ЭЛЕМЕНТАРНОЕ РЕШЕНИЕ ОСНОВНОЙ ЗАДАЧИ ПЕДАГОГИЧЕСКИХ ИЗМЕРЕНИЙ

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

В работе представлен технологически простой, но достаточно эффективный метод решения основной задачи педагогических измерений — определение в метрической шкале отношений уровня подготовленности испытуемых и уровня трудности тестовых заданий.

Ключевые слова: теория педагогических измерений, инвариантные преобразования показателей теста, элементарный метод решения основной задачи педагогических измерений.

Введение

Математическая теория педагогических измерений (IRT) основывается на идее описать процесс тестовых испытаний математическим законом, позволяющим вычислить вероятность правильного ответа испытуемого на задание теста. В теории Раша этот закон представлен логистической функцией (очень близкой к нормальному закону распределения, но заметно удобнее его для применения), что позволяет достаточно надёжно и обоснованно решить основную задачу педагогического тестирования — измерение уровней подготовленности испытуемых и трудности тестовых заданий.

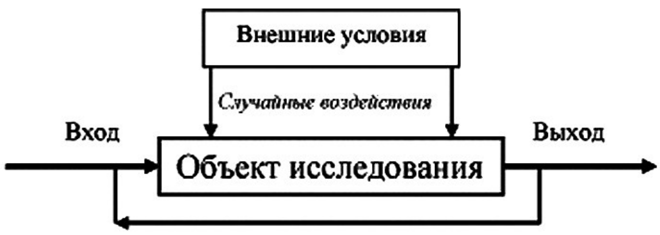
Существенным недостатком, ограничивающим широкое применение такого подхода в практике, является усложнённая интерпретация результатов измерения и заметная привязанность к специальным приложениям электронной обработки результатов тестирования.

Предложенный в работе метод существенно устраняет указанные недостатки, позволяя несложными преобразованиями исходных баллов получать достаточно надёжные и внятно интерпретируемые искомые оценки. Теоретические положения и технология решения поставленной задачи сопровождаются примером с необходимыми расчётами и анализом результатов.

Место МТИ в системе педагогического тестирования

В начале работы кратко обсудим сферу внимания и некоторые проблемы применения методов МТИ в системе педагогического тестирования. Такое обсуждение поможет лучше понять смысл заложенной в таблице тестовых результатов внутренней неопределённости. Именно на этом понимании строится логика элементарного решения основной задачи педагогических измерений. Для обсуждения этого вопроса используем известную кибернетическую модель с «чёрным ящиком». Под «чёрным ящиком» понимается некоторый объект исследования, внутреннее устройство которого неизвестно. Однако, если поведение «черного ящика» отражается на наблюдаемых экспериментальных данных, то существует и попытка предсказать его поведение в вероятностном смысле.

Если приложить эту модель на тестирование, то под входом можно понимать участников теста (испытуемые и тестовые задания) и их свойства, под выходом — педагогическую интерпретацию результатов тестовых испытаний. Обратная связь корректирует вход рекоменда-



Обратная связь

циями по разработке тестовых заданий и формированию групп испытуемых. Под внешними условиями обычно понимаются как принципиально неконтролируемые, так и неучитываемые в эксперименте факторы. Под объектом исследования можно понимать сам процесс тестирования с выделением у этого объекта предметов исследования.

В частности, предметом исследования математической теории измерений являются метрические свойства теста. С прикладной точки зрения эта формальная теория с математическими методами исследований нацелена на разработку измерительного инструментария по оцениванию свойств участников теста. Единственной экспериментальной основой для таких исследований служит таблица чисел с результатами тестирования.

В данной работе не рассматривается сама модель с «чёрным ящиком». Мы привели эту модель лишь для того, чтобы подчеркнуть, что указанные в ней элементы являются внешними по отношению к объекту исследования и не являются предметом исследования математической теории измерений. Но оставить совершенно без внимания эти внешние влияния тоже не удаётся. Дело в том, что сама по себе таблица данных не позволяет получить однозначного решения. На её основе можно получить только множество эквивалентных решений. И уже выбор того единственного и окончательного решения определяется внешними условиями или, как мы будем говорить дальше, системой отсчёта.

Попытка вскрыть содержание «черного ящика» математическими методами (МТИ) строится через введение латентных показателей (вместо реальных участников теста) и математическое моделирование самого процесса тестовых испытаний. В данной работе используется альтернативная (по отношению к модели Раша) математическая модель, в которой латентные показатели уровня трудности тестовых заданий в и уровня подготовленности испытуемых q входят в виде отношения $\gamma = \beta/\theta$. Именно этим отношением, и только им, определяется вероятность правильного ответа испытуемым на тестовое задание $P = 0{,}5^{\gamma}$.

