

## АНАЛИЗ ТАБЛИЦ РЕЗУЛЬТАТОВ ПЕДАГОГИЧЕСКИХ ИЗМЕРЕНИЙ ПОСРЕДСТВОМ ПРОГРАММЫ MINISTER. ЧАСТЬ 1

Галина Смирнова,  
Алексей Смирнов

Славянский-на-Кубани государственный педагогический институт  
smirnova\_g\_i@mail.ru

По результатам анализа определённого набора данных программа Minister может построить 124 таблицы, которые скомпонованы в 34 группы. Наиболее значимые таблицы программа разбивает на шесть блоков. В данной статье начинается описание таблиц, представленных в первом из шести блоков. В частности, рассматривается таблица 3.2 под названием «Градуированная оценка» (Rating (partial credit) scale).

*Ключевые слова:* таблицы результатов расчёта (Output Tables menu), градуированная оценка (Rating (partial credit) scale), уровень трудности ответа (step difficulty), пороговое значение Раша-Эндрича (Rasch-Andrich threshold), градуация шкалы (step calibration), дифференцирующая способность задания (item discrimination), график вероятности ответов испытуемых на задания теста (the probability curves).

### Таблицы программы MINISTER

Перечень таблиц результатов расчёта программы Minister описан в указателе (Index), ознакомиться с которым можно, обратившись к пункту меню «Помощь» — «Содержание» — «Таблицы результатов расчёта» («Help» — «Contents» — «Output Tables Index») (рис. 1).

При выборе соответствующей таблицы из пункта меню «Таблицы результатов расчёта» (рис. 2) создаётся временный файл, который автоматически вписывается в текущей отчёт анализа данных. Для этого в контрольном файле надо указать TABLES= или TFILE=1. Таблица 0 (Table 0), в которой указывается авторство и версия программы, всегда записывается в отчёт автоматически.

Теория

12/0000

1

Подробнее см.:  
Г. Смирнова, А. Смирнов.  
Анализ данных в программе Minister//Педагогические измерения. 2010. № 2. С. 81–93.

**ПЕД**  
**измерения**

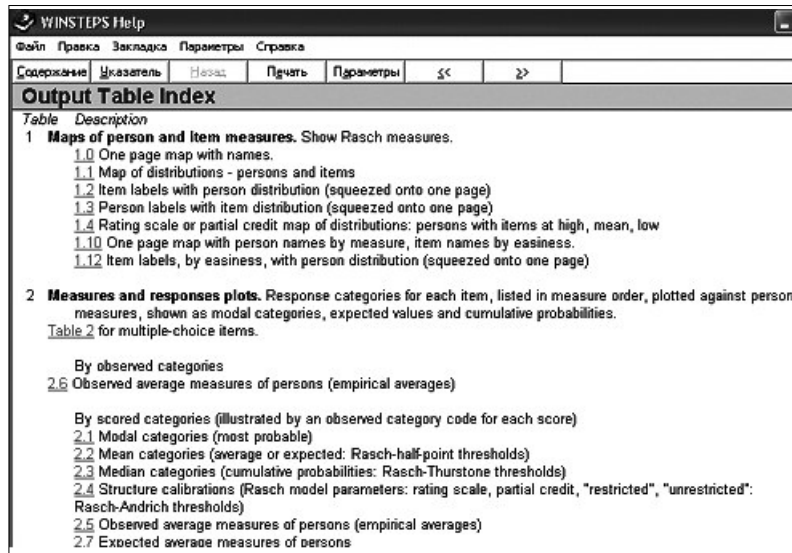


Рис. 1. Меню пункта «Таблицы результатов расчёта» (Output Tables menu)

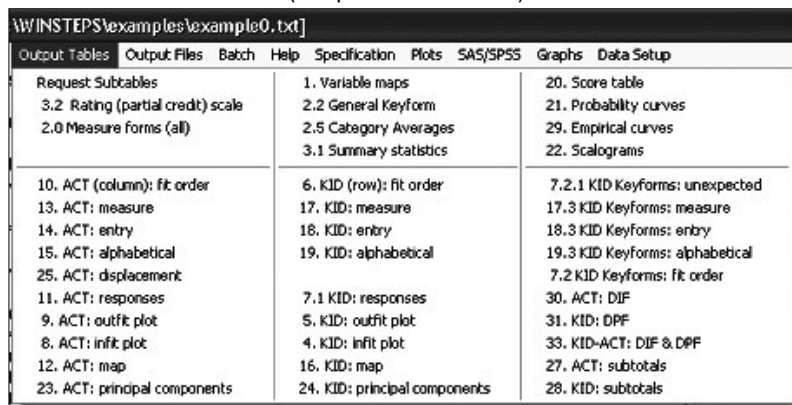


Рис. 2. Таблицы результатов расчёта

2

В качестве примера в статье, как и в предыдущих работах по данной теме, рассмотрены данные файла example0.txt, расположенного в Справке (Help) – Содержание (Content) – Примеры контрольных файлов и файлов данных (Examples of Control and Data files) – Пример0.txt (Example0.txt).

