него элементарных учебных единиц, и это число входит в показатель степени вероятности правильного выполнения одной такой единицы.

Аналогичные рассуждения можно привести и при сравнении уровня подготовленности испытуемых.

Такие представления о процессе тестирования, с одной стороны, ограничивают область применения разрабатываемой системы сферой педагогических из-

1

Формулы R1 и R2, выраженные через шанс правильного ответа, имеют

более простой вид:
 $h = \exp(b-d)$, $b_A - b_B = \ln(h_A/h_B)$.

мерений или других областей с подобной логикой выполнения тестовых заданий. С другой стороны, дают основания для более внятной интерпретации результатов тестирования и к возможности проведения прогностических расчётов на основе более глубокого профессионального анализа природы педагогического тестирования. Если использовать терминологию С.А. Айвазяна¹, то разработка альтернативного подхода основывается на «моделировании механизма явления», в отличие от «формальной статистической фотографии» в системе Раша.

Приведём исходные определения альтернативной системы измерений. Относительные уровни подготовленности испытуемого обозначаем через θ , трудности тестового задания — через β .

Определение А1. Испытуемый А в k раз лучше испытуемого В подготовлен к произвольному

тестовому заданию, т.е. $\frac{\theta_A}{\theta_B} = k$,

если испытуемый А выполнит k заданий с тем же успехом (с той же совместной вероятностью), с каким испытуемый В выполнит только одно задание.

Иными словами, если p_A и p_B вероятности правильного выполнения некоторого задания испытуемыми А и В, а испытуемый А в k раз подготовлен лучше испытуемого В, то $P_A^k = P_B$.

Определение А2. Тестовое задание А в k раз труднее задания В, т.е. $\frac{\beta_A}{\beta_B} = k$, если любой испытуемый выполнит k заданий трудно-

сти В с тем же успехом (с той же совместной вероятностью), с каким выполнит одно задание трудности А.

Иными словами, если p_A и p_B вероятности правильного выполнения некоторым испытуемым заданий А и В, а задание А в k раз труднее задания В, то $P_B^k = P_A$.

Вывод математической модели А1 из этих определений приведён в работе². Там же даются определения вероятностных показателей уровня подготовленности испытуемого $u = 0,5^{1/\theta}$ и уровня трудности тестового задания $v = 0,5^\beta$. В вероятностных переменных математическая модель принимает вид:

$$\log_{0,5} P = \log_{0,5} u \times \log_{0,5} v. \quad (A3)$$

Введение дополнительных вероятностных показателей u и v — не только удобный математический приём. Их введение снимает проблему единых единиц измерения латентных качеств приписывая им естественный в рамках математической теории вероятностный характер. В этом виде все основные переменные математической модели (А3) имеют единую вероятностную «меру», определённую на интервале (0; 1).

Математический закон (А3) также даёт удобную возможность интерпретировать значения показателей θ и β как коэффициенты эластичности (чувствительности). Действительно, дифференцируя это равенство по u или v , имеем:

$$\frac{dP}{P} = \beta \cdot \frac{du}{u} \quad (A4)$$

Отсюда, в частности, следует, что относительный рост вероят-

1

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

2

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

ностного уровня подготовленности испытуемого и на 1% увеличивает относительную вероятность правильного ответа на тестовое задание с уровнем трудности θ уменьшится на $1/\theta\%$.

В заключении подраздела не сложно привести очевидные следствия альтернативной модели.

Следствие А1. Одновременное изменение трудности тестового задания и подготовленности испытуемого в k раз не изменит вероятность правильного выполнения этого задания. Действительно, обозначим $\beta_2 = k \cdot \beta_1$, $\theta_2 = k \cdot \theta_1$, тогда

$$P_2 = 0,5^{\beta_2/\theta_2} = 0,5^{k\beta_1/k\theta_1} = 0,5^{\beta_1/\theta_1} = P_1.$$

Следствие А2. Увеличение трудности тестового задания в k раз эквивалентно уменьшению подготовленности испытуемого тоже в k раз. Действительно, обозначим $\beta_2 = k \cdot \beta_1$, тогда

$$P = 0,5^{\beta_2/\theta_1} = 0,5^{k\beta_1/\theta_1} = 0,5^{\beta_1/\theta_2},$$

$$\text{где } \theta_2 = \theta_1/k.$$

Графические иллюстрации математических моделей

Сопоставление измерительных систем можно провести и графически

русской иллюстрацией математических моделей (R1) и (A1). Именно эти математические законы играют ключевую роль в преобразовании наблюдаемых значений счёта количества правильных ответов на метрические шкалы той или иной измерительной системы. Приведём графики в интервальной шкале и шкале отношений — рисунки 2 и 3 соответственно, чёрной линией изображена логистическая кривая, серой линией изображена альтернативная кривая. Для корректного сопоставления графиков и демонстрации их в стандартизированном виде в качестве аргументов X выбраны следующие величины:

— для интервальной шкалы $X_R = 1,7 \cdot (b - d)$ и $X_A = \ln(\theta/\beta) = \ln\theta - \ln\beta$;

— для шкалы отношений $X_R = \exp(1,7 \cdot (b - d)) \sim V/D$ и $X_A = \theta/\beta$.

Коэффициент 1,7 в модели Раша приводит логистическую кривую (A1) к нормализованной функции нормального распределения (с нулевым средним значением и единичной дисперсией). В системе измерений Раша такие преобразования исходных данных осуществляются процедурой шкалирования.

