

О НЕКОТОРЫХ ИНСТРУМЕНТАЛЬНЫХ ПРОБЛЕМАХ КОМПЬЮТЕРНОГО ТЕСТИРОВАНИЯ И ПУТЯХ ИХ ПРЕОДОЛЕНИЯ

А.И. Шемаров, Е.Г. Гриневич

Одной из важнейших составляющих современного инновационного образовательного процесса является процесс контроля знаний студентов, магистрантов, слушателей. Наиболее востребованной в современной педагогической практике является форма индивидуального компьютерного тестирования. В статье предлагаются методы оценки возможных инструментальных погрешностей, возникающих в процессе компьютерного тестирования, и определяются возможные пути, позволяющие нивелировать в значительной мере эти погрешности.

Современные требования к образовательному процессу предполагают его интенсификацию процесса обучения, формирование профессиональных компетенций обучаемых в процессе обучения, практикоориентированность. Достижение этих результатов невозможно без применения современных информационных технологий. Одной из важнейших составляющих современного инновационного образовательного процесса является процесс контроля знаний студентов, магистрантов, слушателей, осуществляемый как итоговый по окончании изучения модуля или дисциплины в целом, так и промежуточный, позволяющий преподавателю получать оценку эффективности изложения дисциплины и оценить усвоение материала обучающимися. Очевидно, что правильная организация процесса контроля знаний напрямую влияет на качество обучения. Использование компьютерных технологий для проведения аттестации, как и любых иных машинноориентированных технологий, могут сопровождаться появлением погрешностей, как инструментальных, так и методических. В ряде статей авторы пытались оценить возможные инструментальные погрешности, возникающие в процессе компьютерного тестирования, и определить пути, позволяющие нивелировать их в значительной мере.

Традиционные подходы к проведению процедуры контроля знаний в высших учебных заведениях (экзамены, зачеты, контрольные работы, коллоквиумы, рефераты и т.д.) имеют определенные недостатки. В первую очередь они являются весьма трудо и время затратными. А это в свою очередь неприемлемо на современном этапе развития высшей школы, непосредственно предполагающем интенсификацию труда профессорско-преподавательского состава учреждений образования. Во-вторых, принятая методика приема экзаменов с использованием билетов, включающих обычно 2-3 вопроса, не позволяет оценить полноту освоения материала и зачастую провоцирует списывание. В настоящее время, устранению этих трудностей в значительной мере способствует такая форма контроля знаний, как тестирование. При этом наиболее современной и востребованной на данный момент в педагогической практике является форма индивидуального компьютерного тестирования.

В статье [1] рассматриваются возможные инструментальные погрешности компьютерных систем тестирования.

Основой для проверки статистических погрешностей служили следующие гипотезы и допущения. Первое – вероятность появления вопроса, включаемого в тестовое задание (тестовую последовательность), должна соответствовать равномерному закону распределения. Второе – не должно быть корреляции между вопросами в тестовой последовательности, что существенно может снизить качество тестирования. Третье – вероятность влияния вопроса тестовой последовательности на оценку тестирования относительно её среднего уровня соответствует нормальному закону распределения, то есть набор вопросов каждого тестового задания имеет приблизительно одинаковый уровень сложности. Исключив методические проблемы, возникающие при проведении тестирования, при анализе формирования тестовых последовательностей инструментальных систем компьютерного тестирования были выявлены следующие недостатки:

- используемые системами генераторы случайных чисел не позволяют обеспечить равномерное распределение вопросов в тестовых заданиях;

- неравномерность распределения вопросов в тестовых последовательностях может приводить к искажению результатов тестирования, как в сторону уменьшения оценки, так и в сторону увеличения оценки тестируемых, что приводит к искажению результатов экзаменационных испытаний;

- небольшое отклонение средневзвешенных оценок от среднего значения по каждому вопросу может свидетельствовать о недостаточной подготовке студентов к тестовым испытаниям, при этом, существенные пробелы в знаниях могут быть нивелированы при тестировании за счет усреднения результирующей оценки по всем вопросам тестового задания, что, в свою очередь позволило авторам выдвинуть гипотезу о возможности прохождения тестовых испытаний путем простого угадывания ответа на вопрос теста тестируемым.