Как видно из рис. 2, меню содержит шесть блоков, в которых отображены названия доступных таблиц для обрабатываемых данных. Рассмотрим первую таблицу в первом блоке.

### Таблица 3.2 «Градуированная оценка» (Rating (partial credit) scale)

Наведите курсор на данный пункт меню и нажмите его для того, чтобы данные<sup>2</sup> были обра-

ботаны и сформированы во временный файл, который будет выведен на экран в виде таблицы (табл. 1). Имя этого файла 03-859ws.txt. Расшифруем название файла: «03» — означает порядковый номер таблицы в указателе (Index), «859» — уни-

кальный номер для данного анализа, «ws.txt» расшифровывается как «Текстовый файл программы Winsteps» («Winsteps text file»).

Средние арифметические результатов измерения описывают градации ответов по уровню трудности.

**Теория**

Таблица 1

Таблица 3.2. Научные предпочтения (03-859ws.txt)  
 Исходные данные: 75 испытуемых, 25 заданий  
 Измеряемые данные: 75 испытуемых, 12 заданий с тремя градациями ответов в по уровню трудности.  
 Критерия характеристика градаций ответа. Модель = «R»

Метка градации	Сумма градаций	Наблюдяемое мое значение	%	Среднее OBSVD	Ожидается значение	INFIT MNSQ	OUTFIT MNSQ	Проверка структуры	Измеряемые градации	
0	0	667	33	-1.30	-1.30	.96	.95	NONE	(-2.04)	00 dislike
1	1	757	37	-.08	-.09	.90	.78	-.82	.00	01 neutral
2	2	609	30	1.40	1.41	1.09	1.33	.82	(2.04)	02 like

В оригинале таблица выглядит следующим образом:

TABLE 3.2 LIKING FOR SCIENCE (Wright & Masters p. 20U859ws.txt Oct 9 10:54 2002  
 INPUT: 75 PUPILS, 25 ACTS MEASURED: 75 PUPILS, 12 ACTS, 3 CATS WINSTEPS 3.36 SUMMARY OF CATEGORY STRUCTURE. Model="R"

CATEGORY	OBSERVED	OBSVD	SAMPLE	INFIT	OUTFIT	STRUCTURE	CATEGORY			
LABEL	SCORE	COUNT	% AVRGE	EXPECT	MNSQ	MNSQ	MEASURE			
0	0	667	33	-1.30	-1.30	.96	.95	NONE	(-2.04)	00 dislike
1	1	757	37	-.08	-.09	.90	.78	-.82	.00	01 neutral
2	2	609	30	1.40	1.41	1.09	1.33	.82	(2.04)	02 like

---  
 AVERAGE MEASURE is mean of measures in category.

Создание таблицы контролируется кодами STEPT3=, STKEEP=, MRANGE=. «Средние арифметические результатов измерения» (average measure) и статистики адекватности данных модели (fit statistics) показывают, как отвечают испытуемые на ответы в данном задании.

Рассмотрим подробнее термин «средние арифметические

результатов измерения» (average measure). «Средние арифметические результатов измерения» для каждого из трёх ответов на тестовое задание являются средними значениями уровня подготовленности испытуемых при ответе на данное тестовое задание. Это эмпирическое понятие. Оно не принадлежит к параметрам модели Раша.

ПЕД	
	измерения

Каждый последующий уровень трудности («step difficulty»), пороговое значение Раша-Эндрича (Rasch-Andrich threshold), шаг калибровки (step calibration) и другие отображают значения логитов каждой из последующих градаций. В этом заключается полезность и необходимость модели. И разность логитов уровня подготовленности испытуемых, и трудности заданий образуют значения параметра модели Раша.

