

СРАВНЕНИЕ ОТКРЫТЫХ И ВЫБОРОЧНЫХ ТЕСТОВ

И.П.Карпова, к.т.н, доцент кафедры «Вычислительные системы и сети»

Тел.: (495) 916-8909, E-mail: karpova_ip@mail.ru

*Московский государственный институт электроники и математики
miem.edu.ru*

Статья посвящена сравнению выборочных тестов и тестов со свободно-конструируемой формой ответа. Описаны некоторые методы по оценке свободно-конструируемых ответов. Приведены результаты экспериментов по тестированию, проведён анализ этих результатов. Показано, что при прочих равных условиях выборочный тест даёт более высокую оценку, чем открытый тест.

Ключевые слова: контроль знаний, тестирование, свободно-конструируемые ответы, выборочные ответы.

1. Введение

В настоящее время довольно активно ведется дискуссия о целесообразности применения тестов для контроля знаний обучаемых. Существуют как сторонники этой формы, так и её ярые противники. Но даже сторонники широкого применения тестирования не могут не согласиться, например, с тем, что после внедрения ЕГЭ учителя не столько учат своих учеников, сколько "натаскивают" их на сдачу ЕГЭ. Получается, что знание предмета и умение сдать по этому предмету тест – не совсем одно и то же. Но обучаемому-то в конце концов нужны именно знания, а не умение правильно расставить галочки в тесте.

Из этого не следует вывод о том, что от тестов надо совсем отказаться. Отрицательное отношение к тестам часто обусловлено наличием неудачных тестов и примитивных автоматизированных систем тестирования. Но многолетняя практика применения тестов показала, что они обладают целым рядом достоинств, таких как объективность, массовость, технологичность и др.

Тест позволяет качественно оценить структуру и эффективно измерить уровень знаний, умений, навыков и представлений [1]. Уровень знаний выявляется при анализе ответов каждого ученика на все задания теста. Естественно, чем больше правильных ответов, тем выше тестовый балл испытуемых. Общая оценка уровня знаний тестируемого определяется по оценкам отдельных ответов. Таким образом, оценка каждого ответа тестируемого является существенной для оценки теста в целом.

В связи с этим немаловажным представляется ответ на вопрос: насколько оценка зависит от формы ответа? Представляется очевидным, что ответ открытой формы является для тестируемого более сложным, чем ответ закрытой формы, поэтому и оценка должна быть ниже. Но я не собиралась ставить опыты для подтверждения этого факта до тех пор, пока мне на глаза не попала статья [2], в которой косвенно затрагивается этот вопрос.

Авторы этой статьи провели эксперименты для подтверждения гипотезы о том, что мономорфные и полиморфные итоговые тесты равноценны, т.е. обеспечивают одинаковую величину первичных баллов респондентов. (Под мономорфными авторы подразумевают тесты, составленные из закрытых заданий множественного выбора с одним правильным вариантом; полиморфные тесты включают как закрытые, так и открытые задания.) Результаты экспериментов, приведённые в статье, эту гипотезу подтвердили, но, к сожалению, авторы делают из этого следующий вывод: полиморфные тесты (включающие вопросы открытой формы) просто не нужны. Зачем, мол, утруждать составителя тестов, если закрытые тесты дают тот же результат?.. И в соответствии с упомянутой статьёй получается

следующее: экспериментально подтверждено, что открытые задания не изменяют характеристик теста. Поэтому не следует их использовать вообще.

К сожалению, статья [2] отражает достаточно распространённое мнение о том, что при тестировании вполне можно обойтись без ответов открытой формы. Распространённость этого мнения подтверждают и многочисленные публикации в Internet, и мой личный опыт общения с коллегами на различных конференциях. Многие из них стараются на практике не использовать открытые ответы. Объясняется это, обычно, не столько сложностью анализа свободно-конструируемых ответов, сколько представлением о достоверности результатов, полученных на закрытых тестах.

Что же касается упомянутой статьи [2], то у внимательного читателя не могут не возникнуть сомнения в достоверности результатов, полученных её авторами. Во-первых, условия проведённых экспериментов были таковы: тесты содержали по 50 вопросов, из только 5 – открытой формы. 10% – это слишком мало, чтобы обнаружить серьёзные изменения в результатах. Во-вторых, тестирование проводилось в течение двух дней, т.е. у тестируемых была возможность улучшить свои знания после первого тестирования, если они не были удовлетворены своими результатами.

Всё это подтолкнуло меня к тому, чтобы поставить аналогичные эксперименты, только при более строгих условиях.

