

Педагогические**2011****№ 4****измерения**

Зарегистрирован
Министерством Российской
Федерации по делам печати,
телерадиовещания и средств
массовых коммуникаций.
Свидетельство о регистрации
средства массовой
информации ПИ № 77-15870
от 7 июля 2003 г.

Директор проекта
Елена ШИШМАКОВА

Главный редактор
Вадим АВАНЕСОВ

**Ответственный
секретарь**
Светлана ЛЯЧИНА

Зав. производством
Любовь КУЧМИЕВА

Дизайн
Ольга ДЕНИСОВА

Вёрстка
*Александр
ВОЛХОНСКИЙ*

Корректор
Татьяна ДЕНИСЬЕВА

Технолог
Артём ЦЫПАНКОВ

Производство
Александр КЛЯЗМИН

Адрес: 109341, Москва,
ул. Люблинская, д. 157, корп. 2
Тел./факс: (495) 345-52-00,
345-59-00
E-mail:
narodnoe@narodnoe.org

Учредитель
НИИ школьных
технологий

© НИИ школьных технологий

Содержание номера:**Методология**

Олег Деменчёнок

Погрешность баллов Единого государственного
экзамена

3

Вадим Аванесов

Три источника становления метрической системы
Георга Раша (RM)

18

**Л.В. Зайцева,
Н.О. Прокофьева**

Модели и методы адаптивного контроля знаний

30**Теория**

Юрий Каргин

Элементарное решение основной задачи
педагогических измерений

50

**Н.А. Печенежский,
А.А. Маслов**

Модель системы компьютеризированного
адаптивного тестирования

68

Игорь Каневский

Мнимые плюсы и реальные минусы ЕГЭ

74

Игорь Каневский

О «стандартах» и «контроле качества»
преподавания математики в вузе

84

Методика

Наталья Ганина

Задания на установление правильной последовательности в курсе «Общая химия»

91

Надежда Бачурина

Русский язык: «Задания в тестовой форме». Темы: «Причастие» и «Деепричастие»

97

Наталья Богацкая

Русский язык в младшей школе. Задания в тестовой форме по теме: «Части слова»

106

Ксения Девяткина

Методическая разработка по теме: «Two American Presidents». Тексты и задания в тестовой форме

109

Вадим Аванесов.

Числа. Текст и задания в тестовой форме

115

Вячеслав Стуканов

Основы теории автомобильных двигателей: «Процессы газообмена». Задания в тестовой форме

118

Рекомендации авторам по подготовке текстов для публикации в журнале «Педагогические измерения»

126

ТРИ ИСТОЧНИКА СТАНОВЛЕНИЯ МЕТРИЧЕСКОЙ СИСТЕМЫ ГЕОРГА РАША (RM)

Вадим Аванесов

testolog@mail.ru

Система Г. Раша в данной статье анализировалась с позиций педагогической теории измерений¹. RM представляет собой соединение математико-статистической теории, вычислительных программ и практики их применения, в сочетании со статистическими методами обоснования получаемых результатов (fit analysis). Педагогический статус метрической системы Г. Раша признан неопределённым, практически отсутствует педагогический язык этой системы. Определены педагогические условия применимости RM.

Исследованы истоки становления RM. Их несколько, в данной статье представлены только три:

- идея самого Г. Раша о преодолении зависимости результатов испытуемых от уровня трудности заданий теста посредством т.н. расщепления параметров в предложенной им математической модели тестового процесса;
- психологическая (поведенческая) теория Курта Левина;
- психометрическая теория Л.Л. Гутмана.

Ключевые слова: педагогические измерения, тест, тестовое задание, задание в тестовой форме, Rasch Measurement, тезаурус.

Проблемная ситуация

Во многих странах мира применением метрической системы Г. Раша в психологических, педагогических, медицинских и т.п. исследованиях сейчас занимаются преимущественно математики и программисты-вычислители, технические специалисты, мышление и язык которых заметно отличаются от языка и логики педагогической науки. Для большинства прикладных наук, к которым надо отнести и педагогические измерения, ситуация такого рода является скорее обычной, чем исключительной.

