

## РАНЖИРОВАНИЕ ТРУДНОСТИ ТЕСТОВЫХ ЗАДАНИЙ С УЧЕТОМ УГАДЫВАНИЯ

*Елена Геннадьевна Лазарева, кандидат физико-математических наук, доцент,  
Тел: 8 913 847 3561, e-mail: lazareva@math.tsu.ru*

*Национальный исследовательский Томский государственный университет  
<http://math.tsu.ru/>*

*Ирина Георгиевна Устинова, кандидат технических наук, доцент,  
Тел. 8 905 089 9508, e-mail: igu@tpu.ru*

*Национальный исследовательский Томский политехнический университет  
<http://portal.tpu.ru>*

*Одним из факторов адекватности оценки, полученной в процессе тестирования, является возможность угадать правильные ответы. Мы предлагаем способ учесть влияние этого фактора на выборочную трудность тестовых заданий при «мягком» оценивании с помощью поправки «на угадывание». Наш метод позволяет ранжировать разнотипные тестовые задания по трудности с учетом угадывания.*

*Ключевые слова: результаты тестирования, дихотомическое оценивание, мягкое оценивание, трудность задания, исправленная трудность, поправка на угадывание.*

В настоящее время оценивание уровня подготовленности студентов в тестовой форме является объективной реальностью. Существует достаточное количество критериев эффективности обучающих математических тестов [1], но в данной статье мы бы



*Е.Г. Лазарева*

хотели исследовать вопрос о трудности задания теста и выяснить, как изменится трудность задания и результаты тестирования в целом с учетом поправки на угадывание правильного ответа. Трудность заданий представляет собой число, которое равно относительной частоте того, что задание выполнено не верно [2]. В классической теории тестирования трудность задания вычислялась как относительная частота правильных ответов на вопрос [3]. Очевидно, что чем ближе полученный результат к 1, тем большее число испытуемых справилось с этим заданием и, следовательно, тем легче это задание. Поэтому в дальнейшем показатель трудности задания стали вычислять как относительную частоту того, что задание решено не верно. Показатель трудности, по сути, указывает на то, что с рассматриваемым заданием тестируемые справились или не справились, но не дает представления, почему, и не учитывает, какой процент испытуемых просто угадали верный ответ. Соотношение верно/неверно выполнено задание является основным соотношением при нахождении вероятности правильного ответа, предложенным Г. Рашем [4]. По его мнению, следует опираться только на «наблюдаемые параметры», то есть на количество верных ответов. Однако как показывает практика, часто при тестировании встречается угадывание ответов. Это отражается на престиже тестового метода оценки знаний. Поэтому мы поставили задачу получить поправку на угадывание и рассмотреть, каким образом эта поправка повлияет на трудность задания. Известные методики вычисления «поправки на угадывание» [5, 6] имеют дело с однородными тестовыми заданиями, в которых лишь один вариант из  $m$  возможных верен, и вероятность угадать его равна



*И.Г. Устинова*

$\frac{1}{m}$ . В ситуации «мягкого» оценивания вероятность угадать ответ мы заменили математическим ожиданием оценки при случайном ответе. Полученная нами поправка на угадывание позволяет ранжировать задания по трудности угадывания, а также фактической трудности с учетом угадывания.

### **1. Условия проведения тестирования**

Тестирования по дисциплине «Математический анализ» проводились в 2013-2014 учебном году. Студентам первого курса Физического факультета Томского госуниверситета еженедельно в компьютерном классе предлагался один тест, состоящий из 7 или 8 заданий, с ограничением времени (25 минут) и автоматической проверкой. При тестировании мы использовали среду «Айрен» [7]. С содержанием заданий студенты могли познакомиться заранее (аналогичный тест выставлялся в свободном доступе в сети Интернет), но задания, предлагавшиеся в компьютерном классе, содержали другие параметры (числовые значения) или другие начальные условия (функции, множества и т.д.). Некоторые задания не изменялись (в основном это касалось теоретических вопросов). В результатах тестирования эти задания выделяются – их решали правильно большая часть студентов.