Проверка гипотезы о наличии существенных пробелов в знаниях была проверена при проведении экзаменационных испытаний с помощью группы специальных вопросов, внедренных в тест [2].

Основными предпосылками для проведения эксперимента стали следующие допущения.

При недостаточном усвоении материала по дисциплине тестируемые на базе общих знаний могут путем удаления заведомо неверных ответов сократить количество правильных ответов, из которых осуществляется выбор. Таким образом, существенно повышается вероятность правильного ответа. Если количество правильных ответов сокращается до двух или даже одного, то исходя из вероятностных оценок, тестируемый будет получать положительную оценку, выбирая ответы наугад, не имея минимального уровня знаний по дисциплине.

Цель работы оценить, насколько часто студенты прибегают к угадыванию при выборе правильного ответа и выработка рекомендаций по устранению этой ситуации при компьютерном тестировании.

Для проведения эксперимента была сформирована база вопросов для контроля знаний, в которой вопросы были разделены в следующей пропорции: 95% вопросов (основная база вопросов) составили вопросы для проверки знаний, полученных при изучении учебного материала (теоретическая, практическая составляющая курса и материал управляемой самостоятельной работы студентов), и 5% составили вопросы-«пустышки» (дополнительная база вопросов), имеющие псевдонаучный смысл, не имеющие правильного ответа или смысла априори. Обе базы вопросов включали в себя вопросы различных типов: вопросы в закрытой форме (множественный выбор), вопросы на соответствие, вопросы типа «Верно/Неверно». Время прохождения теста было ограничено периодом времени прохождения теста, определяющимся опубликованными данными, по которым при ответе на вопрос, не требующий использования расчетов и справочной литературы, отводилось от 30 до 45 секунд. Технологическое время исключалось из времени, необходимого для прохождения теста. Целью тестирования являлась проверка гипотезы об угадывании правильных ответов. Индикаторами являлись ответы на вопросы-пустышки и аномально низкое время прохождения теста.

Примеры вопросов-«пустышек» представлены ниже:

Вопрос E01 – «Какая часть ядра ОС В-31 была включена в ядро ОС Windows XP?». ОС В-31 в природе не существует, но прослеживается связь с маркировкой вооружения, используемого в армии США и, следовательно, находящаяся «на слуху» благодаря средствам массовой информации. В качестве ответов на вопрос были предложены три варианта ответа:

- 1). 10%. Утвердительно ответил 1 тестируемый из 29 (3%).
- 2). 25%. Утвердительно ответило 19 тестируемый из 29 (66%).
- 3). 50%. Утвердительно ответило 5 тестируемых из 29 (17%).

Не дали ответ на вопрос 4 тестируемых студента из 29 (14%), что в первом приближении можно считать правильным ответом, если вопрос вообще рассматривался тестируемым;

Вопрос E03 – «Эффективность ОС измеряется в - ». Не существует общеупотребительных методик сравнения различных ОС, не говоря уже об единицах измерения, но возможно в вводной лекции были упомянуты фамилии «отцов-основателей» вычислительных машин и кибернетики. В качестве ответов на вопрос были предложены три ответа:

- 1). В Винерах. Утвердительно ответили 3 тестируемых из 31 (10%).
- 2). В Тьюрингах. Утвердительно ответило 4 тестируемый из 31 (13%).
- 3). В Гилбертах. Утвердительно ответило 20 тестируемых из 31 (65%) – возможно слышали о смесях задач, на которых проверяется производительность вычислительных систем, комплексов и сетей.

Не дали ответ на вопрос 4 тестируемых студента из 31 (12%);

Вопрос E07 – «Плотность загрузки операционной системы определяется в ...». Здесь вообще представлена благоглупость, но в таких похожих на обще-

употребительные величины типа кг/мм² и т.д. В качестве ответов на вопрос были предложены три ответа:

1). В байтах на мм² кристалла процессора. Утвердительно ответили 9 тестируемых из 28 (32%).

2). В пикселях на байт. Утвердительно ответило 11 тестируемых из 28 (39%).

3). В командах на мм² кристалла процессора. Утвердительно ответило 3 тестируемых из 28 (11%).

Не дали ответ на вопрос 5 тестируемых студента из 28 (18%).