Теория Раша основана на положении, что испытуемые, отвечающие на более трудные ответы, должны иметь более высокий уровень подготовленности. В этой теории утверждается, что «каждая градация ответа должна быть наиболее предсказуемой на протяжении всей латентной переменной». Однако и противоположные теории имеют право на существование. Например, для того чтобы увеличить дифференцирующую способность задания (item discrimination), можно изменить шкалу измерений, хотя визуально такая шкала ничем не будет отличаться от прежней.

Отношение понятий «средние арифметические результатов измерения», «уровень трудности», «дифференцирующая способность задания» является комплексным. То есть:

*Уровень трудности* =  $\log$  ((низшая градация) / (высшая градация)) + (среднее изме-

рение между двумя градациями) — нормализатор<sup>3</sup>.

Нормировано так, что Сумма (шаг шкалы) = 0.

Таким образом:

1) высшая градация с наибольшей частотой соотносится с низшей (наиболее низкой) на определённый уровень трудности (и/или дифференцирующей способности задания);

2) и наибольшее среднее значение данных измерений для обеих градаций — наибольшее (самое положительное) шага градации (и/или дифференцирующей способности задания);

3) поскольку уровень трудности заданий измеряется как составная градация (набор), взаимосвязь между этими двумя градациями подвержена влиянию связи с любой другой градацией. Из этого следует, что даже если градация явно не видна, её влияние отразится на величине шага всей шкалы измерения.

Для каждого шага градации  $k$ , где уровень подготовленности испытуемого —  $\theta_j$  и уровень трудности задания —  $\beta_j$ , справедливо:

*Среднее измерение* =  $\Sigma(\theta_j - \beta_j)$  / число шагов для данной градации. Это не является оценкой параметра.

График вероятности ответов испытуемых на задания теста (the probability curves) показывает, как выбирается испыты-

3 В оригинале формула выглядит следующим образом:

Step difficulty =  $\log$  ((count in lower category) / (count in higher category)) + (average of the measures across both categories) — normalizer

емыми данный ответ, то есть график прогнозирует работу ответа, при условии, что для данного примера она удовлетворительна.

Логика подобных размышлений проста: если «средние арифметические результатов измерения» (average measure) и статистики адекватности (fit statistics) некорректны для одного примера, то почему они

должны быть таковыми для других примеров? Если же они выглядят достаточно хорошо для конкретного примера, тогда график задания сможет показать нам приемлемость для будущих примеров. Если они неверны на данном этапе, то, увидев это, можно предотвратить проблему в будущем.

А) Пример для дихотомических заданий:

**Теория**

Таблица 2

```

SUMMARY OF MEASURED STRUCTURE
FOR GROUPING "0", MODEL "R", ACT NUMBER:      12  GO TO MUSEUM

ACT          MEASURE OF  -1.07  ADDED TO MEASURES
-----
|CATEGORY  OBSERVED|OBSVD SAMPLE|INFIT  OUTFIT| COHERENCE|ESTIM|
|LABEL SCORE COUNT %|AVRGE EXPECT|  MNSQ  MNSQ| M->C  C->M|DISCR|
-----
|  1    1      13  18|  -0.38  0.01|  0.83  0.52|  75%  23%|      |01 neutral
-----
|  2    2      61  82|  1.12  1.03|  0.78  0.85|  85%  98%|  1.23|02 like
-----
AVERAGE MEASURE is mean of measures in category.
M->C = Does Measure imply Category?
C->M = Does Category imply Measure?
    
```

Метка градации (CATEGORY LABEL) — это число градаций в данных после подсчёта.

Сумма градации (CATEGORY SCORE) — это порядковое значение градации, которое используется компьютером для предварительных расчётов, а так же в таблице 2.

Наблюдаемое значение (OBSERVED COUNT) и % — показывает, сколько раз данная градация была включена в вычисления (к примеру, для тех

тестовых заданий и испытуемых, которые не являются экстремальными).

Среднее OBSVD (OBSVD AVERAGE) — это среднее значение измерений, которое было смоделировано для правильного ответа в данной градации. Ожидается, что среднее значение будет увеличиваться с увеличением шага градации. Отклонения от данного утверждения помечаются «\*». То есть это указание на описание примера,

**ПЕД**  
**измерения**

который не соответствует параметру Раша. Если для каждого шага градации  $k$ , существует испытуемый  $\theta_j$  и тестовое задание  $\beta_i$  тогда:

$$\text{Среднее значение} = \frac{\sum(\theta_j - \beta_i)}{\text{число шагов градации}}$$

### Отклонения шкалы измерений

Вернёмся к отклонениям. Среди специалистов нет единого мнения по поводу статуса рейтинговой шкалы (или доверительного интервала), которые позиционируются как «отклонения» («disorder»). Существует два вида отклонений:

1) Отклонения «средних арифметических результатов измерения» («average measures») для градаций, которые могут быть вызваны неверной формулировкой (рис. 3).

