

Анализ данных
национальной оценки
учебных достижений

Analyzing Data
from a National
Assessment of
Educational
Achievement

National Assessments of Educational Achievement

VOLUME 4

Analyzing Data from a National Assessment of Educational Achievement

Gerry Shiel
Fernando Cartwright

Vincent Greaney and
Thomas Kellaghan, Series Editors



WORLD BANK GROUP

Национальная оценка учебных достижений

Джерри Шил, Фернандо Картрайт

Анализ данных национальной оценки учебных достижений

Винсент Грини и Томас Келлаган, редакторы серии

КНИГА 4



Москва
Логос
2015

УДК 378
ББК 74.58
Ш58



*Издание подготовлено при поддержке
Международного банка реконструкции и развития
в рамках Российской программы содействия образованию
в целях развития – READ (Russia Education Aid for
Development)*



*Перевод подготовлен по инициативе Центра
международного сотрудничества по развитию
образования Российской академии народного
хозяйства и государственной службы
при Президенте Российской Федерации*

Шил, Джерри

Ш58 Анализ данных национальной оценки учебных достижений /
Д. Шил, Ф. Картрайт; ред. серии В. Грини, Т. Келлаган; перевод
с английского на русский язык под научной редакцией В.И. Звонни-
кова и М.Б. Челышковой. – М.: Логос, 2015. – 336 с. – (Национальная
оценка учебных достижений)

ISBN 978-5-98704-836-8

Показаны сложности и потенциал анализа данных, полученных в ходе проведения широкомасштабной национальной оценки. Рассмотрены вопросы генерации информации по тестовым заданиям и оценкам, установления связей между тестовыми оценками и образовательными и социальными факторами. Для того чтобы в полной мере раскрыть то, что сообщают данные о качестве, равенстве и иных аспектах достижений в системе образования, предлагается применять различные методики, содержащиеся в первой части книги. Во второй части книги описана основная аналитическая методика – современная теория тестирования (Item Response Theory – IRT). Представлено специально разработанное и удобное в использовании программное обеспечение для IRT под названием «Анализ заданий и тестов» (Item and Test Analysis – IATA).

Для специалистов, участвующих в проведении национальной оценки учебных достижений обучающихся в средних и других учебных заведениях. В первую очередь представляет интерес для аналитиков, обеспечивающих интерпретацию данных оценки с целью дальнейшего обобщения для их практического использования в ходе решения задач развития и улучшения качества образования. Может использоваться при разработке инструментария проведения национальной оценки и обработки полученных данных.

УДК 378
ББК 74.58

ISBN 978-5-98704-836-8

© Международный банк
реконструкции и развития /
Всемирный банк, 2015
© Логос, 2015

1818 H Street NW
Washington DC 20433
Телефон: 202-473-1000
Internet: www.worldbank.org

Все права защищены
1 2 3 4 18 17 16 15

Книга подготовлена сотрудниками Всемирного банка при участии сторонних авторов. Представленные в книге факты, их трактовка и выводы не обязательно отражают точку зрения Всемирного банка, исполнительных директоров Всемирного банка или правительств государств, которые они представляют. Всемирный банк не гарантирует точности данных, приведенных в книге. Национальные границы, цвета, обозначения и прочая информация на картах, включенных в настоящее издание, не являются выражением мнения Всемирного банка относительно правового статуса какой-либо территории либо поддержки или признания границ.

Информация, представленная в настоящей книге, не является и не должна считаться ограничением или отказом от льгот и привилегий Всемирного банка, и все такие льготы и привилегии в прямой форме сохраняются за Всемирным банком.

Авторское право и разрешение на воспроизведение



Настоящая книга предоставляется по лицензии Creative Commons Attribution 3.0 IGO (CC BY 3.0 IGO) <http://creativecommons.org/licenses/by/3.0/igo>. При наличии лицензии Creative Commons Attribution вы вправе свободно копировать, распространять, передавать и адаптировать настоящую книгу, в том числе в коммерческих целях, с соблюдением следующих условий.