Тесты, предлагавшиеся студентам, отличаются разнообразием типов заданий и способов представления материала, что, по нашему мнению, способствует обучающей функции процесса тестирования [8]. «Мягкое оценивание» мы применяли для того, чтобы заметить появление верных ответов даже в случаях, когда студент не до конца понял тему. Эти факторы привели к большой разнородности тестов и вариативности оценок (оценки нельзя свести к схеме «верно-неверно»). За полностью правильный ответ на задание студент получает 100 баллов, за неполностью правильный – в соответствии с правилами «мягкого» оценивания. Результат тестирования студента – среднее арифметическое баллов за все задания теста. Правила мягкого оценивания в системе «Айрен» будут описаны ниже. Определяются также и общие результаты тестирования – средний балл за тест, трудность каждого задания. Эти результаты важны как для определения того, насколько испытуемым понятны темы, предлагавшиеся в тесте, так и для оптимизации тестов при их дальнейшем использовании.

Мы решили заранее, что студент, набравший более 50 баллов в результате тестирования, прошел тест успешно и можно считать, что он достаточно освоил текущий материал. Однако если студент угадывал правильные ответы и давал верные ответы на задания с меньшей трудностью, то гипотетически он мог получить 50 баллов или больше, не понимая материала. Наша цель – выяснить, насколько изменятся общие результаты тестирования, если ранжировать трудность заданий с учетом возможности угадывания ответов. Учитывая новизну задачи, нам пришлось ограничиться исследованием результатов одного теста.

### **2. Исправление трудности при дихотомическом оценивании**

В [5], с.156, поправка на угадывание вводится следующим образом: трудность задания  $U = 100 \left( 1 - \frac{k}{N} \right)$  ( $k$  – количество верно решивших задание,  $N$  – количество всех тестируемых) заменяется исправленной трудностью (в процентах)

$$U = 100 \left( 1 - \frac{1}{N} \left( k - \frac{N-k}{m-1} \right) \right) = 100 \left( 1 - \frac{k}{N} \right) \frac{m}{m-1}. \quad (1)$$

Предлагаем вероятностное обоснование такой замены.

Пусть  $m$  – количество всевозможных вариантов ответов на тестовое задание закрытого типа,  $k$  – число правильно ответивших на это задание,  $N$  – общее число отве-

чавших. Показатель  $p_n = \frac{k}{N}$  – *легкость*  $n$ -го задания,  $q_n = \left(1 - \frac{k}{N}\right)$  – *выборочная трудность*  $n$ -го задания. Для нахождения *теоретической трудности*  $Q_n$  решения  $n$ -го задания с учетом угадывания, найдем  $M[Q_n]$  в случае, когда ответ угадывается всеми испытуемыми. Для этого введем в рассмотрение индикатор события ( $i \in \{1, \dots, N\}$ )

$$x_i = \begin{cases} 1, & \text{если задание решено верно } i\text{-м студентом,} \\ 0, & \text{если задание решено неверно } i\text{-м студентом.} \end{cases}$$

Тогда при условии равновероятности выбора различных ответов и независимости ответов испытуемых имеем:  $P[x_i = 1] = \frac{1}{m}$ ,  $P[x_i = 0] = \frac{m-1}{m}$ ,  $M[x_i] = \frac{1}{m}$ ,

$$Q_n = 1 - \frac{1}{N} \sum_{i=1}^N x_i,$$

$$M[Q_n] = 1 - \frac{1}{N} \sum_{i=1}^N M[x_i] = 1 - M[x_i] = 1 - \frac{1}{m}.$$

Исходя из условия: чем ближе выборочная трудность к математическому ожиданию трудности задания  $M[Q_n]$ , тем ближе к угадыванию данные ответы, получаем исправленную трудность  $\tilde{q}_n = \left(1 - \frac{k}{N}\right) \cdot \frac{m}{m-1}$ , что соответствует (1). При этом  $\left(1 - \frac{k}{N}\right) \cdot \frac{m}{m-1} = 1$  – граничный случай, т.е. при  $\tilde{q}_n > 1$  испытуемые решили бы задачу лучше, если бы просто угадывали.