Введённые показатели в и q измеряются в шкале отношений с нейтральной единицей и непосредственно не связаны с показателями модели Раша, уже измеряемыми в интервальной шкале логитов. Увязка показателей в и q в единый параметр g, с одной стороны, даёт возможность их измерения в единой шкале. Но с другой стороны, и создаёт внутреннюю неопределённость в решении задачи описания результатов тестирования. Приведём небольшой пример имитационного тестирования.

Пусть проведено два независимых тестовых испытания. В первом случае группа из двух испытуемых выполняет трудный набор из двух заданий. Во втором случае слабая группа из двух испытуемых выполняет средний набор из двух заданий. Качественные характеристики уровней подготовленности испытуемых и трудности тестовых заданий мы приводим из следующих соображений: если значения показателей больше нейтральной единицы, то соответствующий уровень выше среднего, и наоборот. В данном случае нейтральная единица служит некоторым ориентиром, эталоном в количественной оценке показателей.

В таблицах представлены результаты расчётов вероятностей правильных ответов и наиболее вероятный исход таких испытаний при дихотомическом описании:

Расчёт вероятностей правильных ответов

Первое испытание		
	$\beta = 1$	$\beta = 4$
$\theta = 0{,}5$	0,25	0,004
$\theta = 2$	0,707	0,25

Второе испытание		
	$\beta = 0{,}5$	$\beta = 2$
$\theta = 0{,}25$	0,25	0,004
$\theta = 1$	0,707	0,25

Наиболее вероятный исход испытаний

Первое испытание		
	$\beta = 1$	$\beta = 4$
$\theta = 0{,}5$	0	0
$\theta = 2$	1	0

Второе испытание		
	$\beta = 0{,}5$	$\beta = 2$
$\theta = 0{,}25$	0	0
$\theta = 1$	1	0

Как рассчитанные, так и наблюдаемые результаты испытаний в обеих группах эквивалентны, и никакие математические преобразования не способны их различить. Результаты испытаний одинаковые не только в среднем, но и в деталях. Совершенно очевидно, что можно предложить и другие наборы параметров β и θ, приводящие к таким же эквивалентным результатам. Таким образом, совершенно разные исходные тестовые ситуации могут дать совершенно одинаковые наблюдаемые значения и, как следствие, совершенно одинаковые результаты математического анализа этих данных. И никакие внутренние приёмы не определят всех свойств исследуемых систем, не вскроют существующие различия.

На самом деле указанная неопределённость заложена в самом процессе тестовых испытаний, она порождается взаимодействием испытуемых и тестовых заданий. На математическом языке эту неопределённость можно толковать как существование в рассматриваемой математической системе дополнительной степени свободы. Для снятия этой неопределённости и предложено введение «системы отсчёта».

Хотя рассматриваемые группы различны, т.к. у них различны абсолютные значения входных показателей, отношения этих значений подчиняются единым свойствам. Вот некоторые свойства, которые отражают отношения показателей в отдельной группе, но эти отношения одинаковы для обеих групп:

- второй испытуемый в четыре раза лучше подготовлен первого;
- второе тестовое задание в четыре раза труднее первого;
- первый испытуемый в два раза слабее первого задания и в восемь раз слабее второго задания;
- второй испытуемый в два раза сильнее первого задания и в два раза слабее второго задания.

Несложно обнаружить и отношения показателей между рассматриваемыми группами — значение каждого входного показателя первой тестовой группы в два раза превышает соответствующие значения показателей второй группы.

Рассуждения, приведённые в этом подразделе, не частные. Они следуют из свойств математической модели. Различия в абсолютных значениях и независимость относительных результатов от системы отсчёта следуют из инвариантности проведённых с ними преобразований.

Элементарный метод решения задачи педагогических измерений

Как сам метод, так и технологию его применения будем иллюстрировать на модельном примере. Десять учащихся (нумеруются индексом i) прошли испытания по пяти заданиям (нумеруются индексом j). Результаты показаны в матрице значений X_{ij} — строки соответствуют испытуемым, столбцы соответствуют заданиям:

$$\begin{matrix} 1 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{matrix}$$

Сразу сделаем одно замечание. Десятый испытуемый ни на одно из заданий верно не ответил. Он никак количественно не проявил свой уровень подготовленности. Его нулевой результат может трактоваться как очень низкий уровень подготовленности, ниже, чем разрешающие возможности набора тестовых заданий. В схеме Раша такие данные исключаются из дальнейшего анализа. Формальной причиной этому служит возникающая неопределённость в расчётах исходных показателей. В данном подходе этих ограничений нет, как и формальных требований по исключению таких результатов из дальнейшего анализа.

