НЕКОТОРЫЕ ПОНЯТИЯ МАТЕМАТИЧЕСКОЙ ТЕОРИИ ПЕДАГОГИЧЕСКИХ ИЗМЕРЕНИЙ

Галина Смирнова
Славянский-на-Кубани государственный педагогический институт
smirnova_g_i@mail.ru

Данная работа является продолжением цикла статей, освещающих вопросы создания тезауруса педагогических измерений. Рассматриваются такие понятия теории педагогических измерений как идея измерений по модели Раша, единица измерений, проверка данных, смещение, сложение данных, современная теория педагогических измерений.

Ключевые слова: *идея измерений, единица измерений, проверка данных, смещение, сложение данных, современная теория педагогических измерений.*

Измерения по модели Г.РАША

Любые данные не могут быть объективны, если они не прошли обработку по модели Раша. Ведь измерения, проведённые посредством этой модели, могут служить показателем «научности» исследуемых данных. Специалисты в области общественных наук пробовали разработать единицу измерений по модели Раша для того, чтобы результаты измерения были признаны и в других сферах. Однако уверенность в том, что измерения по модели Раша неприменимы в общественных науках, а также в образовании, настолько велика, что все разработки обречены на недопонимание. Существует такое мнение, что измерения по модели Раша возможны не только в общественных науках, но и в других областях, и как следствие, они несовершенны. Однако специалисты в области общественных наук уже давно успешно применяют измерения по модели Раша в своей практике, но не осознают этого.

Осознание того, что измерения по модели Раша могут применяться в общественных науках, пришло к Luce и Turkey. Они заметили, что измерения по модели Раша являются ничем иным, как сравнением между откликами случайным образом составленных пар объектов и двумя случайными классами. Закон L.L. Thurstone содержал в себе данные, которые могут служить примерами измерений по модели Раша. Идеи измерения по модели Раша развивали Bredley и Terry в 1952 году, а так же Раш, в период с 1958 по 1977 гг.

Измерения по модели Раша вытекают из принципа «specific objectivity». Этот принцип G.Rasch и другой принцип — «estimation sufficiency» R.A. Fisher — два решения одной проблемы. Andersen в 1977 году показал, что только тот процесс измерения, который включает в себя «specific objectivity» и, как следствие, измерения по модели Раша, может иметь обоснованные статистические данные для своих параметров. Из этого следует, что обоснованные статистические данные являются первопричиной и они необходимы для измерения по модели Раша.

Некоторые авторы связывают подобные измерения с работой G.Rasch: Keats, R.A. Fischer и Brogden. Другие авторы — Perline, Wright и Vainer в 1977 году разработали два практических основания равенства неметрической многомерной шкалы (Kruskal, 1964, 1965 гг.) и измерений по модели Раша. Wright и Stone в 1979 году показали, как применять измерения по модели Раша в психологическом тестировании. Wright и Masters в 1982 году привели примеры удачного использования измерений на номинальной шкале и частично доверительном интервале.

В противовес этим публикациям существовало мнение, что продвижение, объяснение и приведение примеров успешного применения измерений по модели Раша в общественных науках, в частности при психологическом тестировании и преподавании являются не чем иным, как либо случайностью, либо практической ошибкой.

Единица измерения

L.L. Thurstone писал: «Линейный континуум, который подразумевается для всех измерений, всегда обобщает ... Все измерения подразумевают наличие некоего измерительного свойства...».

Г.Раш показал, что если вероятность правильного ответа испытуемого определять по формуле

$$P = \frac{\exp(b-d)}{G}, \qquad (1)$$

где

$$G = [1+\exp(b-d)], \qquad (2)$$

где b — значение уровня подготовленности испытуемого, а d — уровня трудности задания, и если событие АВ заключается в том, что испытуемый A правильно ответил на задание, но испытуемый B — неправильно, то тогда событие ВА заключается в том, что испытуемый B правильно ответил на задание, а испытуемый A — неправильно. Тогда расстояние между испытуемыми A и B на шкале будет определяться посредством формулы:

$$b_A - b_B = \log N_{AB} - \log N_{BA}, \quad (3)$$

где N_{AB} — это количество правильных ответов испытуемого A и неправильных испытуемого B и N_{BA} — это количество правильных ответов испытуемого B и неправильных испытуемого A на подмножестве всех заданий.