Ссылка на источник – Ссылки на данную книгу должны быть оформлены в следующем формате: Джерри Шил и Фернандо Картрайт. 2015 год. Анализ данных национальной оценки учебных достижений. Книга 4 из серии «Национальная оценка учебных достижений» под редакцией Винсента Грини и Томаса Келлагана. Вашингтон, Федеральный округ Колумбия: Всемирный банк. doi:10.1596/978-0-8213-9583-7. Лицензия: Creative Commons Attribution CC BY 3.0 IGO.

Переводы – Если вы переводите данное издание, просим включать следующую оговорку со ссылкой на источник: «Настоящий перевод не был выполнен Всемирным банком и не должен рассматриваться как официальный перевод Всемирного банка. Всемирный банк не несет ответственность за содержание и любые неточности настоящего перевода».

Адаптации – Если вы адаптируете данное издание, просим включать следующую оговорку со ссылкой на источник: «Настоящая работа является адаптацией оригинального издания Всемирного банка. Взгляды и мнения, выраженные в данной адаптации, являются единоличной ответственностью автора или авторов данной адаптированной работы, которые не подлежали одобрению Всемирным банком».

Согласие третьих лиц – Всемирный банк не обязательно владеет каждым компонентом содержания настоящей книги. Соответственно, Всемирный банк не дает гарантий, что использование любого принадлежащего третьим лицам индивидуального компонента настоящей книги не нарушит прав таких третьих лиц. Вы единолично несете риск предъявления любых исков и претензий в результате такого нарушения. Если вы намерены повторно использовать какой-либо компонент настоящей работы, вы

отвечаете за установление необходимости получения разрешения у владельца авторских прав в отношении такого повторного использования, а также за получение такого разрешения. Примеры таких компонентов могут включать, в том числе, таблицы, показатели или рисунки.

Все остальные запросы, связанные с получением прав и лицензий, необходимо направлять в отдел публикаций и просвещения Всемирного банка по адресу: 1818 H Street NW, Washington DC 20433, USA; факс: 202-522-2625; адрес электронной почты: pubrights@worldbank.org.

ISBN (печатный вариант): 978-0-8213-9583-7

ISBN (электронная версия): 978-0-8213-9584-4

DOI: 10.1596/978-0-8213-9583-7

Данные каталогизации перед публикацией Библиотеки Конгресса были запрошены.

Дизайн обложки: Naylor Design (Вашингтон, федеральный округ Колумбия)

Анализ заданий и тестов (IATA) © 2015 год. Фернандо Картрайт. Использовано с разрешения. Для повторного использования требуется получение дополнительного разрешения. Microsoft, Access, Excel, Office, Windows и Word являются либо зарегистрированными товарными знаками, либо товарными знаками корпорации Microsoft Corporation в Соединенных Штатах Америки и/или иных странах. SPSS является зарегистрированным товарным знаком компании IBM. WesVar является зарегистрированным товарным знаком компании Westat.

Выражение признательности

Научные консультанты, переводчики, редакторы Издательской группы «Логос» выражают признательность Московскому представительству Международного банка реконструкции и развития, особенно Тиграну Шмису и Дмитрию Чугунову, за помощь и поддержку в подготовке и выпуске этой книги и всей серии книг «Национальная оценка учебных достижений». Мы благодарим В.А. Мау, ректора Российской академии народного хозяйства и государственной службы при Президенте Российской Федерации, В.А. Болотова, академика Российской академии образования, и И.Д. Фрумина, научного руководителя Института образования Национального исследовательского университета «Высшая школа экономики», которые направляли наши усилия на достижение целей, имеющих исключительное значение для реформирования российской системы образования. Высокой оценки заслуживает вклад Центра международного сотрудничества по развитию образования, выступившего инициатором перевода книг по оценке учебных достижений на русский язык.



