Магранова Ю.В.
(выпуск 2004 г., специалист, науч. рук. д-р социол. наук Толстова Ю.Н.)

ТЕОРИЯ ТЕСТИРОВАНИЯ КАК ОСНОВА ОЦЕНИВАНИЯ УРОВНЯ ЗНАНИЙ В СОВРЕМЕННОЙ СИСТЕМЕ ОБРАЗОВАНИЯ

Тестирование как способ педагогического контроля в современном учебном процессе

Важнейшим компонентом системы образования и частью учебного процесса является педагогический контроль, имеющий место на всех стадиях процесса обучения. До сих пор результатом педагогического контроля безоговорочно считается оценка успеваемости учащихся. Нынешнюю российскую систему образования нередко обвиняют в субъективности. Знания одних и тех же учащихся различными преподавателями оцениваются по-разному, и расхождение в значении отметок для одной и той же группы учащихся оказывается весьма значительным. Необъективные оценки воспринимаются учащимися как несправедливые и являются главной причиной возникновения конфронтации в учебном процессе [2].

В настоящее время в сфере педагогического контроля все больше осознается необходимость введения в практику обучения количественных методов оценки знаний учащихся. В связи с такой потребностью общества в получении независимой, «объективной» информации об учебных достижениях обучающихся одновременно с традиционной системой оценки и контроля результатов обучения в России начинает складываться новая система — тестирование. В педагогической диагностике под *тестированием* понимаются методы, с помощью которых результаты учебного процесса могут быть измерены (максимально сопоставимы), обработаны и интерпретированы с целью использовать результаты измерений в педагогической практике (подробнее с этим можно ознакомиться в [1]).

Если еще около 15 лет назад тесты практически не применялись в сфере российского образования, то сейчас мы можем наблюдать их широкое распространение в качестве средства педагогического контроля. Около 40 лет вплоть до 80-х гг. использование педагогических тестов в нашей стране было запрещено. Это привело к торможению теоретического и методического развития тестов и тестовых методов. Несмотря на то, что тестовая система становится в России все более популярной, сам тестовый аппарат и его методы развиты в нашей стране очень слабо [7].

Немало руководителей учебных заведений считает, что их преподаватели в состоянии «придумать» за короткое время сколько угодно «тестов». На самом же деле можно придумать сколько угодно заданий в тестовой форме (а это еще не тесты). *Педагогическим тестом* называется система заданий специфической формы, определенного содержания, возрастающей трудности — система, создаваемая с целью объективно оценить структуру и измерить уровень подготовленности учащихся, студентов. Поэтому до начала тестирования каждое задание обязательно должно быть подвергнуто предварительной эмпирической проверке. В процессе проверки многие задания (обычно больше половины) не выдерживают предъявляемых к ним требований и потому не включаются в тест [1].

Таким образом, возрастающая популярность применения тестов в образовании, с одной стороны, и слабое применение математического аппарата разработки и проверки тестов, с

другой, делают актуальным обращение к современной теории тестирования как способу оценивания знаний учащихся.

Теория педагогического тестирования

В теории тестирования выделяют два подхода к оценке уровня подготовленности учащихся: классический и современный. В данной работе акцент делается на современную теорию тестирования, называемую на Западе Item Response Theory (IRT). Это связано с обнаружением некоторых серьезных недостатков в ходе применения классической теории тестирования (с классической теорией тестирования можно ознакомиться в [6] и [7]). Во-первых, в основе теории лежит трудоемкий и неудобный метод параллельных форм, предполагающий двукратное тестирование учеников. Во-вторых, в классической теории имеются затруднения, связанные с определением уровня знаний учащихся. Кроме того, используется порядковая шкала. Это не позволяет сравнить попарно различие индивидуальных баллов испытуемых.

Средством преодоления отмеченных недостатков классической теории является создание таких тестов, в которых индивидуальный балл испытуемых не зависел бы от состава совокупности тестируемых, изменяющегося числа заданий и их трудности. Требуется разработать такую стратегию тестирования, чтобы каждый учащийся отвечал на задания, которые ближе других соответствуют уровню его знаний. Именно такой подход к созданию педагогических тестов и к интерпретации результатов их выполнения представлен в IRT, получившей широкое развитие в 50—60-е гг. в ряде западных стран. IRT является частью более общей теории — латентно-структурного анализа (ЛСА).

Идея ЛСА основана на предположении, что наблюдаемое поведение (например, ответы индивидов на вопросы теста или анкеты) есть внешнее проявление некоторой скрытой (латентной) характеристики, присущей индивидам. Задача метода заключается в том, чтобы, изучив наблюдаемое поведение индивидов, вывести эту скрытую характеристику и разделить (классифицировать) индивидов по сходству (равенству) ее значений ([4], [5]).

