PI_2_2010.qxd 21.06.2010 23:02 Page 22

ТЕСТОВЫЕ ЗАДАНИЯ С ВЫБОРОМ НЕСКОЛЬКИХ ПРАВИЛЬНЫХ ОТВЕТОВ КАК СРЕДСТВО ПОВЫШЕНИЯ ТОЧНОСТИ ПЕДАГОГИЧЕСКОГО ИЗМЕРЕНИЯ

Олег Деменчёнок
Восточно-Сибирский институт МВД России
AskSystem@yandex.ru

Предложена технология оценки нижнего предела стандартной ошибки уровня подготовленности испытуемого на основе анализа количества и способа оценивания тестовых заданий. Описана практическая реализация методики в электронной таблице Microsoft Excel. Показано, что эффективным средством минимизации погрешности педагогического измерения является переход от заданий с выбором одного правильного ответа к заданиям с выбором нескольких правильных ответов.

Ключевые слова: тест, уровень подготовленности, уровень трудности задания, погрешность измерения, математические модели педагогических измерений, Partial Credit Model.

Результаты измерений из-за погрешностей всегда несколько отличаются от истинного значения измеряемой величины. Результатами контроля знаний (или профессиональной компетентности) на основе математических моделей педагогических измерений являются не истинные, а лишь наиболее вероятные значения уровней подготовленности обучаемых и уровней трудности заданий. Погрешность искажает тестовый балл, снижает точность педагогического измерения.

Для практического использования результатов измерений желательна минимизация погрешности до некоторого приемлемого уровня; в идеальном случае — до уровня, при котором погрешностью можно пренебречь. По мнению автора, одним из эффективных средств решения этой задачи могут стать тестовые задания с выбором нескольких правильных ответов.

Основные составляющие погрешности

В статистике различают три основных вида ошибок: грубые, систематические и случайные.

Грубые ошибки возникают вследствие просчёта при вычислении тестового балла или неправильной регистрации результата.

Систематические ошибки однонаправлено либо преувеличивают, либо преуменьшают результаты измерений. Например, случайное угадывание испытуемым правильных ответов увеличивает результат по сравнению с истинным значением тестового балла, т.е. вызывает систематическую ошибку.

Случайная погрешность меняется по величине и по знаку, от измерения к измерению. Источниками случайных ошибок в случае тестового контроля могут стать:

- ограниченность числа заданий. Ограниченный набор заданий не всегда достаточен для полной проверки уровня и структуры знаний. Возникающие ошибки репрезентативности в сочетании с фрагментарностью знаний части обучаемых могут привести к зависимости тестового балла от того, какие именно задания предложены конкретному испытуемому;
- низкая дифференцирующая способность распространённой сейчас дихотомической системы оценки правильности ответа на каждое задание — правильно или неправильно, 1 или 0. Неполные или неточные ответы квалифицируются как незнание ответа, что не всегда оправдано. Правильный ответ, оцениваемый максимальным баллом, не всегда соответствует известным критериям оценки «отлично»[1]. Напомним, это точное и прочное знание материала в заданном объёме, исчерпывающее и логически стройное его изложение, умение обосновывать принятые решения, обобщать материал и др.;
- случайными можно считать также ошибки ввода данных; ошибки, вызванные неверным истолкованием условия выполнения заданий теста и т.п.

Для повышения точности измерений следует стремиться к всемерному снижению как систематических, так и случайных погрешностей.

Снижение систематической погрешности

Основной путь снижения вызываемой угадыванием систематической погрешности — рациональное применение форм тестовых заданий. В.С. Аванесов рекомендует переходить от заданий с выбором одного правильного ответа к заданиям

1 *Буланова-Топоркова М.В. и др.* Педагогика и психология высшей школы. Ростов-на-Дону: Феникс, 2002. 544 с.

с выбором нескольких правильных ответов[2], с числом ответов до 8–12, там, где можно подобрать эффективные дистракторы. Такие задания устойчивы к угадыванию правильного ответа благодаря своей форме и числу дистракторов, что устраняет необходимость коррекции тестового балла.

Тестовое задание с несколькими правильными ответами — задание, в котором правильных ответов может быть несколько, например:

1. ПРОСТЫЕ ЧИСЛА:

☑ 2	☐ 15
☑ 7	☑ 19
☐ 9	☐ 28
☑ 11	☑ 31

В данном задании каждый из ответов выбирается независимо от остальных. Вероятность случайно сделать правильный выбор для любого из элементов равна 0,5, так как нужно угадать, какой из двух возможных вариантов правильный.

По теореме умножения вероятностей независимых событий вероятность угадывания правильного ответа задания определяется произведением вероятностей угадывания для всех k вариантов ответа:

$$P = \frac{1}{2} \cdot \frac{1}{2} \cdots \frac{1}{2} = \frac{1}{2^k}. \qquad (1)$$

При $k = 4$ вероятность угадывания равна $^1/_{16} \approx 0{,}06$, при $k = 6$ вероятность $^1/_{64} \approx 0{,}016$, при $k = 10$ вероятность менее 0,001.