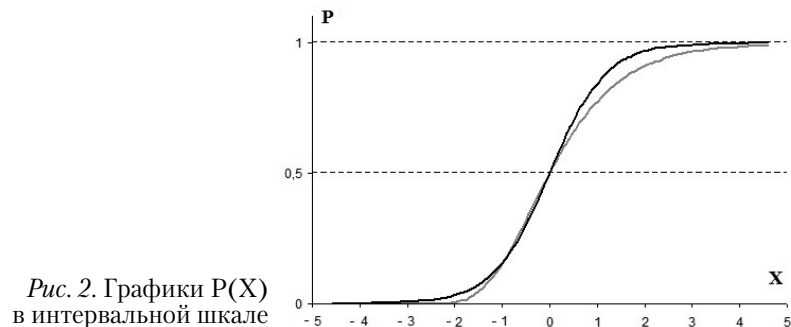


Рис. 2. Графики $P(X)$ в интервальной шкале

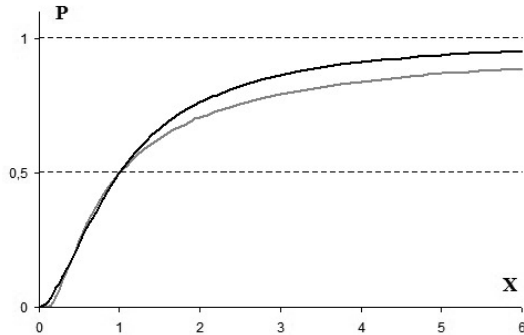


Рис. 3. Графики $P(X)$ в шкале отношений

Проведём краткий анализ графиков. Качественно поведение графиков эквивалентно. Графики всегда монотонны, имеют единые области определения, единые асимптоты. В нейтральной точке, когда подготовленности испытуемого совпадают с трудностью тестового задания, вероятность правильного ответа равна 0,5. К качественным различиям можно отнести симметричность логистической кривой относительно линии $X = 0$ в интервальной шкале и отсутствие подобной симметрии альтернативной кривой. Есть и некоторые количественные различия, наиболее заметные в области $X \approx 1$ логит. Оценим эти различия. Если уровень подготовленности испытуемого b больше уровня трудности тестового задания d на один логит (что соответствует отношению $\theta/\beta = e \approx 2,72$), то вероятности правильного ответа соответственно равны: $P_R \approx 0,85$, $P_A \approx 0,78$.

В измерительной системе Раша графические представления не просто иллюстрируют математическую модель измерительной системы, но и позволяют наглядно сопоставлять результаты обработки тестовых данных. В частно-

сти, графики различных заданий в измерительной системе Раша представляют собой набор т.н. параллельных логистических кривых. Проекция нейтральной точки каждого графика на единую ось логитов даёт уровень трудности каждого задания. В примере, изображённом на рисунке 4, задания **а**, **б** и **в** имеют соответственно уровни трудности 0, -1 и -1,5 логит.

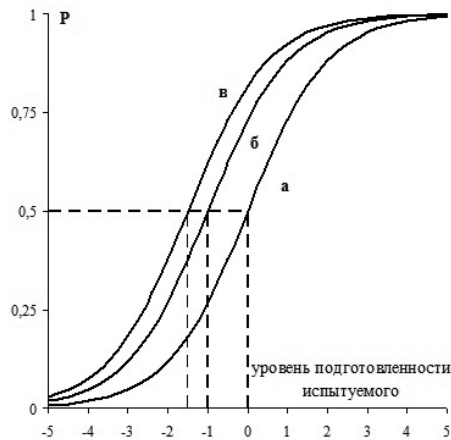


Рис. 4. Графики трёх заданий в системе Раша

В альтернативной системе подобные графики удобно приводить для вероятностных показа-

телей u и v участников теста. На рисунке 5 приведён аналогичный пример с графиками $P(u)$ для трёх заданий различного уровня трудности. Опорное задание «а» с уровнем трудности $v = 0,5$ ($\beta = 1$). Задание «б» легче опорного почти в два раза с уровнем трудности $v = 0,7$ ($\beta \approx 0,51$). Ещё более лёгкое задание «в», легче опорного почти в три раза, с уровнем трудности $v = 0,8$ ($\beta \approx 0,32$). В зависимости от уровня трудности графики деформированы, но имеют общие граничные точки $P(0) = 0$ и $P(1) = 1$. Уровень трудности каждого задания можно определить по графику как проекцию опорной точки при $u = 0,5$ ($\theta = 1$) на вертикальную ось ожидаемой вероятности правильного ответа P , т.е. $v = P(0,5)$. Последнее равенство следует и из самой альтернативной модели.

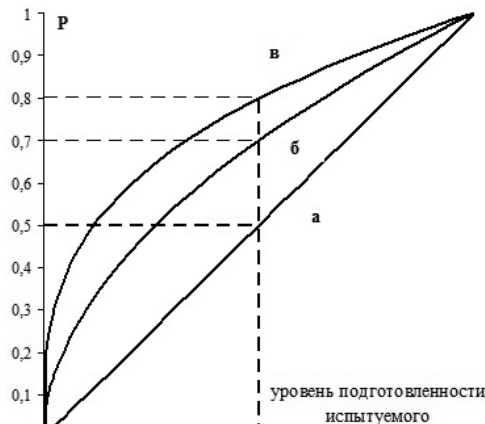


Рис. 5. Графики трёх заданий в альтернативной системе

На рисунках 4 и 5 представлены графики тестовых заданий

в различных системах измерения, в различных единицах измерения, и непосредственно сравнивать уровни трудности заданий некорректно. Однако для более удобного сопоставления самих рисунков мы выбрали такие уровни трудности тестовых заданий **а**, **б** и **в**, которые дают близкие рассчитываемые доли правильных ответов нейтральным испытуемым. В этом смысле рисунки 4 и 5 демонстрируют единые свойства теста из разных измерительных систем.