СОДЕРЖАНИЕ

ПРЕДИСЛОВИЕ К РУССКОМУ ИЗДАНИЮ	19
ПРЕДИСЛОВИЕ	27
ОБ АВТОРАХ И РЕДАКТОРАХ	29
БЛАГОДАРНОСТИ	31
СПИСОК СОКРАЩЕНИЙ	33
ВВЕДЕНИЕ	35

Часть I

Введение в статистический анализ данных национальной оценки

Джерри Шил

ГЛАВА 1. БАЗА ДАННЫХ ДЛЯ ПРОВЕДЕНИЯ АНАЛИЗА	43
Сохранение файлов с CD-диска на ваш жесткий диск или сервер	45
Инструментарий исследования	47
Выборочные весовые коэффициенты	49
Статистический пакет для социальных наук SPSS	51
Статистический пакет для оценивания вариаций WesVar	55

ГЛАВА 2. ИССЛЕДОВАНИЕ ДАННЫХ НАЦИОНАЛЬНОЙ	
ОЦЕНКИ С ИСПОЛЬЗОВАНИЕМ SPSS	57
Меры центральной тенденции	58
Меры разброса	58
Меры положения	59
Меры формы распределения	60
Исследование множества данных с использованием SPSS	62
ГЛАВА 3. ВВЕДЕНИЕ В WESVAR	71
Установка файла данных в WesVar	71
Добавление подписей к переменным	72
Вычисление показателей дескриптивной статистики в WesVar	73
Расчет среднего балла и его стандартной ошибки	79
Вычисление средних оценок и стандартных ошибок для подгрупп в генеральной совокупности	82
ГЛАВА 4. СРАВНЕНИЕ УЧЕБНЫХ ДОСТИЖЕНИЙ ДВУХ И БОЛЕЕ ГРУПП	85
Поверка различия между двумя средними баллами	85
Исследование различий между тремя и более средними баллами	91
ГЛАВА 5. ВЫЯВЛЕНИЕ УЧАЩИХСЯ С ВЫСОКИМИ И НИЗКИМИ ДОСТИЖЕНИЯМИ	99
Приближенные баллы, соответствующие ключевым национальным процентильным рангам	99
Оценка процента учащихся в подгруппах с использованием национальных процентильных рангов	105
ГЛАВА 6. СВЯЗЬ МЕЖДУ ПЕРЕМЕННЫМИ: КОРРЕЛЯЦИЯ И РЕГРЕССИЯ	111
Корреляция	111
Регрессия	119
Корреляция и причинно-следственная связь	135

ГЛАВА 7. ПРЕДСТАВЛЕНИЕ ДАННЫХ С ИСПОЛЬЗОВАНИЕМ ГРАФИКОВ И ДИАГРАММ	139
Диаграммы	140
Линейные графики	
с доверительными интервалами	146
Графики для представления динамики показателей	149
ПРИЛОЖЕНИЕ I.A. АНАЛИЗ ДАННЫХ NAEА: СТРУКТУРА КАТАЛОГА ФАЙЛОВ	153
ПРИЛОЖЕНИЕ I.B. АНАЛИЗ ДАННЫХ NAEА: ПОДПАПКИ И ФАЙЛЫ	155
ПРИЛОЖЕНИЕ I.C. ОТКРЫТИЕ ФАЙЛА SPSS В WESVAR	161

Часть II

Анализ заданий и тестов

Фернандо Картфайт

ГЛАВА 8. ВВЕДЕНИЕ В IATA	173
Установка IATA	173
Данные оценки	174
Данные, выдаваемые IATA	187
Интерпретация результатов IATA	189
Данные выборки	190
Последовательности действий и интерфейсы анализа IATA	191
Навигация по последовательностям действий в IATA	195
ГЛАВА 9. АНАЛИЗ ДАННЫХ ПРЕДЪЯВЛЕНИЯ ПИЛОТНОГО ТЕСТА	197
Этап 1: загрузка данных по ответам	199
Этап 2: загрузка ключа ответов	201

Этап 3: анализ спецификаций	202
Этап 4: анализ заданий	206
Этап 5: размерность теста	217
Этап 6: дифференцирующее функционирование заданий (DIF)	224
Этап 7: анализ шкалы	231
Этап 8: выбор тестовых заданий	235
Этап 9: стандарты пороговых значений	241
Этап 10: просмотр и сохранение результатов	241