Проведём первые расчёты и проанализируем их результаты. По таблице результатов X_{ij} несложно рассчитать доли правильных ответов испытуемыми на все задания теста $X_i = {}^1/_5 \times \Sigma X_{ij}$ и доли правильных ответов на задания всеми испытуемыми $X_j = {}^1/_{10} \times \Sigma X_{ij}$. Эти доли служат исходными значениями для получения оценок уровня подготовленности испытуемых и уровня трудности тестовых заданий. Далее, из $10 \times 5 = 50$ максимально возможных баллов испытуемая группа учащихся набрала только 20 баллов.

Общая доля правильных ответов ниже среднего, и равна $X={}^{20}/_{50}=0{,}4$. Этот результат можно получить и усредняя исходные доли: $X={}^1/_5 \times \Sigma X_j = {}^1/_{10}\times \Sigma X_i = 0{,}4$. Свойство таблицы данных, выраженное последним равенством, используется далее для оценки надёжности полученных решений. Один учитель, анализируя значение общей доли правильных ответов, может охарактеризовать такой результат «слабой подготовкой учащихся», другой объяснит его «высокой трудностью заданий». Кто из них прав?

Поиск ответа на этот вопрос[1] не даст определённого результата без дополнительных внешних условий, снимающих внутреннюю неопределённость набора наблюдаемых результатов. Действительно, наблюдаемые значения есть результат взаимодействия двух факторов — подготовленность группы учащихся и трудность набора тестовых заданий. И результаты тестирования зависят не отдельно от уровня подготовленности учащихся и уровня трудности тестовых заданий, а от отношения этих уровней.

Детализированная оценка именно этих отношений есть и цель, и возможность математического анализа результатов теста. С формальной точки зрения оба объяснения эквивалентны, а различия в оценках связаны с исходной позицией исследователя в выборе дополнительного условия относительно свойств участников тес-

1 Поиск ответа на этот вопрос аналогичен поиску решения уравнения с двумя неизвестными, например, $x + y = 0{,}8$. В общем виде это уравнение имеет множество частных решений. Вот некоторые из них. Если положить $x = 0{,}5$, то получим $y = 0{,}3$ или, наоборот, при $y = 0{,}5$ имеем $x = 0{,}3$. Возможны и другие исходные предположения, например, о равенстве x и y. Тогда имеем решения $x = y = 0{,}4$.

та, с выбором «системы отсчёта». Важно другое, если оба преподавателя оставляют неизменными отношения итоговых оценок, то их результаты эквивалентны.

Приведённые рассуждения можно лаконично представить в виде следующего утверждения: существуют такие системы отсчёта, дающие разные решения основной задачи педагогических измерений, относительно которых отношения итоговых оценок остаются неизменными.

Для более строгого обоснования этого утверждения следует обратиться к математической модели теста[2]. Если обозначить через β_j — уровень трудности j-го тестового задания, а θ_i — уровень подготовленности i-го испытуемого, то вероятность правильного ответа P определяется отношением этих уровней $\gamma = \beta/\theta$ по формуле $P = 0{,}5^{\gamma}$. Математический вид закона задаёт инвариантные преобразования данной модели γ = const. Или, переход от одной системы отсчёта к другой одновременным изменением показателей β и θ в k раз никак не отражается на результатах анализа теста.

Значения показателей β и θ в общем случае изменяются от нуля до бесконечности и имеют только относительный содержательный смысл. Т.е. мы можем говорить только об уровне подготовленности испытуемого относительно тестового задания, относительно теста в целом, относительно другого испытуемого, но никак не об уровне подготовленности, самом по себе. Аналогичные рассуждения справедливы и для относительного уровня трудности тестового задания.

Указанные свойства относительности позволяют рассматривать тест как измерительную систему с одним общим эталоном, относительно которого будут оцениваться показатели и уровней подготовленности испытуемых, и уровней трудности тестовых заданий. Об этой возможности измерять единой мерой, казалось бы, несоизмеримые качества указывали и классики IRT, о чём хорошо известно читателям журнала «Педагогические измерения». Но если в вопросе о разработке универсальной меры для всех педагогических тестов, единого эталона для «измерения знаний», пока много нерешенных проблем, то право определения эталона для конкретного теста лежит за его исследователем.

Вполне естественно за такой внутренний эталон выбрать нейтральный (единичный) уровень. Для обозначения такого внутреннего эталона теста мы используем термин «опорный». Совершенно неважно, им будет идеализированный испытуемый или идеализированное тес-

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

товое задание. Эти эталоны однозначно взаимосвязаны — опорный испытуемый правильно выполнит опорное задание с вероятностью 0,5.

Теоретически, в качестве эталона можно выбрать любого испытуемого или любое тестовое задание. И уже относительно него проводить оценки и интерпретацию результатов. Но вряд ли это целесообразно.

Другим вариантом решения этой проблемы может служить поставка эталона извне. Мы об этом варианте уже говорили выше. Но такое решение проблемы лежит вне сферы действий МТИ.