2. Постановка задачи

Тест со свободно-конструируемыми ответами (открытая форма ответа) является наиболее естественной, т.е. близкой к классическому обучению и привычной формой контроля знаний, в том числе, автоматизированного контроля знаний. Тестируемый получает вопрос (задание), решает это задание или формулирует ответ, а система оценивает этот ответ. Конечно, оценка открытого ответа является достаточно сложной для программной реализации, отчасти поэтому самой часто используемой формой тестовых вопросов являются выборочные (закрытые).

С моей точки зрения единственным достоинством закрытой формы ответа является простота реализации; всё остальное – недостатки. В реальной жизни человек при решении задачи генерирует её решение. В отличие от тестов закрытой формы он не имеет списка готовых ответов, один из которых – правильный. В закрытом тесте этап генерации решения просто отсутствует: тестируемый сначала отбрасывает самые неправдоподобные варианты, а из оставшихся выбирает тот, который больше похож на правильный ответ. Это совершенно другая работа. Закрытая форма ответа имеет и другие недостатки:

- Тестируемый может запомнить неправильные варианты ответа как правильные. У тестируемого не всегда есть желание (и возможность) подробно проанализировать свои результаты и выяснить, на какие вопросы теста были даны неправильные ответы.
- В закрытом тесте существует ненулевая вероятность угадывания правильного ответа. Сторонники закрытой формы тестирования ссылаются на расчёты, которые показывают, что эта вероятность для случая, например, "5 вариантов – из них один правильный", достаточно низкая (0,2). Для учёта этой вероятности даже применяют специальные модели (например, трёхпараметрическую модель Бирнбаума, [3]). Но для таких расчётов предполагается, что тестируемый осуществляет случайный выбор. На самом деле выбор ответа редко осуществляется случайным образом. Тестируемый обычно хоть что-то знает о предмете теста: он посетил несколько лекций, что-то почитал (полистал), что-то слышал, умеет логично рассуждать. Поэтому наличие вариантов для него – прекрасная подсказка, которую он использует для исключения наименее правдоподобных вариантов.

Последний тезис породил следующее предположение: тестирование с закрытой формой ответов при прочих равных условиях покажет более высокие результаты, чем тестирование со свободно-конструируемыми ответами. Для проверки этого предположения автором была проведена серия экспериментов по тестированию студентов.

3. Условия проведения экспериментов

Тестирование проводилось в мае 2008 года по предмету "Базы данных" при следующих условиях:

- Оба теста – закрытый и открытый – содержали по 35 вопросов с одинаковыми формулировками.
- Вопросы были составлены таким образом, чтобы формулировка одного вопроса не являлась подсказкой ответа на другой вопрос.
- Тестирование проводилось в один день, сначала – открытое, затем – закрытое.

Прежде чем обсуждать результаты экспериментов, остановимся на тех методах анализа правильности ответов тестируемых, которые применялись в используемой для тестирования системе контроля знаний TS [4].

4. Методика оценивания ответов

4.1. Методика оценивания выборочного ответа

В общем случае перечень вариантов ответов для выборочной формы может содержать более одного правильного ответа. Поэтому выборочный ответ оценивался по формуле:

$$\delta_{ми} = \frac{K_A}{L_E + K'}, \quad (1)$$

где L_E – общее количество правильных вариантов,

K_A – количество правильных вариантов, вошедших в ответ тестируемого;

K' – количество неправильных вариантов, вошедших в ответ.

Полученная таким образом оценка находится в интервале $[0,1]$ и затем пересчитывается в привычную шкалу (2, 3, 4, 5), которая для простоты далее будет называться пятибалльной шкалой.

Примечание: если для расчётов характеристик теста и оценок тестируемых применяется IRT (Item Response Theory, [5]), то используются дихотомические тесты. Но результаты IRT можно обобщить на дифференцированные оценки [6].

4.2. Методика оценивания открытого ответа

Ответ свободно-конструируемой формы в системе контроля знаний TS может относиться к одному из следующих видов:

- Числовое выражение. Это выражение вычисляется и проверяется на равенство эталонному значению E с учётом погрешности ϵ . Если ответ A находится в интервале $(E - \epsilon \leq A \leq E + \epsilon)$, то он признаётся правильным (оценка "5"), иначе – неправильным (оценка "2").
- Текст. Он подвергается синтаксическому анализу, грамматические ошибки игнорируются, возможно наличие нескольких эталонов. Рассчитывается степень сходства ответа и каждого эталона, в качестве окончательной оценки берётся максимальное из полученных значений и пересчитывается в пятибалльную шкалу.
- Множество элементов. В этом случае оценка определяется по формуле (1).
- Список элементов. В этой ситуации правильность ответа зависит также от порядка элементов. Оценка списка проводится в два этапа:
 - 1) Список оценивается как множество (1), получается степень сходства δ_1 .
 - 2) Отбрасываются элементы, не входящие в эталонный ответ. Полученный список оценивается так:

$$\delta_2 = 1 - \frac{K_i}{K_n}$$

где K_i - количество инверсий списка относительно эталонного списка;

K_n - максимальное количество инверсий списка длиной n :

$$K_n = \frac{n(n-1)}{2}.$$

Окончательно оценка δ рассчитывается так: $\delta = \delta_1 * \delta_2$.