Иногда учитываются частично правильные ответы. В.С. Аванесов предлагает при двухбалльной оценке за правильное выполнение задания снимать один балл за одну допущенную ошибку, за две и более ошибки — снимать оба балла. При трёхбалльной оценке — снимать один балл за одну допущенную ошибку и снимать два балла за вторую допущенную ошибку. Очевидно, что учёт частично правильных ответов снижает устойчивость к угадыванию.

Поэтому целесообразно оценить величину этого снижения.

Анализ устойчивости к угадыванию при учёте частично правильных ответов

Используя формулу Бернулли[3], несложно получить выражение вероятности угадывания с одной ошибкой P_1 и двумя ошибками P_2 для задания с несколькими правильными ответами:

$$P_1 = C_k^{k-1} p^{k-1} (1-p)^1 = \frac{k!}{1!(k-1)} \left(\frac{1}{2}\right)^{k-1} \left(1 - \frac{1}{2}\right)^1 = \frac{k}{2^k}, \qquad (2)$$

2 *Аванесов В.С.* Применение тестовых форм в Rasch Measurement//Педагогические измерения, 2005, 4. С. 3–20.

3 *Вентцель Е.С.* Теория вероятностей. М.: Высшая школа, 2001. 576 с.

$$P_2 = C_k^{k-2} p^{k-2} (1-p)^2 = \\ = \frac{k!}{2!(k-2)} \left(\frac{1}{2}\right)^{k-2} \left(1-\frac{1}{2}\right)^2 = \\ = \frac{k(k-1)}{2^k}, \qquad (3)$$

где p = 0,5 — вероятность угадывания при выборе одного из элементов ответа.

Очевидно, что безошибочное угадывание ответа на двухбалльное задание увеличивает тестовый балл на 2 балла, угадывание с одной ошибкой — всего на один балл. Теоретическое среднее значение увеличения тестового балла за счёт угадывания равно математическому ожиданию дискретной случайной величины[4]:

$$\Delta x = \sum_{i=0}^{x_{\max}-1} P_i \cdot \Delta x_i, \qquad (4)$$

где Δx — увеличение тестового балла за счёт угадывания; $x_{\max}$ — количество баллов, начисляемое за безошибочное выполнение тестового задания; P_i и Δx_i — вероятность угадывания правильного ответа с i ошибками и соответствующее увеличение тестового балла.

При двухбалльной оценке уравнение (4) примет вид:

$$\Delta x = \frac{1}{2^k} \cdot 2 + \frac{k}{2^k} \cdot 1 = \frac{k+2}{2^k}. \qquad (5)$$

Если максимальная оценка за задания равна трём баллам, то математическое ожидание увеличения тестового балла за счёт угадывания:

$$\Delta x = \frac{1}{2^k} \cdot 3 + \frac{k}{2^k} \cdot 2 + \frac{k(k-1)}{2^k} \cdot 1 = \\ = \frac{k(k-1) + 2k + 3}{2^k}. \qquad (6)$$

Для сопоставимости влияния угадывания в тестовых заданиях, оцениваемых разным количеством баллов, целесообразно перейти к относительным величинам:

$$\frac{\Delta x}{x_{\max}} \cdot 100\%.$$

Из формул видно, что влияние угадывания зависит от числа вариантов ответа k. Эта зависимость для тестовых заданий с выбором нескольких вариантов ответа, оцениваемых тремя баллами (допускается 1–2 ошибки), двумя (допускается 1 ошибка) или одним баллом (ошибки не допускаются) представлена графически на рис. 1. На этом же рисунке приведён график для задания с выбором одного ответа.

Тестовое задание с выбором нескольких вариантов ответа отличается самой высокой устойчивостью к угадыванию. Учёт частично правильных ответов, допускающий одну ошибку, менее устойчив к угадыванию. Однако даже в этом случае устойчивость к угадыванию существенно выше, чем у задания с выбором одного ответа. Задание с несколькими правильными ответами, оцениваемое тремя баллами, устойчивее к угадыванию, чем задание с выбором одного ответа, если количество вариантов ответа не менее восьми.

4 *Айвазян С.А., Мхитарян В.С.* Прикладная статистика и основы эконометрики. М.: ЮНИТИ, 1998. 1022 с.