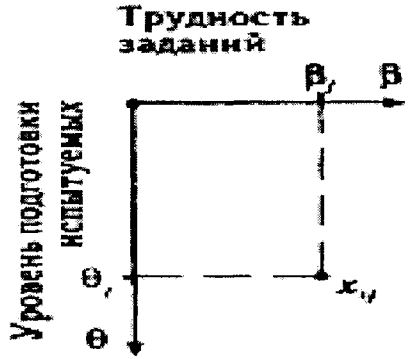

Рис. 1. Взаимодействие множеств латентных параметров

В IRT латентные качества трактуются как способности испытуемых или как уровни подготовки по предмету в зависимости от целей измерения, которые выдвигаются при создании педагогического теста. В рамках основного предположения устанавливается связь между латентными параметрами испытуемых и наблюдаемыми результатами выполнения теста [8].

Элементы первого множества — это значения латентного параметра, определяющего уровень подготовки N испытуемых θ_i (i = 1, 2, ..., N). Второе множество образуют значения латентного параметра β_j (j = 1, 2, ..., n), равные трудностям *n* заданий теста. Первым предположение о том, что параметры θ и β можно оценить в одной шкале, ввел датский математик Г.Раш. Значение параметра θ_i можно рассматривать как положение i-го испытуемого, а значение β_j,— как положение j-го задания на одной и той же оси переменных θ и β.

Абсолютная величина разности $|\theta_i - \beta_j|$ — это расстояние, на котором находится испытуемый с уровнем подготовки θ_i от задания с трудностью β_j. Если эта разность велика по модулю и отрицательна, то задание бесполезно для измерения уровня знаний i-го ученика. Ученик наверняка не может выполнить его верно. Большие положительные значения этой

разности тоже не представляют интереса ни для процесса контроля, ни для обучения i-го испытуемого. Задание такой трудности давно им освоено, и он наверняка справится с ним успешно при выполнении теста. С точки зрения подхода, предлагаемого в IRT, такие задания неэффективны для оценивания данного значения θ.

Предполагается, что на основе параметров θ и β можно найти вероятность того, что ученик с уровнем подготовки θ_i правильно ответит на задание теста с уровнем трудности β_j.

Для тестовых заданий с одним вариантом ответа имеются основные модели IRT для нахождения такой вероятности. Это однопараметрическая модель Г.Раша и двухпараметрическая модель А.Бирнбаума [8].

Практическое применение математических моделей IRT для измерения уровня знаний тестируемых и уровня трудности тестовых заданий

Рассмотрим применение этих моделей на практике на примере анализа результатов ЕГЭ (индивидуальных тестовых баллов) по математике у самарских школьников (результаты ЕГЭ представлены в Приложении 1). На основе простой случайной выборки при помощи датчика случайных чисел из 14 средних общеобразовательных школ города Самары была выбрана школа № 7. Все приведенные в практической части расчеты основываются на результатах ЕГЭ 2003 по математике у выпускного класса СОШ № 7 города Самары (всего в классе 22 человека). ЕГЭ по математике в 2003 г. включал 22 задания: 13 примеров и 9 задач.

Было проведено исследование того, как на основе математических моделей IRT можно оценить уровень знаний учеников и трудность выполненных ими заданий тестирования. Сначала было продемонстрировано применение модели Раша.

Применение однопараметрической модели Раша

Алгоритм расчета уровня подготовленности учеников по школьной дисциплине «Математика» (параметр θ) и

уровня трудности тестовых заданий по математике, представленных на ЕГЭ 2003 (параметр β), можно разбить на ряд этапов.

Этап 1. На первом этапе производится подсчет долей правильных и неправильных ответов каждого испытуемого на все задания теста. Доля правильных ответов i-го ученика находится по формуле:

$$p_i = \frac{X_i}{n},$$

где: X_i – количество заданий, правильно сделанных i-м учеником; i = 1, 2, ..., N; n – число заданий в тесте.

Доля неправильных ответов $q_i = 1 - p_i$, i = 1, 2, ..., N.

Доля правильных ответов p_i для изучаемой нами совокупности варьируется от 0,14 (для учеников, сделавших правильно 3 задания из теста с 22 заданиями) до 0,73 (для учеников, сделавших правильно 16 заданий из теста с 22 заданиями).

Этап 2. На втором этапе производится предварительная оценка значений параметра, характеризующего уровень подготовки учеников тестируемой группы. Начальные значения параметра оцениваются в логитах. Логит уровня подготовки i-го ученика θ_i^0 находят по формуле:

$$\theta_i^0 = \ln \frac{p_i}{q_i},$$

где p_i и q_i – доли правильных и неправильных соответственно ответов i-го ученика на задания теста.

В соответствии с долями правильных и неправильных ответов на задания теста для каждого ученика было рассчитано индивидуальное начальное значение логитов уровня подготовки испытуемых θ_i^0. Теоретически начальные значения θ и β могут меняться в интервале $(-\infty, +\infty)$. Но практически при $\theta_i - \beta_i < -5$ значения P_{ij} (P_{ij} – вероятность того, что респондент со способностью θ_i даст правильный ответ на тест трудности β_j) близки к нулю. Аналогичная пограничная

ситуация наблюдается, когда $\theta_i - \beta_j > 5$, тогда P_{ij} очень близка к единице. В исследуемом примере самому низкому начальному значению логита уровня знаний соответствует уровень подготовки ученика, равный «−2,197» логита. Самому высокому значению уровня знаний в исследуемой группе соответствует значение, равное «0,847» логита.