Основная задача педагогических измерений: принятые обозначения

Основная задача педагогических измерений состоит в поиске оптимальных оценок показателей подготовленности испытуемых и трудности заданий по наблюдаемым результатам тестирования. Строгое аналитическое решение этой задачи в общем виде для любой из рассматриваемых моделей измерения имеет трудности¹.

Сегодня выход из этой ситуации находят в поиске приближённых решений, основанных на тех или иных предположениях, допущениях и затем статистическом обосновании полученных решений.

Рассматривая некоторые простые алгоритмы приближённых решений, будем использовать следующие обозначения:

— $X_{ij} = 0; 1$ — один из двух возможных вариантов ответа i -м

испытуемым на j -ое тестовое задание. Матрица этих элементов, состоящая из N строк по количеству испытуемых и M столбцов по количеству тестовых заданий, представляет собой наблюдаемые данные результатов тестирования;

$$- X_i = \frac{1}{M} \sum_j X_{ij} \text{ — наблюдае-}$$

мые доли правильных ответов i -м испытуемым на все тестовые

задания, $X_i = \frac{1}{N} \sum_i X_{ij}$ — наблюдаемые доли правильных ответов всеми испытуемыми на j -ое тестовое

задание, $X_i = \frac{1}{N \cdot M} \sum_{ij} X_{ij} =$

$$= \frac{1}{N} \sum_i X_i = \frac{1}{M} \sum_j X_j \text{ — наблюдае-}$$

мая доля правильных ответов всей группы испытуемых на весь набор тестовых заданий ($0 \leq X \leq 1$). Эти результаты получаются счётом и усреднением значений элементов матрицы наблюдаемых данных. Значение общей доли правильных ответов X является основной лаконичной характеристикой наблюдаемых результатов тестирования. Если $X \approx 0,5$, то уровень трудности набора тестовых заданий соответствует подготовленности группы испытуемых. При X близких к значению 1, когда группа испытуемых правильно отвечает на большинство тестовых заданий, можно говорить о несоответствии набора тестовых заданий контролируемой группе учащихся, о низком уровне трудности используемого набора тестовых заданий. При $X \approx 0$ наоборот, только малая доля заданий выполнена правильно, уровень

трудности набора тестовых заданий несоразмерно высок для данной группы испытуемых;

— $0 < P < 1$ — ожидаемые (расчитываемые) в соответствии с выбранной математической моделью вероятности правильного ответа i -м испытуемым на j -е тестовое задание. Математические модели (R1) и (A1) задают правила преобразования наблюдаемых данных в искомые оценки на измерительной шкале, задают метрические свойства измерительной системы;

$$- P_i = \frac{1}{M} \sum_j P_{ij} \text{ — ожидаемые}$$

оценки доли правильных ответов i -м испытуемым на все тестовые

задания, $P_j = \frac{1}{N} \sum_i P_{ij}$ — ожидае-

мые оценки доли правильных ответов всеми испытуемыми на j -е тестовое задание,

$$P = \frac{1}{N \cdot M} \sum_{ij} P_{ij} =$$

$$= \frac{1}{N} \sum_i P_i = \frac{1}{M} \sum_j P_j \text{ — ожидае-}$$

мая оценка доли правильных ответов всей группы испытуемых на весь набор тестовых заданий.

Чем ближе рассчитываемые в рамках математической модели оценки P к наблюдаемым значениям X , тем точнее принятая математическая модель описывает результаты тестирования. Эту мысль можно приложить и к точечным показателям $P_{ij} \approx X_{ij}$, к их групповым усреднениям $P_i \approx X_i$ и $P_j \approx X_j$, так и общим показателям усреднения по всем участникам теста $P \approx X$. Фактически

в этом высказывании заложены и направления в поиске методов решения задачи педагогических измерений и направления в выборе критериев оптимальности полученных оценок.

Для дальнейшего обсуждения приближённых методов решения задачи удобно ввести основное наблюдаемое значение параметра взаимодействия латентных показателей. Это то значение единственного параметра модели, около которого группируются комбинации взаимодействующих наборов латентных пар b и d или θ и β . Это то значение параметра модели g или γ , для которого ожидаемая вероятность правильного ответа P равна наблюдаемой общей доле правильных ответов X .

Для модели Раша (R1) это значение рассчитывается по формуле

$$g_0 \approx \ln \frac{X}{1-X}. \quad (R3)$$

Значение этого показателя даёт ответ на вопрос, на сколько логитов в среднем уровень подготовленности группы испытуемых выше уровня трудности набора тестовых заданий. Если $g_0 \approx 0$, то набор тестовых заданий соответствует группе испытуемых по среднему уровню трудности. Если $g_0 < -0,5$ логит, когда испытуемые правильно отвечают менее чем на 38% тестовых заданий, то трудность набора тестовых заданий завышена для данной группы испытуемых. Для $g_0 > 0,5$, наоборот, набор тестовых заданий достаточно лёгкий для выполнения данной группой испытуемых.