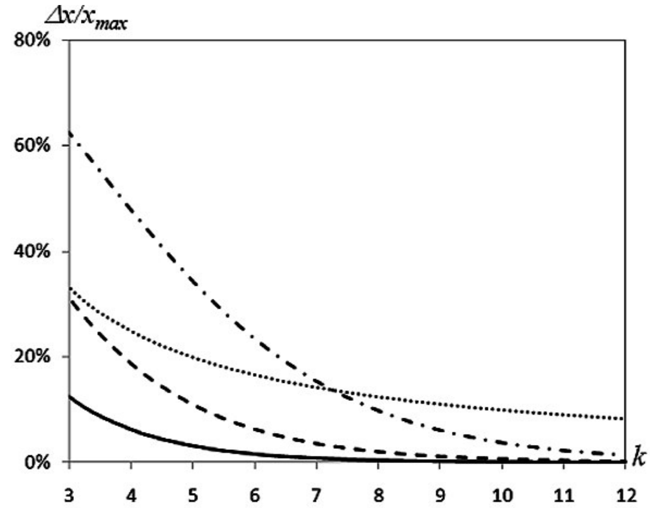

Рис. 1. Влияние угадывания в тестовых заданиях:
———— — выбор нескольких вариантов ответа без ошибок;
— — — — — выбор нескольких вариантов ответа не более чем с одной ошибкой;
– • — • — выбор нескольких вариантов ответа не более чем с двумя ошибками;
• • • • — выбор одного варианта ответа без ошибок

В работе[5] показано, что влиянием угадывания можно пренебречь, если среднее значение вероятности угадывания менее 10%. Это достигается для задания с несколькими правильными ответами:

- оцениваемого одним баллом — при четырёх вариантах ответа ($k = 4$);
- оцениваемого двумя баллами — при $k = 6$;
- оцениваемого тремя баллами — при $k = 8$;
- для задания с одним правильным ответом — при $k = 11$.

Таким образом, переход от заданий с выбором одного правильного ответа к заданиям с выбором нескольких правильных ответов, с числом ответов до 10–12, даже при учёте частично правильных ответов обеспечивает повышение устойчивости к угадыванию.

Следует отметить, что кроме заданий с несколькими правильными ответами весьма устойчивы к угадыванию задания на восстановление последовательности и соответствия, а также задания с кратким свободным ответом[6]. Однако задания на восстановление последовательности и соответствия не вполне универсальны, а зада-

5 *Деменчёнок О.Г.* Влияние угадывания на значение тестового балла: корректировать или устранять?//Педагогические измерения, 2007, 1. С. 56–70.

6 *Деменчёнок О.Г.* Влияние угадывания на значение тестового балла: корректировать или устранять?//Педагогические измерения, 2007, 1. С. 56–70.

ния с кратким свободным ответом отличает сложность синтаксического (тем более — семантического) анализа ответа, невозможность в ряде случаев предусмотреть ввод испытуемыми различных синонимов, различных вариантов написания ответа, частично правильных ответов и т.п.

Снижение случайной погрешности

Тестовые задания с выбором нескольких из предложенных вариантов ответа позволяют различать степень правильности ответа, что повышает информативность результата решения. А чем больше информации, тем точнее наши сведения и меньше ошибка. В теории педагогических измерений количеством информации I [7] называют величину, обратную дисперсии ошибок D, а информационной функцией — соответствующую аналитическую зависимость:

$$I = \frac{1}{D} = \frac{1}{\sigma^2}, \qquad (7)$$

где σ — стандартная ошибка.

Для изучения возможности снижения случайной погрешности необходимо проанализировать модель педагогического измерения.

Математическая модель измерений

Наиболее известной математической моделью педагогических измерений с градацией степени правильности ответа (т.е. с возможностью учёта частично правильных ответов) является модификация модели Раша с произвольными промежуточными категориями выполнения тестового задания, известная в англоязычной литературе как Partial Credit Model (PCM). Эта модель может быть записана в виде[8]:

$$\pi_{ijx} = \frac{e^{\sum_{k=0}^{x}(\theta_i - \beta_{jk})}}{\sum_{l=0}^{x_{\max j}} e^{\sum_{k=0}^{l}(\theta_i - \beta_{jk})}}, \qquad (8)$$

где π_{ijx} — вероятность достижения тестируемым результата x_{ij} (т.е. того, что тестируемый i выполнит ровно x шагов и получит x баллов в задании j); $x = 0, 1...$ $... x_{ij}...x_{\max j}$ — количество шагов; $x_{\max j}$ — максимально возможное количество баллов за задание j; $\beta_{j0} = 0, \ \sum_{n=0}^{0}(\theta_i - \beta_{j0}) = 0.$

Например, для задания, максимально оцениваемого двумя баллами, вероятности получения одного и двух баллов соответственно равны:

$$\pi_1 = \frac{e^{\theta - \beta_1}}{1 + e^{\theta - \beta_1} + e^{\theta - \beta_1} \cdot e^{\theta - \beta_2}}, \qquad (9)$$

---

7 Количество информации — показатель, характеризующий уменьшение неопределённости состояния системы.