Этап 3. На третьем этапе подсчитываются доли правильных p_j и неправильных q_j ответов на каждое задание теста:

$$p_i = \frac{R_j}{N},\ q_j = 1 - p_j,$$

где R_j – количество правильных ответов на j-е задание теста, j = 1, 2, ..., n и n – число заданий в тесте.

Доля правильных ответов R_j для каждого из тестового набора заданий ЕГЭ по математике в изучаемой нами совокупности варьируется от 0,05 (для заданий, которые выполнил один ученик из изучаемой совокупности) до 0,73 (для заданий, которые выполнило 16 учеников из изучаемой совокупности).

Этап 4. На четвертом этапе производится предварительная оценка значений параметра β, характеризующего трудность заданий теста. В качестве меры трудности заданий выбирается единица измерения, называемая логитом. Логит трудности j-го задания равен:

$$\beta_j^0 = \ln \frac{q_j}{p_j},$$

где p_j и q_j – доли правильных и неправильных ответов на j-е задание теста.

В исследуемом примере самому низкому начальному значению логита трудности задания соответствует уровень трудности, равный «−0,981» логита. Самому высокому значению уровня трудности предложенных заданий соответствует значение, равное «3,045» логита.

Этап 5. На пятом этапе подсчитываются средние значения логитов уровня подготовки и логитов трудности заданий теста. Среднее значение уровня знаний испытуемых для множества θ_i^0 (i = 1, 2, ..., N) подсчитывают по формуле:

$$\bar{\theta} = \frac{\sum_{j=1}^{N} \theta_j^0}{N},$$

где θ_i^0 – начальные значения уровня подготовки i-го ученика; N – число учеников в группе.

Среднее значение уровня трудности заданий для множества β_i^0 будет:

$$\bar{\beta} = \frac{\sum_{j=1}^{n} \beta_j^0}{n},$$

где β_i^0 – начальные значения логитов трудности заданий; n – число заданий теста.

Среднее значение уровня подготовки изучаемой выборочной совокупности составило «−0,488» логита уровня подготовки. Среднее значение уровня трудности задания данного теста составило «0,532» логита трудности.

Этап 6. После завершения пятого этапа оценки каждого из параметров θ и β будут выражены в интервальной шкале, но с разными значениями средних и разными стандартными отклонениями. На шестом этапе начальные значения логитов уровней подготовки и трудности заданий теста переводятся в единую интервальную шкалу стандартных оценок. Стандартизация достигается с помощью ряда специальных преобразований, для осуществления которых вычисляются:

- дисперсия по множеству значений θ_i^0 (i = 1, 2, ..., N)

$$V = \frac{\sum_{j=1}^{N} (\theta_i^0)^2 - N(\bar{\theta})^2}{N - 1};$$

- дисперсия по множеству β_i^0 (j = 1, 2, ..., n)

$$U = \frac{\sum_{j=1}^{n}(\beta_j^0)^2 - n(\bar{\beta})^2}{n-1};$$

• поправочные коэффициенты

$$X = \sqrt{\frac{1 + U / 2{,}89}{1 - UV / 8{,}35}}, \quad Y = \sqrt{\frac{1 + V / 2{,}89}{1 - UV / 8{,}35}}.$$

Оценки параметров и и β в единой интервальной шкале находятся по формулам:

$$\theta_i = \bar{\beta} + X\theta_i^0, \quad \beta_j = \bar{\theta} + Y\beta_j^0,$$

где все обозначения прежние, а параметры θ и β имеют оценки θ_i и β_j в стандартной интервальной шкале [8].

Таблица 1

Расчеты показателей для вычисления поправочных коэффициентов

Показатель	Значение	Показатель	Значение
V	0,6253	U	1,9032
X	1,3908	Y	1,1910

Роль двух последних формул в развитии IRT трудно переоценить, хотя на первый взгляд они имеют узкую практическую направленность. Эти формулы позволяют преодолеть ряд существенных недостатков классической теории тестов, поскольку с их помощью можно получить объективные оценки параметров испытуемых и заданий, не зависящие друг от друга и выраженные в единой интервальной шкале.

Значения расчетов по данным показателям для нашей выборочной совокупности приведены в табл. 1.

Таким образом, на основе одной из математических моделей IRT (в данном случае модели Раша) были получены

стандартные оценки уровня подготовки рассматриваемой группы испытуемых (табл. 2) и стандартные оценки параметра трудности заданий ЕГЭ по математике для этих испытуемых (табл. 3).

$\theta_i = 0{,}532 + 1{,}3908\,\theta_i^0$, $b_i = (-0{,}4882) + 1{,}1910\,b_i^0 = 1{,}1910\,b_i^0 - 0{,}4882$.

Тест считается удачно сбалансированным по трудности заданий, если:

$$\sum_{j=1}^{n}\beta_j^0 = 0.$$

В нашем исследовании данный параметр равен 3,188 > 0. Это означает, что в предложенном выпускникам тесте ЕГЭ по математике наблюдалось избыточное количество трудных заданий.