1

Аванесов В.С.
Основы педагогической
теории измерений // Педагогические измерения. №1. 2004. С. 15–21.

Все эти специалисты, каждый по-своему, ищут своё место в педагогической системе, помогают продвижению перспективного метода RM. Но одновременно с этим вводят не всегда подходящие элементы своего языка в теорию педагогических измерений. Между тем, в образовательной сфере действительное развитие теории и методики RM зависит, как минимум, от двух условий: от становления педагогического языка метрической системы Г. Раша и от готовности педагогической науки инкорпорировать достоинства этой системы в состав педагогического знания. В России сейчас нет ни того, ни другого.

Всё, что пишется по этой теме на русском языке, пока не обработано научным языком, а потому становится фрагментарным, путанным и непригодным для исследовательской работы. Именно по этой причине возникла ситуация, когда пользователи компьютерных программ по RM в России появились, а научных публикаций по теории педагогических измерений по RM слишком мало, их очень не хватает. Примерно такая же ситуация сложилась несколько раньше в России и с математической теорией педагогических измерений (Item Response Theory)².

Заметным тормозом и препятствием на пути нормального

развития педагогических измерений стал бездарный, в научном отношении, т.н. единый государственный экзамен, который с маниакальным упорством одиннадцать лет насаждался в стране вопреки научной и общественной критике.

Три источника метрической системы Г. Раша

В качестве трёх основных источников возникновения метрической системы Г. Раша в этой работе выделяются:

1. Идея объективного измерения, нацеленная на преодоление зависимости оценки результатов испытуемых от уровня трудности заданий. Эта идея имеет очень древнюю, тысячелетнюю историю³. Впервые проблема объективности психолого-педагогических измерений была поставлена основателем статистической (классической) теории тестов⁴ Ч. Спирманом. В исследовании 1904 года он назвал объективность главным свойством тестов⁵. В работе 1907 года Ч. Спирман писал: «Объективные тестовые результаты практически недостижимы, но приближения к ним вполне возможны»⁶.

И только в середине XX века благодаря математической модели Г. Раша проблему неза-

Методология

2

Аванесов В.С.

Item Response Theory: Основные понятия и положения. Статья первая. Педагогические измерения. № 2. 2007. С. 3–28;

Аванесов В.С.

Истоки и основные понятия математической теории педагогических измерений (Item Response Theory). Статья вторая. Педагогические измерения. № 3. 2007. С. 3–36.

3

Аванесов В.С.

Из глубины веков. Управление школой, № 12, март 1999 г. С. 4. <http://testolog.narod.ru>

4

Spearman C.

General intelligence, objectively determined and measured. American Journal of Psychology. 1904a. 15. 201–293.

5

Spearman C.

The proof and measurement of association between two things. American Journal of Psychology, 1904b. 15. 72–101.

Spearman C.

1907. Demonstration of formulae for true measurement of correlation. Am. J. of Psychology. 18, 160–169.

ПЕД
измерения

6

Spearman C.
Correlation from faulty
data, British J. of
Psychology. 1910. 3.
271–295.

7

Guttman L.L.
The problem of altitude
and opinion measure-
ment// In S. A. Stouffer
and others. (Eds.).
Measurement and
Prediction. New York:
Wiley. 1950.

8

Andrich D.
An Elaboration of the
Guttman scaling with
Rasch models for meas-
urement. Sociological
Methodology. 1985. The
Jossey-Bass Social and
Behavioral Science Series.
P. 33–80.

9

Деменчёнок О.Г.
Погрешность баллов
единого государственного
экзамена // Педагогиче-
ские измерения. №4.
2011.

10

Аванесов В.С.
Объективность педаго-
гических измерений // Педагогические измерения. № 3. 2008. С. 3–40.

висимости результатов тестирования испытуемых от меры трудности заданий удалось решить математически. Этот успех знаменателен и тем, что предложенным им методом удалось решить одновременно и обратную задачу — определить уровень трудности заданий независимо от уровня подготовленности испытуемых тестируемой группы.