8 *Wright B.D., Masters G.N.* Rating Scale Analysis: Rasch Measurement. Chicago: Mesa Press, 1982. 204 p.

$$\pi_2 = \frac{e^{\theta-\beta_1} \cdot e^{\theta-\beta_2}}{1+e^{\theta-\beta_1}+e^{\theta-\beta_1} \cdot e^{\theta-\beta_2}}. \quad (10)$$

Если максимальная оценка задания равна трём баллам, имеем:

$$\pi_1 = \frac{e^{\theta-\beta_1}}{1+e^{\theta-\beta_1}+e^{\theta-\beta_1} \cdot e^{\theta-\beta_2}+e^{\theta-\beta_1} \cdot e^{\theta-\beta_2} \cdot e^{\theta-\beta_3}}, \quad (11)$$

$$\pi_2 = \frac{e^{\theta-\beta_1} \cdot e^{\theta-\beta_2}}{1+e^{\theta-\beta_1}+e^{\theta-\beta_1} \cdot e^{\theta-\beta_2}+e^{\theta-\beta_1} \cdot e^{\theta-\beta_2} \cdot e^{\theta-\beta_3}}. \quad (12)$$

$$\pi_3 = \frac{e^{\theta-\beta_1} \cdot e^{\theta-\beta_2} \cdot e^{\theta-\beta_3}}{1+e^{\theta-\beta_1}+e^{\theta-\beta_1} \cdot e^{\theta-\beta_2}+e^{\theta-\beta_1} \cdot e^{\theta-\beta_2} \cdot e^{\theta-\beta_3}}. \quad (13)$$

Аналогичным образом можно применить уравнение Partial Credit Model (8) для анализа заданий с большим количеством градаций степени правильности ответа.

Для оценки стандартной ошибки измерения уровня подготовленности испытуемого i используется формула[9]:

$$\sigma_{\theta i} = \frac{1}{\sqrt{\sum_{j=1}^{m}\left(\sum_{l=1}^{x_{\max j}} l^2 \cdot \pi_{ijl} - \left(\sum_{l=1}^{x_{\max j}} l \cdot \pi_{ijl}\right)^2\right)}}, \quad (14)$$

где m — число тестовых заданий.

Для заданий, максимально оцениваемых двумя и тремя баллами, уравнение (14) принимает вид:

$$\sigma_{\theta i} = \frac{1}{\sqrt{\sum_{j=1}^{m}\left(\pi_{ij1}+4\pi_{ij2}-\left(\pi_{ij1}+2\pi_{ij2}\right)^2\right)}}, \quad (15)$$

$$\sigma_{\theta i} = \frac{1}{\sqrt{\sum_{j=1}^{m}\left(\pi_{ij1}+4\pi_{ij2}+9\pi_{ij3}-\left(\pi_{ij1}+2\pi_{ij2}+3\pi_{ij3}\right)^2\right)}}. \quad (16)$$

Оценки стандартной ошибки измерения уровня подготовленности испытуемого для анализа заданий с большим количеством градаций степени правильности ответа находятся аналогичным образом.

Для дихотомических заданий ($x_{\max j} = 1$) формула (14) существенно упрощается:

$$\sigma_{\theta i} = \frac{1}{\sqrt{\sum_{j=1}^{m}\left(\pi_{ij1}-\pi_{ij1}^{\ 2}\right)}} = \frac{1}{\sqrt{\sum_{j=1}^{m}\pi_{ij1}(1-\pi_{ij1})}} = \frac{1}{\sqrt{\sum_{j=1}^{m}P_{ij}(1-P_{ij})}}, \quad (17)$$

где $p_{ij1} = P_{ij}$ — вероятность получения тестируемым i одного балла в задании j, что для дихотомического задания соответствует вероятности правильного ответа испытуемого i на задание j.

Вероятность правильного ответа для дихотомического задания может быть найдена по модели Георга Раша[10]:

$$P_{ij} = \frac{e^{\theta_i-\beta_j}}{1+e^{\theta_i-\beta_j}} = \frac{1}{1+e^{-(\theta_i-\beta_j)}}. \quad (18)$$

9 Там же: *Wright B.D., Masters G.N.* Rating Scale Analysis: Rasch Measurement. Chicago: Mesa Press, 1982. 204 p.

10 Rasch G. Probabilistic models for some intelligence and attainment tests. Copenhagen, Denmark: Danish Institute for Educational Research, 1960.

Анализ стандартной ошибки измерения уровня подготовленности

Сравним стандартные ошибки измерения уровня подготовленности испытуемых по результатам выполнения двух тестов. Первый тест включает дихотомические задания (оцениваемые 0 или 1 баллом), а второй — политомические тестовые задания, выполнение которых допускает несколько различающихся по баллам категорий ответа (например, полностью верный ответ — 2 балла, частично верный ответ — 1 балл, неверный ответ — 0 баллов).