Этап 7. На седьмом этапе оценивается стандартная ошибка измерения $S_e(\theta_i)$, которая вычисляется для каждого значения θ_i (i = 1, 2, ..., N):

$$S_e(\theta_i) = \frac{X}{\sqrt{p_i(n - X_i)}} = \frac{X}{\sqrt{np_i(1 - p_i)}} = \frac{X}{\sqrt{np_iq_i}}.$$

Этап 8. На восьмом этапе оценивается стандартная ошибка измерения $Se(\beta_j)$, которая вычисляется для каждого значения β_j.

$$S_e(\beta_j) = \frac{Y}{\sqrt{p_j(N - R_j)}} = \frac{Y}{\sqrt{Np_j(1 - p_j)}} = \frac{Y}{\sqrt{Np_jq_j}}.$$

Проделанные нами вычисления позволяют определить, какова вероятность того, что ученик с определенным уровнем подготовки выполнит задание теста правильно. В табл. 4 представлены значения функции вероятности 22 заданий теста ЕГЭ по математике P(j) для учеников с уровнем подготовки от − 5 до +3 логитов. Значения вычислялись на основе однопараметрической модели Раша по формуле:

Таблица 2

Стандартные оценки уровня подготовки испытуемых

i	Вектор уровня подготовки испытуемых θ_i в логитах	i	Вектор уровня подготовки испытуемых θ_i в логитах
1	1,71	13	−2,5243
2	−0,0323	14	0,5316
3	0,5316	15	−0,0323
4	−1,3965	16	−0,6468
5	0,5316	17	0,5316
6	−1,3965	18	0,5316
7	0,5316	19	−0,6468
8	1,0955	20	-0,6468
9	−0,6468	21	0,5316
10	0,5316	22	−1,3965
11	−1,3965	23	−1,3965
12	1,71		

$$P_j(\theta) = \frac{e^{1,7a_j(\theta-\beta_j)}}{1 + e^{1,7a_j(\theta-\beta_j)}}.$$

На основе этих значений строятся характеристические кривые заданий теста. Анализ их взаимного расположения позволяет наметить пути дальнейшего совершенствования теста и сформировать систему заданий, наиболее эффективных для оценки уровня подготовки каждого испытуемого

Таблица 3

Стандартные оценки трудности заданий теста

j	Вектор уровня трудности заданий β_j, логитов	j	Вектор уровня трудности заданий β_j, логитов
1	−1,3959	12	−1,6564
2	−1,3959	13	−0,2711
3	−1,3959	14	2,2542
4	−1,6564	15	3,1378
5	−1,6564	16	−0,4882
6	0,1783	17	2,2542
7	0,4195	18	1,7102
8	−0,2711	19	−0,7053
9	−0,4882	20	1,3032
10	−0,9262	21	3,1378
11	−1,1547	22	2,2542

выборки. На рис. 2 представлены характеристические кривые 22 заданий теста ЕГЭ по математике, основанные на значениях табл. 4. Для заданий с одинаковыми функциями P(j) строилась одна общая характеристическая кривая.

Процесс совершенствования теста начинается с удаления лишних заданий, нарушающих нормальный характер распределения значений β. Далее разработчику необходимо обратить внимание на случаи наложения характеристических кривых и избавиться от лишних заданий, которые ничего не дают для теста как совокупности работающих заданий возрастающей трудности. В нашем примере друг на друга накладываются характеристические кривые заданий 6 и 18.

Это означает, что при использовании данного теста в дальнейшем одно из этих заданий надо убрать, так как оно не несет никакой ценности для оценивания уровня знаний тестируемых. Следующий важный шаг при коррекции теста связан с выделением «пустых» интервалов оси θ, где нет характеристических кривых. В тест необходимо добавить зада-

Таблица 4

Значения функции $P_j(\theta)$ на основе модели Раша

№ зад.	Значения θ											
	-5	-4,5	-4	-3,5	-3	-2,5	-2	-1	0,5	1	2	3
1	0,002	0,005	0,012	0,027	0,061	0,133	0,264	0,662	0,962	0,983	0,997	0,999
2	0,002	0,005	0,012	0,027	0,061	0,133	0,264	0,662	0,962	0,983	0,997	0,999
3	0,002	0,005	0,012	0,027	0,061	0,133	0,264	0,662	0,962	0,983	0,997	0,999
4	0,003	0,008	0,018	0,042	0,092	0,192	0,358	0,753	0,975	0,989	0,998	1,000
5	0,003	0,008	0,018	0,042	0,092	0,192	0,358	0,753	0,975	0,989	0,998	1,000
6	0,000	0,000	0,001	0,002	0,004	0,010	0,024	0,119	0,633	0,802	0,957	0,992
7	0,000	0,000	0,001	0,002	0,005	0,011	0,025	0,124	0,645	0,809	0,959	0,992
8	0,000	0,001	0,002	0,004	0,010	0,022	0,050	0,225	0,788	0,897	0,979	0,996
9	0,000	0,001	0,003	0,006	0,014	0,032	0,071	0,295	0,843	0,926	0,986	0,997
10	0,001	0,002	0,005	0,012	0,029	0,064	0,139	0,469	0,919	0,964	0,993	0,999
11	0,001	0,003	0,008	0,018	0,042	0,092	0,192	0,565	0,943	0,975	0,995	0,999
12	0,003	0,008	0,018	0,042	0,092	0,192	0,358	0,753	0,975	0,989	0,998	1,000
13	0,000	0,001	0,002	0,004	0,010	0,022	0,050	0,225	0,788	0,897	0,979	0,996
14	0,000	0,000	0,000	0,000	0,000	0,000	0,001	0,004	0,048	0,106	0,394	0,780
15	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,001	0,011	0,026	0,126	0,442
16	0,000	0,001	0,003	0,006	0,014	0,032	0,071	0,295	0,843	0,926	0,986	0,997
17	0,000	0,000	0,000	0,000	0,000	0,000	0,001	0,004	0,048	0,106	0,394	0,780
18	0,000	0,000	0,000	0,000	0,000	0,001	0,002	0,010	0,113	0,230	0,621	0,900
19	0,001	0,002	0,004	0,009	0,020	0,045	0,100	0,377	0,886	0,948	0,990	0,998
20	0,000	0,000	0,000	0,000	0,001	0,002	0,004	0,020	0,203	0,374	0,766	0,947
21	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,001	0,011	0,026	0,126	0,442
22	0,000	0,000	0,000	0,000	0,000	0,000	0,001	0,004	0,048	0,106	0,394	0,780