Второй источник становления метрической системы Г. Раша — это психологическая теория Курта Левина о поведении личности, в зависимости от соотношения уровня развития способностей и уровня противодействия. Но эту гипотезу о связи работ Г. Раша с исследованиями К. Левина можно считать пока неисследованной.

Третьим источником можно назвать психометрическую теорию уровня и структуры подготовленности испытуемых, автором которой является Л.Л. Гутман. Основы этой теории впервые были опубликованы в его исследованиях по шкалограммному анализу, в 1950 году⁷. Одним из первых на связь работ Л.Л. Гутмана и Г. Раша аргументированно указал Д. Эндрич⁸.

Сложилось так, что к моменту создания Г. Рашем своей модели две названные теории были уже опубликованы в западной печати.

Идея объективированного измерения

Хотя метрическую систему Г. Раша обычно называют объективной, правильнее было бы говорить о ней иначе, как о системе объективированного педагогического измерения. Потому что в любом измерении всегда имеют место погрешности, иногда недопустимо большие, как это имеет место в баллах ЕГЭ⁹. В педагогических измерениях нередки и субъективные элементы. Как субъективные элементы, некоторые такие элементы не устранимы или трудно устранимы. Целенаправленное уменьшение субъективных элементов есть процесс продвижения к объективности¹⁰.

Поскольку полностью объективных педагогических измерений не бывает, было бы правильнее развивать теорию и методы объективированного измерения. Идея объективированного измерения уровня подготовленности испытуемых и уровня трудности заданий уже давно обсуждается в теории и практике педагогического оценивания. Там неоднократно ставилась задача преодоления зависимости результатов тестирования от уровня трудности заданий теста.

Рассмотрим простые примеры. Сто испытуемых отвечают на лёгкое задание. Результатом

стало 90 % правильных ответов, или иначе доля правильных ответов (р) в данной группе равна 0,900. Из чего можно вывести, что испытуемые хорошо подготовлены. Но если этой же группе испытуемых дать трудное задание, которое способны правильно выполнить только 10% наиболее подготовленных испытуемых, то станет очевидным вывод, что основная масса испытуемых подготовлена плохо.

Где истина? В ситуации взаимозависимости результатов испытуемых от трудности заданий истинный ответ на выборочных совокупностях недостижим. Использование же генеральных совокупностей испытуемых слишком затратно.

Теперь, по аналогии, возьмём две группы испытуемых, хорошо и слабо подготовленных, отвечающих на одно и то же задание. При высоком уровне подготовленности первой группы задание становится лёгким, при низком уровне второй группы испытуемых задание становится трудным. И здесь истинная мера трудности задания остаётся неизвестной.

Как видно из приведённых примеров, оценка индивидуальных результатов испытуемых зависит не только от используемых заданий, но и от состава группы. Если испытуемый оказывался слабее подготовлен, чем другие в данной группе, то он получал низкий рейтинг (за-

нимал низкое место). Если тот же испытуемый попадал в слабую группу испытуемых, то с похожим результатом он получал высокий рейтинг¹¹.

Отмеченные недостатки ряд авторов относят к т.н. классической (статистической теории) тестов¹². Но дело не столько в этой теории, сколько в несоответствии предметов двух теорий и их методов.

Предмет классической теории — разработка качественных тестов, анализ тестовых результатов и статистических свойств заданий теста. В этой теории большое внимание уделялось расчётам мер центральной тенденции, показателям вариации, связи и зависимости тестовых результатов от множества признаков.

Соответственным был и упомянутый статистический аппарат, к которому надо ещё обязательно добавить корреляционный, регрессионный и факторный анализ, имеющих в своей основе общее — исследование вариации результатов зависимых и независимых переменных величин.

Личностная информация здесь определялась через групповые статистики, что было чётко осознано в психометрической литературе как существенный недостаток тестирования. Отсюда возникло научно-педагогическое движение в пользу личностно-ориентированной и

Методология

11

По этому поводу хорошо выразился B.D.Wright. «My ability depended not only on which items I took but on who I was and the company I kept!» См.: Sample-free Test Calibration and Person Measurement. <http://www.rasch.org/me mo1.htm>

12

Так она раньше называлась.