Для проведения пробных расчётов примем тест из 20 заданий. Распределение уровней трудности заданий β для пробных расчётов примем близким к равномерному:

— 20 дихотомических заданий равномерно распределены по уровню трудности от –5 до 5 (шаг равен $^{10}/_{19}$, т.е. $\beta_1 = -5$; $\beta_2 = -4{,}47$; $\beta_3 = -3{,}95$ ... $\beta_{19} = 4{,}47$; $\beta_{20} = 5$);

— для 20 политомических заданий примем, что максимальная оценка равна двум баллам; уровни трудности первого шага меньше уровней трудности второго шага на один логит, а распределение уровней трудности равномерное:

- $\beta_{1\,1} = -5$; $\beta_{1\,2} = -4$ ($\beta_{1\,2}$ — уровень трудности второго шага в 1-м задании);
- $\beta_{2\,1} = -4{,}53$; $\beta_{2\,2} = -3{,}53$;
- $\beta_{3\,1} = -4{,}05$; $\beta_{3\,2} = -3{,}05$;

. . . .

- $\beta_{19\,1} = 3{,}53$; $\beta_{19\,2} = 4{,}53$;
- $\beta_{20\,1} = 4$; $\beta_{19\,2} = 5$.

Все расчёты выполним в электронной таблице Microsoft Excel. Это позволит любому желающему перепроверить полученные результаты или провести оценку погрешности измерений для теста любой другой структуры.

Сначала введём исходные данные (см. рис. 2) — идентификаторы заданий (строка 2) и испытуемых (столбец А), используя для простоты нумерацию. Примем, что уровни подготовленности испытуемых θ меняются от –5 до 5 с шагом 0,5 (столбец В). Ранее принятые уровни трудности отдельных шагов β_1, β_2 ... β_{20} поместим в ячейки B3:V3.

Далее рассчитаем стандартную ошибку для уровня подготовленности испытуемых по формуле (17). Для этого в ячейку C4 запишем формулу = (1/(1+EXP(C$3-$B4)))*
*(1-1/(1+EXP(C$3-$B4))) и скопируем содержимое C4 в диапазон ячеек C4:V24. Затем просуммируем количество информации по каждому испытуемому: поместим в ячейку W4 = СУММ(C4:V4) и скопируем эту формулу в диапазон ячеек W4:W24. Осталось извлечь квадратный корень и найти обратную величину: в ячейку

C4	f_x	=(1/(1+EXP(C$3-$B4)))*(1-1/(1+EXP(C$3-$B4)))								
	A	B	C	D	E	...	U	V	W	X
2			Задание 1	Задание 2	Задание 3	...	Задание 19	Задание 20	I	σ
3			-5	-4,47	-3,95	...	4,47	5		
4	Студент 1	-5	0,25	0,23	0,19	...	0,00	0,00	1,07	0,96
5	Студент 2	-4,5	0,24	0,25	0,23	...	0,00	0,00	1,30	0,88
6	Студент 3	-4	0,20	0,24	0,25	...	0,00	0,00	1,48	0,82
...	...	...	...	...	...	...	...	...	...	...
24	Студент 21	5	0,00	0,00	0,00	...	0,23	0,25	1,08	0,96

Рис. 2. Ввод исходных данных теста с дихотомическими заданиями

X4 запишем =1/W4^0,5 и скопируем содержимое X4 в диапазон X4:X24.

Для политомических заданий расчётные формулы составляются на основе уравнения (8). Так, для первого задания, которое оценивается двумя баллами, в ячейки C4 и D4 свободного листа запишем на основе выражений (9–10) формулы вероятностей получения одного и двух баллов:

- =EXP($B4-C$3)/(1+EXP($B4-C$3)+EXP($B4-C$3)*EXP($B4-D$3))
- =EXP($B4-C$3)*EXP($B4-D$3)/(1+EXP($B4-C$3)+EXP($B4-C$3)*EXP($B4-D$3))

Затем скопируем содержимое C4:D4 в диапазон C4:AP24.

Осталось найти погрешности измерений. Для этого запишем входящие в уравнение (15) слагаемые в диапазон C29:AP49. Так, в ячейку C29 введём =C4+4*D4-(C4+2*D4)^2, что соответствует выражению $\pi_{11\,1} + 4\pi_{11\,2} - (\pi_{11\,1} + 2\pi_{11\,2})^2$. Затем скопируем содержимое C29 в диапазон C29:C49. Слагаемые остальных заданий вводятся аналогично.