Этот способ (Д-метод) более подробно изложен в работах [7, 8].

Полученная интегральная оценка δ также находится в интервале [0,1] и пересчитывается в пятибалльную шкалу.

5. Результаты экспериментов

В эксперименте участвовали студенты 4-го курса МФТИ и 3-го курса МИЭМ (2 потока с разных факультетов), изучающие курсы "Базы данных" и "Распределенные базы данных". Ограничений по времени установлено не было, но испытуемые тратили на открытый тест от 26 до 54 минут, а на закрытый – от 15 до 33 минут. Итоговые результаты эксперимента приведены в табл. 1.

Институт, поток	Количество тестируемых	Средняя оценка по открытому тесту	Средняя оценка по выборочному тесту	Разница в оценках
МФТИ, ФНТИ	9	3.01	3.84	0,83
МИЭМ, гр. С	15	3.07	3.92	0,85
МИЭМ, гр. К	11	2.98	3.89	0.91
Средняя разница между оценками двух тестов				0.86

Таблица 1. Результаты тестирования

Как видно из табл. 1, тест с выборочными ответами даёт более высокие оценки: средняя разница между результатами открытого и закрытого тестирования составила 0.86 балла. (Диаграмма распределения оценок приведена в Приложении).

Разница между оценками за тесты у одних и тех же студентов составила от 0.41 до 1.52 балла. Разницу в интервале [0,7; 1,2] показали 66% тестируемых (23 человека из 35). В Приложении приведён пример вопроса в открытой и закрытой формах, ответы на этот вопрос и результаты оценки ответов.

Результаты показали высокую корреляцию между оценками одного и того же тестируемого на один и тот же вопрос в разных тестах. Из статистики, приведённой в табл. 2, видно, что в 59% случаев тестируемый получал в обоих тестах на один и тот же вопрос одинаковые оценки. В то же время, в трети случаев (34%) тестируемые смогли улучшить результат ответа на вопрос в закрытой форме по сравнению с этим же вопросом в открытой форме. Таким образом, экспериментально было подтверждено предположение о том, что тестируемые успешно используют варианты ответов закрытого теста как подсказки.

Соотношение оценок за тесты	Количество ответов	В процентах от общего количества
Вопрос получил в открытом тесте более низкую оценку, чем в закрытом	840	34%
в том числе: открытый тест – 2, закрытый тест – 5	441	
Вопрос получил в открытом тесте такую же оценку, как в закрытом	1443	59%
в том числе:		
– обе оценки – 2	623	
– обе оценки – 5	754	
– обе оценки находятся в интервале от 2 до 5	66	
Вопрос получил в открытом тесте более высокую оценку, чем в закрытом	167	7%
в том числе: открытый тест – 2, закрытый тест – 5	68	

Таблица 2. Распределение оценок по вопросам

6. Заключение

Полученные результаты подтверждают высказанное ранее предположение о том, что при прочих равных условиях выборочный тест показывает более высокие результаты, чем тест со свободно-конструируемыми ответами. Разница в оценках составляет около одного балла (при использовании стандартной "пятибалльной" шкалы).

Из всего вышесказанного можно сделать вывод: при проведении тестирования желательно учитывать наличие эффекта повышения оценки при использовании только вопросов закрытой формы. Результаты можно интерпретировать и следующим образом: выборочный тест не отражает реального уровня знаний тестируемого. Потому что люди, перед которыми за полчаса-час до этого тестирования поставили те же самые вопросы, не смогли на них ответить.

Очевидно, что совсем отказаться от выборочных вопросов при тестировании нереально. Но, если тест призван проверять именно уровень знаний обучаемых, а не их умение логически мыслить, а зачастую и угадывать, то выборочные вопросы стоит использовать как можно реже: 10-15% от общего числа вопросов в тесте, и только тогда, когда сложно переформулировать вопрос для применения другой формы ответа. Необходимо более широко использовать вопросы на соответствие и на упорядочение.