ПЕД
измерения

критериально-ориентированной интерпретации тестовых результатов — вместо нормативно-ориентированной интерпретации.

Необъективность объективной психолого-педагогической информации о каждом испытуемом и о каждом задании — вот главная причина^{13,14}, которая побудила Г. Раша искать математико-статистический метод устранения зависимости результатов одного показателя от другого показателя. В такую возможность мало кто верил, но это оказалось возможным.

Судя по всему, понятие «объективированное» измерение, применительно к RM, в научной оборот впервые введено автором этой статьи. В англоязычной психометрической литературе практически все авторы до сего дня утверждают об объективности метрической системы RM, кроме самого её автора, Г. Раша. При том, что он очень хорошо понимал, что получаемые по его модели данные существенно объективируют результаты измерения, но не являются, всё-таки, объективными в философском смысле этого понятия.

Г. Раш долго искал подходящую характеристику своему подходу, пока не остановился на понятии «специфически объективный метод измерения»¹⁵. Это объяснялось несоответствием философского понятия

математическому методу. Метод не позволял получать данные, отвечающие философским требованиям.

Именно из-за этого, а также из-за того, что данное Г. Рашем определение не имело ни аналогов, ни опоры в научной системе знаний, автор данной статьи вместо «специфически объективных» результатов ввёл другое, уже имевшееся в философии, более подходящее понятие — «объективированное знание». Исходя из мысли, что если есть такое знание, значит должен быть и соответствующий метод. Метрическая система Г. Раша вполне отвечала этому ходу мысли. Психолого-педагогическая система объективированных измерений и есть главный итог применения RM в педагогической науке.

Предметом метрической системы Г. Раша стало построение такой математической модели тестового процесса, которая позволяла бы преодолеть объективно существующую зависимость упомянутых показателей подготовленности испытуемых и меры трудности, а также предсказать исход соревнования каждого испытуемого при попытке ответить на любое задание теста.

Это открывало затем путь к новой технологии тестового процесса, в которой существенное место занимали статистические методы обоснования каче-

13

Rasch G.
Probabilistic Models for
Some Intelligence and
Achievement Tests.
Copenhagen: Danish
Institute for Educational
Research. Expanded edition 1983. Chicago:
MESA Press; 1960.

14

Rasch G.
On specific objectivity: an
attempt at formalising the
request for generality and
validity of scientific statements. Danish Yearbook
Philos. 1977; 14: 58–94.

15

Аванесов В.С.
Объективность педагогических измерений // Педагогические измерения № 3, 2008. С. 3–40.

ства педагогических измерений, компьютеры и компьютерные программы для вычисления параметров испытуемых, заданий и формальных характеристик качества RM.

Модель Курта Левина

Вторым источником появления модели математической модели Г. Раша стала метафора процесса тестирования, которую Г. Раш высказывал неоднократно. Он рассматривал тестирование как противоборство испытуемого с заданием, исход которого можно предсказать, в зависимости от потенциала сил, участвующих в этом. Эта метафора, скорее всего, имела своим источником теорию К. Левина.

В метафоре Г. Раша испытуемый пытается успешно ответить на задание, преодолевая трудности, стоящие на пути правильного решения. Если испытуемый подготовлен больше, чем требует то-

го задание для успешного решения, то победит испытуемый. Он получит, например, один балл, если успех оценивается в дихотомической шкале. В противоположном случае испытуемый получает ноль.

Подход Г. Раша подтверждается анализом его текстов. Он создавался, похоже, не на пустом месте. К тому времени на Западе уже была известна т.н. теория силового взаимодействия американского психолога Курта Левина, переехавшего в США из Германии после утверждения там фашистского режима.