В работе[3] подробно даны определения понятий «опорный испытуемый», «опорное тестовое задание» и описано введение вероятностных параметров рассматриваемой здесь модели педагогического теста. В данной работе этот фрагмент представим кратко.

Введение понятия опорных участников тестирования позволяет ввести следующие вероятностные эквиваленты, альтернативные относительным показателям β и θ: $v = 0{,}5^{\beta}$ — вероятностный показатель уровня трудности тестового задания, с такой вероятностью его выполнит опорный испытуемый; $u = 0{,}5^{1/\theta}$ — вероятностный показатель уровня подготовленности испытуемого, с такой вероятностью испытуемый выполнит опорное задание. Значения вероятностных показателей изменяются от нуля до единицы. В новых переменных математическая модель принимает вид: $\log_{0,5}P = \log_{0,5}v \times \log_{0,5}u$.

Введение дополнительных вероятностных показателей v и u обусловлено следующими соображениями. *Первое.* Эти показатели несут внятный вероятностный смысл, что существенно упрощает процесс интерпретации результатов тестовых измерений. Ниже мы проводим такую интерпретацию на примере. *Вторая причина* следует из математических соображений. Переход от исходных относительных показателей β и θ к их вероятностным эквивалентам v и u существенно упрощает математические преобразования и поиск методов решения задачи педагогических измерений.

В новых переменных инвариантные преобразования можно сформулировать следующим образом: если для каждой пары «испытуемый — тестовое задание» значения вероятностных показателей равны u и v, то их одновременные преобразования по закону u^{k} и $v^{1/k}$ никак не отразятся на результатах их анализа. Эти преобразования следуют из свойств логарифма и ещё раз подтверждают высказанное выше утверждение.

Наблюдаемыми результатами педагогического тестирова-

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

ния являются доли правильных ответов испытуемых на каждое задание теста (X_j) и доли правильных ответов каждого испытуемого на все задания теста (X_i). Тогда в приведённой терминологии основную задачу педагогических измерений можно свести к следующей задаче — по наблюдаемым долям правильных ответов X_j и X_i оценить скрытые (латентные) значения показателей v_j и u_i (или β_j и θ_i).

В общем виде эта задача не решена. Представим здесь пусть не точный, пусть не строго обоснованный, но зато достаточно простой, надёжный и апробированный метод её решения. Этот метод мы обозначили как «элементарный» из следующих соображений. Технология его применения действительно состоит из простых и наглядных вычислительных процедур. Для того, чтобы самостоятельно получить искомые оценки уровней подготовленности участников теста и/или уровней трудности тестовых заданий показателей, достаточно знаний элементарной математики (школьный уровень). И если использовать термин «элементарный», как синоним слову «простой», то такое обозначение вполне уместно.

Можно предложить и другое обоснование. Этот метод использует только элементарные (в математической терминологии) преобразования. Этот метод не использует ни статистических преобразований, ни других методов из специальных разделов математики. Единственным исключением можно считать применение простейших методов вычислительной математики, но и они соответствуют школьному уровню. Сам процесс вычислений предполагает лишь элементарный набор вычислительных инструментов — достаточно инженерного калькулятора или простейшего вычислительного приложения на персональном компьютере.

Элементарный метод решения основной задачи педагогических измерений основывается на четырёх положениях:

1. Следствие из свойств математической модели и определения опорных испытуемых и заданий.

Если в качестве опорного установить задание с относительным уровнем трудности $\beta = 1$, и предоставить любому испытуемому ответить на набор опорных заданий, то в пределе бесконечно большого количества таких заданий мы получим, что доля правильных ответов X_i совпадает с вероятностным показателем уровня подготовленности этого испытуемого: $X_i = u_i$. Этот вывод следует из того факта, что при $\beta = 1$ вероятность правильного ответа испытуемым совпадает с вероятностным показателем уровня подготовленности испытуемого. При достаточно большом количестве испытаний значение доли правильных ответов приближается к вероятности этого события (сходимость по вероятности).

Математическое обоснование этого положения сводится к определению вероятности. Пусть имеется N опорных заданий с уровнем трудности $\beta = 1$. Тогда испытуемый с уровнем подготовленности θ правильно ответит на каждое такое задание с вероятностью, совпадающей с вероятностным показателем уровня его подготовленности: $P = u = 0{,}5^{1/\theta}$. Если предоставить этому испытуемому выполнить очень большое число опорных тестовых заданий и подсчитать долю X правильных ответов на них, то она будет примерно равна теоретической вероятности u. При N, стремящемся к бесконечности, мы получим $X = u$ — статистическое определение вероятности.

Совершенно несложно проиллюстрировать это положение на классической модели «игральная кость». Если принять шестигранную кость за опорную, то для опорного испытуемого у этой кости закрашено три грани. Тогда при подбрасывании кости вероятность того, что сверху окажется закрашенная грань (благоприятный исход события), будет равна 0,5.