Литература

1. Аванесов В.С. Определение качества знаний школьников с помощью тестов. Материалы семинара "Педагогические технологии тестирования в интернете и локальных сетях". – URL: <http://www.botik.ru/~mupc/27.11.2002/mogilev2/mogilev2.htm> (дата обращения: 05.12.2009).
2. Дятлова К.Д., Михалева Т.Г. Исследование влияния разнообразия форм тестовых заданий на статистические характеристики теста. // Вопросы тестирования в образовании. – №4. – 2006. – с. 65-75.
3. Дружинин В.Н. Экспериментальная психология: Учебное пособие. – М.: ИНФРА-М, 1997. – 256 с.
4. Документация по системе контроля знаний TS. <http://rema44.ru/razrabotki/> (дата обращения: 05.12.2009).
5. Аванесов В.С. ITEM RESPONSE THEORY: основные понятия и положения. – URL: <http://testolog.narod.ru/Theory59.html> (дата обращения: 05.12.2009).
6. Карданова Е.Ю. О применимости политомической модели Г. Раша к тестовым заданиям различных форм, оцениваемым политомически. // Вопросы тестирования в образовании. – №4. – 2006. – с. 44-56.
7. Карпов В.Э., Карпова И.П. Язык описания системы контроля знаний. – Компьютеры в учебном процессе, 2000, №4. – URL: <http://rema44.ru/resurs/> (дата обращения: 05.12.2009).
8. Карпова И.П. Анализ ответов обучаемого в автоматизированных обучающих системах. – Информационные технологии, 2001, № 11. – с.49-55. URL: <http://rema44.ru/resurs/> (дата обращения: 05.12.2009).

THE COMPARISON OF OPEN AND SELECTIVE TESTS

Irina Karpova, Institute of Electronic and Mathematic, Moscow, Cand.Sc., docent (karpova_ip@mail.ru).

Abstract: The comparison of selective tests and tests with free-designed form of the answer is discussed in this paper. Some formal methods of free-designed answers analysis in tutoring systems are described. The results of experiments on testing are proposed, the analysis of these results is carried out. It is shown that with other things being equal the selective test produces the higher estimation, than the open test.

Keywords: control of knowledge, testing, selective answers, free-designed answers.

Приложение

На рис. 1 приведена диаграмма распределения оценок за тесты. Данные на диаграмме упорядочены по оценкам тестируемых за открытый тест.

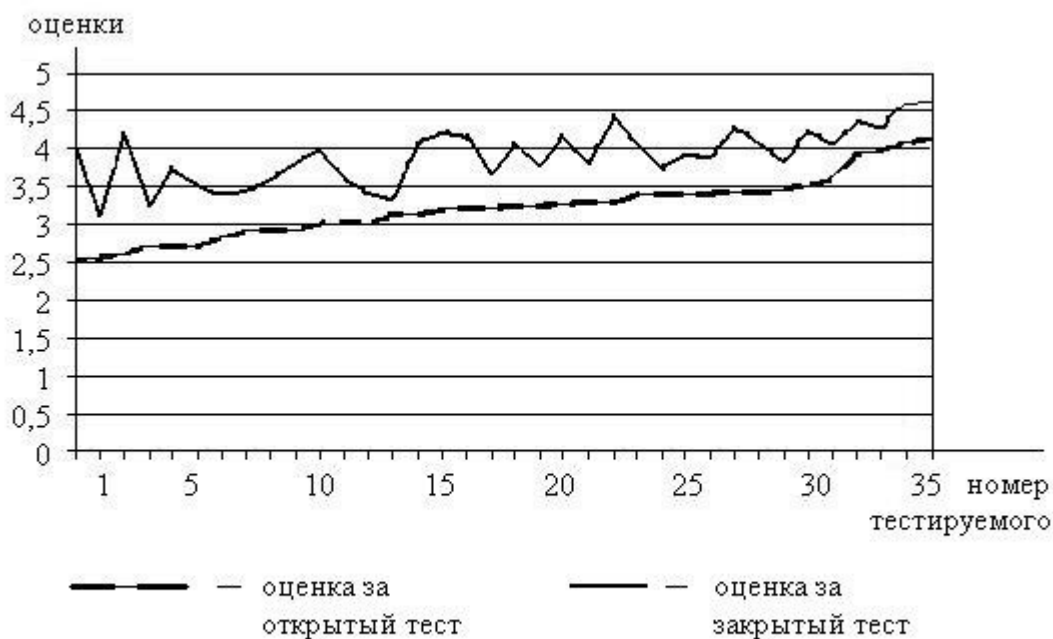


Рис. 1. Диаграмма оценок за открытый и закрытый тесты

На рис. 2 приведена диаграмма распределения разницы оценок за два теста в процентах от общего числа тестируемых. В интервале $[0,7; 1,2]$ находится разница в оценках у 66% тестируемых (23 человека из 35).

