

ОЦЕНКА ДИФФЕРЕНЦИРУЮЩЕЙ СПОСОБНОСТИ ОДНОГО КЛАССА ТЕСТОВЫХ ЗАДАНИЙ С ПРЕДОПРЕДЕЛЕННЫМИ НАБОРАМИ ОТВЕТОВ

Ю.М. Бруштейн, С.В. Окладникова

Проанализированы типы тестовых заданий, используемые в тестовых системах. Показано, что среди них абсолютно преобладают задания с текстовой формой представления содержания вопросов и ответов. Подробно рассмотрены характеристики класса заданий с predetermined наборами ответов, из которых нужно выбрать все верные (количество верных ответов в задании может либо указываться, либо нет). Для этого класса даны оценки априорных вероятностей успешного выполнения заданий при различных уровнях знаний тестируемых. Отдельно рассмотрены вопросы, связанные с эффективностью расхода времени на тестирование.

В настоящее время тестовые материалы находят широкое применение не только при оценке знаний студентов, но и при профессиональной аттестации специалистов, конкурсном приеме на работу и пр. При этом в ряде случаев время на тестирование серьезно ограничивается, что ужесточает требования к дифференцирующей способности (ДС) отдельных тестовых заданий (ТЗ) и теста в целом. Наверное, наиболее ярким примером таких ограничений являются мини-тесты на компетентность, которые используются на некоторых сайтах для допуска к участию в голосовании, обсуждения на форумах и пр.

В общем случае проблема управления качеством тестов включает ряд направлений, в том числе оценку ДС различных типов ТЗ в увязке с расходом времени на их выполнение; выбор между использованием типов ТЗ для теста при ограниченном времени на тест в целом; управление сложностью восприятия тестов и пр. Ниже мы ограничимся только первым направлением и для определенности рассмотрим лишь один класс ТЗ.

Анализ массива тестовых материалов, накопленных в Астраханском государственном университете (АГУ) в период за 2004–2007 гг., показал, что ТЗ с predetermined наборами ответов (ПНО) абсолютно преобладают (не менее 97 %). Среди них большинство (около 90 %) составляют ТЗ, связанные с выбором заданного количества верных ответов (чаще всего – одного) из списка, одновременно предъявляемого тестируемому.

Для ТЗ этого класса в 84 % случаев используются чисто текстовые формулировки вопроса и ответов (рис. 1).

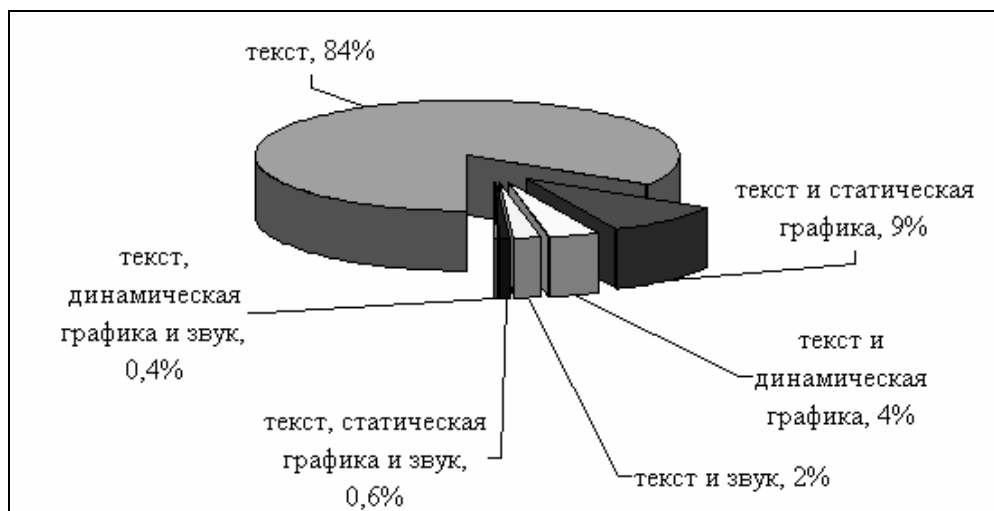


Рис. 1. Соотношение видов наполнения ТЗ в общем массиве тестовых материалов, накопленных в АГУ за 2004–2007 гг.