В табл. 1 указано математическое ожидание  $M[Q_n]$  для каждого задания теста при «жестком», то есть дихотомическом оценивании, а также поправочный коэффициент  $m/(m-1)$ , выборочная трудность  $q_n = \left(1 - \frac{k}{N}\right)$  и исправленная трудность  $\tilde{q}_n$  ( $M[Q_n] = 1$  в заданиях с открытым ответом). Задания расположены в порядке возрастания трудности  $q_n$ , указан номер задания в тесте.

Таблица 1

Расчёт исправленной трудности теста при дихотомическом оценивании

Названия заданий	1.Символы	2.Операции	5.Отображения	3.Множества-1	4.Множества-2	7.Графики	6.Биекции
$m$	24	24	6	–	64	–	64
$M[Q_n]$	0,9583	0,9583	0,8333	1	0,9844	1	0,9844
$m/(m-1)$	1,0435	1,0435	1,2	–	1,0159	–	1,0159
$q_n$	0,1493	0,1642	0,6418	0,6567	0,7313	0,7313	0,8656
$\tilde{q}_n$	0,1558	0,1713	0,7702	0,6567	0,7429	0,7313	0,8793

Из табл. 1 видно, что коэффициент  $m/(m-1)$  при дихотомическом оценивании очень близок к 1 и мало влияет на трудность задания. Только в задании «Отображения» трудность значительно выросла.

С другой стороны, по данным, полученным нами в результате тестирований, получаем: при «мягком» оценивании первый тест был зачтен 28 студентам из 68, при

жестком – 21, т.е. 25% успешно ответивших не получили бы зачет. Для второго теста этот показатель составляет 23%, для третьего теста – 10%.

Поэтому мы сочли неправильным исследовать результаты «жесткого» оценивания – они слишком отличаются от реальных результатов и дают мало возможностей для учета угадывания.

### **3. «Мягкое» оценивание. Нахождение математического ожидания баллов за ответы к тестовым заданиям при случайных ответах**

Идея наших рассуждений повторяет рассуждения при «жестком» оценивании. Индикатор события мы заменяем рядом распределений дискретной случайной величины  $x_i$ , которая зависит от типа тестового задания. Затем нужно найти для каждого задания  $M[x_i]$ , используя теоретическую трудность  $Q_n = 1 - \frac{1}{N} \sum_{i=1}^N x_i$ , получить

$$M[Q_n] = 1 - \frac{1}{N} \sum_{i=1}^N M[x_i] = 1 - M[x_i], \text{ и, наконец, вычислить } \tilde{q}_n = \left(1 - \frac{k}{N}\right) \cdot \frac{1}{M[Q_n]}.$$

При этом в случае «мягкого» оценивания и негомогенного теста мы должны учесть различие заданий и все правила «мягкого» оценивания.

#### **Правила «мягкого» оценивания**

Мягкое оценивание применяется в «Айрен» в заданиях четырех типов: множественный выбор (когда среди предложенных есть несколько правильных ответов), соответствие, классификация, упорядочение. В исследуемом тесте заданий на упорядочение нет.

Правило мягкого оценивания в задании «множественный выбор»: если правильных вариантов ответов в задании –  $k$  штук, то за выбор каждого правильного ответа дается

$(1/k) \cdot 100$  баллов. Если выбран хотя бы один неправильный ответ, то студент получает 0 баллов за задание.

Правило мягкого оценивания в задании на соответствие: если нужно поставить в соответствие  $k$  данным условиям некоторые заключения, то за каждое правильное заключение дается  $(1/k) \cdot 100$  баллов, а за каждое неправильное отнимается  $(1/k) \cdot 100$  баллов. Если полученный в результате балл равен 0 или отрицательный, то студент получает 0 баллов за задание.