C4	f_x	=EXP($B4-C$3)/(1+EXP($B4-C$3)+EXP($B4-C$3)*EXP($B4-D$3))									
	A	B	C	D	E	F	...	AO	AP	AQ	AR
2			Задание 1		Задание 2		...	Задание 20		I	σ
3			-5	-4	-4,53	-3,53	...	4	5		
4	Студент 1	-5	0,42	0,16	0,35	0,08	...	0,00	0,00	1,81	0,74
5	Студент 2	-4,5	0,45	0,27	0,43	0,16	...	0,00	0,00	2,38	0,65
6	Студент 3	-4	0,42	0,42	0,45	0,28	...	0,00	0,00	2,92	0,59
...	...	...	...	...	...	...	...	...	...	...	...
24	Студент 21	5	0	1	0	1	...	0,42	0,42	1,81	0,74

Рис. 3. Расчёт вероятностей получения одного и двух баллов в тесте с политомическими заданиями

Далее просуммируем количество информации по каждому испытуемому: поместим в ячейку AQ4 =СУММ(C29:AO29) и скопируем эту формулу в диапазон ячеек AQ4:AQ24. Теперь рассчитаем стандартную ошибку измерения уровня подготовленности, для чего в ячейку AR4 введём =1/AQ4^0,5 и скопируем эту формулу в диапазон AR4:AR24.

Результаты свидетельствуют о нелинейном характере зависимости стандартной ошибки измерения σ от уровня подготовленности θ (рис. 4). Стандартная ошибка минимальна при уровне подготовленности, равном нулю; по мере удаления от центра шкалы стандартная ошибка увеличивается. Характер зависимости одинаков для всех графиков, что объясняется одинаковым характером распределения уровней трудности заданий. Даже простое сравнение графиков наглядно выявляет преимущество политомического оценивания в точности педагогического измерения. Тест из политомических заданий, оцениваемых двумя баллами, позволяет снизить стандартную

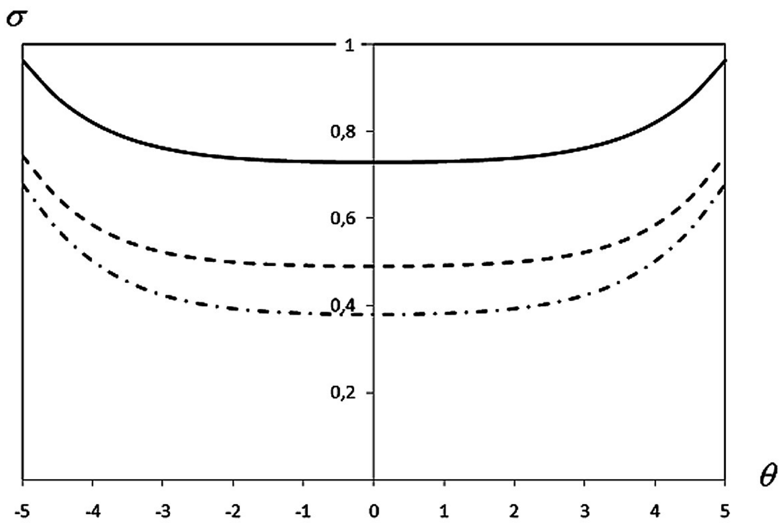

Рис. 4. Зависимость стандартной ошибки измерения от уровня подготовленности:

———— — тест из дихотомических заданий;

— — — — — тест из политомических заданий, оцениваемых двумя баллами;

– • — • — — тест из политомических заданий, оцениваемых тремя баллами ошибку определения уровня подготовленности испытуемых на 33% по сравнению с дихотомическим оцениванием. Для теста из политомических заданий, оцениваемых тремя баллами, снижение ещё больше — почти 50%.

Технология оценки нижнего предела погрешности

Самый точный измерительный инструмент даёт минимальную ошибку при измерении любого объекта. Следовательно, лучшим следует признать тот набор заданий, при котором стандартная ошибка минимальна для испытуемого с любым уровнем подготовленности. Формально это условие можно записать в виде:

$$\max\left(\sigma_i(\theta_i,\beta_j)\right)\to\min \quad \text{при} \quad i=1.. \qquad (19)$$

По сути, выполнение условия (19) означает достижение нижнего предела погрешности измерения уровня подготовленности испытуемых. Добиться большей точности без применения адаптивных технологий тестирования невозможно.

Попробуем сформулировать технологию оценки нижнего предела погрешности. Данная задача относится к классу задач математического программирования, что позволяет рекомендовать следующую последовательность решения[11]:

1. Создание формы для ввода условий задачи. Рекомендуемый вид такой формы представлен на рис. 2 и 3.

2. Ввод исходных данных: уровней подготовленности испытуемых и уровней трудности заданий. По мнению автора, на этом этапе решения целесообразно принять равномерное распределение уровней в интервале [–5; 5].

3. Ввод зависимостей математической модели: формулы (14) для оценки стандартной ошибки измерения уровня подготовленности каждого испытуемого, а также зависимостей для промежуточных вычислений.

4. Назначение целевой функции в соответствии с условием (19).