ГЛАВА 10. ПРОВЕДЕНИЕ ПОЛНОГО АНАЛИЗА ДАННЫХ ПРЕДЪЯВЛЕНИЯ ОКОНЧАТЕЛЬНОГО ТЕСТА 245

Этап 1: настройка параметров анализа	246
Этап 2: основные результаты анализа	248
Этап 3: анализ дифференцированного функционирования заданий	248
Этап 4: шкалирование	250
Этап 5: выбор тестовых заданий	256
Этап 6: установление стандартов выполнения	258
Этап 7: сохранение результатов	267

ГЛАВА 11. АНАЛИЗ РОТИРУЕМЫХ БУКЛЕТОВ 269

Этап 1: загрузка данных	269
Этап 2: спецификации анализа	271
Этап 3: результаты анализа заданий	272

ГЛАВА 12. АНАЛИЗ ЗАДАНИЙ С ДИФФЕРЕНЦИРОВАННОЙ ШКАЛОЙ БАЛЛОВ 275

Этап 1: загрузка данных	275
Этап 2: спецификации анализа	277
Этап 3: результаты анализа заданий	278

ГЛАВА 13. СРАВНЕНИЕ ОЦЕНОК 283

Этап 1: настройка параметров анализа	285
Этап 2: связывание общих заданий	289

Этап 3: перешкалирование связанных результатов	294
Этап 4: назначение стандартов выполнения	296
ГЛАВА 14. СПЕЦИАЛИЗИРОВАННЫЕ МЕТОДЫ IATA	301
Связывание данных задания	302
Выбор оптимальных тестовых заданий	305
Разработка и назначение стандартов выполнения	308
Анализ данных по ответам с параметрами якорных заданий	312
ГЛАВА 15. КРАТКИЙ ОБЗОР ПОШАГОВЫХ ПРИМЕРОВ АНАЛИЗА IATA	319
II.A. СОВРЕМЕННАЯ ТЕОРИЯ ТЕСТИРОВАНИЯ	323
СПИСОК ЛИТЕРАТУРЫ	333
ТЕКСТОВЫЕ ВСТАВКИ	
6.1. Переменные в стандартной регрессии	121
УПРАЖНЕНИЯ	
1.1. Проведение анализа для получения показателей описательной статистики в SPSS и сохранение файлов	52
2.1. Использование команды «Эксплаторный анализ» в SPSS для одной зависимой переменной (один уровень)	63
2.2. Использование команды «Эксплаторный анализ» в SPSS для одной зависимой переменной (более одного уровня)	68
3.1. Получение показателей описательной статистики в WesVar	74
3.2. Расчет среднего балла и его стандартной ошибки в WesVar	80
3.3. Расчет средних баллов и стандартных ошибок в WesVar для четырех регионов	82
4.1. Оценка различий между двумя средними баллами	87
4.2. Оценка и анализ различий между тремя и более средними баллами	92
5.1. Расчет национальных процентильных баллов	100
5.2. Расчет процентильных баллов по регионам	102

5.3. Перекодировка переменной в процентильные категории с использованием WesVar	105
5.4. Расчет процента учащихся с баллом ниже ключевого порогового значения национального процентиля и стандартных ошибок по регионам	107
6.1. Построение диаграммы рассеивания в SPSS	113
6.2. Расчет коэффициента корреляции, национальный уровень	116
6.3. Проведение регрессионного анализа в WesVar с одной независимой переменной (непрерывной)	123
6.4. Проведение регрессионного анализа в WesVar с одной независимой переменной (дискретной)	127
6.5. Оценка коэффициентов корреляции	129
6.6. Проведение регрессионного анализа в WesVar более чем с одной независимой переменной	131
7.1. Построение столбиковой гистограммы для отображения результатов по уровням подготовки учащихся, национальные данные	140
7.2. Построение линейчатой диаграммы для отображения результатов по уровням подготовки учащихся, национальные данные	143
7.3. Графическое отображение 95-процентных доверительных интервалов для серии средних баллов	146
7.4. Представление динамики показателей на графике	150