Не вызывает сомнений, что при проведении различных вариантов тестирования, даже хорошо подготовленные студенты иногда угадывают правильный вариант ответа. Как показывает практика, особенно для дисциплин гуманитарного и естественнонаучного цикла, вопросы с множественным выбором при формировании тестов составляют подавляющее большинство. В статье было проведено исследование значимость влияния, которое оказывает угадывание правильного ответа тестируемым на его результирующую оценку.

Результаты эксперимента показали наличие значительного количества попыток угадывания правильного результата. Существующая система проведения тестирования существенно нивелирует знания студентов, так как делает возможным получить положительную оценку без, сколько бы то ни было, существенной подготовки к экзамену. Этот результат, полученный в ходе проведения эксперимента, позволяет сделать вывод о необходимости внесения изменений в систему оценивания знаний и разработке новых алгоритмов оценивания знаний, позволяющих выявлять попытки угадывания правильных ответов и активно противодействовать этому.

В работе [3] приводятся результаты исследований, утверждающие, что учащиеся, имеющие более высокие учебные достижения, также и более точны в оценке уровня своих знаний, лучше прогнозируют собственную успешность при контроле знания. Наоборот, учащимся с низкими достижениями свойственна «сверхуверенность» в собственных знаниях, поскольку такие ученики переоценивают уровень своих знаний.

Всегда можно выделить группу учащихся, не обладающих достаточным уровнем знаний и пытающихся пройти тест на удачу, по вероятностному принципу, тем более, что зачастую допускается возможность нескольких попыток прохождения теста.

Для формирования более объективной оценки знаний учащихся предлагается построить нелинейную шкалу оценки их знаний [4]. Наиболее просто это достигается путем введения степенных зависимостей общей оценки от относительного количества правильных ответов. При этом каждый следующий правильный ответ имеет больший вес, чем предыдущий. Оценка тестового испытания можно вычислить по следующей формуле:

$$A = A_{\min} + (A_{\max} - A_{\min}) \times \left(\frac{n}{N}\right)^{\alpha}, \quad (1)$$

где A_{\min} - значение минимальной оценки;
 A_{\max} - значение максимальной оценки;
 N – общее число вопросов в тесте;
 n – число вопросов, на которые был дан правильный ответ;
 α - коэффициент в интервале $[1..5]$, определяющий нелинейность значения оценки. Значение коэффициента α выбирается эмпирически.
 Графики значений оценки $A = f(n, \alpha)$ для различных значений коэффициента α представлены на рисунке ($N = 50, A_{\min} = 1, A_{\max} = 10$).