Согласно прилагаемой схеме¹⁶ (рис. 1), успех в преодолении препятствия по теории К. Левина может прийти либо вследствие усиления действия, либо вследствие ослабления противодействия. У К. Левина к тому времени (1951) уже была разработана общая психологическая концепция успеха в поведении личности, выраженная языком математики.

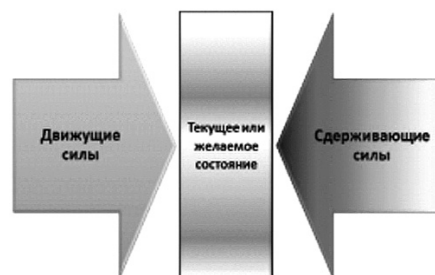


Рис. 1. Схема действия и противодействия

Методология

16

http://change-management.ru/index.php?option=com_content&view=article&id=62:force-field-analysis&catid=36:2010-05-24-09-48-22&Itemid=108

$$\text{Успех} = f(P, S),$$

где успех означает достижение намеченного личностного результата, понимаемого в широком смысле этого слова, f — функция, P — личность (person), S — ситуации, опять-таки понимаемые в широком смысле.

Применительно к модели Г. Раша, которая стала известна в 1953 году, поведение личности её автор толковал узко, как результат одного испытуемого под номером i , при ответе на задание под номером j . После замены понятий «поведение» на «результат тестирования», «личности» — на «испытуемого» и «ситуации» S — на ситуацию противоборства испытуемого с заданием, получалось, что в данной модели под успехом можно понимать правильный ответ испытуемого i по заданию j .

Концептуально модель Г. Раша приняла такой вид:

$$P_{ij} = f(\theta_i, \beta_j),$$

которую надо понимать так: вероятность правильного ответа испытуемого под номером i , отвечающего на задание под номером j зависит только от двух параметров — уровня подготовленности испытуемого (θ_i) и от уровня трудности задания β_j . Г. Раш полагал, что влиянием остальных факторов можно пренебречь, если тестирование

проводится без нарушений. Тогда данные будут соответствовать этой модели.

Если же данные не соответствуют модели влияния главных факторов, то это значит, что некие побочные факторы (часто неизвестные, т.е. латентные) влияют больше, чем эти два основных фактора. Тогда следует вывод, что имеющиеся данные не годятся для его модели.

На этом моменте многие оппоненты Г. Раша оживляются, утверждая, что, может быть, модель не годится для собранных данных. Он, однако, не переставал утверждать и объяснять, что его модель относится не к разряду объясняющих те или иные результаты, а к разряду предписывающих — каковыми они должны быть в рамках его понимания сути тестового процесса, а следовательно, и математической модели. И если какие-то данные не подходят, то это значит только одно — измерение по модели Г. Раша не получилось¹⁷.

И действительно, если испытуемые списывают, угадывают правильные ответы или допускают другие ухищрения, несовместимые с настоящим тестированием, то на такие данные модель Г. Раша не может распространяться. Они не применимы в таких случаях.

Об этом приходится нередко напоминать, когда RM ис-

17
Вот что написал Давид Эндрич: «...when the Rasch model is employed, the objective is to obtain data which fit the model. Andrich, D. (2004). Controversy and the Rasch model: a characteristic of incompatible paradigms? Medical Care, 42, 1–16.

18
Аванесов В.С. Применение тестовых форм в Rasch Measurement // Педагогические измерения. 2005.

пользуется для разработки теста, где представлены преимущественно задания с выбором одного правильного ответа из трёх-четырёх ответов, где вероятность угадывания правильного ответа достаточно высокая¹⁸.

В основе метрической системы Г. Раша находится математическая модель расчёта вероятности правильного ответа каждого отдельно взятого испытуемого, под номером i , на любое задание проектируемого¹⁹ теста под номером j .

$$P_j \{X_{ij} = 1 | \beta_j\} = \frac{\exp(\theta_i - \beta_j)}{1 + \exp(\theta_i - \beta_j)}. \quad (1)$$

Эту формулу часто записывают в строку

$$P_j \{X_{ij} = 1 | \beta_j\} = \exp(\theta_i - \beta_j) / (1 + \exp(\theta_i - \beta_j)), \quad (1a)$$

где $X_{ij} = 1$, если ответ любого испытуемого i на j -е задание правильный; θ_i — уровень знаний, латентная переменная; β_j — уровень трудности j -го задания теста.