Ситуация преобладания ТЗ с ПНО характерна не только для АГУ, но и в целом по России. Это связано с рядом факторов: низкая трудоемкость разработки таких ТЗ; достаточность для разработчиков владения только общеизвестными текстовыми редакторами; нежесткие требования к аппаратным средствам как на стадии разработки ТЗ, так и при их использовании; отсутствие необходимости в применении специального программного обеспечения; более простой контроль соответствия ТЗ формальным требованиям тестирующей системы и пр. Приведенные факты показывают актуальность темы статьи в отношении именно ТЗ с ПНО в текстовой форме.

В отношении оценки ДС отдельных ТЗ и их совокупностей ключевое значение имеет вероятность случайного выбора тестируемыми верных ответов (одного или большего количества). Эти вопросы рассматривались в ряде работ. В частности, в диссертационном исследовании И.П. Карповой¹ принималось, что выбор тестируемыми ответов из ПНО является полностью случайным. Такой подход достаточно типичен, но реальная ситуация обычно иная – какие-то знания у тестируемых есть. Поэтому выбор ответов из ПНО не является полностью случайным. Ниже мы попытаемся рассмотреть также и этот случай.

Для простоты будем считать, что результат прохождения одного ТЗ может быть оценен в бинарной форме – как «правильно» или «неправильно» (без возможностей промежуточных оценок). При этом в общем случае для ТЗ возможны «ошибки диагностики уровня знаний». Ошибка первого рода соответствует случаю, когда на ТЗ дается верный ответ при отсутствии знаний или их уровне ниже некоторого критического («плохие» знания). Ошибка второго рода соответствует варианту, когда уровень знаний выше критического («хорошие» знания), но тестируемый дал неверный ответ (например, из-за сложности восприятия вопроса или ПНО ТЗ, по некоторым случайным причинам и пр.).

Наиболее распространенный (первый) вариант ТЗ, когда в формулировке вопроса оговаривается, что один и только один ответ правильный (или неправильный). Тестируемый должен его выбрать – путем или ввода номера ответа или (чаще) щелчком мышью в квадрате около ответа. При этом вероятность полностью случайного выбора единственного верного ответа ($P_{e.o.}$)²

$$P_{e.o.}(\%) = 100\% / V \quad (1)$$

где V – количество возможных вариантов выбора в ПНО.

В данном случае

$$V = N, \quad (2)$$

где N – количество ответов в ПНО.

На первый взгляд кажется, что для уменьшения вероятности случайного угадывания было бы разумным увеличивать общее количество ответов в ПНО ТЗ. Однако такое решение имеет ряд минусов. Во-первых, при большом количестве ответов в ПНО они могут все сразу просто не поместиться на экране мониторе, что усложнит работу тестируемого (особенно в условиях дефицита времени). Во-вторых, придумать значительное количество правдоподобных ответов для некоторых вопросов трудно, а наращивание их количества за счет малоправдоподобных – не будет вести к увеличению ДС ТЗ. В-третьих, сложность восприятия содержания ТЗ в зависимости от количества ответов обычно растет быстрее, чем по линейному закону (особенно, если какие-то ответы близки по своим формулировкам).

Рассмотрим в качестве альтернативы одному ТЗ с большим количеством ответов (например, 10) два ТЗ с количеством ответов, меньшим в два раза (в данном случае – с 5). Вероятность случайного угадывания студентом правильного ответа в первом случае составляет 10 %, а для двух ТЗ с 5 ответами в каждом – 4 % (по теореме умножения вероятностей совместных событий). Таким образом, вторая альтернатива в отношении ДС оказывается явно предпочтительней.