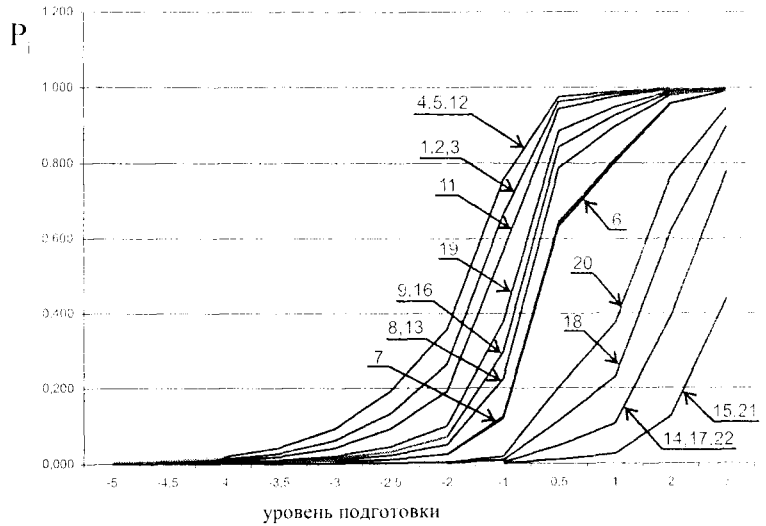

Рис. 2. Характеристические кривые заданий теста

ния, соответствующие по трудности выделенным интервалам на оси латентной переменной θ. В идеале характеристические кривые должны заполнять более или менее равномерно практически весь интервал (-5; +5) шкалы логитов. В исследуемом примере характеристические кривые P(j) заполняют ось и достаточно равномерно.

Преимущества применения модели Раша

Подведем некоторые итоги практического применения модели Раша для оценивания уровня знаний учащихся на основе тестового контроля. Тестовый контроль на примере анализа результатов ЕГЭ по математике школьников города Самары с помощью модели Раша обеспечил ряд положительных возможностей.

1. Возможность дифференциации учеников по уровню подготовленности, а заданий по уровню трудности.

2. Приведение значений уровня подготовленности учеников и уровня трудностей заданий к единой интервальной

ной шкале логитов. Логарифмические оценки уровня знаний и уровня трудности заданий дали возможность сравнить их.

3. Возможность непосредственно сопоставить любое задание с любым испытуемым, и на основе такого сопоставления вычислить вероятность получения правильного ответа.

4. Возможность оценить трудности тестовых заданий вне зависимости от выборки испытуемых.

5. Возможность оценить уровень знаний учащихся вне зависимости от используемого набора тестовых заданий.

6. Простота расчетов по сравнению с другими математическими моделями IRT. Это обусловлено введением только одного параметра уровня знаний для каждого испытуемого и только одного параметра трудности для каждого задания.

7. Благодаря простой структуре модели возможны удобные вычислительные процедуры для многоаспектной проверки адекватности модели: для всего набора тестовых результатов, для каждого испытуемого, для каждого задания и для каждого конкретного ответа. На основе характеристических кривых заданий теста можно осуществить процесс усовершенствования теста. В практическом примере данной работы анализ характеристических кривых показал, что задания 6 и 8 ничего не дают для теста как совокупности работающих заданий возрастающей трудности.

Применение двухпараметрической модели Бирнбаума

Нами было рассмотрено применение однопараметрической модели для анализа результатов применения IRT. Однако что будет, если в расчеты ввести еще один параметр. И насколько будут различаться данные, полученные по однопараметрической и двухпараметрической моделям. Для ответа на эти вопросы продемонстрируем практическое применение модели Бирнбаума для анализа результатов ЕГЭ той же совокупности тестируемых.

Для вычисления условной вероятности правильного выполнения j-го задания теста испытуемыми с различными значениями и в случае двухпараметрической модели А.Бирнбаума используется формула:

$$P_j\{x_{ij}=1|\beta_j\}=\{1+\exp[-1,7a_i(\theta-\beta_i)]\}^{-1},$$

где кроме прежних обозначений вводится новое a_i для 2-го параметра j-го задания теста.