Как видно из формулы, параметры функции связаны между собой операцией вычитания $\theta_i - \beta_j$. Чем больше значение разности, тем выше вероятность правильного ответа испытуемого i на задание j .

График зависимости вероятности правильного ответа (Y) от значений разности параметров (испытуемых и заданий $\theta_i - \beta_j$) представлен на рис. 2.

В этой модели вероятность определённого ответа определяется как функция от двух переменных величин: уровня подготовки испытуемых и уровня трудности заданий. Это, по сути, редуцированная мо-

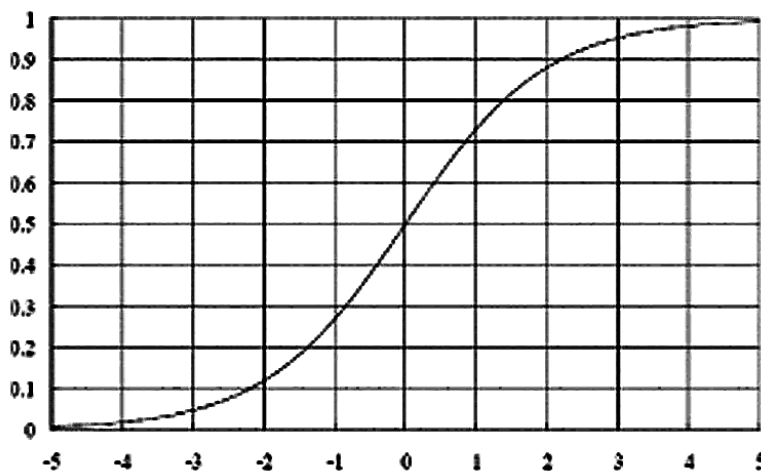


Рис. 2

Методология

19

См. Аванесов В.С. Item response theory: Основные понятия и положения // Педагогические измерения. № 2. 2007. С. 3–28.

ПЕД
измерения

20

Психометрики старой школы не допускали наличие испытуемых с нулевым уровнем интеллекта, справедливо полагая, что какой-то уровень этого качества у любого человека всё-таки есть. Надо только уметь оценить этот уровень. Из этого убеждения, а также из идеи непрерывной вариации этого свойства среди социума там возникла идея континуальности измеряемого свойства, не имеющего ни начала, ни конца. Подробнее см., напр., Guilford J.P. *Psychometric Methods*. McGraw Hill, 1954.

21

Обычно в проектируемом тесте бывает порядка 60 заданий, с учётом вероятности отбраковки примерно половины потенциально некачественных заданий. Здесь даётся, из дидактических соображений, краткий пример результата тестирования по десяти заданиям.

22

Аванесов В.С.
Проблема качества педагогических измерений// Педагогические измерения. №2. 2004. С. 3–27.

дель более широкого подхода к оценке поведенческих актов личности американского психолога Курта Левина.

Научный вклад Л.Л. Гутмана

Третьим источником метрической системы Г. Раша был шкалограммный анализ Л.Л. Гутмана, в котором оценивались уровень и структура подготовленности испытуемых.

В качестве показателя уровня подготовленности Л.Л. Гутман использовал сумму баллов испытуемого i , полученную им по результатам выполнения заданий теста. Эти результаты представляются как вектор строка матрицы результатов тестирования. Чем больше у испытуемого правильных ответов на задания теста, тем выше уровень его подготовленности. Ноль баллов означает отсутствие какой-либо подготовленности, а случай получения всех ответов правильными рассматривается как высший уровень подготовленности испытуемого²⁰.

При разработке показателя структуры подготовленности испытуемого Л.Л. Гутман опирался на введённое им понятие ошибочных элементов в вектор строке тестовых баллов испытуемого i . Ошибочным им назывался элемент, который не укладывался в понятие правиль-

ного профиля испытуемого. Правильным назывался профиль, в котором все нули следовали за всеми единицами.