В данном примере<sup>4</sup> автор сознательно организовал неверный порядок данных в градаци-

ях. Это видно из неправдоподобных результатов «средние арифметические результатов измерения» («average measures») или «эмпирических средних» («observed averages»), где средний уровень подготовленности испытуемого вычисляется в каждой градации. На неверный порядок данных также указывают значение статистики адекватности (fit statistics). Общеизвестно, что такие понятия как «измерение структуры шкалы» («scale structure measures»), «шаг калибровки» («step calibrations»), «уровень трудности» («step difficulties»), «пороговое значение Раша-Андрича» («Rasch-Andrich thresholds»), «допустимые ошибки» («deltas») и другие требуют порядка в данных. А если данный порядок нарушен, то это сразу проявляется в их значениях.

2) Отклонения «шага шкалы» или «пороговой величины Раша-Эндрича» предполагает наиболее редкие промежуточ-

FIM LEVEL	COUNT	AVERAGE MEASURE	INFIT MNSQ	OUTFIT MNSQ	STEP CALIBRATN
1 (2)	88	-1.97	1.47	1.41	NONE
2 (1)	96	2.18	.54	.69	-2.08
3	101	-.95	1.05	1.02	-1.49
4	168	-.25	.91	.99	-1.24
5	210	.80	.97	.87	.08
6	146	2.14	.66	.75	1.87
7	101	3.02	.83	.86	2.86

Рис. 3. Пример отклонения «средних арифметических результатов измерения» для градаций

<sup>4</sup> Linacre J.M. Category Disorder vs. Step Disorder, Rasch Measurement Transactions. 1999. 13:1 p. 675, «FIM™ Level».

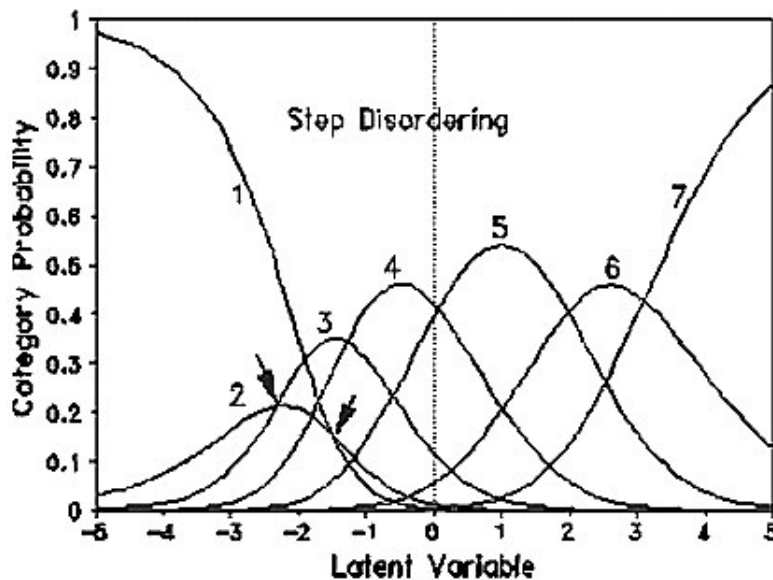


Рис. 4. Пример отклонения «шага калибровки» или «порогового значения Раша—Эндрича»

ные градации, т.е. они взаимодействуют друг с другом в узких интервалах на шкале латентной переменной (рис. 4).

В следующем примере FIM-градации<sup>5</sup> организованы корректно, однако частота второго уровня уменьшена из-за некоторых действий в процессе получения данных (удаление экстремальных тестовых заданий, испытуемых). Несмотря на это, «средние арифметические результатов измерения» и «статистики адекватности» удовлетворительны. Беспорядок наблюдается в «шаге шкалы», он отразился на градации, близкой ко второй. Согласно значению, шаг шкалы очень редкий. Эта ред-

кость шага отразилась на графике, согласно которому градация 2 не является модальной для этих данных. Шаг шкалы не указывает, которое из этих двух значений должно быть исправлено: уровней 1 и 2, или уровней 2 и 3, и, соответственно, какие из них должны быть исключены из градации согласно их месторасположению (рис. 5).