Данные расчёты необходимы для того, чтобы, используя модель Раша, можно было вычислить трудность d любого задания:

$$P_{AB} = P_A(1 - P_B) = \frac{\exp(b_A - d)}{G_A G_B}, \quad (4)$$

и

$$P_{BA} = P_B(1 - P_A) = \frac{\exp(b_B - d)}{G_A G_B}, \quad (5)$$

таким образом, d, G_A and G_B, сводят на нет значение $\frac{P_{AB}}{P_{BA}} = \exp(b_A - b_B)$, обнуляя

$$\log \frac{P_{AB}}{P_{BA}} = b_A - b_B = \log \frac{N_{AB}}{N_{BA}}. \quad (6)$$

Так вычисляется дистанция между испытуемыми, которая не зависит от значения d.

С тех пор как d не появлялось в этой формуле, значение дистанции между A и B моделировалось таким образом, чтобы оно статистически было равно любому значению сложности задания d.

Любой из двух подходящих наборов данных соответствует модели Раша и может быть использован для последующих вычислений и интерпретации результатов. Счёт и сложение – это именно те характеристики, которые соответствуют модели Раша и никакой другой. Когда мы приняли, что счёт необходим для измерений, во всяком случае для нашего случая, тогда мы приняли и модель Раша в качестве математического объяснения того, что мы делаем при измерениях, а также как оправдывающее обстоятельство при математических расчётах.

Определение термина ***«единица измерений»*** таково: единицей измерения является определённый процесс, который может быть повторен без изменений на различных этапах измерительного континуума; для измерений по методу Раша необходима дополнительная операция, которая сохраняет значение единицы измерения, и эта операция — сложение.

Проверка данных, смещение

Какой вид должны иметь данные для того, чтобы соответствовать модели Раша? Наиболее разумный ответ таков: «Вполне достаточно решать именно те практические задачи, которые могут быть исследованы с помощью модели Раша, поэтому вполне достаточно иметь определённое количество инвариантных данных, которые можно собрать вручную».

Как мы можем обосновать степень инвариантности модели Раша, когда речь идёт о конкретном наборе данных? Один из методов таков: точно определить подмножество заданий, которое нас интересует, но одновременно оно должно быть независимого от набранного балла отдельного испытуемого, который только что прошел тестирование (N_{AB} N_{BA}), и далее нужно сравнить сумму баллов данного подмножества с эквивалентной ей дистанцией между испытуемыми.

Расстояние между испытуемыми A и B инвариантно, то есть неизменно, и состоит из отдельных заданий, следующих друг за другом, составляя единое пространство, которое может быть исследовано с помощью модели Раша.

Наиболее общий способ проверки и записи соответствующих данных – это составление для каждого ответа равенства $x = 0$ или 1 и вычисление разности:

$$y = x - Ex = x - P, \qquad (7)$$

где

$$P = \frac{\exp(b-d)}{[1+\exp(b-d)]}. \qquad (8)$$

Это значение вычисляется с использованием текущих оценок уровня знаний испытуемого b и уровня сложности задания d и ожидаемым значением Ex при проведении наблюдений за x

$$Ex = P \qquad (9)$$

и далее собирается в одно подмножество все полученные значения.

Если $(b_1 - b_0)$ определяется как протяжённость, которая формируется из подмножества заданий, выраженных значением b_1, недостаточного для формирования единого целого пространства, состоящего из набора заданий, который включает в себя значение b_0, то это подмножество последовательно суммируется по $\sum y$:

$$Ey = (b_1 - b_0)\sum \frac{dy}{db}, \qquad (10)$$

здесь суммирование $\sum$ идёт над заданиями указанного подмножества.