Для альтернативной модели (A1) этот показатель рассчитыва-

ется по формуле:

$$\gamma_0 = 1/\log_{0,5} X, \quad (A5)$$

а его значение показывает, во сколько раз в среднем уровень подготовленности группы испытуемых выше уровня трудности набора тестовых заданий. Для $\gamma_0 < \sqrt{2}/2 \approx 0,7$, когда испытуемые правильно отвечают менее чем на 38% тестовых заданий, трудность набора тестовых заданий заметно завышена, а для $\gamma_0 > \sqrt{2}/2 \approx 1,41$ занижена для данной группы испытуемых.

Исходные решения задачи педагогических измерений

Как и прежде, поиск приближённых решений будем проводить в сопоставлении измерительных систем Раша и альтернативной.

В измерительной системе Раша искомые показатели уровней подготовленности испытуемого b и трудности тестового задания d входят в систему измерения в виде разности $g = b - d$, и ожидаемые оценки доли правильных ответов определяются только значениями этой разности (R1). Для начальной оценки уровня подготовленности испытуемых предполагается, что все тестовые задания имеют нейтральный (нулевой) уровень трудности $d = 0$. Тогда по долям правильных ответов i -м испытуемым на все задания теста X_i из приближённого урав-

нения $X_i \approx P_i = \frac{\exp b_i}{1 + \exp b_i}$ и определяются начальные (исходные) значения уровней подготовленности испытуемых как логарифмы соответствующих потенциалов:

$b_i \approx \ln \frac{X_i}{1 - X_i}$. Аналогично предполагается, что все испытуемые имеют нейтральный (нулевой) уровень подготовленности $b = 0$, рассчитываются и начальные уровни трудности j -х тестовых заданий: $d_j \approx \ln \frac{1 - X_j}{1 - X_j}$.

В альтернативной системе измерений, как и в системе Раша, для начальной оценки уровня подготовленности испытуемых предполагается, что все тестовые задания имеют единичный (опорный) уровень трудности $\beta = 1$. Если X_i — доли правильных ответов i -м испытуемым на все задания теста, то из приближённого уравнения $X_i \approx P_i = 0,5^{1/\theta_i} = u_i$ следует, что начальной оценкой вероятностного показателя уровня подготовленности испытуемого является наблюдаемая доля правильных ответов. Аналогично: при $\theta = 1$ начальными оценками вероятностного показателя уровня трудности тестового задания являются доли правильных ответов всеми испытуемыми $v_j \approx X_j$. Пере-

ход к относительным показателям осуществляется по формулам:

$$\theta_i = 1/\log_{0,5} u_i; \beta_j = \log_{0,5} v_j. (A6)$$

Этот этап достаточно распространён и в педагогической практике. Иногда учителя, минимизируя усилия на расчётную часть, для оценивания знаний своих учеников опираются непосредственно на подсчитанные доли правильных ответов X_i . В этом случае они фактически опираются на начальные оценки u_i альтернативной измерительной системы и (скорее неосмысленно) предполагают, что используемый набор тестовых заданий в среднем имеет нейтральный уровень трудности. Если на этом этапе перейти по правилам (A6) к относительным показателям θ_i , то можно получить дополнительную и более конкретизированную информацию об отношении уровней подготовленности испытуемых.

Продемонстрируем алгоритм поиска исходных решений на небольшом примере. Пусть задана таблица наблюдаемых результатов тестирования $N = 8$ испытуемых по набору из $M = 10$ тестовых заданий (таблица 4).

Таблица тестовых данных

Таблица 4

		Задания										Доли X_i
		1	2	3	4	5	6	7	8	9	10	
Испытуемые	1	1	1	1	1	0	1	1	0	1	0	0,70
	2	1	1	0	1	1	0	1	0	0	0	0,50
	3	1	1	1	1	1	0	0	0	0	0	0,50
	4	1	0	1	0	1	1	0	0	0	0	0,40
	5	0	1	1	0	0	0	0	0	1	0	0,40
	6	1	1	1	0	0	0	0	0	0	0	0,30
	7	1	1	0	0	0	0	0	0	0	0	0,20
	8	1	0	0	0	0	0	0	0	0	0	0,10
Доли X_j		0,875	0,750	0,625	0,375	0,375	0,250	0,250	0,125	0,125	0,125	$X=0,3875$

Первый результат можно привести для всей таблицы по наблюдаемой общей доле правильных ответов X . Из (R3) и (A5) имеем следующие оценки значения основного параметра взаимодействия латентных качеств $g_0 \approx -0,459$; $\gamma_0 \approx -0,731$. Отсюда, по системе Раша, имеем завышенный уровень (на 0,459 логит) трудности набора тестовых заданий для данной группы испытуемых. В альтернативной системе этот вывод обосновывается другими расчётами, и сам вывод делается в относительных единицах. Для данной группы испытуемых уровень трудности набора тестовых заданий завы-

шен примерно в $1/0,731 \approx 1,37$ раза.

По приведённым выше расчётным формулам несложно получить исходные оценки b и d в системе Раша, u , θ и v , β — в альтернативной системе. В таблицах 5 и 6 приведены результаты таких расчётов. В последних столбцах приведены средние арифметические значения для показателей b и d в интервальной шкале логитов; средние арифметические значения вероятностных показателей подготовленности испытуемых u и v ; средние геометрические значения для относительных показателей подготовленности θ и β в шкале отношений.