РИСУНКИ

2.1. Нормальное распределение, демонстрирующее стандартные отклонения	60
2.2. Примеры распределения с правосторонней, левосторонней асимметрией и отсутствием асимметрии	61
3.1. Добавление подписей к переменным в WesVar	73
6.1. Положительная и отрицательная корреляция	113
6.2. Линия регрессии и уравнение регрессии на диаграмме рассеяния	121
I.C.1. Агрегирование данных в SPSS	163
I.C.2. Введение переменных в файл SPSS	166
I.C.3. Список доступных переменных в файле с данными WesVar	167
I.C.4. Создание весовых коэффициентов в WesVar	168
I.C.5. Репликационные весовые коэффициенты, созданные в WesVar	169

8.1. Примеры корректного и некорректного форматирования данных	176
8.2. Изначальный выбор языка и опциональная регистрация в IATA	193
8.3. Главное меню IATA	194
8.4. Окно с инструкциями и кнопки навигации интерфейса задач в IATA	196
9.1. Последовательность действий «Анализ данных по ответам»	198
9.2. Интерфейс загрузки данных по ответам	200
9.3. Данные по заданиям для данных по ответам PILOT1	203
9.4. Анализ спецификаций для данных PILOT1	204
9.5. Результаты анализа заданий для данных PILOT1, MATHC1019	207
9.6. Результаты анализа заданий для данных PILOT1, MATHC1027	213
9.7. Результаты анализа заданий для данных PILOT1, MATHC1075	214
9.8. Результаты анализа заданий для данных PILOT1 после удаления MATHC1075	216
9.9. Размерность тестов и заданий для данных PILOT1, MATHC1019	218
9.10. Результаты оценки размерности задания для данных PILOT1, MATHC1035	220
9.11. Результаты оценки размерности задания для данных PILOT1, MATHC1002	223
9.12. Результаты анализа DIF для данных PILOT1 по гендерному признаку, MATHC1046	225
9.13. Результаты анализа DIF для данных PILOT1 по гендерному признаку, MATHC1035	227
9.14. Результаты анализа DIF для данных PILOT1 по гендерному признаку, MATHC1042	228
9.15. Результаты анализа DIF для данных PILOT1 по языку, на котором учащиеся говорят дома, MATHC1006	230
9.16. Интерфейс анализа и установки шкалы	232
9.17. Результаты выбора заданий для данных PILOT1, 50 заданий	237
9.18. Результаты выбора заданий для данных PILOT1, 79 заданий	240
9.19. Просмотр результатов анализа данных для PILOT1	242
10.1. Спецификации анализа для данных CYCLE1	247

10.2. Результаты анализа DIF для данных CYCLE1 по местонахождению школ, MATHC1043	250
10.3. Распределение уровней (IRT-балл) и тестовой информации, данные CYCLE1	252
10.4. Сравнение идеальной тестовой информации нормального распределения	253
10.5. Распределение и сводные статистические показатели для нового шкалированного балла (NAMscore), данные CYCLE1	255
10.6. Выбор задания, данные CYCLE1	257
10.7. Интерфейс установленных по умолчанию стандартов выполнения, данные CYCLE1	260
10.8. Интерфейс стандартов выполнения, RP = 50 %, данные CYCLE1	262
10.9. Данные закладок, RP = 50 %, данные CYCLE1	264
10.10. Интерфейс стандартов выполнения с установленными вручную пороговыми значениями, данные CYCLE1	267
11.1. Ответы учащихся, данные PILOT2	270
11.2. Спецификации анализа, ротируемые буклеты, данные PILOT2	271
11.3. Результаты анализа заданий, данные PILOT2, MATHC2003	273
12.1. Ключи ответов и метаданные, данные PILOT2	276
12.2. Спецификации анализа для ротируемых буклетов с заданиями с дифференцированной шкалой ответов, данные PILOT2	277
12.3. Результаты анализа заданий, данные PILOT2, MATHC2003	278
12.4. Функция ответов на задания с дифференцированной шкалой ответов, данные CYCLE2, MATHSA001, оценка = 2	279
13.1. Последовательность действий анализа данных по ответам со связыванием	285
13.2. Справочные данные по заданиям из CYCLE1 для связи с данными CYCLE2	287
13.3. Результаты анализа заданий для данных CYCLE2, MATHSA005, оценка = 1	288
13.4. Результаты связывания общих заданий, CYCLE2 и CYCLE1	290
13.5. Результаты связывания общих заданий, CYCLE2 и CYCLE1, MATHC1052	292
13.6. Результаты тестов CYCLE2, выраженные по шкале CYCLE1 (NAMscore)	295
13.7. Определение стандартов выполнения, данные CYCLE2	298