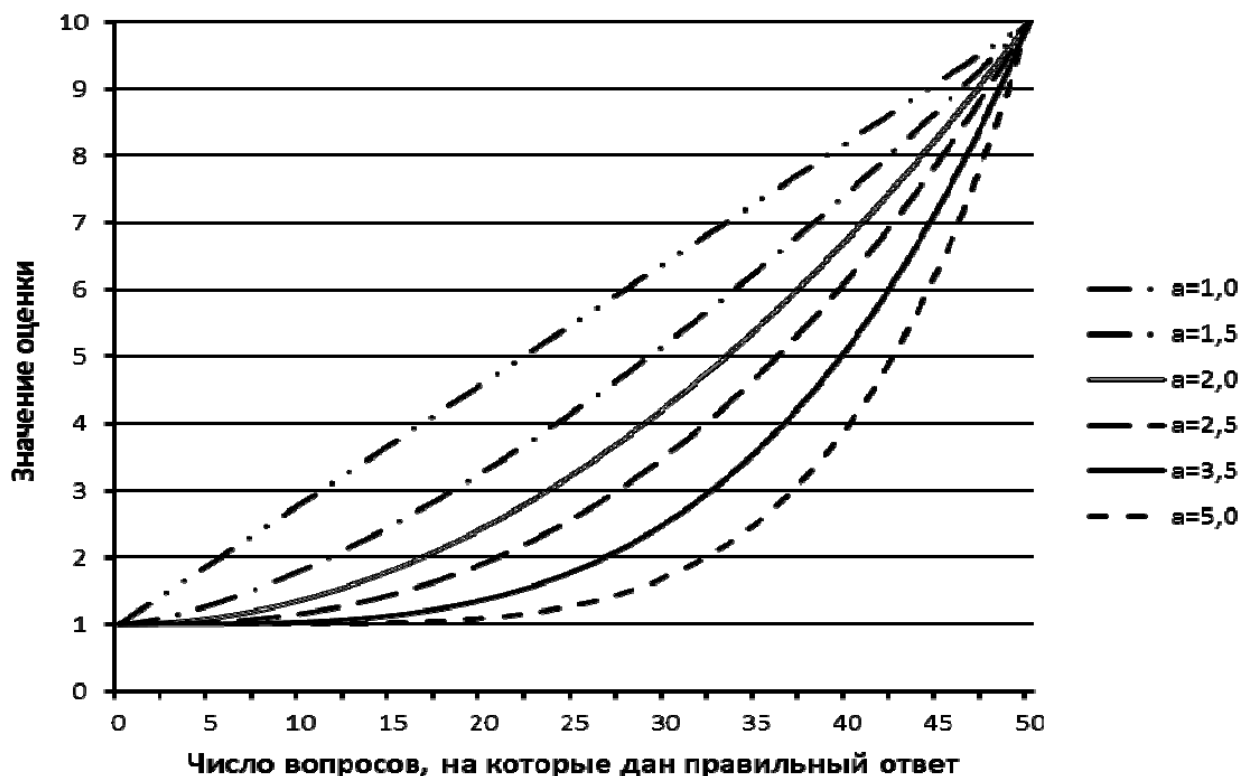


Рис. 1. Функция $A = f(n, \alpha)$, при $N = 50, A_{\min} = 1, A_{\max} = 10$

Как видно из представленных на рисунке 1 графиков, оценка по тесту нелинейно зависит от доли правильных ответов. Предложенная зависимость позволяет учитывать долю правильных ответов на вопросы, но не позволяет дифференцировать оценки для учащихся, угадывающих ответы.

Для того, чтобы учесть влияние угадывания правильных ответов на общую оценку, формула 1 была модифицирована. Был введен коэффициент η , который авторы назвали коэффициентом уровня притязаний, который был поставлен его в прямую зависимость от количества вопросов, на которые был дан правильный ответ тестируемым (формула 2). Тестируемые обязательно должны быть предупреждены о введении такого коэффициента.

Использование формулы 2 для вычисления оценок позволяет дифференцировать оценку, полученную за одинаковое количество ответов, в зависимости от общего количества ответов на вопросы, данного тестируемым.

$$\begin{aligned}
 A &= A_{\min} + (A_{\max} \times \eta - A_{\min}) \times \left(\frac{n}{N \times \eta}\right)^\alpha = \\
 &= A_{\min} + (A_{\max} \times \eta - A_{\min}) \times \left(\frac{n}{m}\right)^\alpha,
 \end{aligned}
 \tag{2}$$

где $\eta = m / N$ - коэффициент уровня притязаний;
 m - количество вопросов, на которые дан ответ.

Графики значений оценки $A = f(n, \eta)$ для различных значений коэффициента η представлены на рисунке 2 ($N = 50$, $A_{\min} = 2$, $A_{\max} = 10$, $\alpha = 2,5$).