### Продолжение описания таблиц программы MINISTER

Ожидаемое значение (SAMPLE EXPECT) — это ожидаемое значение среднего для данного

Теория

#### 5

FIM (Functional Independence Measure) — мера функциональной независимости: Jamie Amundson, Amy Brunner, and Mary Ewers. FIM Scores as an Indicator of Length of Stay and Discharge Destination in CVA Patients: A Retrospective Outcomes Study. (<http://murphylibrary.uwax.edu/digital/jur/2000/amundson-bru...>)

ПЕД  
измерения

FIM LEVEL	COUNT	AVERAGE MEASURE	INFIT MNSQ	OUTFIT MNSQ	STEP CALIBRATN
1	96	-2.81	.90	.96	NONE
2	44	-1.96	.88	.92	-1.49
3	101	-1.03	1.02	.98	-2.33
4	168	-.30	1.07	1.22	-1.29
5	210	.82	.96	.88	.05
6	146	2.30	.75	.82	1.97
7	101	3.27	.87	.89	3.09

Рис. 5. Пример корректно организованных FIM-градаций

примера. Это значение всегда приближается к градации. Это снова описание примера, который не соответствует параметру Раша.

INFIT MNSQ — это среднее значение от квадрата INFIT, соотносённое с ответами в каждой градации. Ожидаемое значение для всех градаций 1.0.

OUTFIT MNSQ — это среднее значение от квадрата OUTFIT, соотносённое с ответами в каждой градации. Ожидаемое значение для всех градаций 1.0. Отметим, что данные показатели очень чувствительны к некорректно сформулированным тестовым заданиям<sup>6</sup>.

### Связь (COHERENCE)

M->C<sup>7</sup> — показывает процентное отношение, в котором ожидаемые результаты каждого шага градации оправдались. Соответствует ли измерение градации?

Коэффициент воспроизводимости результатов измерений Гутмана (Guttman's Coefficient of Reproducibility) — это среднее от M->C, т.е.

$$\text{Воспроизводимость} = \frac{\sum(\text{число} * M->C)}{\sum(\text{число} * 100)}$$

C->M — показывает процентное отношение, в котором количество шагов в данной градации зависело от измерений, соответствующих градаций. Соответствует ли градация измерениям?

Дифференцирующая способность оценки (ESTIM DISCR) — это оценка различий, когда модель задана формой:  $\log\text{-odds} = a_j (\theta_j - \beta_i - F_j)$ .

Остаток (RESIDUAL) (если он отражен в таблице) — это разница между ожидаемыми и полученными результатами измерений для данной градации. Данное значение выражено в % для ожидаемых результатов, иначе полученные результаты = 0. Тогда остаток отражается в таблице. Его значение должно

#### 6

Замечание: программа Winsteps всегда включает MNSQ-значения в таблицу № 3.2. Приближение (approximation) к их стандартным значениям можно получить, используя число шагов градации в качестве степеней свободы, и затем посмотреть результат на графике.

#### 7

Напоминаем, что согласно табл. 2: M->C = Does Measure imply Category? А так же C->M = Does Category imply Measure?



быть  $\geq 1.0$ . Является показателем недостаточной сходимости, нарушений в структуре данных или большого объема данных.

Коды градаций и метки (CATEGORY CODES and LABELS) отражают верное описание данных в CODES=, CFILE= and CLFILE=.

Измерения, соответствующие дихотомическим градациям, не отображаются, но могут быть обработаны программой посредством таблицы «What is a Logit?» и LOWADJ=, а также HIADJ=.

Б) Пример для градуированной шкалы:

**Теория**

Таблица 3

```

SUMMARY OF CATEGORY STRUCTURE. Model="R"
FOR GROUPING "0" ITEM NUMBER: 1 A. EATING

ITEM DIFFICULTY MEASURE OF -.61 ADDED TO MEASURES
-----
| CATEGORY   OBSERVED|OBSVD SAMPLE|INFIT
OUTFIT| |STRUCTURE|CATEGORY|
| LABEL SCORE COUNT %|AVRGE EXPECT| MNSQ MNSQ| |CALIBRATN|
MEASURE|
-----+-----+-----+-----+-----+
| 5 5 5 14| -.51 -.42| .89 .68|| NONE | (-
2.22)| 5 Supervision
| 6 6 9 26| .39 .04| 1.45 1.63|| -.18 | -
.61 | 6 Device
| 7 7 21 60| .73 .86| 1.34 1.32|| .18 | (
1.00)| 7 Independent
-----
AVERAGE MEASURE is mean of measures in category.
-----+-----+
| CATEGORY   STRUCTURE | SCORE-TO-MEASURE | 50% CUM. |
COHERENCE|ESTIM| OBSERVED-EXPECTED |
| LABEL MEASURE S.E. | AT CAT. ----ZONE----|PROBABLTY| M-
>C C->M|DISCR|RESIDUAL DIFFERENCE|
-----+-----+
| 5 NONE | (-2.22) -INF -1.50|
0% 0%| | -1.4% -.1 | 5 Supervision
| 6 -.79 .52 | -.61 -1.50 .28| -1.18 |
31% 66%| 1.22| .2% .0 | 6 Device
| 7 -.43 .39 |( 1.00) .28 +INF | -.04 |
81% 61%| .59| .2% .0 | 7 Independent
-----+
M->C = Does Measure imply Category?
C->M = Does Category imply Measure?
    
```

ПЕД	
	измерения

Значение уровня трудности задания  $-0,61$  добавлено. Это происходит, когда группа состоит из одного тестового задания (the Partial Credit model) и в отчёт добавляется значение уровня трудности этого задания<sup>8</sup>.

Метка градации (CATEGORY LABEL), Сумма значений по градации (CATEGORY SCORE), Наблюдаемое значение (OBSERVED COUNT) и %, Среднее OBSVD (OBSVD AVERAGE), Ожидаемое значение (SAMPLE EXPECT), INFIT MNSQ, OUTFIT MNSQ аналогичны таблице 2.

Проверка структуры (STRUCTURE CALIBRATN) — мера перехода от одной градации к другой по нисходящей. Это и есть оценка параметра модели Раша, так называемое  $F_j$  (это относится к  $F_{ij}$  в выражении  $\beta_i + F_j$  при оценивании параметров «Рейтинговой шкалы», и подобным образом к  $F_{ij}$  в выражении  $\beta_{ij} = \beta_i + F_{ij}$  для «Доверительного интервала»).

Нижняя градация не является важным связующим звеном и поэтому измерения считаются несостоявшимися (NONE). Этот параметр иногда называют «уровень трудности» (Step Difficulty), «шаг калибровки» (Step Calibration), «пороговое значение Раша-Эндрича» (Rasch-Andrich threshold), которые наглядно показывают, на-

сколько это трудно — оценить градацию, вести научное наблюдение за ней. Некоторый беспорядок в данных оценках (то, что они не поднимаются по шкале измерений с увеличением их значения) иногда называют «неупорядоченной ошибкой» («disordered deltas»), которая показывает, что градация относительно редко исследуется, т.е. занимает узкий интервал на шкале латентной переменной и в связи с этим возникают проблемы с рейтингом градации и её определением на шкале. Эти значения аналогичны парным оценкам переходов между градациями. Они являются своеобразными точками, в которых пересекаются смежные градации. Наглядно это представлено на рис. 7.

Оценка градации (CATEGORY MEASURE) — соответствующее значение градации. Скобки ( ) указывают на большее значение градации. Величина в данной графе, равная значению «.25» (или LOWADJ= и HIADJ=), не является экстремальной. То есть можно сделать следующее заключение: «коэффициент усредненного  $x$  предполагает измерение  $y$ » («ratings averaging  $x$  imply measures of  $y$ ») или «измерение  $y$  предполагает усреднение коэффициента  $x$ » («measures of  $y$  imply ratings averaging  $x$ »). Это вытекает из параметров модели Раша.

**8**  
Группа может состоять из одного тестового задания при ISGROUPS=0.  
Подробнее см.:  
Г. Смирнова,  
А. Смирнов. Редактирование данных в программе Minister // Педагогические измерения. 2010. № 4.

Структура измерений (STRUCTURE MEASURE) — здесь к уровню трудности тестового задания прибавляется значение перехода его градаций к последующей нижней градации. Для структуры с одним тестовым заданием — это оценка параметра модели Раша  $\beta_{ij} = \beta_i + F_{ij}$  (это относится к параметру  $\beta_{ij}$  модели «Доверительного интервала» («Partial Credit»). Нижняя градация не имеет предшествующего перехода, поэтому измерения считаются несостоявшимися (NONE). Градация повышается с увеличением её значения, но это может быть ложным утверждением. « $\beta_{gi} + F_{gj}$ » расположено в графике таблицы №2.4, где «g» относится к ISGROUPS= распределению.