Правило мягкого оценивания в задании на классификацию: если нужно расставить  $k$  объектов по нескольким группам, то за каждый объект, отнесенный в правильную группу, дается  $(1/k) \cdot 100$  баллов, а за каждый, отнесенный в неправильную группу, отнимается

$(1/k) \cdot 100$  баллов. Если полученный в результате балл равен 0 или отрицательный, то студент получает 0 баллов за задание.

#### **Построение рядов распределений для каждого задания теста**

Ниже описано построение рядов распределений случайных величин  $x_i$ ,  $i \in \{1, \dots, N\}$ , для заданий рассматриваемого теста.

#### **Задание «Биекции»**

Студентам случайным образом предлагался один из двух вопросов типа «множественный выбор». В первом вопросе 6 вариантов ответов, 7-й вариант – «Среди предложенных ответов нет верного», правильных ответов – 3. Во втором вопросе все аналогично, но 4 правильных ответа.

Значения случайной величины  $x_i$  – баллы, которые мог получить студент за этот вопрос. Вероятности баллов рассчитывались, исходя из равновероятности всех воз-

можных ответов. Результаты расчетов приведены в табл. 2.

Таблица 2

Ряд распределений  $x_i$  для первого задания

$x_i$	100	75	66,(6)	50	33,(3)	25	0
$P_i$	$0.5 \cdot 20/64 + 1/20 + 0.5 \cdot 15/64 + 1/15$	$0.5 \cdot 20/64 + 4 \cdot 1/20$ (второй вопрос)	$0.5 \cdot 15/64 + 1/5$ (первый вопрос)	$0.5 \cdot 15/64 + 2/5$ (второй вопрос)	$0.5 \cdot 7/64 + 3/7$ (первый вопрос)	$0.5 \cdot 7/64 + 4/7$ (второй вопрос)	$\frac{53}{64}$

Здесь

$$P\{x_i = 0\} = 0.5 \cdot (7/64 \cdot (2/7 + 6/7 \cdot 0/5 + 6/7 \cdot 1/3)) + 15/64 \cdot (12/15 + 9/15) + 20/64 \cdot (19/20 + 16/20) + 15/64(1 + 14/15) + 6/64(1 + 6) + 2/64 = 53/64.$$

Математическое ожидание  $M[x_i] = 300/32 = 9,375$ . Заметим, что при «жестком» оценивании в этом задании  $M[x_i] = (1/64) \cdot 100 = 1,5625$ .

#### Задания «Операции», «Символы»

Эти задания устроены одинаково: даны три названия и 4 обозначения. Нужно подобрать правильное обозначение к каждому названию. Тип вопроса – «на соответствие». Вероятности баллов (см. табл.3) рассчитывались, исходя из равновероятности всех возможных ответов. Неполные ответы не считались возможными (исходя из практики, – студенты не дают неполных ответов в задании на соответствие).

Таблица 3

Ряд распределений  $x_i$  для заданий «Операции», «Символы»

$x_i$	100	33,(3)	0
$P_i$	1/24	3/24	5/6

$$M[x_i] = 8,(3) \text{ (4,167 при «жестком» оценивании)}$$

#### Задание «Отображение»

Это задание предполагает выбор одного варианта из шести, поэтому при равновероятных ответах  $M[x_i] = (1/6) \cdot 100 = 16,(6)$  (не отличается от «жесткого» оценивания).

#### Задание «Множества 2»

Студентам случайным образом предлагался один из четырехвопросов типа «множественный выбор». В трех вопросах 6 вариантов ответов, 7-й вариант – «Среди предложенных ответов нет верного», правильных ответов – 2. В четвертом вопросе все аналогично, но один правильный ответ, а именно, «Среди предложенных ответов нет верного». Ряд распределений  $x_i$  приведен в табл. 4.