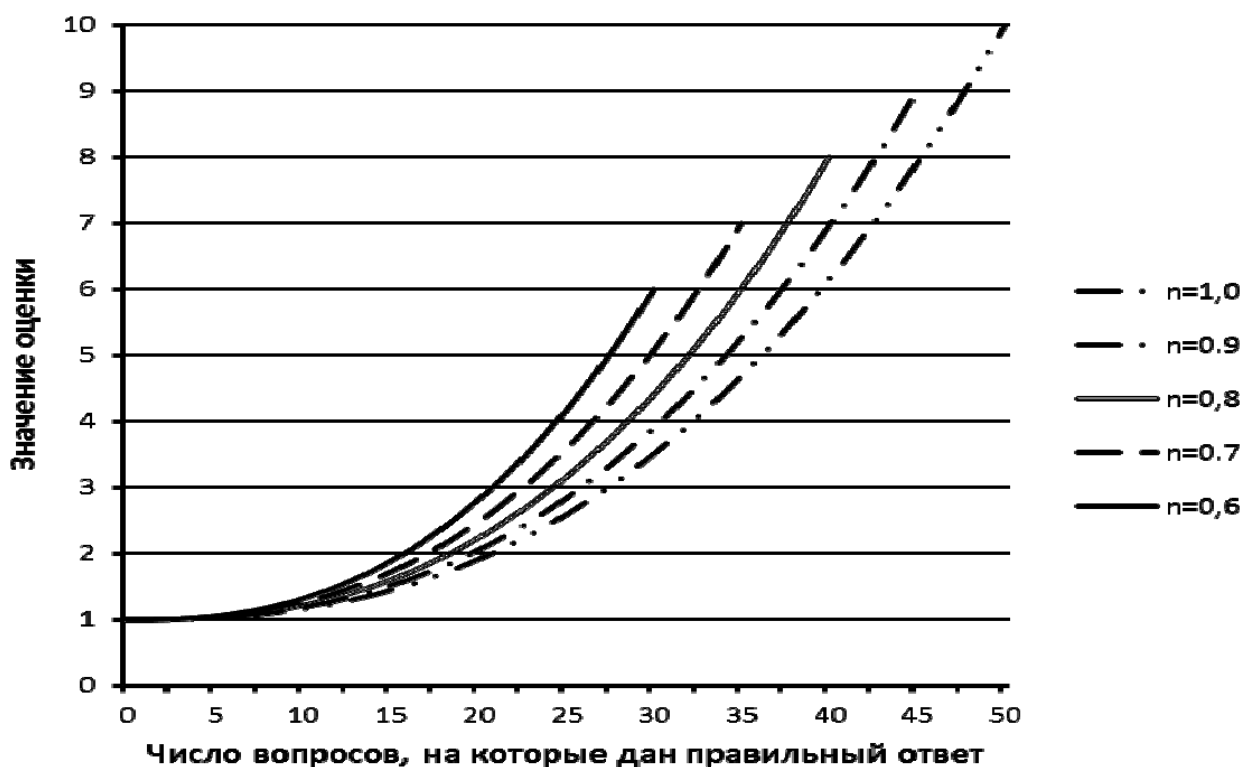


Рис. 2. Функция $A=f(n, \eta)$, при $N = 50$, $A_{\min} = 1$, $A_{\max} = 10$, $\alpha = 2,5$.

Для более удобного использования методики оценки результатов тестирования с использованием коэффициента уровня притязаний оценки тестирования сведены в таблицу при $N = 100$, $A_{\min} = 1$, $A_{\max} = 10$, $\alpha = 2,5$.

Таблица

Отношение правильных ответов к общему количеству вопросов в %	Коэффициент уровня притязаний							Отношение правильных ответов к общему количеству вопросов в %	Коэффициент уровня притязаний						
	0,4	0,5	0,6	0,7	0,8	0,9	1,0		0,4	0,5	0,6	0,7	0,8	0,9	1,0
1%	1	1	1	1	1	1	1	11%	1	1	1	1	1	1	1
2%	1	1	1	1	1	1	1	12%	1	1	1	1	1	1	1
3%	1	1	1	1	1	1	1	13%	1	1	1	1	1	1	1
4%	1	1	1	1	1	1	1	14%	1	1	1	1	1	1	1
5%	1	1	1	1	1	1	1	15%	1	1	1	1	1	1	1
6%	1	1	1	1	1	1	1	16%	1	1	1	1	1	1	1
7%	1	1	1	1	1	1	1	17%	1	1	1	1	1	1	1
8%	1	1	1	1	1	1	1	18%	1	1	1	1	1	1	1
9%	1	1	1	1	1	1	1	19%	1	1	1	1	1	1	1
10%	1	1	1	1	1	1	1	20%	2	1	1	1	1	1	1
21%	2	1	1	1	1	1	1	31%	3	2	2	2	2	2	1
22%	2	2	1	1	1	1	1	32%	3	2	2	2	2	2	2
23%	2	2	1	1	1	1	1	33%	3	2	2	2	2	2	2
24%	2	2	2	1	1	1	1	34%	3	3	2	2	2	2	2
25%	2	2	2	1	1	1	1	35%	3	3	2	2	2	2	2
26%	2	2	2	2	1	1	1	36%	3	3	2	2	2	2	2
27%	2	2	2	2	1	1	1	37%	3	3	2	2	2	2	2
28%	2	2	2	2	2	1	1	38%	4	3	3	2	2	2	2
29%	2	2	2	2	2	1	1	39%	4	3	3	2	2	2	2
30%	2	2	2	2	2	2	1	40%	4	3	3	2	2	2	2
41%		3	3	3	2	2	2	51%			4	4	3	3	3
42%		4	3	3	2	2	2	52%			4	4	3	3	3
43%		4	3	3	2	2	2	53%			5	4	4	3	3
44%		4	3	3	3	2	2	54%			5	4	4	3	3
45%		4	3	3	3	2	2	55%			5	4	4	3	3
46%		4	4	3	3	2	2	56%			5	4	4	3	3
47%		4	4	3	3	3	2	57%			5	5	4	4	3
48%		5	4	3	3	3	2	58%			6	5	4	4	3
49%		5	4	3	3	3	3	59%			6	5	4	4	3
50%		5	4	4	3	3	3	60%			6	5	4	4	4
61%				5	5	4	4	71%					6	5	5
62%				5	5	4	4	72%					6	6	5
63%				6	5	4	4	73%					7	6	5
64%				6	5	4	4	74%					7	6	5
65%				6	5	5	4	75%					7	6	5
66%				6	5	5	4	76%					7	6	6
67%				6	5	5	4	77%					7	6	6
68%				7	6	5	4	78%					8	7	6
69%				7	6	5	5	79%					8	7	6
70%				7	6	5	5	80%					8	7	6
81%						7	6	91%							8
82%						7	6	92%							8
83%						8	7	93%							9
84%						8	7	94%							9
85%						8	7	95%							9
86%						8	7	96%							9
87%						8	7	97%							9
88%						9	8	98%							10
89%						9	8	99%							10
90%						9	8	100%							10