STRUCTURE S.E. — это приближенная стандартная ошибка в измерениях с переходами.

Значение SCORE-TO-MEASURE описано в таблице 21 в разделе распределение «Ожидаемых данных». Эти данные необходимы для определения градации измерения, что является следствием параметра модели Раша.

Компьютерная томография (AT CAT — сокращённое от Computer - Aided Tomography) — это измерения (для тестовых заданий с 0-ми логитами) ожидаемых значений (меток градации), на которые

будет получен правильный ответ с наибольшей вероятностью. Подробнее смотрите на рис. 7.

Скобки () указываются там, где значение шкалы бесконечно большое. Значение, равное .25 (или LOWADJ= и HIADJ=), не является экстремальным.

Область ограничений (ZONE) — это диапазон ожидаемых значений от точки  $1/2$  под градацией до точки  $1/2$  над ней. Измерения в данном диапазоне должны совпадать, в среднем, с наблюдаемыми значениями. Подробнее смотрите на рис. 7.

50%-ная интегральная вероятность (50% CUMULATIVE PROBABILITY) определяет местоположение медианного значения вероятности, то есть является пороговым значением для шкалы Раша-Терстоуна (Rasch-Thurstone), аналогичной оценкам в таких моделях, как «Градуированные ответы» («Graded Response») или «Пропорциональные неравенства» («Proportional odds»). Согласно данным шкалирования заданий, вероятность правильного ответа на градацию, расположенную ниже, равняется вероятности правильного ответа на градации равные или выше данной. Значение 0,5 или 50% общей вероятности является точкой на переменной, от которой начинается интервал. Это утверждение вытекает из параметров модели Раша.

При этих калибровках вероятность наблюдения категорий ниже равняется вероятности наблюдения равных категорий или выше. 5 или 50%-ная совокупная вероятность — пункт на переменной, в которой начинается интервал категории. Это подразумевается параметрами модели Rasch.

Связь (COHERENCE) аналогична табл. 2.

Дифференцирующая способность оценки (когда DISCRIM=Y) (ESTIM DISCR (when DISCRIM=Y)) — это оценка различий, когда модель задана формой:

$$\log\text{-odds} = a_j (\theta_j - \beta_j - F_j).$$

Разность между наблюдаемым и ожидаемым значением (если указано) (OBSERVED — EXPECTED RESIDUAL DIFFERENCE) — это значение разности между наблюдаемым и ожидаемым значением в процессе наблюдения за градацией.

*Значение разницы в % = (наблюдаемое значение — ожидаемое значение) \* 100 / (ожидаемое значение)*

*Разность = наблюдаемое значение — ожидаемое значение*

Это значение отображается в таблице, если оно больше или равно 1%. Это показывает, что оценки по модели Раша не стремятся к их максимальному правдоподобию из-за того, что они не сходятся, зависят или имеют очень большой объем данных. К примеру:

1) пошаговое выполнение программы было прекращено раньше, чем Вы нажали Ctrl+F или соответствующую опцию в меню;

2) пошаговое выполнение программы было прекращено, когда максимальное число итераций достигло порогового для MJMLE=;

3) критерий сходимости в LCONV= и RCONV= имеют недостаточно малое значение для этого набора данных;

4) закреплённые значения (anchor values) (PAFILE=, IAFILE= и/или SAFILE=) не позволяют достигнуть максимально правдоподобных оценок.

Измерение уровня трудности задания суммируется ко всем заданиям? (ITEM MEASURE ADDED TO MEASURES) — это отображается в таблице в том случае, когда рейтинговая шкала (или доверительный интервал) используется для одного тестового задания, то есть когда ISGROUPS=0. Тогда все измерения данной таблицы корректируются для измеряемого тестового задания.