Когда данные соответствуют модели Раша, тогда разность y относительно b равна расхождению в баллах $P\,(1 - P)$, поэтому

$$\frac{dy}{db} = \frac{dP}{dB} = P(1-P) = q, \qquad (11)$$

$$Ey \approx (b_1 - b_0)\sum q, \qquad (12)$$

и

$$(b_1 - b_0) = \frac{\sum y}{\sum q} = g. \qquad (13)$$

Таким образом, простое отношение $g = Ey / Eq$ показывает несовместимость логита для инвариантной шкалы ($b_1 - b_0$) потому, что подмножество задания специфично, а также для g может быть определено математическое ожидание и расхождение

$$Eg = 0 \text{ and } Vg = \frac{1}{\sum q}. \qquad (14)$$

Тогда данные будут соответствовать модели Раша.

Необходимо, чтобы подмножества заданий были не ограничены. Таким образом, определение понятия ***«проверка данных»*** таково: для соответствия модели Раша данные должны иметь определённый вид. Для этого необходимо точно определить подмножество заданий, которое нас интересует, но, одновременно, оно должно быть независимого от набранного балла отдельного испытуемого, который только что прошел тестирование, и далее нужно сравнить сумму баллов данного подмножества с эквивалентной ей дистанцией между испытуемыми.

Понятие *«смещение»* имеет следующую формулировку: когда данные соответствуют модели Раша, тогда разность y относительно b равна расхождению в баллах $P(1 - P)$, и простое отношение $g = Ey / Eq$ показывает несовместимость логита для инвариантной шкалы ($b_1 - b_0$). Когда для g определено математическое ожидание и расхождение, тогда данные будут соответствовать утверждённому единству модели Раша. То есть необходимо, чтобы подмножества заданий были не ограничены. Группы испытуемых могут использовать для проверки пространство, в котором любое задание может быть переформулировано «за» или «против» определённой группы испытуемых. В общем, любое сочетание заданий и испытуемых вплоть до их взаимодействия может влиять на формирование единого целого пространства, которое, в свою очередь, может быть использовано для определения подмножества вычисляемого g.

Сложение данных

Линейные шкалы основаны на операции сложения, которая отвечает на вопрос: «Если испытуемый A имеет более высокий уровень подготовленности b_A, чем испытуемый B с уровнем подготовленности b_B, то насколько необходимо увеличить показатель b_B для того, чтобы характеристики испытуемого B сделать примерно равными характеристикам испытуемого A?» Или, говоря более образно: «Что необходимо прибавить к b_B, чтобы было верно равенство $P_B = P_A$?»

Отвечая на этот вопрос, мы должны осознавать, что единственная ситуация, в которой вероятность свершения данного события близка к единице, это сопоставление испытуемых и заданий определённым образом. В связи с этим меняется формулировка вопроса: «Что должно измениться в ситуации, когда мы узнаем об уровне подготовленности испытуемого путем тестирования его такими заданиями, которые позволяют испытуемому B отвечать правильно на вопросы с такой же вероятностью, что и испытуемый A?» Говоря другими словами: «Что необходимо прибавить к b_B, чтобы было верно равенство $P_{Bj} = P_{Ai}$?»

Или же: «Какое задание j, имеющее сложность d_j, может сделать характеристики испытуемого B одинаковыми с характеристиками испытуемого A при ответе на вопрос i ?»