Исходя из полученных экспериментальных данных, можно сделать допущение, что существующая система проведения тестирования существенно нивелирует знания студентов, так как позволяет получать положительные оценки практически без подготовки к экзамену. Для устранения выявленной проблемы необходимо внедрять алгоритмы, активно выявляющие попытки угадывания правильных ответов. В качестве промежуточной меры можно использовать в каждом вопросе добавочный ответ – «нет правильного ответа», который должен быть истинным для определенного количества вопросов, и позволит снизить вероятность угадывания правильного ответа.

Кардинально решению проблемы может способствовать разработка алгоритмов активного противодействия системе угадывания. Предложенная методика позволит более объективно оценивать знания, а учащихся стимулировать к получению им профессиональных компетенций и неформальной подготовке к контролю знаний.

Список литературы

1. Шемаров, А.И. Анализ влияния статистических погрешностей распределения вопросов на оценки тестируемых в СДО МООДУС / А.И. Шемаров, Е.Г. Гриневич // Информатизация образования. - 2008. - № 2(51). - с. 52-63.

2. Шемаров, А.И. Устранение влияния случайных ответов учащихся на результаты тестирования / А.И. Шемаров, Е.Г. Гриневич // Информатизация образования. -2013. - № 2(71). - с. 66-77.

3. Hacker, D.J. Metacognition in education: A focus on calibration / D.J. Hacker, L. Bol, M.C. Keener // Handbook of metamemory and memory / ed. J. Dunlosky, R.A. Bjork. . N.Y., Psychology Press. - 2008. P. 429–455.

4. Гриневич, Е.Г. Повышение достоверности оценки знаний при проведении тестирования / Е.Г. Гриневич, А.И. Шемаров // Образовательно-инновационные технологии: монография/ [И.В. Абрамова, А.М. Гамаюнова, Е.Г. Гриневич и др.]; под общей ред. Проф. О.И. Кирикова. - Кн. 16. - Москва: Наука: информ; Воронеж: ВГПУ, 2013. - 162 с.

Шемаров Александр Иванович, заведующий кафедрой управления информационными ресурсами факультета инновационной подготовки Института управленческих кадров Академии управления при Президенте Республики Беларусь, кандидат технических наук, доцент, shemarov_ai@pac.by, al_shemarov@mail.ru

Гриневич Елена Георгиевна, старший преподаватель кафедры менеджмента технологий Института бизнеса и менеджмента технологий Белорусского государственного университета, e_grinevich@mail.ru