Теперь сравним эти две альтернативы по расходу времени. Пусть для определенности формулировки всех ответов одинаковы по длине, а формулировки вопросов втрое превосходят длины каждого из ответов. Тогда расход времени при ответе на ТЗ с 10 ответами можно оценить в 13 условных единиц (3 – на вопрос и 10 – на ответ), а при ответе на два ТЗ с 5 ответами – в 16 единиц ($16 = 2 \times (3 + 5)$). Следовательно, расход времени возрастает незначительно. Кроме того, когда ответов много, тестируемый может не воспринять их все сразу и

вынужден будет ответы перечитывать, тратя на это дополнительное время. В результате расход времени для первой альтернативы может быть даже больше, чем для второй.

Примем теперь, что выбор ответа тестируемым является не полностью случайным в силу наличия у него некоторых знаний. Тогда в простейшем случае априорную вероятность верного выбора оценим как

$$P_{\text{в.о.}}(\%) = U + (100 - U)/V \quad (3)$$

где U – процент знаний тестируемым материала от общего объема тестируемого курса. При $U = 0$ получаем формулу (1), а при $U = 100\%$ – вероятность выбора верного ответа из ПНО равна 100 %. График для формулы (3) – соответствует жирной линии на рис. 2.

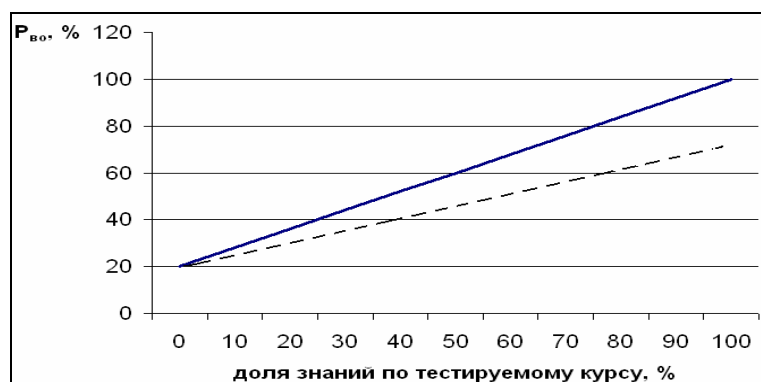


Рис. 2. Вероятность выбора правильного ответа при частичных знаниях (при $N = 5$)

При этом 70 %-ому положительному результату тестирования по серии ТЗ (это обычно нижний порог для «удовлетворительно») при 5-ти ответах на каждое из ТЗ соответствует знаниям на уровне примерно 63 %. С учетом того, что правдоподобность отдельных ответов в ПНО обычно все же разная, оценку «удовлетворительно» по тесту достаточно часто можно получить и при менее чем 50 % знаний по предмету тестирования.

В общем случае формулу (3) можно модифицировать в нескольких направлениях.

(А) Задав нелинейную зависимость выбора верного ответа от процента знаний материала тестируемым, например, в параболическую

$$P_{\text{в.о.}}(\%) = a_0 + a_1 U + a_2 U^2. \quad (4)$$

Как и для формулы (3), будем исходить из того, что

$$P_{\text{в.о.}}(U = 0) = 100/V; \quad P_{\text{в.о.}}(U = 100) = 100 \quad (5)$$

Тогда a_0 – фиксировано, а величина a_2 определяется по a_1

$$a_0 = 100/V; \quad a_2 = (100 - 100/V - a_1)/100 \quad (6)$$

Таким образом, в зависимости от a_1 парабола, соответствующая (5), может быть и выпуклой, и вогнутой – с минимумом на левом конце и максимумом на правом. Оценка величины a_1 в конкретном случае может быть дана путем сравнения результатов компьютерного тестирования достаточно большой группы испытуемых (на серии ТЗ одинаковой сложности) и независимой оценки процента знаний тех же студентов – например, преподавателями. Первый из этих параметров будет соответствовать вертикальной оси графика на рис. 2, а второй – горизонтальной. При этом подбор параметра a_1 для (4) с учетом (6) придется производить по методу наименьших квадратов.