Таблица 5
Результаты оценивания испытуемых (исходные оценки)

Показатели		Испытуемые								Средние значения
		1	2	3	4	5	6	7	8	
R:	$b_i \approx \ln \frac{1-X_i}{1-X_i}$	0,847	0	0	-0,405	-0,405	-0,847	-1,386	-2,197	-0,549
A:	$u_i = X_i$	0,7	0,5	0,5	0,4	0,4	0,3	0,2	0,1	0,3875
	$\theta_i = 1/\log_{0,5} X_i$	1,943	1	1	0,756	0,756	0,576	0,431	0,301	0,733

Таблица 6
Результаты оценивания заданий (исходные оценки)

Показатели		Тестовые задания										Средние значения
		1	2	3	4	5	6	7	8	9	10	
R:	$d_j \approx \ln \frac{1-X_j}{X_j}$	-1,95	-1,10	-0,51	0,51	0,51	1,10	1,10	1,95	1,95	1,95	0,550
A:	$v_j = X_j$	0,88	0,75	0,63	0,38	0,38	0,25	0,25	0,13	0,13	0,13	0,3875
	$\beta_j = \log_{0,5} X_j$	0,19	0,42	0,68	1,42	1,42	2	2	3	3	3	1,279

Полученные исходные оценки явно противоречат принятым для их расчёта исходным поло-

жениям. Действительно, изначально предполагая нейтральные уровни латентных показателей,

даже средние значения исходных оценок существенно отличаются от этих предположений. В нашем примере средний уровень подготовленности группы испытуемых более чем на 0,5 логита ниже нейтрального уровня для схемы Раша и примерно в $1/0,733 \approx 1,36$ раза ниже нейтральной единицы альтернативной схемы. Подобные существенные отклонения присутствуют и в оценках заданий. И очень ошибочны будут, казалось бы, очевидные выводы: в среднем уровень набора тестовых заданий примерно на $0,550 - (-0,549) \approx 1,1$ логит завышен для данной группы испытуемых по системе Раша и примерно в $1,275/0,733 \approx 1,74$ раза по альтернативной шкале отношений. На самом деле такие выводы преувеличены в два раза. Эти очевидные противоречия нивелируются лишь в том случае, когда в среднем уровень трудности набора тестовых заданий близок к уровню подготовленности испытуемых.

Попробуем снять это ограничение о равенстве средних уровней и привести решения в большее соответствие с исходными положениями в следующем подразделе.

Первое приближение решения задачи педагогических измерений коррективной системы отсчёта

Идея первого приближения состоит в «смещении» одного из исходных положений к наблюдаемой общей доле правильных ответов X , которое удобно про-

вести введением понятия «система отсчёта». В основе этого метода решения служат допустимые инвариантные преобразования.

Понятие системы отсчёта обсуждалось в работе¹. Здесь же первое приближение решения задачи ПИ для альтернативной системы измерений обозначается как элементарный метод. В данной работе в качестве примера будем отталкиваться лишь от системы отсчёта «нейтральный набор тестовых заданий».

Напомним, в схеме Раша исходные оценки получены из двух предположений — все уровни d и b нейтральные. Для оценки уровня подготовленности испытуемых в первом приближении первое предположение не меняем и считаем, что все тестовые задания имеют нейтральный (нулевой) уровень трудности $d = 0$. Это предположение фактически определяет выбор системы отсчёта с нейтральным набором тестовых заданий. Тогда имеем следующие оценки уровней подготовленности испытуемых:

$$b_i \approx \ln \frac{X_i}{1 - X_i}. \text{ Предположим, что}$$

все испытуемые имеют такой уровень подготовленности, при котором они правильно отвечают на нейтральное тестовое задание с вероятностью равной общей доле правильных ответов X . С учётом определения (R3) уровень подготовленности этих испытуемых равен g_0 .

Тогда закон (R1) принимает вид $X_j \approx P_j = \frac{\exp(g_0 - d_j)}{1 + \exp(g_0 - d_j)}$ из ко-

1

Каргин Ю. Элементарное решение основной задачи педагогических измерений // Педагогические измерения. 2011. № 4. С. 50–67.

того находим расчётную формулу для оценки уровней трудности тестовых заданий в первом приближении: $d_j \approx \ln \frac{1-X_j}{X_j} + g_0$.

В этой системе отсчёта оценки уровней подготовленности испытуемых не изменяются (таблица 5), определяются только индивидуальной долей правильных ответов на все тестовые задания. Первый вывод анализа резуль-

татов теста о сниженном уровне подготовленности испытуемых примерно на 0,5 логит относительно набора заданий справедлив.

Среднее значение уровня трудности набора тестовых заданий не противоречит исходному предположению – значение среднего уровня трудности тестовых заданий приближается к нейтральному (таблица 7, последний столбец в строке R).

Таблица 7

Результаты оценивания заданий
(нейтральный набор тестовых заданий)

Показатели		Тестовые задания										Средние значения
		1	2	3	4	5	6	7	8	9	10	
R:	$d_j = \ln \frac{1-X_j}{X_j} + g_0$	-2,40	-1,56	-1,97	0,05	0,05	0,64	0,64	1,49	1,49	1,49	0,092
A:	$v_j = X_j^{\gamma_0}$	0,91	0,81	0,71	0,49	0,49	0,36	0,36	0,22	0,22	0,22	0,478
	$\beta_j = \gamma_0 \cdot \log_{0,5} X_j$	0,14	0,30	0,50	1,03	1,03	1,46	1,46	2,19	2,19	2,19	0,935

Без фактически дублирующих комментариев приведём расчётные формулы решения основной задачи ПИ в первом приближении относительно системы отсчёта «нейтральная группа

испытуемых»: $b_i \approx \ln \frac{X_i}{1-X_i} - g_0$,
 $d_j \approx \ln \frac{1-X_j}{X_j}$.