Формула для оценки параметра дифференцирующей способности заданий имеет вид:

$$a_j = \frac{(r_{bis})_j}{\sqrt{1-[(r_{bis})_j]^2}},$$

где r_{bis} – коэффициент бисериальной корреляции, с помощью которого считается корреляция между результатами выполнения каждого задания (дихотомическая шкала) и суммой баллов испытуемых (интервальная шкала) по заданиям теста.

Значения дифференцирующего коэффициента для каждого из 22 заданий представлены в табл. 5.

Значения a_j, близкие к нулю, соответствуют случаю, когда испытуемые с разными уровнями подготовки правильно отвечают на j-е задание с приблизительно равной вероятностью, что противоречит ожидаемым прогнозам разработчиков теста. Эти задания оказываются бесполезными при дифференциации испытуемых группы по оцениваемому параметру, так как они не несут информации об индивидуальных различиях учеников. В нашем примере бесполезными для дифференциации учеников оказываются задания 7, 8, 9 и 17.

Еще более бесполезны задания с отрицательными значениями a_i: на них отвечают правильно с большей вероятностью испытуемые с низким уровнем подготовки, а для знающих учеников с большими значениями и вероятность правильного ответа стремится к нулю. Число заданий в тесте должно сокращаться в первую очередь за счет устранения таких неудачных заданий даже в том случае, когда другие их

Таблица 5

Значения параметра a_j для заданий теста ЕГЭ по математике

№ задания	a_j	№ задания	a_j
1	1,593	12	0,646
2	0,442	13	5,255
3	1,454	14	0,323
4	1,757	15	0,446
5	1,664	16	1,738
6	0,151	17	0,011
7	0,087	18	0,161
8	0,009	19	3,576
9	0,077	20	0,653
10	1,679	21	0,446
11	1,851	22	0,257

характеристики устраивают разработчиков теста. Как правило, такое сокращение приводит к повышению надежности теста. В рассматриваемом нами примере задания с отрицательными значениями a_j отсутствуют.

Отбор заданий с большими значениями a_j является одним из важных принципов при конструировании теста. На практике рекомендуется, как правило, оставлять задания со значениями a_j, лежащими в интервале (0,5; 2,5). Если исходить из данной рекомендации, то при дальнейшем применении данного теста из него надо исключить 12 заданий (задания 2, 6, 7, 8, 9, 10, 14, 15, 17, 18, 21, 22) и предложить вместо них задания, обладающие большей дифференцирующей способностью.

Преимущества применения модели Бирнбаума

Таким образом, применение двухпараметрической модели Бирнбаума для анализа результатов тестирования позволило:

1. Дифференцировать задания по уровню трудности.
2. Определить тестовые задания, не позволяющие достаточно дифференцировать учеников по уровню подготовленности, и заменить их с целью повышения надежности тестирования.

Выбор математической модели IRT: Раш или Бирнбаум

В настоящее время ведутся многочисленные споры о том, какая математическая модель IRT позволяет наиболее точно измерить уровни знаний и уровни трудностей заданий при тестировании в закрытой форме с заданиями в дихотомической шкале. Так, в 2002 г. ГУ «Центром тестирования министерства образования России» было проведено голосование по выбору модели расчета тестовых баллов при проведении централизованного тестирования в 2002 г. В голосовании приняло участие 47 представительств Центра тестирований. Голоса распределились следующим образом: 15 голосов было отдано за модель Раша и 34 голоса за модель Бирнбаума.

Сторонники модели Раша апеллируют к тому, что двухпараметрическая модель Бирнбаума не дает однозначного соответствия между количеством первичных и тестовых баллов выпускника. Одному и тому же количеству верных ответов может соответствовать различный тестовый балл. Сторонники модели Бирнбаума видят ее преимущества в возможности дифференцировать задания по уровню трудности на ее основе. Кроме того, модель Бирнбаума предъявляет гораздо менее жесткие требования к свойствам тестовых заданий в тестировании.

В силу вышесказанного сложно прийти к однозначному выбору той или иной модели для измерения уровня зна-

ний в IRT. Однако считается, что для заданий в закрытой форме предпочтительнее использовать модель Раша. Для теста, содержащего задания с выборочными ответами, лучше применять двухпараметрическую или трехпараметрическую модели Бирнбаума.

Применение современной теории тестирования для оценки уровня знаний учащихся

Таким образом, подведем основные итоги. Использование тестирования в совокупности с традиционной (пятибалльной) системой педагогического контроля позволяет от абстрактных и субъективных оценок перейти к количественным методам измерения знаний. Измерение знаний на основе современной теории тестирования предполагает проведение объективного количественного сопоставления оцениваемого свойства ученика с некоторым эталоном, принятым в качестве единицы измерения. Специальные математические методы и модели измерения теории тестирования обеспечивают переход от «сырых» баллов испытуемых к наиболее правдоподобным оценкам, которые дают оптимальное приближение к истинным компонентам измерения.