(Б) Другое направление модификации (3) – указание некоторой величины случайного выбора неверных ответов при 100 % знаниях тестируемого по курсу. Возможные причины: случайные ошибки из-за спешки; непонимание ТЗ из-за сложности или неудачности формулировок вопроса и/или ответов в ПНО. Приняв в варианте Б линейную зависимость $P_{\text{в.о.}}$ от уровня знаний, получим:

$$P_{\text{в.о.}}(\%) = b_0 + b_1 U; \quad b_0 = P_{\text{в.о.}}(U = 0) = 100/V \quad (7)$$

$$b_1 = \frac{(100 - P_{\text{с.н.}}) - (100/V)}{100} \quad (8)$$

где $P_{с.н.}$ – это вероятность (в %) случайного неверного ответа. На рис. 2 для наглядности вариант для (7), (8) показан пунктиром при $P_{с.н.} = 30\%$. На практике для $P_{с.н.}$ более характерны меньшие значения (обычно порядка нескольких процентов). Ненулевое $P_{с.н.}$, конечно, снижает ДС тестов, так как фактический интервал для серии ТЗ, соответствующий неслучайному характеру выбора испытуемыми сужается до $100\% - P_{в.о.} - P_{с.н.}$. При $P_{с.н.} = 0$ (идеальный вариант) формула (7) переходит в (3).

(В) Возможны, конечно, и комбинации вариантов А и Б.

Для второго типа ТЗ необходимо выбрать комбинацию в виде k ответов ($1 < k < N$) из числа N предъявленных. При этом какого-либо ранжирования ответов (по степени важности, соответствия вопросу и т.д.) не предполагается. В этом случае формула (1) для оценки вероятности полностью случайного выбора (при отсутствии знаний у тестируемого) сохраняется, но в ней вместо (2) будет

$$V = \frac{N!}{k!(N-k)!} \quad (9)$$

Результаты расчетов по формуле (9) представлены в таблице 1 (в ее первом столбце для сравнения даны результаты для формулы (1)).

Таблица 1

**Вероятности случайного выбора правильной комбинации ответов
для одного ТЗ типа 2 при различных сочетаниях количеств
«ответов всего» (N) – «верных ответов» (k), %**

k	1	2	3	4	5	6	7	8	9	10	11
N											
2	50,00										
3	33,33	33,33									
4	25,00	16,67	25,00								
5	20,00	10,00	10,00	20,00							
6	16,67	6,67	5,00	6,67	16,67						
7	14,29	4,76	2,86	2,86	4,76	14,29					
8	12,50	3,57	1,79	1,43	1,79	3,57	12,50				
9	11,11	2,78	1,19	0,79	0,79	1,19	2,78	11,11			
10	10,00	2,22	0,83	0,48	0,40	0,48	0,83	2,22	10,00		
11	9,09	1,82	0,61	0,30	0,22	0,22	0,30	0,61	1,82	9,09	
12	8,33	1,52	0,45	0,20	0,13	0,11	0,13	0,20	0,45	1,52	8,33

При наличии у тестируемого знаний в объеме $U\%$ от общего объема материала могут также использоваться формулы (3); (4), (5), (6) и (7), (8) – с той разницей, что в них V будет определяться по (9), а не по (2).

Для второго типа ТЗ время на ознакомление с ним (вопрос + ответы) лишь немного выше по сравнению с первым вариантом ТЗ (у которого единственный верный ответ). Эта добавка будет связана с тем, что в рамках тестовой системы (например, АИСТ) студенту придется пометать не один ответ, а несколько. Однако (см. табл. 2), вероятность случайного выбора верной комбинации ответов для этого типа ТЗ существенно ниже.