Кривые на рис. 6 показывают, как будут вести себя градации относительно измерений уровня трудности тестового задания. Обычно 0 логитов на графике совпадают с началом измерений уровня трудности тестового задания и являются точкой, в которой измерение

CATEGORY PROBABILITIES: MODES - Structure measures at intersections

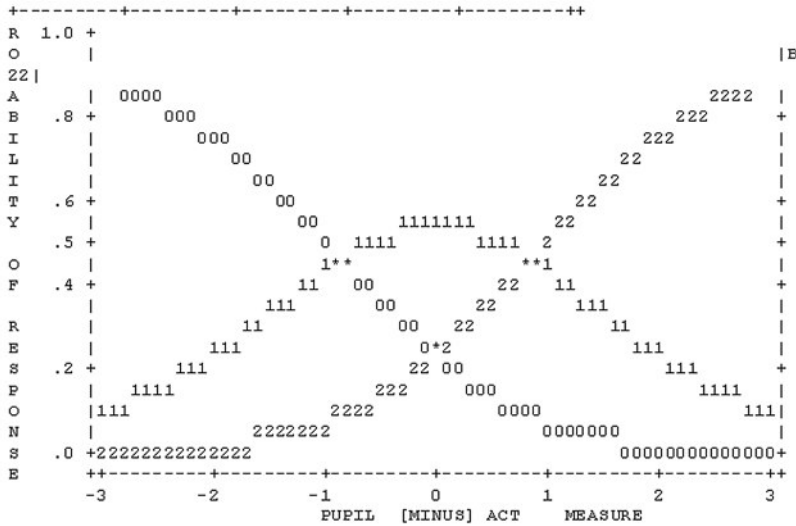


Рис. 6. Градации относительно измерений уровня трудности тестового задания

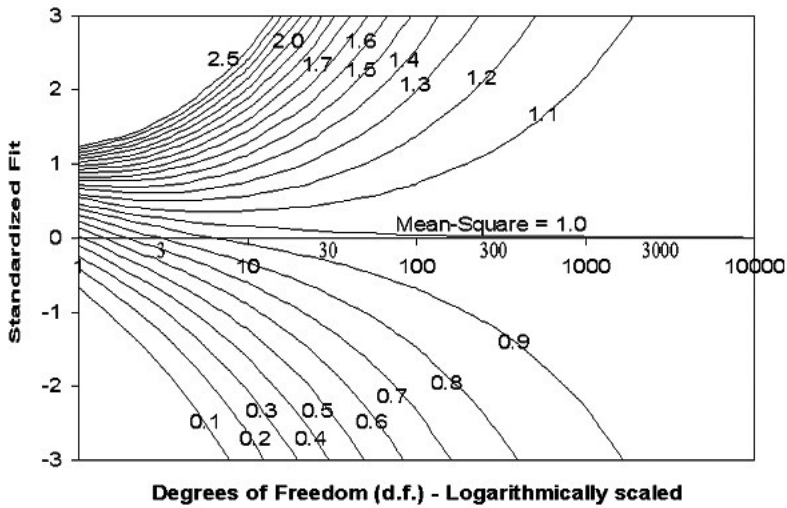


Рис. 7. Расположение градаций на интервальной шкале

Теория

**ПЕД**  
**измерения**

самой высокой и самой низкой градации равновозможны. График должен выглядеть как цепь холмов. Градации, которые не появляются на графике в качестве одного из холмов, являются неупорядоченными для перехода к калибровке. Это противоречие обычно ин-

терпретируется для градации как наиболее вероятный результат.

Ноль (Null, Zero), ненаблюдаемые градации (Unobserved Categories). Используйте код STKEEP=YES и градация 2 будет ненаблюдаемой (no observations) (табл. 4).

Таблица 4

CATEGORY	OBSERVED	OBSVD	SAMPLE	INFIT	OUTFIT	STRUCTURE	CATEGORY
LABEL	SCORE	COUNT	%	AVRGE	EXPECT	MNSQ	MNSQ
						CALIBRATN	MEASURE
0	0	378	20	-.67	-.73	.96	1.16
1	1	620	34	-.11	-.06	.81	.57
2	2	0	0			.00	.00
3	3	852	46	1.34	1.33	1.00	1.64

Градация 2 является частичным (sampling) нулем. Градация 2 поддержана в структуре ответа. Используйте код

STKEEP=NO и градация 2 станет ненаблюдаемой (no observations) (табл. 5).

Таблица 5

CATEGORY	OBSERVED	OBSVD	SAMPLE	INFIT	OUTFIT	STRUCTURE	CATEGORY
LABEL	SCORE	COUNT	%	AVRGE	EXPECT	MNSQ	MNSQ
						CALIBRATN	MEASURE
0	0	378	20	-.87	-1.03	1.08	1.20
1	1	620	34	.13	.33	.85	.69
3	2	852	46	2.24	2.16	1.00	1.47

Градация 2 является неразличимым (unobservable) нулем. Градация устранена из структуры ответа.

На этом описание таблицы 3.2 Градуированная оценка (Rating (partial credit) scale) заканчивается.