В настоящее время работниками российской сферы образования постепенно начинает осознаваться тот факт, что тест — это не просто список заданий с вариантами ответов, который может составить любой преподаватель. Тест — это метод педагогического измерения, состоящий из системы тестовых заданий возрастающей трудности. Все задания теста должны быть четко определены и подвержены обязательной предварительной эмпирической проверке каждого задания до начала тестирования.

Тестирование как способ педагогического контроля начинает складываться в российской системе образования в ответ на проблему объективного измерения знаний. Сейчас активно ведутся попытки применять тестирование не просто на уровне отдельного учебного заведения, но и на уровне федерального масштаба. Зарождается понятие тестирования как системы. Свидетельство этому — введение Централизованного тестирования и эксперименты по введению Единого государственного экзамена (ЕГЭ).

Организации, занимающиеся работой с тестовыми методиками

В настоящее время в нашей стране появилось несколько центров, в которых достаточно профессионально занимаются работой с тестовыми методиками. Среди наиболее активных следует назвать: Центр оценки качества образования Института общего среднего образования РАО, Центр тестирования выпускников общеобразовательных учреждений Российской Федерации, Центр психологического и профессионального тестирования МГУ, Лаборатория аттестационных технологий Московского института повышения квалификации работников образования (МИПКРО), Лаборатория изучения образовательных систем Центра развития образования (г. Санкт-Петербург), Центр аттестации областного института повышения квалификации и переподготовки педагогических кадров (г. Вологда), Научно-информационный центр государственной аккредитации Минобразования России (г. Йошкар-Ола), Исследовательский центр проблем качества подготовки специалистов, Центр аттестации Института развития регионального образования (г. Екатеринбург) и целый ряд других.

Заключение

Вполне понятно, что тестирование, как и любое другое явление, имеет и сторонников, и противников. Не все одобряют введение тестовой системы, аргументируя это тем, что никакие методы измерения не могут заменить преподавателя и его личный опыт. Однако в данной работе не настаивается на универсальности и исключительности тестирования как единственного способа педагогического контроля. Более того, отмечается, что контроль в современной системе образования должен основываться и на оценке знаний, по-

лученной с помощью тестирования, и на оценке знаний, данной преподавателем.

Что же касается тестирования, то в работе на примере результатов ЕГЭ самарских школьников было продемонстрировано применение двух основных математических моделей современной теории тестирования для дихотомических тестов. Первая модель — однопараметрическая модель Г.Раша, определяющая уровень знаний тестируемых и уровень трудности тестовых заданий. Вторая модель — двухпараметрическая модель А.Бирнбаума, выявляющая, помимо уровня знаний и трудности заданий, дифференцирующую способность заданий при измерении различных значений уровня знаний.

Как и во всех странах, где используется современная теория тестирования, у российской системы образования нет четкого ответа на вопрос о том, какая же модель дает более точные оценки уровня знаний учащихся. На Западе обе модели широко распространены. В России для оценивания результатов тестирования тоже используется и модель Раша, и модель Бирбаума. Модель Раша — при анализе результатов ЕГЭ, модель Бирнбаума — при анализе результатов Централизованного тестирования.

В данной работе были показаны преимущества и недостатки каждой из моделей. Модель Раша дала возможность оценить трудность тестовых заданий вне зависимости от выборки испытуемых, а уровень знаний учащихся — вне зависимости от используемого набора тестовых заданий. Кроме того, наличие в модели Раша только одного параметра уровня знаний для каждого испытуемого и только одного параметра трудности для каждого задания обеспечивает простоту расчетов по сравнению с другими математическими моделями современной теории тестирования. Если говорить о плюсах модели Бирнбаума, то она позволяет дифференцировать задания по уровню трудности, а также определить и заменить задания, неспособные достаточно дифференцировать учеников по уровню подготовленности.

Конечно, при применении каждой модели обнаруживаются свои недостатки. Методика Раша предъявляет жесткие требования к свойствам используемых тестовых заданий. Методика Бирнбаума не дает однозначного соответствия между количеством первичных и тестовых баллов обучаемого. Однако, несмотря на это, использование теории тестирований в педагогическом контроле дает ряд значительных преимуществ, которые и были продемонстрированы в работе. В ходе практического применения математических моделей современной теории тестирования, были отмечены следующие ее плюсы:

- Возможность дифференциации учеников по уровню подготовленности, а заданий — по уровню трудности.
- Приведение значений уровня подготовленности учеников и уровня трудностей заданий к единой интервальной шкале логитов. Логарифмические оценки уровня знаний и уровня трудности заданий дали возможность сравнить их.
- Возможность непосредственно сопоставить любое задание с любым испытуемым, и на основе такого сопоставления вычислить вероятность получения правильного ответа.
- Возможность применения удобных вычислительных процедур для многоаспектной проверки адекватности модели тестирования: для всего набора тестовых результатов, для каждого испытуемого, для каждого задания и для каждого конкретного ответа. На основе характеристических кривых заданий теста можно осуществить процесс усовершенствования теста.

Плюсы применения современной теории тестирования

Таким образом, исходя из положительных аспектов, выявленных при применении теории тестирования на конкретном примере с самарскими школьниками, можно отметить плюсы использования современной теории тестирования для российской системы образования в целом:

• Точность измерений. Оценка, получаемая с помощью теста, более дифференцирована. Высокая точность измерения обеспечивается большей градацией оценки.