14.1. Выбор оптимальных тестовых заданий, данные CYCLE1	307
14.2. Данные по заданиям для CYCLE3 с параметрами якорных заданий	314
14.3. Результаты анализа заданий с параметрами якорных заданий, данные CYCLE3, MATHC2047	315
П.А.1. Распределения уровней подготовки для респондентов, ответивших верно, и респондентов, ответивших неверно на одно тестовое задание (легкость = 0,50, средний уровень подготовки верно ответивших респондентов = 0)	326
П.А.2. Распределения уровней подготовки для респондентов, ответивших верно, и респондентов, ответивших неверно на одно тестовое задание (легкость = 0,50, средний уровень подготовки верно ответивших респондентов = 0,99)	327
П.А.3. Распределения уровней подготовки для респондентов, ответивших верно, и респондентов, ответивших неверно на одно тестовое задание (легкость = 0,60, средний уровень подготовки верно ответивших респондентов = 0,40)	329

ТАБЛИЦЫ

1.1. Тест по математике: распределение заданий по содержательным разделам и когнитивным процессам	48
1.2. Краткое описание анкет	49
5.1. Процентные доли учащихся с баллом ниже порогового значения национального 25-го перцентиля по регионам	110
5.2. Процентные доли учащихся с баллом не ниже порогового значения национального 75-го перцентиля по регионам	110
8.1. Переменные, генерируемые или используемые IATA для описания уровня подготовки и результатов выполнения теста учащегося	181
8.2. Переменные в файле с данными по заданиям	183
8.3. Раздел выборки файла с данными по заданиям	184
8.4. Раздел выборки файла с данными по заданиям для задания с дифференцированной шкалой ответов	186
8.5. Таблицы с данными, создаваемые IATA	188
8.6. Символы трафика в IATA и их значение	189
8.7. Задачи в IATA и последовательности действий, в которых они используются	195
9.1. Анализ дистракторов для задания MATHC1019 из данных PILOT1	210

такая шкала соответствовала шкале нормального распределения. Переменная *theta* представляет собой уровень подготовки учащегося. Такая же модель описывает задания с дифференцированными ответами, где $P(u \geq x)$ представляет собой любой балл, больший или равный заданному частичному баллу x . В случае дифференцированных ответов каждый ненулевой балл будет иметь набор параметров.

Хотя все параметры взаимодействуют для описания статистических свойств задания, параметр **a** преимущественно соответствует расстоянию между средними значениями распределений верно и неверно ответивших респондентов. Параметр **b** в основном соответствует степени легкости задания; а параметр **c** соответствует вероятности, с которой учащийся из распределения неверно ответивших будет ошибочно включен в распределение верно ответивших (например, учащийся угадал верный ответ).