Первое приближение в альтернативной системе измерений строится аналогично. В системе отсчёта с нейтральным набором тестовых заданий сначала предполагаем, что все тестовые задания имеют единичный уровень трудности $\beta = 1$. Тогда имеем следующие, совпадающие с исход-

ными, расчётные формулы для оценки уровней подготовленности испытуемых – $u_i \approx X_i$ и $\theta_i \approx 1/\log_{0,5} X_i$. Предполагая далее, что все испытуемые имеют такой уровень подготовленности, при котором они правильно отвечают на нейтральное тестовое задание с вероятностью, равной общей доле правильных ответов X, определяем уровень их подготовленности γ_0 (A5).

Тогда закон (A1) принимает вид $X_j \approx P_j = 0,5^{\beta_j/\gamma_0}$ из которого находим расчётные формулы для оценки уровней трудности тестовых заданий в первом приближении: $v_j \approx X_j^{\gamma_0}$ и

$\beta_j \approx \gamma_0 \cdot \log_{0,5} X_j$ (таблица 7, строки А). Нетрудно убедиться в том, что в среднем полученные оценки не противоречат исходным предположениям.

Для системы отсчёта с нейтральной группой испытуемых расчётные формулы имеют вид: для вероятностных показателей — $u_i \approx X_i^{\gamma_0}$, $v_j \approx X_j$ и для относительных показателей — $\theta_i \approx 1/\gamma_0 \log_{0,5} X_i$ и $\beta_j \approx \log_{0,5} X_j$.

Следует отметить, что переходы от одной системы отсчёта к другой не изменяют метрических свойств решений, а лишь дают основания для интерпретации результатов.

Первое приближение решения задачи ПИ обладает рядом достоинств. Первое. Оно снимает одно из основных ограничений к свойствам участников теста — примерное равенство среднего уровня трудности набора тестовых заданий уровню подготовленности группы испытуемых. Второе. Это решение с вычислительной стороны лишь незначительно сложнее исходного и легко найдётся вручную. Корректировка исходного решения слагаемым g_0 в системе Раша и множителем 0 в альтернативной системе приводит к взаимному соответствию измерительные шкалы латентных показателей. И, наконец, третье. Выбор исследователем системы отсчёта, относительно которой строится решение, создаёт дополнительные возможности для внятной интерпретации результатов тестирования.

Второе приближение решения задачи педагогических измерений корректировкой средних оценок и их рассеяния. Качество решений

Если первое приближение решения основной задачи педагогических измерений направлено на разработку вычислительных алгоритмов оценки значений латентных показателей с контролем только средних значений, то второе приближение призвано уточнить его учётом распределений этих показателей. Удачным примером такого решения в системе Раша является алгоритм PROX¹. Здесь, основываясь на предположении, что уровни подготовленности испытуемых и трудности заданий распределены по двухпараметрическому нормальному закону, исходные оценки согласованно преобразуются к стандартизированному виду. Эти более жёсткие требования к распределениям участников теста, с одной стороны, ограничивают область эффективного применения алгоритма, с другой стороны — приводят к заметному повышению качества обработки результатов тестирования при условии их выполнения.

Примером поиска решений задачи ПИ с учётом распределений искомых показателей в альтернативной схеме является предложенный в работе² аналитический метод. В основе этого метода лежит возможность аналитического решения задачи с описанием распределения искомых показателей двухпараметрической фун-

1

Wright B.D., Stone M.H. Measurement Essentials. (Глава 16);

Демещёнок О. Реализация алгоритма PROX в MS EXCEL // Педагогические измерения. 2008. № 3. С. 64–71.

2

Каргин Ю. Аналитический метод решения основной задачи педагогических измерений // Педагогические измерения. 2011. № 2. С. 54–76.

кцией распределения (бета распределение).

В таблицах 8 и 9 приведены результаты таких расчётов для рассматриваемого в работе примера. Учитывая рекомендации анали-

тического метода, мы приняли следующий закон распределения $p(m = 0,484; n = 1)$ для вероятностного показателя трудности тестовых заданий v .

Таблица 8
Результаты оценивания испытуемых (второе приближение)

Показатели		Испытуемые								Средние значения
		1	2	3	4	5	6	7	8	
R:	b_i	1,755	0,550	0,550	-0,027	-0,027	-0,655	-1,422	-2,575	-0,231
A:	$u_i = 0,5^{1/\theta_i}$	0,866	0,715	0,715	0,604	0,604	0,457	0,261	0,049	0,534
	$\theta_i = \frac{X_i}{m \cdot (1 - X_i)}$	4,818	2,065	2,065	1,377	1,377	0,885	0,516	0,229	1,192

Таблица 9
Результаты оценивания заданий (второе приближение)

Показатели		Тестовые задания										Средние значения
		1	2	3	4	5	6	7	8	9	10	
R:	d_i	-2,98	-1,92	-1,19	0,09	0,09	0,83	0,83	1,89	1,89	1,89	0,139
A:	$v_j = 0,5^{\beta_j}$	0,91	0,79	0,66	0,31	0,31	0,13	0,13	0,01	0,01	0,01	0,326
	$\beta_j = \frac{1 - X_j}{X_j}$	0,14	0,33	0,60	1,67	1,67	3	3	7	7	7	1,733

Вычислительные процедуры второго приближения ещё более усложняются, по сравнению с первым приближением, но остаются ручными и прозрачными для исследователя. В этом смысле они удобны. Однако, важен и другой вопрос — вопрос качества решений.