Измерения по модели Раша определяют, что $P_{Bj} = P_{Ai}$, тогда

$$d_B - d_j = d_A - d_i. \qquad (16)$$

Подобное «сложение» отвечает на вопрос, почему B похоже на A тогда, когда:

$$b_B + (b_A - b_B) = b_A. \qquad (17)$$

И это «сложение» доведено до логического конца, когда испытуемый B отвечает на задание j, которое обладает следующей сложностью:

$$d_i - d_j = b_A - b_B, \qquad (18)$$

оно легче, чем задание i, то есть задание j имеет следующую сложность:

$$d_j - d_i = b_A - b_B. \qquad (19)$$

поэтому

$$b_B + (b_A - b_B) = b_B + (d_i - d_j) = b_A \qquad (20)$$

и

$$P_{Bj} = P_{Ai}. \qquad (21)$$

Правильность подобного «сложения» доказана тогда, когда ответ испытуемого B на задания подобные заданию j оценен верно и статистически равен ответу испытуемого A на задания подобные заданию i. Фактически, это сравнение проверяется соответствующим подробным анализом данных.

Таким образом, определение понятия «сложение данных» таково: линейные шкалы основаны на операции сложения, которая отвечает на вопрос: «Если испытуемый A имеет более высокий уровень подготовленности b_A, чем испытуемый B с уровнем подготовленности b_B, то сколько «знаний» необходимо добавить к b_B для того, чтобы характеристики испытуемого B сделать примерно равными характеристикам испытуемого A?» Или, говоря более образно: «Что необходимо прибавить к b_B, чтобы было верно равенство $P_B = P_A$?»

Счёт и измерение

В общественных науках традиционно закрепилось понятие, что исследования основываются на подсчёте правильных ответов (подсчитывать и располагать их в порядке согласно показанным возможностям) и затем использовать полученные данные и однообразные их преобразования в качестве измерений. Когда вопросы содержат только два варианта ответа, тогда мы подсчитываем правильные ответы. Когда же вопросы содержат более двух вариантов ответов, тогда подсчитываем от «минимальной» до «максимальной» (от «наихудшей» до «наилучшей», от «самой слабой» до «самой сильной») категории, которые являются наиболее правильными. В общественных науках практически все численные данные представлены в подобной форме скорее всего потому, что их просто собирать.

Если бы существовал какойлибо прогресс в представлении количественных данных в общественных науках, то данный вид подсчёта данных был бы востребован. Однако он существует как приложение. Подсчёт в таком виде подразумевает измерительный процесс, не никакой-нибудь, а специфический. Подсчёт подразумевает процесс, который выводит счёт в категорию необходимых и обоснованных процедур.

Сегодня подсчёт — это уникальная обоснованная статистическая процедура оценки измерений при использовании модели Раша. С тех пор как измерения на основе модели Раша приобрели статус совместного измерения, данные являются надёжными для подобного построения и счёт на практике является первым шагом для дальнейшего измерения. Все, что нам необходимо сделать, это использовать это приложение для наших измерений и анализа, контролируя полученные величины, ведь с ними мы работаем на основе модели Раша. Когда наши данные будут сформированы таким образом, что их можно будет использовать, тогда мы сможем воспользоваться полученными числовыми величинами для построения линейных шкал Thurstone и с их помощью обозначить измерения Luce и Turkey.

Таким образом, мы изложили правила, работающие преимущественно в общественных науках и образовании. Для того, чтобы использовать полученные при помощи счета данные, мы просто суммируем правильные ответы, так как наша «достаточно хорошая» система статистической обработки 90-х годов проверяет наши латентные утверждения на наличие данных, с которыми нам придётся работать, и их полезность и возможность использования в других не менее сложных моделях, чем модель Раша. Необходимо запомнить, что среди всех известных способов математической обработки данных, которые могут трансформировать обыкновенные данные подсчёта в данные для измерений, только модель Раша делает это наиболее точно, используя сложение и выдавая итоговые данные в числовом значении.

С тех пор как измерения на основе модели Раша приобрели статус измерения, суммирование представляется первым шагом для дальнейшего измерения. Подсчёт данных — это обоснованный статистический процесс оценки измерений при использовании модели Раша.