В нормальных случаях правильно организованного тестирования испытуемый обычно даёт правильные ответы на лёгкие задания и неправильные ответы — на трудные задания. Вот пример правильного профиля испытуемого, отвечавшего на десять заданий: 1 1 1 1 1 0 0 0 0 0²¹. Он знает правильные ответы на первые пять лёгких заданий и не знает правильные ответы на последующие пять трудных заданий.

Подход Л.Л. Гутмана уже однажды освещался в нашем журнале, но в связи с проблемой расчёта коэффициента надёжности результатов испытуемых²². Здесь этот подход затрагивается ещё раз, но теперь уже в другом контексте — исследования метрической системы Г. Раша.

После упорядочения двоичных элементов матрицы тестовых результатов 1/0 (испытываемые — по убыванию баллов, задания — по возрастанию трудности) элементы матрицы принимают вид, напоминающий треугольную матрицу. В левой верхней части (выше главной диагонали) матрицы оказываются преимущественно единицы, а в правой части, ниже главной диагонали матрицы, — нули. Это и есть один из существ-

венных признаков успешности проектируемого педагогического теста.

Каждая матрица тестовых результатов состоит из векторов-строк и векторов-столбцов. По строкам такой матрицы располагаются испытуемые, по столбцам — задания. Совокупность элементов каждой строки матрицы образует, на языке математики, кортеж, а на языке психометрики — профиль испытуемого. Аналогично, совокупность элементов каждого задания, по множеству испытуемых, образует профиль задания.

Профессионально сделанный тест представляет собой систему заданий возрастающей трудности. В этом случае все получаемые профили испытуемых в данной матрице можно разделить на правильные, частично правильные (с ошибками) и совсем неправильные. Так, например, для испытуемых среднего уровня подготовленности, в случае предъявления десяти заданий, можно обнаружить такие типы профилей:

Правильный профиль:

1 1 1 1 1 0 0 0 0 0

Частично правильный профиль

0 1 1 1 1 1 0 0 0 0

Совсем неправильный профиль

0 0 0 0 0 1 1 1 1 1

Ошибкой считается каждый элемент, стоящий не на своём месте. Например, у первого ис-

пытуемого сумма полученных баллов равна 5. При идеальной структуре теста и такой же структуре знаний эти пять баллов должны были располагаться на первых пяти местах, если помнить, что тест представляет систему заданий возрастающей трудности. Отклонение от этого идеала может быть вызвано незнанием, или неудачной попыткой угадать правильный ответ.

В частично неправильном профиле заметен ноль в первом задании, самом лёгком из проектируемого теста. Очевидно, этот результат требует объяснения. Если на первом задании появляются нули несколько раз, при похожих профилях, то это, скорее всего, означает плохую формулировку содержания задания. Её надо менять. Впрочем, могут быть и иные толкования.

Педагогические условия применимости модели Г. Раша

Правильное использование моделей Г. Раша предполагает выполнение ряда исходных педагогических условий:

1. Содержание заданий проектируемого теста должно качественно отражать содержание учебного предмета (учебной дисциплины), по которой проводится педагогическое измерение.

Методология

2. Форма заданий не должна предоставлять ни малейшей возможности угадывания правильного ответа. С точки зрения данного условия, все исследования по разработке теста, имеющего задания с выбором одного правильного ответа из двух-пяти ответов, нуждаются в обосновании. Надо статистически доказать пригодность таких данных, где вероятность угадывания заметна и вполне предсказуема. Это значит, что должны применяться задания открытой формы, что делал и сам Г. Раш.

Это требование, однако, почти повсеместно не выполняется, вследствие чего валидность результатов существенно уменьшается. Критерий пригодности (или непригодности данных, что зависит от толкования), сформулированный Г. Рашем и его последователями, выполняет важную регулятивную функцию, не позволяя снижать валидность результатов измерения относительно применяемой модели.