Эту задачу для измерительной системы Раша на примере из цитируемой выше работы Б. Райта и М. Стоуна¹ рассматривал О. Деменчёнок². В подразделе «Анализ качества моделей» им проведены оценки показателей качества мо-

делей по четырём независимым критериям и для нескольких методов расчёта параметров модели педагогических измерений. В частности, по его расчётам, метод PROX улучшает качества решений от двух до более десяти процентов, исходя из выбранного критерия оценки этого качества. Проведём и мы подобный анализ качества полученных выше решений, выбрав в качестве критерия другой вариант.

В основу метода оценки качества полученных решений положим приближённые равенства:

1
Wright B.D., Stone M.H. Measurement Essentials. (Глава 18).

2
Деменчёнок О. Подбор параметров модели педагогических измерений // Педагогические измерения. 2008. № 1. С. 27–48.

$P_i \approx X_i$ и $P_j \approx X_j$. Чем точнее выполняются эти равенства, т.е. чем ближе рассчитанные на основе полученных решений оценки долей правильных ответов каждым испытуемым на все задания теста P_i и всеми испытуемыми на каждое задание теста P_j наблюдаемым долям X_i и X_j , тем выше качество полученных решений. Точность выполнения этих приближённых равенств можно оценить значениями и надёжностью коэффициентов линейной регрессии k_1 и k_2 уравнений: $P_i = k_1 \cdot X_i$ и $P_j = k_2 \cdot X_j$. Чем ближе коэффициенты k_1 , k_2 и соответствующие коэффициенты детерминации R^2 (мера адекватности уравнения регрессии) к единице, тем качественней построена параметрическая модель. При желании не сложно оценить и достоверность значений коэффициентов k_1 и k_2 . В пределе, при равенстве

единицы коэффициентов регрессии и коэффициентов детерминации, приближённые равенства преобразуются в точные $P_i \approx X_i$ и $P_j \approx X_j$.

Необходимые расчёты оценки качества решений несложно провести в среде табличного процессора Microsoft Excel. Первый шаг состоит в построении точечного графика $P(X)$, который сам по себе достаточно наглядно отражает качество полученных решений. Затем по построенному ряду данных следует добавить линейную линию тренда с параметрами: пересечение кривой с осью Y в точке 0; показывать уравнение на диаграмме; поместить на диаграмму величину достоверности аппроксимации (R^2).

В таблице 10 приведены результаты таких расчётов для построенных в данной работе решений.

Таблица 10

Показатели качества решений

Метод расчёта	Система Раша				Альтернативная система			
	k_1	R_1^2	k_2	R_2^2	k_1	R_1^2	k_2	R_2^2
Начальное приближение	0,792	0,980	0,827	0,992	0,806	0,966	0,856	0,981
Первое приближение	0,955	0,936	0,966	0,994	0,983	0,926	0,992	0,991
Второе приближение	1,102	0,986	1,083	0,983	0,978	0,995	1,022	0,989

Заметное повышение качества решений при переходе от начального к первому приближению для рассматриваемого примера вполне предсказуемо. Это следует из описанных выше противоречий в предположениях начального приближения, существенно проявляющихся с ростом различия в средних уровнях подготовленности группы испытуемых и труд-

ности набора заданий. Достаточно высокие показатели качества решений первого приближения, достаточно простые алгоритмы их построения и внятная интерпретация самих решений позволяют рекомендовать именно этот подход к широкому применению в педагогической практике.

Если обратиться к показателям качества решений второго

приближения, то следует отметить, что алгоритм PROX не дал ожидаемого положительного эффекта. Более того, первое приближение может даже более удачное. Такой результат легко объясняется тем фактом, что распределение участников теста достаточно далеки от нормального. Т.е. сами исходные предположения применения алгоритма PROX для рассматриваемого набора тестовых данных не выполняются. Если второе приближение альтернативной системы и даёт некоторый положительный результат, что отражается повышением коэффициентов детерминации, то и этот эффект связан только с тем, насколько удачно выбрана функция распределения $p(m; n)$.

Продолжая сопоставление измерительных систем, есть все основания утверждать о целесообразности проведения педагогических измерений в шкале отношений и в проблеме методов построения решений.

Некоторые особенности проведения педагогических измерений метрических шкалах

Отличительным признаком метрической шкалы является возможность установить единицу измерения. Обычно единицы измерения задаются исследователем, это условно принятые меры, в их выборе присутствует произвол. Иногда единицы измерения связывают с именем учёного (например, единицы измерения температуры — градус Цельсия и градус Кельвина), иногда они принимаются из соображений удобства

или совместного соглашения (например, единица измерения длины — метр, принятый эталон, примерная длина шага взрослого человека).

Другим признаком, уже различающим метрические шкалы на интервальную или пропорциональную, является существование естественного нулевого начала отсчёта. В интервальной шкале такой точки нет, и исследователь её должен сам определять; в пропорциональной шкале она есть и существует независимо от исследователя. Например, в интервальной шкале Цельсия нулевым началом отсчёта температур принята температура таяния льда. Эта температура не означает отсутствие у тела тепла или температуры. Это значение условное, оно не уникальное, оно лишь приписывается к температуре перехода «вода — лёд» привычного для человека вещества. В пропорциональной шкале Кельвина нулевое начало отсчёта задано естественным образом. В этой шкале существует точка абсолютного нуля температуры. Это такое недостижимое состояние вещества, при котором должно прекратиться всякое хаотическое движение молекул вещества.