5. Ввод ограничений и граничных условий. Следует ограничить пределы изменения уровней трудности заданий интервалом [–5; 5], а для политомических заданий указать, что уровень трудности каждого последующего шага должен быть больше уровня трудности шага предыдущего.

6. Поиск решения — поиск минимума нелинейной целевой функции при наличии ограничений численными методами.

11 *Гельман В.Я.* Решение математических задач средствами Excel. СПб.: Питер, 2003. 240 с.

Поиск нижнего предела погрешности измерения

Для этого в ячейку X26 листа с данными дихотомических заданий (см. рис. 2) введём формулу =МАКС(X:X24), которая находит максимальное значение стандартной ошибки измерения σ. Далее следует подобрать уровни трудности всех заданий таким образом, чтобы минимизировать значение X26. Встроенное в электронную таблицу Microsoft Excel средство *Поиск решения* легко справляется с подобными задачами. В окне *Поиска решения* указываем целевую ячейку X26, диапазон изменяемых ячеек C3:V3 (в этих ячейках хранятся значения уровней трудности всех заданий) и отмечаем направление поиска — *равной минимальному значению* (рис. 5). Кроме того, ограничим пределы изменения уровней трудности заданий интервалом [–5; 5]. Для увеличения точности решения можно с помощью кнопки *Параметры* увеличить принятое по умолчанию предельное число итераций и уменьшить относительную погрешность численного решения. Поиск решения запускается кнопкой *Выполнить*.

Поиск параметров политомических заданий проводился аналогично. Дополнительно было указано, что уровень трудности каждого последующего шага должен быть больше уровня трудности шага предыдущего.

Результаты поиска решения сведены в табл. 1. С помощью таких наборов заданий можно практически выровнять ошибку для всего интервала измерения. Стандартные ошибки измерения находятся в сравнительно узких интервалах значений (см. рис. 6):

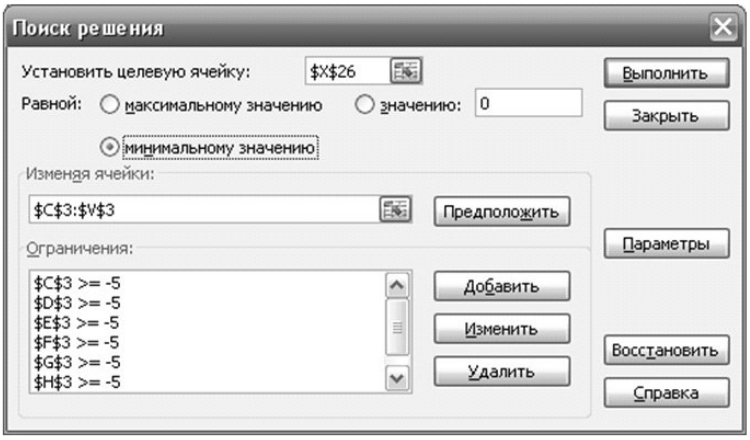

Рис. 5. Ввод параметров поиска решения

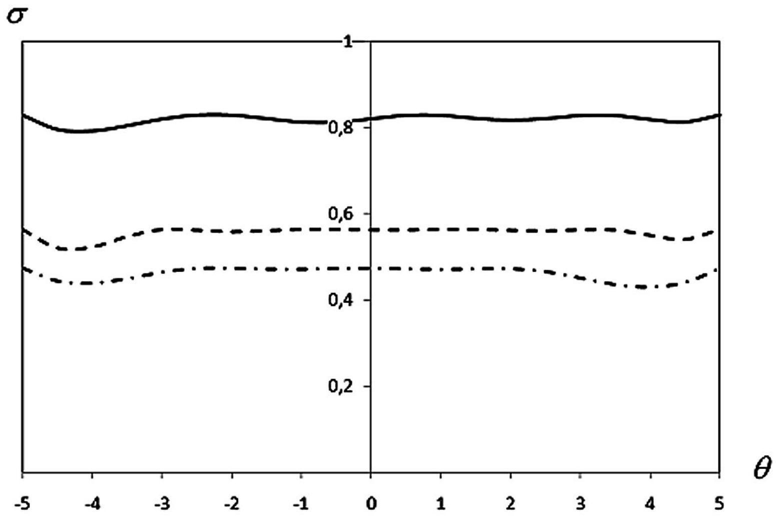

Рис. 6. Зависимость стандартной ошибки измерения от уровня подготовленности испытуемого:

———————— — тест из дихотомических заданий;

— — — — — — — — тест из политомических заданий, оцениваемых двумя баллами;

– • — • — • — — тест из политомических заданий, оцениваемых тремя баллами

- для теста из дихотомических заданий σ = 0,79...0,83;
- для теста из оцениваемых двумя баллами политомических заданий σ = 0,52...0,57;
- для теста из оцениваемых тремя баллами политомических заданий σ = 0,43...0,47.