3. Число дистракторов для заданий педагогического теста желательно иметь в пределах примерно от пяти до 10, если это задания с выбором нескольких правильных ответов. В каждом задании используются только работающие дистракторы, т.е. если их выбирают испытуемые. Для выявления и удаления нера-

ботающих дистракторов необходима предварительная эмпирическая апробация каждого проектируемого педагогического теста.

4. Распределение заданий по трудности должно соответствовать распределению испытуемых по уровню подготовленности.

Уровень трудности части заданий должен быть адекватен соответствующей части испытуемых. В противном случае появляются так называемые экстремальные задания — очень лёгкие, на которые правильно отвечают все испытуемые или, наоборот, слишком трудные задания, на которые не отвечает правильно никто. Такие задания удаляются из теста, нацеленного на применение в данной совокупности.

Распределение заданий по трудности должно быть более или менее равномерным, не допуская разрывов между ними более чем в 0,5 логита.

Диапазон варьирования заданий по трудности не должен искусственно сужаться. В идеале он должен находиться в пределах диапазона различий испытуемых по уровню подготовленности. Обычно этот диапазон бывает меньше. Точно так же бывает меньше и стандартная ошибки уровня трудности заданий, поскольку заданий всегда меньше, чем испытуемых.

Соотношение числа испытуемых к числу заданий — не менее 5–10 испытуемых — должно приходиться на одно задание проектируемого теста. Только в этом случае повышается устойчивость выборочных статистик и становятся приемлемыми результаты факторного анализа.

С точки зрения педагогического измерения в тесте не нужны два и большее число заданий одинаковой трудности.

Система заданий возрастающей трудности, превраща-

ясь в тест, становится операциональным выражением измеряемой переменной величины.

5. При выполнении заданий на результаты испытуемых не должны влиять заметным образом никакие иные факторы, кроме уровня подготовленности испытуемых и уровня трудности заданий. Коррупционные искажения, списывания и иные нарушения этики педагогического контроля не имеют ничего общего с применением метрической системы Г. Раша.

Методология

Методология

РЕКОМЕНДАЦИИ авторам по подготовке текстов для публикации в журнале «Педагогические измерения»

Статьи для публикации просьба присылать в редакторе Word, шрифт 12, с указанием имени и фамилии, названия образовательного учреждения и адреса электронной почты.

НАЗВАНИЕ СТАТЬИ (пишется прописными буквами).

Под названием статьи: имя и фамилия автора пишутся полностью, справа. Название вуза пишется полностью, справа.

Адрес электронной почты автора пишется справа.

Статье предшествует небольшая (до 10 строк) аннотация. Расстояние между строчками в аннотации — один интервал, отступ слева и справа — по 2,5 см. Аннотация заканчивается перечислением примерно пяти-деяти ключевых слов.

Рекомендуемый примерный объём статьи — до 30 страниц. Текст статьи рекомендуется разделить несколькими подзаголовками, из примерного расчёта по одному заголовку на две-три страницы текста.

Сноски делать постраничные, шрифт в носках Times New Roman, 11 размер.

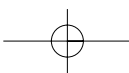
Жирный шрифт не использовать. Расстояние между строчками в тексте статьи — полтора интервала.

Отступы в статье справа и слева — по 2,5 см.

Абзацный отступ — обычный, 1,27 или 1,25 см.

Статью высылать по адресу: testolog@mail.ru

Для своевременного получения бесплатного авторского экземпляра журнала в конце статьи указывается почтовый адрес автора с указанием индекса, фамилии, имени и отчества.



Полнотекстовые номера журнала «Педагогические измерения» публикует Научная электронная библиотека: elibrary.ru

Подписано в печать 25.12.2011

Формат 70×90/16. Бумага офсетная. Печать офсетная. Печ. л. 8. Усл.-печ. л. 9,33.

Заказ № 2201

Издатель: НИИ школьных технологий
109341, Москва, ул. Люблинская, д. 157, корп. 2.
Тел.: (495) 345-52-00.

Отпечатано в типографии НИИ школьных технологий
Тел.: (495) 972-59-62