Так как процесс IRT является итеративным и вычислительно трудоемким, разные пакеты программного обеспечения могут выставлять немного отличающиеся друг от друга оценки и требовать разного количества времени на выполнение вычислений. Алгоритм оценивания IATA обычно является более устойчивым для выборок различных размеров и заметно быстрее, чем другие программы оценивания IRT. В то время как другие методы используют итеративные аппроксимирующие алгоритмы для выполнения этапа оценивания параметров задания, IATA вычисляет параметры задания алгебраически, используя следующие уравнения:

$$a = -(\mu_{\text{верный}}^* / (-1 + p^* + p^* \mu_{\text{верный}}^{*2})) / 1,7(1 + q / (q + q_{\text{верный}}))$$

$$b = (\mu_{\text{неверный}} + \mu_{\text{верный}}^* - (2 * \sigma^2 * \text{LOG}(q^* / p^*)) / (\mu_{\text{неверный}} + \mu_{\text{верный}}^*)) / 2$$

$$c = q / (q + q_{\text{верный}}),$$

где

$$p^* = (1 - (1 - p) / (1 - c));$$

$$q^* = q + q_{\text{верный}};$$

$$\mu_{\text{верный}}^* = (-\mu_{\text{неверный}} * (1 - p^*)) / p^*;$$

$$\sigma^2 = 1 - (p * \mu_{\text{верный}}^* - (q^*) * \mu_{\text{неверный}})^2 + p * \mu_{\text{верный}}^* + (q^*) \mu_{\text{неверный}}^* ;$$

$\mu_{\text{верный}}$ = средний уровень подготовки учащихся, которые ответили верно;

$\mu_{\text{неверный}}$ = средний уровень подготовки учащихся, которые ответили неверно;

p = доля учащихся, которые ответили верно;

q = доля учащихся, которые ответили неверно;

$q_{\text{верный}}$ = доля учащихся, которые не смогли ответить на вопрос, но угадали верный ответ (этот статистический показатель должен рассчитываться округлением нижней асимптоты эмпирической функции ответов на задание). Обратите внимание, что если параметр s ограничен значением ноль (а такое ограничение действительно рекомендуется вводить во множестве оценочных ситуаций), то $q_{\text{верный}}$ вычислять не нужно.

Новые оценки параметров используются в каждом цикле вычислений для получения обновленных функций подготовки для каждого студента с помощью методов, описанных Бэйкером и Ким (Baker and Kim 2004). Хотя алгоритм все же требует проведения многих итеративных циклов для получения окончательной оценки, ошибкоустойчивость предшествующих уравнений на этапе оценивания параметров задания значительно снижает время вычисления и повышает стабильность расчетов.

ПРИМЕЧАНИЕ

1. В IRT уровень подготовки учащегося описывается по шкале (часто называемой *theta*), аналогичной Z-балльной шкале: теоретический средний уровень подготовки равен 0, а стандартное отклонение равно 1. Большинство учащихся обычно имеют баллы между -2 и 2, и менее одного учащегося из тысячи будут иметь баллы меньше -3 (или больше 3).



СПИСОК ЛИТЕРАТУРЫ

- Anderson P. and G. Morgan. 2008. *Developing Tests and Questionnaires for a National Assessment of Educational Achievement*. Washington, DC: World Bank.
- Baker F.B. and S.-H. Kim. 2004. *Item Response Theory: Parameter Estimation Techniques*. 2nd ed. New York: Marcel Dekker.
- Bullock J.G., D.P. Green and S.E. Ha. 2010. "Yes, but What's the Mechanism? (Don't Expect an Easy Answer)." *Journal of Personality and Social Psychology* 98 (4): 550–58.
- Crocker L. and J. Algina. 2006. *Introduction to Classical and Modern Test Theory*. Pacific Grove, CA: Wadsworth.
- Cronbach L.J. 1970. "Test Validation." In *Educational Measurement*, 2nd ed., edited by R.L. Thorndyke, 443–507. Washington, DC: American Council on Education.
- De Ayala R.J. 2009. *The Theory and Practice of Item Response Theory*. New York: Guilford Press.
- DeMars C. 2010. *Item Response Theory*. New York: Oxford University Press.
- Dumais J. and J.H. Gough. 2012a. "School Sampling and Methodology." In *Implementing a National Assessment of Educational Achievement*, edited by V. Greaney and T. Kellaghan, 57–106. Washington, DC: World Bank.

———. 2012b. “Weighting, Estimating, and Sampling Error.” In *Implementing a National Assessment of Educational Achievement*, edited by V. Greaney and T. Kellaghan, 181–257. Washington, DC: World Bank.