Рассмотрим менее подготовленного испытуемого. Для него у опорной кости закрашено только две грани. Тогда вероятность того, что при её подбрасывании сверху окажется закрашенная грань, рассчитывается и равна отношению числа закрашенных граней к общему числу граней $^2/_6 = ^1/_3$ (это значение аналогично рассчитанной вероятности u, с такой вероятностью испытуемый правильно выполнит опорное задание теста). Если провести испытания с набором из N опорных костей и рассчитать долю благоприятных исходов X (т.е. вычислить отношение благоприятных исходов к общему числу испытаний), то получим значение, близкое к $^1/_3$. При очень большом наборе опорных костей значение X будет приближаться к рассчитанной вероятности $^1/_3$.

Аналогичные рассуждения можно провести и для группы опорных испытуемых, которая выполнит любое тестовое задание с долей правильных ответов $X_j = v_j$.

2. Определения опорного набора тестовых заданий и опорной группы испытуемых. Эти определения являются второстепенными и их введение необходимо для демонстрации элементарного метода. Приведём эти определения. Набор тестовых заданий называется опорным, если опорный испытуемый выполнит этот набор со средним арифметическим значением вероятностей правильного ответа равным 0,5. Группа испытуемых называется опорной, если она выполняет опорное задание со средним арифметическим значением вероятностей правильного ответа равным 0,5.

Для иллюстрации этого положения продолжим использовать игральную кость. Теперь для опорного испытуемого опорным набором может быть следующий набор из трёх костей: у первой кости закрашено две грани, у второй — три, у третьей — закрашено четыре грани. Средняя вероятность благоприятного исхода при подбрасывании этих костей равна 0,5. Несложно привести и другие примеры опорных наборов состоящих из большого числа разноокрашенных костей.

Для менее подготовленного испытуемого опорный набор уже будет и менее окрашенным: у первой кости закрашена одна грань, у второй — две, у третьей — три грани. Или другой вариант: у первой кости все грани закрашены, две другие вообще не закрашены. В обоих случаях средняя вероятность благоприятного исхода при подбрасывании этих костей равна $^1/_3$.

3. Примем следующее предположение: если при решении задачи педагогических измерений заменить набор опорных заданий опорным набором заданий, то такая замена не отразится на результатах. Аналогичные рассуждения справедливы и для набора испытуемых.

Логика такого предположения аналогична следующей: доля благоприятных исходов при подбрасывании набора опорных костей примерно равна доле благоприятных исходов при подбрасывании опорного набора костей.

Безусловно, набор опорных элементов не эквивалентен опорному набору элементов, и такая замена нетождественна. Это предположение, а не утверждение. Но по нашим оценкам, проведённым с помощью имитационного тестирования, такое утверждение оправданно, и изменение результатов тестирования от такой замены лежит в области принципиально существующих погрешностей. Более детально мы этот вопрос не исследовали.

4. Свойство инвариантности преобразований. Мы выше указывали, что можно предложить множество опорных наборов. Это положение снимает такой произвол. Оно устанавливает правила составления единственного опорного набора исходя из выбранной системы отсчёта.

Преобразования наблюдаемых значений в искомые оценки должны оставлять неизменными и отношения уровней трудности тестовых заданий для любого испытуемого, и отношения уровней подготовленности испытуемых для любого тестового задания. Требование неизменности, или инвариантности, таких оценок и есть принципиальное требование к преобразованиям в элементарном методе решения задачи.

Инвариантность преобразований на примере игральных костей можно проиллюстрировать следующим образом. Если для опорного испытуемого у первой кости площадь окрашенной поверхности в два раза больше, чем у второй кости, то такая же пропорция сохранится и для любого другого испытуемого. Другое дело, что у менее подготовленного испытуемого площади этих окрашенных поверхностей будут меньше.

Прежде чем показать примеры расчётов, приведём алгоритм решения основной задачи педагогических измерений элементарным методом:

1. По дихотомической таблице данных X_{ij} определяем доли правильных ответов испытуемыми X_i, на тестовые задания X_j, всего теста X.

2. Выбираем систему отсчёта. Она позволит наложить ограничение, снимающее неопределённость в таблице наблюдаемых данных.

3. Исходя из требований системы отсчёта, инвариантно преобразуем доли правильных ответов испытуемыми на все задания теста X_i и/или доли правильных ответов на задания теста всеми испытуемыми X_j в искомые оценки, выраженные вероятностными показателями u и v. Показатель преобразований k рассчитываем или по предложенной ниже аналитической формуле, или находим более точно численными методами. Этот шаг алгоритма выполняется на основе четырёх, изложенных выше, положений.