Несложный анализ показывает, что способ оценивания результата выполнения тестового задания оказывает большое влияние на точность педагогического измерения. Переход от дихотомического оценивания к политомическому значительно уменьшает стандартную ошибку уровня подготовленности испытуемого:

- для теста из двухбалльных заданий ошибка снижается на 33%;
- для теста из трёхбалльных заданий ошибка снижается на 46%.

Это объясняется различием информационной ценности результатов выполнения заданий (рис. 7). Так, среднее значение количества информации теста из дихотомических заданий равно 1,49, а теста из оцениваемых двумя баллами политомических заданий — 3,19 (т.е. в 2,1 раза больше). Количество информации теста из оценивае-

Таблица 1

Уровни трудности заданий, обеспечивающие минимизацию ошибки измерения

№ задания	Уровни трудности заданий					
	дихотомические задания	политомические задания, оцениваемые двумя баллами		политомические задания, оцениваемые тремя баллами		
i	i	β_i 1	β_i 2	β_i 1	β_i 2	β_i 3
1	–4,73	–4,57	–4,36	–4,94	–3,37	4,56
2	–4,73	–4,57	–4,36	–4,88	–4,30	4,56
3	–4,72	–4,57	–4,36	–4,94	–3,38	4,56
4	–4,72	–4,57	–4,36	–4,94	–3,35	4,56
5	–4,72	–4,57	–4,36	–4,94	–3,37	4,56
6	–2,76	–2,31	–2,28	–4,87	–4,31	4,56
7	–2,76	–2,44	–1,71	–4,94	–3,39	4,56
8	–0,90	–2,16	–0,62	–4,94	–3,37	4,56
9	–0,90	–1,62	0,02	–4,76	–4,29	4,56
10	–0,44	–0,86	0,42	–1,43	1,11	3,98
11	–0,44	–0,38	1,12	–1,43	1,11	3,98
12	1,93	0,18	1,88	–1,43	1,11	3,98
13	1,93	0,87	2,53	–1,43	1,11	3,98
14	1,94	1,51	3,10	–1,43	1,11	3,98
15	2,40	2,05	3,34	–1,43	1,11	3,98
16	4,93	2,58	3,79	–1,43	1,11	3,98
17	4,93	4,62	5,14	–1,43	1,11	3,98
18	4,93	4,62	5,14	–1,43	1,11	3,98
19	4,93	4,62	5,14	–1,43	1,11	3,98
20	4,93	4,59	5,00	–1,43	1,11	3,98

мых тремя баллами политомических заданий ещё выше — 4,70, что превышает количество информации при дихотомическом оценивании в 3,2 раза.

Выводы

1. Разработанная технология позволяет оценить нижний предел стандартной ошибки уровня подготовленности испытуемого на основе анализа количества и способа оценивания тестовых заданий с использованием математической модели Раша и Partial Credit Model.

2. Эффективным средством минимизации погрешности педагогического измерения является переход от заданий с выбором одного правильного ответа к заданиям с выбором нескольких правильных ответов. Такой переход способен много-

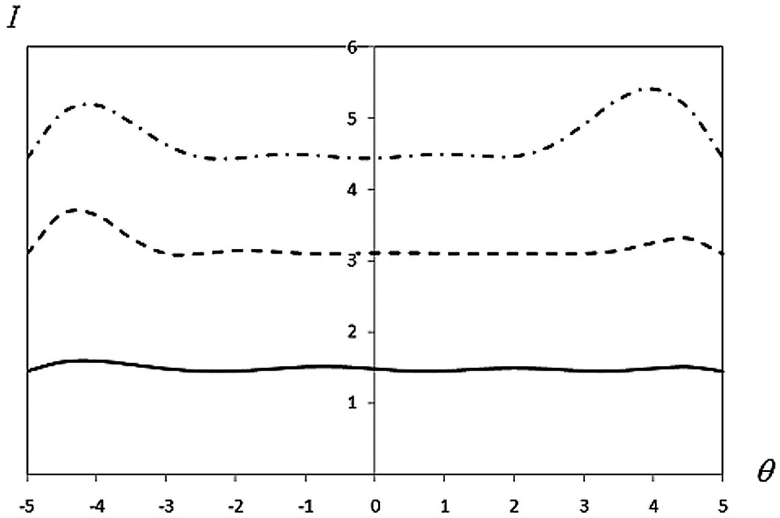

Рис. 7. Зависимость количества информации от уровня подготовленности испытуемого:
———————— — тест из дихотомических заданий;
— — — — — тест из политомических заданий, оцениваемых двумя баллами;
– • — • — — тест из политомических заданий, оцениваемых тремя баллами

кратно снизить вызываемую угадыванием систематическую ошибку и существенно (на 30% и более) снизить случайную ошибку.

3. Полученные результаты могут служить основой для оптимизации количества и параметров заданий с целью снижения ошибки педагогического измерения.