Fan X. 1998. “Item Response Theory and Classical Test Theory: An Empirical Comparison of Their Item/Person Statistics.” *Educational and Psychological Measurement* 58(3): 357–81.

Freeman C. and K. O'Malley. 2012. “Data Preparation, Validation and Management.” In *Implementing a National Assessment of Educational Achievement*, edited by V. Greaney and T. Kellaghan, 107–79. Washington, DC: World Bank.

Goldstein H. and R. Wood. 1989. “Five Decades of Item Response Modelling.” *British Journal of Mathematical and Statistical Psychology* 42 (2): 139–67.

Greaney V. and T. Kellaghan. 2008. *Assessing National Achievement Levels in Education*. Washington, DC: World Bank.

——, eds. 2012. *Implementing a National Assessment of Educational Achievement*. Washington, DC: World Bank.

Haladyna T.M. 2004. *Developing and Validating Multiple-Choice Test Items*. 3rd ed. Mahwah, NJ: Lawrence Erlbaum Associates.

Hambleton R.K., H. Swaminathan and H.J. Rogers. 1991. *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage.

Hutcheson G. and N. Sofroniou. 1999. *The Multivariate Social Scientist*. London: Sage.

Karantonis A. and S.G. Sireci. 2006. “The Bookmark Standard Setting Method: A Literature Review.” *Educational Measurement: Issues and Practice* 25 (1): 4–12.

Kellaghan, T., and V. Greaney. 2001. *Using Assessment to Improve the Quality of Education*. Paris: United Nations Educational, Scientific, and Cultural Organisation, International Institute for Educational Planning.

Kellaghan T., V. Greaney and T.S. Murray. 2009. *Using the Results of a National Assessment of Educational Achievement*. Washington, DC: World Bank.

- Lord F.M. and M.R. Novick. 1968. *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley.
- Martin M.O., I.V.S. Mullis and P. Foy (with J. F. Olson, E. Erberber, C. Prewschoff and J. Galia). 2008. *TIMSS 2007 International Science Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Mislevy R.J. 1992. *Linking Educational Assessments: Concepts, Issues, Methods, and Prospects*. Princeton, NJ: Educational Testing Service.
- Mitzel H.C., D.M. Lewis, R.J. Patz and D.R. Green. 2001. "The Bookmark Procedure: Psychological Perspectives". In *Setting Performance Standards: Concepts, Methods, and Perspectives*, edited by G. J. Cizek, 249–81. Mahwah, NJ: Lawrence Erlbaum Associates.
- OECD (Organization for Economic Co-operation and Development). 2007. *PISA 2006: Science Competencies for Tomorrow's World. Volume I: Analysis*. Paris: OECD.
- Raudenbush S.W. and A.S. Bryk. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods*, 2nd ed. Thousand Oaks, CA: Sage.
- Snijders T.A.B. and R.J. Bosker. 1999. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. Thousand Oaks, CA: Sage.

Научное издание

Национальная оценка учебных достижений

Шил Джерри
Картрайт Фернандо

**АНАЛИЗ ДАННЫХ НАЦИОНАЛЬНОЙ
ОЦЕНКИ УЧЕБНЫХ ДОСТИЖЕНИЙ**

Выпускающий редактор *Ю.П. Голобокова*
Редакторы *Ю.П. Голобокова, Д.А. Басипова*
Корректор *А.А. Нотик*
Верстка и дизайн *А.М. Моисеева*

*Издательская группа «Логос» выражает благодарность руководству
и сотрудникам Бюро переводов «Альба» за консультации
и помощь в переводе и оформлении настоящего издания*

Подписано в печать 11.06.2015. Формат 70×100/16,
Печать офсетная. Бумага офсетная. 21 печ. л.
Тираж 1000 экз. Заказ

Издательская группа «Логос»
111024, г. Москва, ул. Авиамоторная, д. 55, корп. 31, офис 305
Тел. (495) 981-51-12; 955-78-30
Электронная почта: universitas@mail.ru
Дополнительная информация на сайте: www.logosbook.ru