4. Рассчитываем относительные показатели. Если в качестве эталона выбран нейтральный уровень, то по найденным вероятностным показателям определяем их относительные эквиваленты θ и β. Если выбран другой эталон, то показатели q и в можно пересчитать под новый эталон по инвариантным правилам.

5. Для проверки расчётов и/или получения дополнительных эквивалентных решения, можно пересчитать искомые оценки из другой системы отсчёта. Рекомендуем не опускать этот шаг.

6. Рассчитать отношения показателей θ и β. Именно они обладают свойствами внутренней объективности. Их значения не зависят от выбора системы отсчёта. При интерпретации результатов следует опираться на эти значения.

7. Оценить надёжность результатов. Под этим шагом следует понимать внутреннюю оценку надёжности результатов. Т.е через сопоставление решений из разных систем отсчёта или другими приёмами в рамках элементарного метода.

Пример расчёта

Таким образом, под системами отсчёта мы понимаем такую «позицию исследователя», которая своими дополнительными условиями относительно свойств участников теста снимает существующую в таблице эмпирических данных неопределённость.

Приведём две такие системы.

1. Первую систему отсчёта будем обозначать как «Нейтральная группа испытуемых». В этой системе отсчёта предполагается, что средний уровень подготовленности группы испытуемых равен $u_{\text{ср}} = 0{,}5$. Если перейти к относительному показателю, то это значение соответствует нейтральной единице измерительной шкалы $\theta = 1$. Правила преобразования исходных баллов в искомые оценки уровней подготовленности u должны удовлетворять свойству инвариантности.

2. Система отсчёта «Нейтральный набор тестовых заданий». В этой системе отсчёта средний уровень трудности тестовых заданий равен $v_{\text{ср}} = 0{,}5$. Относительный показатель соответствует нейтральной единице $\beta = 1$. Преобразования исходных баллов в искомые оценки уровней подготовленности u должны быть инвариантными.

Проведём анализ данных в этих системах отсчёта. Система отсчёта — нейтральная группа испытуемых. В первых строках таблиц 1 и 2 обозначены тестовые задания и испытуемые, наблюдаемые доли правильных ответов.

Формируем опорную группу испытуемых, т.е. так преобразуем наблюдаемые доли правильных ответов группы испытуемых, чтобы группа выполнила опорное задание со средним арифметическим значением вероятностей правильного ответа равным 0,5. Условию инвариантности удовлетворяют преоб-

Таблица 1

Показатели тестовых заданий (нейтральная группа испытуемых)

Номер тестового задания, j	**1**	**2**	**3**	**4**	**5**
Доля правильных ответов X_j	0,7	0,5	0,4	0,3	0,1
$v_j = X_j$	0,7	0,5	0,4	0,3	0,1
$\beta_j = \log_{0,5} v_j$	0,51	1	1,32	1,74	3,32
β_j / β_5	0,15	0,30	0,40	0,52	1

Таблица 2

Показатели испытуемых (нейтральная группа испытуемых)

Номер тестового задания, i	**1**	**2**	**3–8**	**9**	**10**
Доля правильных ответов X_i	0,8	0,6	0,4	0,2	0
$u_i = X_i^k$	0,85	0,70	0,52	0,32	0
$\theta_j = \log_{0,5} u_j$	4,38	1,91	1,07	0,61	0
θ_i / θ_1	1	0,44	0,24	0,14	0

разования $u_i = X_i^k$. Подбираем необходимое значение показателя преобразований: $k_u \approx 0{,}710$. По полученным оценкам вероятностных показателей u рассчитываем относительные показатели $\theta = 1/\log_{0,5} u$. Результаты этих расчётов представлены в третьей и четвёртой строках табл. 2.

Предполагая, что опорная группа испытуемых выполняет набор тестовых заданий с тем же результатом, что и группа опорных испытуемых, получаем следующие оценки вероятностных $v_j = X_j$ и относительных $\beta = \log_{0,5} v$ показателей уровня трудности тестовых заданий. Результаты этих расчётов представлены в третьей и четвёртой строках табл. 1.

Интерпретацию результатов начнём с задания 2. Его не следует путать с опорным заданием. Если у опорного, идеализированного задания, относительный уровень трудности по определению равен единице, то у реального задания 2 значение $\beta = 1$ является приближённой оценкой. В работе[4] предложены расчётные формулы, по которым можно оценить и надёжность этого измерения. При $P = 0{,}5$ и десяти испытуемых имеем стандартное отклонение $\sigma_\beta \approx 0{,}13$, т.е. можно предположить, что истинное значение в лежит в интервале от 0,87 до 1,13. Если оценить такие интервалы для соседних по уровню трудности тестовых заданий, то мы увидим, что области значений этих показателей не перекрываются с данной областью. Т.е., хотя относительная погрешность данного измерения и достаточно велика — более 10%, само задание достаточно надёжно идентифицируется в имеющемся наборе тестовых заданий.