• Объективность результатов. Исключается влияние субъективных факторов на определение отметки (строгость или либеральность преподавателя, характер взаимоотношений учителя и учащегося и др.).

• Быстрота и легкость проверки результатов для больших групп учащихся.

• Возможность сравнения результатов тестирования для различных классов, школ, районов, городов.

• Наличие одинаковых для всех учащихся правил проведения педагогического контроля и адекватной интерпретации тестовых результатов.

• Возможность выявления структуры знаний каждого ученика для дальнейшего изменения методики обучения.

• Возможность включения в тесты заданий на все темы изученного материала.

Список использованной литературы

1. Аванесов В.С. Композиция тестовых заданий. М.: Центр тестирования, 2002. 162 с.

2. Афонина Л.А. Критериально-ориентированное тестирование как эффективное средство измерения и оценки учебных достижений учащихся средних образовательных учреждений: Дис. канд. пед. наук. Саратов, 2000.

3. Сайт Информационных технологий в образовании (2003). И. Д. Руднинский. О реализации экспертных технологий, алгоритмов прямого тестирования и модели нечеткого оценивания в интегрированной автоматизированной системе контроля знаний. XIII Международная конференция-выставка «Информационные технологии в образовании». http://ito.edu.ru/2003/ VI/VI-0-2187.html.

4. Толстова Ю.Н. Одномерное шкалирование: тестовая традиция в социологии (построение индексов, шкала Лайкерта, латентно-структурный анализ) // Социология: 4М. Вып. № 8. 1997. С. 54—65.

5. Толстова Ю.Н. Измерение в социологии. М.: Инфра-М, 1998. 224 с.

6. Челышкова М.Б. Организация контроля учебной деятельности студентов в условиях педагогического сотрудничества: Дис. канд. пед. наук. Киев, 1990.

7. Челышкова М.Б. Разработка педагогических тестов на основе современных математических моделей. М.: МИСИС, 1995. 195 с.

8. Челышкова М.Б. Теория и практика конструирования педагогических тестов: Учеб. пособие. М.: Логос, 2002. 432 с.

■■■

Приложение 1

Индивидуальные тестовые баллы выпускников СОШ №7 г. Самары по результатам ЕГЭ в 2003 г.

	СОШ № 7, г. Саратов	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	B1	B2	B3	B4	B5	B6	B7	B8	B9
50	Мырсин Сергей Петрович	1	1	1	1	1	1	0	1	1	1	1	1	1	0	0	1	0	1	1	0	0	1
51	Алексеев Константин Сергеевич	0	1	0	1	0	0	0	0	1	1	1	0	1	1	0	1	1	0	0	0	0	0
52	Колобов Андрей Геннадьевич	1	1	1	1	1	1	0	1	0	1	0	0	0	0	0	1	0	0	1	0	0	0
53	Хлебухов Дмитрий Сергеевич	0	1	0	0	1	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0
54	Филиппова Анна Михайловна	0	1	1	1	1	0	1	0	1	1	1	1	0	0	0	0	0	1	0	0	0	0
55	Мартынов Николай Александрович	0	1	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
56	Северский Владимир Михайлович	1	1	1	1	1	1	0	1	0	1	1	1	0	0	0	1	0	0	1	0	0	0
57	Подгорнов Сергей Александрович	1	0	1	1	1	0	1	0	1	1	1	1	1	0	0	1	0	0	1	1	0	0
58	Маркина Татьяна Александровна	1	1	1	0	0	0	0	0	1	0	1	0	0	0	0	1	0	0	0	0	0	0
59	Юрашевич Светлана Сергеевна	1	1	1	1	1	0	1	0	0	1	1	1	1	0	0	1	0	0	1	0	0	0
60	Никифорова Анастасия Ивановна	1	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0
61	Горшунова Галина Юрьевна	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	0	0	0	1	1	1	0
62	Зимирова Марина Владимировна	0	0	0	0	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0
63	Бухальцева Анна Сергеевна	1	0	0	1	1	1	1	0	0	1	1	1	1	0	0	1	0	0	1	1	0	0
64	Бирюков Дмитрий Александрович	1	1	1	1	0	0	0	1	1	0	1	0	1	0	0	0	1	0	0	0	0	0
65	Бондаренко Ольга Сергеевна	0	1	0	0	1	1	0	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0
66	Карцева Елена Васильевна	1	1	1	1	1	0	1	0	0	1	1	1	1	0	0	1	0	0	1	0	0	0
67	Дробенюк Татьяна Александровна	1	1	1	1	1	0	1	0	0	1	1	1	1	0	0	1	0	0	1	0	0	0
68	Волкова Екатерина Александровна	1	0	1	0	0	0	0	1	1	0	0	1	0	0	0	0	0	0	1	0	0	0
69	Грошева Екатерина Михайловна	0	0	0	1	1	0	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0
70	Миронов Петр Александрович	0	0	1	1	1	1	0	0	0	1	1	1	1	0	0	1	0	0	1	1	0	0
71	Сергеева Екатерина Александровна	1	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
72	Бочина Дарья Анатольевна	1	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	1