Третий практически важный тип ТЗ – количество верных ответов (j) в формулировке вопроса ТЗ не указывается, и тестируемый должен его определить сам. В общем случае $0 \leq j \leq N$ (и далее мы рассматриваем именно этот случай), но на практике обычно $1 \leq j \leq N-1$. Для оценки вероятности полностью случайного выбора правильной комбинации ответов (при отсутствии знаний у тестируемого) может быть также использована формула (1), но с

$$V = \sum_{j=0}^N \left(\frac{N!}{j!(N-j)!} \right) \quad (10)$$

Заметим, что $\sum_{j=0}^N \left(\frac{N!}{j!(N-j)!} \right) = 2^N$, но формула (10) по нашему мнению более ясно отражает

идеологию перебора всех вариантов.

Таблица 2

**Зависимость вероятности случайного правильного выбора
комбинации верных ответов от числа ответов (N),
когда число верных ответов не указывается в вопросе, %**

N	2	3	4	5	6	7	8	9	10	11	12
P _{в.о.}	25,00	12,50	06,25	3,13	1,56	0,78	0,39	0,20	0,10	0,05	0,02

Для третьего типа ТЗ также возможны аналоги формул (3); (4), (5), (6) и (7), (8) с вычислением V по (10). При этом если у тестируемых какие-то знания по курсу есть, то вероятности правильных комбинаций ответов будут значительно выше, чем в таблице 2. Если принять зависимость $P_{в.о.} = f(U)$ по формуле (3), то, начиная уже со значений U в несколько десятков процентов, величина $P_{в.о.}$ будет определяться в основном U, а не N.

Для третьего типа ТЗ время на ознакомление с его содержанием и выбор комбинации ответов формально те же, что и для второго типа.

С целью оценки преимуществ 3-го типа ТЗ в отношении ДС по сравнению со 2-м приведем таблицу отношений вероятностей случайного выбора комбинаций верных ответов при $N = 10$, $k = 2..9$, $U = 0..100\%$; $P_{с.н.} = 0$ (последнее значение соответствует идеальному варианту). В обоих случаях нами была использована формула (3).

Таблица 3

**Отношение вероятностей выбора комбинаций верных ответов тестируемым
для ТЗ типов 3 и 2 для параметров, указанных в тексте статьи, %**

U k->	2	3	4	5	6	7	8	9
0	4,39	11,72	20,51	24,61	20,51	11,72	4,39	0,98
10	84,07	93,84	96,73	97,40	96,73	93,84	84,07	53,09
25	94,02	97,85	98,88	99,11	98,88	97,85	94,02	77,15
50	97,92	99,27	99,62	99,70	99,62	99,27	97,92	91,00
75	99,30	99,76	99,87	99,90	99,87	99,76	99,30	96,80
90	99,76	99,92	99,96	99,97	99,96	99,92	99,76	98,91
100	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00

Таким образом, если принять модель зависимости выбора верной комбинации ответов по формуле (3), то 3-й тип ТЗ по отношению ко 2-му имеет существенные преимущества лишь при низких процентах знаний материала тестируемыми лицами. Отметим, что при росте $P_{с.н.}$ значения отношений в таблице 3 будут изменяться в направлении **100**.

Понятно, что если вместо модели выбора по формуле (3) будет принята иная зависимость, то результаты в таблице 3 изменятся.

Итак, при одном и том же времени на ознакомление с ТЗ (вопрос и НПО) 2-й и 3-й типы ТЗ позволяют уменьшить вероятности случайного выбора комбинаций верных ответов тестируемых, однако это уменьшение значительно лишь для низких уровней знаний – если принять модель выбора по формуле (3).

Поэтому использование ТЗ 2-го и 3-го типов (вместо 1-ого) позволяет снизить общее количество ТЗ в тесте (а следовательно, и время, нужное на его прохождение) либо повысить ДС теста (если количество ТЗ нельзя уменьшить ниже какого-то предела из-за необходимости достаточно полного охвата материала курса).

¹ Карпова И.П. Исследование и разработка подсистемы контроля знаний в распределенных автоматизированных обучающих системах: Дис. ... канд. техн. наук. М., 2002.; <http://diss.rsl.ru>.

² Глас Дж., Стенли Дж. Статистические методы в педагогике и психологии. М.: Прогресс, 1976. 495 с.