В качестве примера опишем некоторые свойства задания 5. Уровень трудности этого задания примерно в 3,32 раза выше опорного (или второго) задания. Вероятностный уровень трудности $v = 0{,}1$. Как можно истолковать эти значения? Например, так: опорный испытуемый правильно выполнит 3,32 опорных заданий с той же совместной вероятностью $v = 0{,}1$, что и одно самое трудное пятое задание.

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

Аналогичные рассуждения можно проводить и для испытуемых, оценивая уровень их подготовленности относительно опорного задания. Например, первый испытуемый правильно выполнит опорное задание с вероятность 0,85, а второй только с вероятностью 0,7. Или, первый испытуемый в $^{4,38}/_{1,91} \approx 2,3$ раза лучше подготовлен к тесту, чем второй испытуемый. Можно и так, первый испытуемый в $^{4,38}/_{3,32} \approx 1,3$ раз «сильней» пятого задания. Ещё раз отметим, хотя десятый испытуемый по всем показателям и оценивается на неизмеримо низком уровне, его вклад в общем анализе результатов ненулевой.

В последних строках табл. 1 и 2 представлены отношения соответствующих показателей к наибольшему значению. Эти строки не только дают дополнительную информацию о распределении показателей внутри группы, но и являются независимыми относительно выбора системы отсчёта оценками.

Проведём аналогичные расчёты относительно системы отсчёта — нейтральный набор тестовых заданий. Результаты этих расчётов представлены в табл. 3 и 4.

При построении этого решения использованы преобразования: $v_j = X_j^k$, показатель преобразований $k_v \approx 0,724$; $u_i = X_i$.

Если анализ результатов теста в системе отсчёта с нейтральной группой испытуемых

Таблица 3

Показатели тестовых заданий (нейтральный набор заданий)

Номер тестового задания, j	**1**	**2**	**3**	**4**	**5**
Доля правильных ответов X_j	0,7	0,5	0,4	0,3	0,1
$v_j = X_j^k$	0,77	0,61	0,52	0,42	0,19
$\beta_j = \log_{0,5} v_j$	0,37	0,72	0,96	1,26	2,41
β_j / β_5	0,15	0,30	0,40	0,52	1

Таблица 4

Показатели испытуемых (нейтральный набор заданий)

Номер тестового задания, i	**1**	**2**	**3–8**	**9**	**10**
Доля правильных ответов X_i	0,8	0,6	0,4	0,2	0
$u_i = X_i$	0,8	0,6	0,4	0,2	0
$\theta_i = \log_{0,5} u_i$	3,11	1,36	0,76	0,43	0
θ_i / θ_1	1	0,44	0,24	0,14	0

со средним уровнем подготовленности u = 0,5, оценивает средний уровень трудности тестовых заданий как завышенный и равный доле правильных ответов на тест v = 0,4, то при переходе к системе отсчёта с нейтральным набором заданий изменяется и позиция исследователя. Теперь исследователь предполагает, что тестовые задания имеют средний уровень сложности v = 0,5, а испытуемые недостаточно подготовлены к тестовым заданиям и имеют средний уровень подготовленности, равный доле правильных ответов на тест u = 0,4. Однако, и в том, и другом случае средние доли правильных ответов на тест равны экспериментально наблюдаемому значению 0,4.

Теперь первый испытуемый правильно выполнит опорное задание с вероятность 0,8, а второй только с вероятностью 0,6. Снижение этих показателей связано с тем, что опорное задание в этой системе отсчёта сместилось и позиционируется между вторым и третьим заданиями теста. Однако, и в этой системе отсчёта первый испытуемый в $^{3,11}/_{1,36} \approx 2,3$ раза лучше подготовлен к тесту, чем второй испытуемый, первый испытуемый в $^{3,11}/_{2,41} \approx 1,3$ раза «сильней» пятого задания.

Одинаковые результаты расчётов, приведённые в последних строках таблиц, являются проверочными и подтверждают, что две системы отсчёта эквивалентны и используемые преобразования исходных баллов инвариантны.

Теоретически значения показателя преобразований k в рассмотренных решениях должны быть одинаковыми. Имеющиеся незначительные различия есть следствие применения численных расчётов для их определения. Строгой аналитической формулы для определения этого показателя мы найти не можем. Иногда удовлетворительным приближением для решения этой задачи может служить формула $k \approx (2X + 1)/(4(1 - X))$, где X — доля правильных ответов на тест. Можно также указать следующие закономерности: если доля правильных ответов $X < 0,5$, то $k < 1$; если $X > 0,5$, то $k > 1$; если $X = 0,5$, то $k = 1$. Т.е., если доля правильных ответов на тестовые задания равна 0,5, то значения вероятностных показателей можно оценить просто долями правильных ответов: $v_j = X_j$, $u_i = X_i$.