Таблица 4

Ряд распределений  $x_i$  для задания «Множества 2»

$x_i$	100	50	0
$P_i$	$3/4 \cdot 1/64 + 1/4 \cdot 1/64$	$3/4 \cdot 7/64 + 2/7$ (только 3 вопроса из 4)	$63/64 \cdot 1/4 + 61/64 \cdot 3/4$

$$M[x_i] = 175/64 = 2,734375. \text{ (} M[x_i] = (1/64) \cdot 100 = 1,5625 \text{ при жестком оценивании)}$$

#### 4. Расчет поправок на угадывание при «мягком» оценивании

Для расчета поправок перейдем от математического ожидания в баллах математическому ожиданию в долях, что соответствует поправке на угадывание при дихотомическом оценивании. Пусть теперь  $q_n$  – выборочная трудность  $n$ -го задания при

«мягком» оценивании, то есть  $q_n = (1 - a/100)$ , где  $a$  – средний балл испытуемых в процентах за задание,  $\tilde{q}_n = q_n \cdot \frac{1}{M[Q_n]}$  – исправленная трудность  $n$ -го задания. Порядок следования заданий в табл.5 соответствует порядку в табл.1.

Таблица 5

Расчет исправленной трудности теста при «мягком» оценивании

Названия заданий	1.Символы	2.Операции	5.Отображения	3.Множества-1	4. Множества-2	7.Графики	6.Биекции
$M[q_n]$	0,91666	0,91666	0,8333	1	0,97265625	1	0,90625
$q_n$	0,119403	0,139304	0,6418	0,6567	0,686567	0,7313	0,74005
$\tilde{q}_n$	0,130258	0,151969	0,7702	0,6567	0,705868	0,7313	0,816607

### 5. Выводы и рекомендации

Цель нашего исследования состояла в получении исправленной трудности заданий при дихотомическом оценивании и при «мягком» оценивании, а также в сравнении показателей трудности заданий в зависимости от типа оценивания и от введения поправки на угадывание.

Результатом исследования является упорядочение тестовых заданий по трудности. Такое упорядочение используется для составления тестов (от простых заданий – к трудным) [6], а также для коррекции заданий с целью нормировки их трудности.

В исследованном нами тесте задания были расположены в следующем порядке: 1) Символы, 2) Операции, 3) Множества-1, 4) Множества-2, 5) Отображения, 6) Биекции, 7) Графики.

При дихотомическом оценивании без учета угадывания ранжирование по выборочной трудности задания дало следующую перестановку заданий теста: (1, 2, 5, 3, 4, 7, 6). Однако с учетом угадывания эта перестановка имеет вид: (1, 2, 3, 7, 4, 5, 6). Как видно, только 7-е задание оказалось не на «своем» месте. При мягком оценивании без учета угадывания ранжирование по выборочной трудности задания дало такую же перестановку заданий теста, что и при жестком оценивании. С учетом угадывания: (1, 2, 3, 4, 7, 5, 6).

Мы считаем, что «мягкое» оценивание в тесте с разнотипными заданиями точнее показывает содержательную трудность заданий. Так, при дихотомическом оценивании задания 4 и 7 получили одинаковую выборочную трудность 0,7313, но при «мягком» трудность задания 4 оказалась ниже (см. табл.1, табл.5), так как в этом задании несколько правильных вариантов ответа, и у студентов была возможность получить «часть» от 100 баллов за задание.