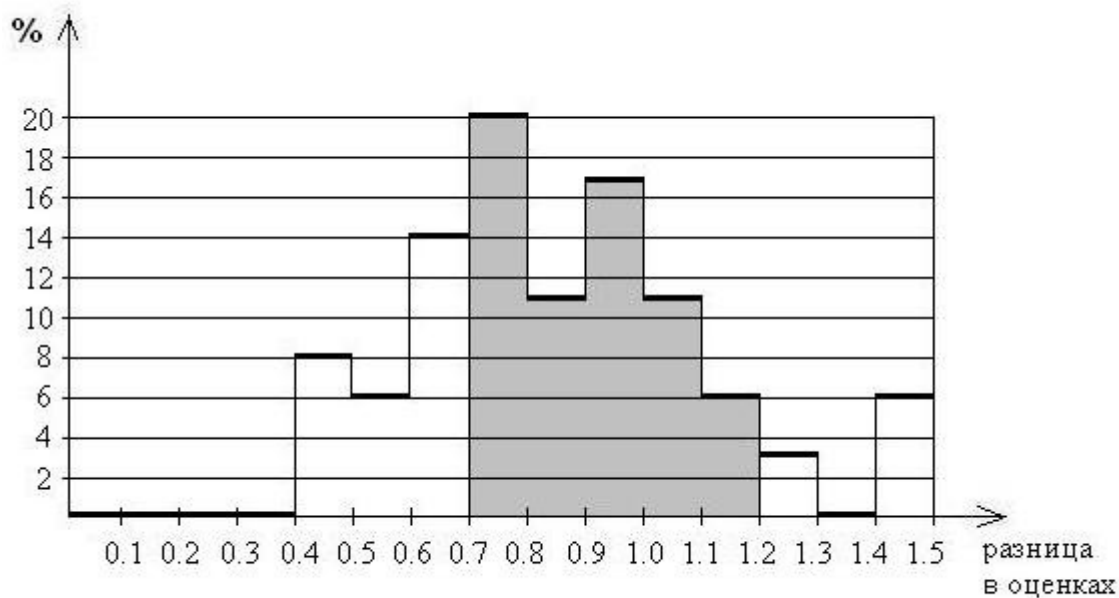


Рис. 2. Диаграмма разницы оценок за два теста в процентах от числа тестируемых

В качестве конкретного примера распределения оценок на свободно-конструируемый вопрос приведём оценки за ответы на вопрос, представленный на рис. 3.

Открытая форма Вопрос 1: Какие виды блокировок по степени параллельности работы вы знаете? Эталон 1: {разделяемые, монопольные} Эталон 2: {разделяемые, исключаяющие}
Закрытая форма Вопрос 1: Какие виды блокировок по степени параллельности работы вы знаете? разделяемые монопольные отложенные явные автоматические

Рис. 3. Пример вопроса открытой и закрытой форм

В табл. 3 приведены ответы и оценки этих ответов, полученные в ходе открытого тестирования; а в табл. 4 – ответы и оценки закрытого тестирования.

Варианты ответов	Количество	Оценк а	Всего
(разделяемые, исключаяющие); (разделенная, исключаяющая); (раздельные, монопольные)	8	5	11
монопольная	2	3,5	
разделяемые, автоматические	1	3	
(гонки, тупики); (полная, распределенная); (по множеству данных, по степени доступности, по способу установки); (блок на запись, блок на чтение) и др.	24	2	24
Средняя оценка за ответ		2,8	

Таблица 3. Распределение оценок за ответы на Вопрос 1 в открытой форме

Варианты ответов	Количество	Оценка	Всего
(1, 3); (3, 1)	10	5	26
(1, 3, 4); (1, 3, 5)	6	4	
(1, 3, 4, 5); (1); (3)	5	3,5	
(1, 4); (3, 5); (1, 5)	5	3	
(1, 2, 4); (1, 2, 5); (1, 4, 5)	4	2,75	9
(1, 5, 2, 4)	1	2,6	
(4); (2)	2	2	
(4, 5); (5, 2)	2	2	
Средняя оценка за ответ		3,7	

Таблица 4. Распределение оценок за ответы на Вопрос 1 в закрытой форме

Как видно из табл. 3 и 4, соотношение правильных и неправильных ответов на вопрос открытого теста 11/24, а на вопрос закрытого теста – 26/9, т.е. около 40% тестируемых получили более высокие оценки на закрытом тесте.