В данном примере мы использовали инвариантные преобразования только для одной группы показателей модели. В первом случае мы корректировали доли правильных ответов испытуемых X_i и получали искомые оценки u; во втором — из X_j получали оценки v. Такой подход связан с выбором указанных выше систем отсчёта. Теоретически допустимы и другие типы корректировок, в частности симметричная, как это делается в схеме Раша. Совершенно очевидно, что и эти преобразования будут инвариантны, если β/θ = const. В общем виде элементарный метод решения задачи педагогических измерений допускает множество систем отсчёта и, как следствие, множество вариантов преобразования наблюдаемых баллов X_i и X_j. Приведём и обозначим три наиболее очевидных варианта; X_i^k; X_j (нейтральная группа испытуемых) ↔ X_i; X_j^k (нейтральный набор тестовых заданий) ↔ $X_i^{Цk}$; $X_j^{Цk}$ (симметричный).

Несложно получить результаты и при симметричной корректировке исходных баллов. Однако, на наш взгляд, этот вариант имеет и недостатки — он одновременно смещает от центра опорные характеристики участников теста и размывает само понятие система отсчёта (табл. 5 и 6).

Но и в этом случае первый испытуемый в $^{3,67}/_{1,6} \approx 2{,}3$ раза лучше подготовлен к тесту, чем второй, и в $^{3,67}/_{2,81} \approx 1{,}3$ раза «сильней» пятого задания.

И наконец, приведём ещё один вариант решения, полученный достаточно простыми численными расчётами. Если для оценки искомых показателей использовать симметричные инвариантные преобразования $v_j = X_j^{Цk}$, $u_i = X_i^{Цk}$ и использовать критерий равенства среднего значения рассчитанных вероятностей правильного ответа на-

Таблица 5

Показатели тестовых заданий (симметричные преобразования)

Номер тестового задания, j	**1**	**2**	**3**	**4**	**5**
Доля правильных ответов X_j	0,7	0,5	0,4	0,3	0,1
$v_j = X_j^{\sqrt{k}}$	0,74	0,56	0,46	0,36	0,14
$\beta_j = \log_{0,5} v_j$	0,44	0,85	1,12	1,47	2,81
β_j/β_5	0,15	0,30	0,40	0,52	1

Таблица 6

Показатели испытуемых (симметричные преобразования)

Номер тестового задания, i	**1**	**2**	**3–8**	**9**	**10**
Доля правильных ответов X_i	0,8	0,6	0,4	0,2	0
$\upsilon_i = X_i^{\sqrt{k}}$	0,83	0,65	0,46	0,26	0
$\theta_i = \log_{0,5} u_i$	3,67	1,60	0,89	0,51	0
θ_i/θ_1	1	0,44	0,24	0,14	0

Таблица 7

Показатели тестовых заданий (численное решение)

Номер тестового задания, j	**1**	**2**	**3**	**4**	**5**
Доля правильных ответов X_j	0,7	0,5	0,4	0,3	0,1
$\nu_j = X_j^{\sqrt{k}}$	0,73	0,55	0,45	0,35	0,14
$\beta_j = \log_{0,5}\nu_j$	0,45	0,87	1,15	1,51	2,88
β_j / β_5	0,15	0,30	0,40	0,52	1

Таблица 8

Показатели испытуемых (численное решение)

Номер тестового задания, i	**1**	**2**	**3–8**	**9**	**10**
Доля правильных ответов X_i	0,8	0,6	0,4	0,2	0
$\upsilon_i = X_i^{\sqrt{k}}$	0,82	0,64	0,45	0,25	0
$\theta_i = \log_{0,5}u_i$	3,58	1,56	0,87	0,50	0
θ_i / θ_1	1	0,44	0,24	0,14	0

блюдаемому значению $X = 0,4$, то получим показатель преобразования $k \approx 0,753$. Это значение практически совпадает со значением $k = 0,75$, рассчитанным по приближённой аналитической формуле. Такое хорошее совпадение скорее случайность. Однако если удастся найти адекватную аналитическую зависимость $k(X)$, то элементарный метод решения задачи педагогических измерений преобразуется в элементарные формулы для расчёта искомых показателей.

В табл. 7 и 8 представлены результаты численного решения и анализа исходных данных. Совершенно очевидно, что и это решение эквивалентно предыдущим.

Заключение

В данной работе представлен простой приближённый метод решения основной задачи педагогических измерений — по наблюдаемым долям правильных ответов получить адекватные оценки латентных показателей участников теста. Показано, что для данного набора эмпирических данных можно предложить различные наборы решений. Различия в оценках объясняются выбором системы отсчёта. Эквивалентность полученных наборов решений следует из инвариантности используемых преобразований.