Что касается поправки на угадывание, ее применение обосновано при обоих типах оценивания, но при «мягком» оценивании она дает больший эффект. Например, мы видим, что относительно «нетрудное» как при дихотомическом, так и при «мягком» оценивании 5-е задание с учетом угадывания становится значительно труднее. Действительно, студентам нужно выбрать один ответ из шести в этом задании, но они ошибаются не намного реже, чем в заданиях, где есть несколько правильных ответов (задание 4): 0,6418 и 0,686567 в табл.5. Это означает, что испытуемые не понимают или не знают чего-то, что от них требуется, т.е. 5-е задание труднее других, сходных по выборочной трудности. Обратим внимание на 6-е задание, которое при введении поправки на угадывание и «мягком» оценивании получило значительное увеличение трудности: от 0,74005 до 0,816607. Это отдалило 6-е задание от 7-го, хотя выборочная трудность обоих заданий при «мягком» оценивании без учета угадывания близка: 0,7313 и 0,74005.

Особо стоит обратить внимание на 7-е задания, которое всегда оказывалось не на «своем» месте. По-видимому, это говорит о недостаточной содержательной трудности последнего задания теста. Переставлять это задание не имеет смысла, т.к. по логике со-

держания теста оно является обобщающим и заключительным. Следовательно, стоит обдумать, как повысить его содержательную трудность, не утратив при этом обучающей направленности теста.

### Заключение

Мы считаем, что предложенный нами метод ранжирования трудности тестовых заданий с учетом угадывания может применяться к разнородным тестам с целью улучшения их структуры по возрастанию трудности заданий. При этом «мягкое» оценивание точнее показывает содержательную трудность заданий, чем «жесткое», поэтому нужно ориентироваться именно на результаты мягкого оценивания. Мы считаем полезным производить статистическую обработку результатов тестирований и вычисление поправки на угадывание в заданиях различных типов, т.к. эта информация дополняет результаты, полученные при тестировании, и позволяет анализировать и улучшать содержание тестов.

### Литература

1. *Ustinova I., Lazareva E.* Performance criteria of learning math tests // Proceedings of International Conference on Interactive Collaborative Learning (ICL), 2015. P. 686-689. 978-1-4799-8706-1.15.;31.00 ©2015 URL: [http://www.weef2015.eu/Proceedings\\_WEEF2015/start.html](http://www.weef2015.eu/Proceedings_WEEF2015/start.html)(дата обращения 28.02.2016)ю
2. Bortz&Döring. Forschungsmethoden und Evaluation. – Heidelberg: Springer-Verlag. 2005.
3. *Аванесов В.С.* Трудность теста и тестовых заданий // Теория и методика педагогических измерений. URL: <http://testolog.narod.ru/Theory7.html> (дата обращения: 28.02.2016).
4. *Аванесов В.С.* Применение тестовых форм в Rasch Measurement // Педагогические измерения. URL: <http://testolog.narod.ru/Theory49.html> (дата обращения: 28.02.2016).
5. *Майоров А.Н.* Теория и практика создания тестов для системы образования. – М: «Интеллект-центр», 2002. 296 с.
6. *Ким В.С.* Тестирование учебных достижений. Изд-во: Уссурийский гос. пед. ин-т, 2007. 215 с. URL: <http://uss.dvfu.ru/struct/ui/kim/monograph/index.html> (дата обращения: 28.02.2016).
7. Программа тестирования знаний «Айрен». URL: <http://irenproject.ru/> (дата обращения: 28.02.2016).
8. *Лазарева Е.Г., Устинова И.Г.* Обучающие приемы компьютерных тестов по высшей математике. В сборнике: Некоторые актуальные проблемы современной математики и математического образования. Герценовские чтения – 2015: материалы 68-ой научной конференции. – 2015. С. 157-162.

### Ranking of difficulty of test items considering guessing

*Elena Gennadyevna Lazareva, PhD, Associate Professor*

*Irina Georgievna Ustinova, PhD, Associate Professor*

*One of the factors of the test assessment adequacy is the ability to guess the correct answers. We offer a way to consider this factor's impact on the sample difficulty of the test tasks during "soft" estimating using amendment on the "guessing". Our method allows us to rank different types of tests by difficulty considering guessing.*

*Keywords – results of tests, a dichotomous evaluating, soft evaluating, difficulty of a task, a corrected difficulty of the task, correction for guessing.*