Существование произвола в метрических шкалах проявляется и в допустимых правилах преобразований, которые не нарушают структуру шкалы. Для интервальной шкалы они сводятся к допустимым линейным преобразованиям (сдвиг и деформация или прибавление числа и умножение на число), для измерений в шкале отношений к преобразованию по-

добия (деформация, умножение на число).

Примером преобразований в интервальной шкале может служить правила преобразования значения температуры тела измеренного в шкале Цельсия в шкалу Фаренгейта, состоящие в умножении исходной цифры на $9/5$ и прибавлении числа 32. Т.е. единица измерения в шкале Цельсия в 1,8 раза «шире», а нулевая точка в шкале Цельсия соответствует 32 градусам Фаренгейта. На бытовом уровне этой шкалой пользуются в США, и никаких неудобств при этом никто не испытывает, дело только в привычке.

В качестве примера преобразования результатов измерений в шкале отношений можно привести перевод значения массы тела, измеренного в килограммах, в фунты умножением исходного значения на число $\approx 0,49$. Т.е. единица измерения килограмм примерно в 2 раза «тяжелее» фунта.

В педагогических измерениях примером интервальной шкалы служит шкала логитов измерительной системы Раша. В этой системе единица измерения и нулевая точка отсчёта определяются по результатам обработки тестовых данных и не являются уникальными. Возможность проводить допустимые преобразования с измерениями в шкале логитов используются, в частности, при оптимизации исходных решений методом PROX.

В альтернативную систему заложена пропорциональная шкала измерения относительных показателей с естественным нулём. Единица измерения, как и в си-

стеме Раша, определяется в процессе обработки наблюдаемых результатов теста и связывается с опорным участником теста, а допустимые преобразования деформации масштаба лежат в основе элементарного метода поиска решения задачи ПИ.

Свойства измерительной шкалы не ограничиваются её формальными метрическими характеристиками. С практической точки зрения важно убедиться в выполнении и понять смысл ещё одного важного свойства метрической шкалы — свойства равенства интервалов или отношений между измеряемыми величинами. Поясним это свойство на примере. Пусть первое тело нагрелось от 10°C до 11°C , а второе — от 20°C до 21°C . Одинаково ли нагрелись эти тела, одинаково ли они стали теплее? Предполагается что да, каждое из них стало теплее на одну сотую часть разности температур кипящей и замерзающей воды.

Подобные рассуждения справедливы и в пропорциональной шкале Кельвина. Нагревая первое тело с 10 К до 20 К, мы действительно в два раза его нагреваем. Этот вывод не изменится и при нагреве второго тела с любой начальной температуры, например от 100 К до 200 К. В обоих случаях средняя кинетическая энергия движения молекул увеличится в два раза.

Рассмотрим, как выполняются эти свойства в педагогических измерениях. Интервальная шкала логитов — статистическая шкала, градуированная по логистической функции, очень близкой к фун-

кции равномерного распределения. Факт расположения эмпирических точек подготовленности испытуемых вдоль нормальной кривой можно отнести к следствию предельной теоремы Ляпунова (действительно, можно считать, что случайная величина «уровень подготовленности испытуемого» формируется под влиянием множества случайных факторов).

Отметим, что подобием шкале логитов является широко известная шкала стенов Р.Б. Кеттела с единичным интервалом в 0,5 стандартного отклонения. Содержательная интерпретация свойства равенства интервалов в таких шкалах весьма затруднительна. Приведём пример. Пусть первый испытуемый повысил уровень своей подготовленности с 0 до 1 логита, а второй — с 3 логитов до 4 логитов. В этом случае вероятность правильного ответа на нейтральное задание увеличится для первого испытуемого на 0,23 (с 0,5 до 0,73), а для второго испытуемого — на 0,03 (с 0,95 до 0,98). Одинаково ли эти испытуемые повысили свой уровень подготовленности? Ответ «да», если исходить из формального определения «уровень подготовленности», и скорее «нет», если опереться на мнение педагога-практика.

В пропорциональной шкале измерения относительных показателей уровня подготовленности испытуемых и уровня трудности заданий содержание свойства равенства отношений интерпретируется достаточно наглядно. Если первый и второй испытуемые повысили свой уровень подготовленности в два раза, то это

означает, что теперь каждый из них с прежней совместной вероятностью правильно выполнит в два раза больше тестовых заданий. И не важно, на каком начальном уровне подготовки находились испытуемые.

И в заключение отметим: пропорциональная шкала более совершенная хотя бы потому, что её можно использовать как интервальную. Обратный переход если и возможен, то он потребует дополнительных обоснований. Другое дело, насколько удобно использовать интервальные единицы в шкале отношений. Приведём пример применения интервальных единиц к измерениям массы тел (в пропорциональной шкале). Если один пакет с яблоками тяжелее другого на килограмм, то содержание этой информации вполне понятно и может быть актуальным. Если один грузовик тяжелее другого на один килограмм, то эта информация совершенно не актуальна, даже если точность проведённых измерений позволяет делать такой вывод.

Однако если один пакет с яблоками тяжелее другого в два раза и один автомобиль тяжелее другого в два раза, то эти результаты измерений информативны и актуальны. Т.е. применение интервальных единиц в пропорциональной шкале весьма ограничено, оно может быть целесообразным только в том случае, когда совокупность измеренных параметров находится в весьма ограниченной области, когда расстояние между измеренными величинами сопоставимо со значениями этих величин.