

РЕАЛЬНЫЙ ПУТЬ К СОЗДАНИЮ САМОНАСТРАИВАЮЩЕЙСЯ КОМПЬЮТЕРНОЙ ТЕСТОВОЙ СИСТЕМЫ

В. П. Гусаков, Е. М. Редикарцева

Северо-Казахстанский Государственный университет

В настоящее время в вузах Республики Казахстан тестирование широко используется в качестве измерительного инструмента как для объективной оценки знаний в сессионный период, так и для контроля качества обучения во время аттестации и аккредитации вузов. Одним из преимуществ тестового контроля является возможность его автоматизации. В связи с этим создано и продолжает пополняться множество компьютерных тестовых систем. Практически в каждом крупном вузе Республики Казахстан созданы свои компьютерные тестовые системы. Вот уже несколько лет успешно функционирует республиканский центр компьютерного тестирования, главная задача которого – проведение централизованного вступительного экзамена в вузы страны на основе комплексного тестирования.

Попытаемся классифицировать существующие на сегодняшний день в нашей стране и за её пределами компьютерные тестовые системы и отметить присущие каждому классу достоинства и недостатки. В основу приведённой ниже классификации положен вид (вариант) теста, предлагаемого испытуемому.

Системы, имитирующие бланковое тестирование. Характерной особенностью таких систем является использование жёстко заданного варианта теста. Достоинством такой системы является то, что все испытуемые находятся в равных условиях. Подобные системы целесообразно использовать на вступительных экзаменах в вуз либо при дистанционном тестировании. Однако, как показывает практика, если такая система используется для контроля знаний в студенческой группе в рамках компьютерного класса, когда испытуемые, в силу расположения компьютеров, находятся близко друг от друга, очень велика вероятность искажения результатов по причине списывания, коллективных ответов и т. п. Возможна модификация системы, когда вопросы и/или варианты ответов в каждом задании перемешиваются. В этом случае возможность искажения результатов существенно уменьшается, хотя и не исключается. Неоспоримым достоинством систем данного класса является применимость научно обоснованных методов интерпретации результатов и статистической обработки данных (вычисление показателей трудности, различительной способности, валидности тестовых заданий) с целью последующего повышения качества теста [1].

Системы, позволяющие генерировать различные варианты теста для каждого испытуемого на основе случайного выбора из большой базы тестовых заданий. В нашем университете подобная система тестирования применяется для приёма текущих экзаменов, при этом, согласно разработанному положению, число вопросов в тестовой базе по каждой дисциплине не долж-

но быть меньше, чем число лекционных часов, умноженное на шесть. При использовании такой системы коллективные ответы и списывание друг у друга практически исключены. Также существуют методики статистической обработки данных в таких системах [2]. Но в этом случае очевидно, что различным испытуемым предлагаются неравнозначные по трудности тесты. Таким образом, традиционные методы интерпретации результатов, основанные на кумулятивности (т. е. когда суммируется число заданий теста, на которые испытуемый ответил правильно, и на основе этого показателя выставляется оценка), оказываются не совсем пригодными. Поэтому требуется разработка и обоснование новых методов. Например, при суммировании числа успешно выполненных заданий можно складывать не единицы, а веса заданий – некоторые коэффициенты, приписанные каждому заданию в соответствии с его уровнем трудности. Веса могут быть расставлены либо посредством экспертных оценок [3], либо с использованием статистических оценок, полученных по итогам предварительного тестирования [1]. Наилучшим, на наш взгляд, приписывание весов будет в том случае, если оно проведено на основе экспертных оценок и впоследствии скорректировано в соответствии с накопленными статистическими данными.

Системы, использующие нетрадиционные методики тестирования. К этому классу мы относим системы, в основу которых положено использование так называемых «нечётких» тестов [4], тестов без неправильных и правдоподобных вариантов ответов [5], а также адаптивные системы тестирования [6].

«Нечёткие» тесты позволяют более точно (по сравнению с традиционными тестами) определить уровень знания и понимания предмета. Это достигается за счёт того, что в отличие от «чётких» тестов, где неполный или неточный ответ оценивается как неправильный, в «нечётком» тесте варианты ответов тестового задания считаются более или менее правильными. Степень правильности определяется значением так называемой функции принадлежности $F(X)$ варианта ответа X , принимающей значения из некоторого отрезка. Чем правильней ответ, тем больше значение этой функции.

Тест без неправильных и правдоподобных вариантов ответов представляет собой гибрид задания закрытой формы и задания на установление соответствия. Создаётся база вопросов и ответов, из которой случайным образом выбираются задания для тестирования. «Лишние» варианты ответов находятся случайной выборкой из базы ответов на оставшиеся вопросы. Достоинством данной системы является то, что не происходит подсознательного запоминания испытуемыми неправильных вариантов ответов.

Адаптивные системы тестирования разрабатываются с целью оптимизации процесса тестирования – за наименьшее время достаточно точно определить уровень знаний испытуемого. Существенным отличием адаптивного тестирования от остальных видов является то, что вариант теста жёстко привязан к конкретному испытуемому. Различные варианты могут отличаться не

только содержанием (составом заданий, уровнем трудности), но и количеством заданий, включённых в тест. Выбор каждого предъявляемого тестового задания зависит от того, как испытуемый справился с предыдущими заданиями, поэтому данная технология тестирования невозможна без использования компьютеров.

Основной недостаток, зачастую присущий системам, базирующимся на нетрадиционных методиках тестирования – это недостаточная разработанность и/или научная обоснованность математических моделей и алгоритмов, на основе которых они функционируют. Кроме того, при создании систем адаптивного и «нечёткого» тестирования необходимо выполнение большого объёма кропотливой работы по созданию качественной тестовой базы.

Тем не менее, на наш взгляд, создание компьютерных систем оценки знаний на базе нетрадиционных методик тестирования, а также построение математических моделей и алгоритмов их функционирования являются одним из перспективных направлений информатизации образования.

Рассмотрим один из возможных алгоритмов функционирования адаптивной (самонастраивающейся) системы тестирования, предназначенной для выявления уровня знаний студентов. Целесообразно, на наш взгляд, в качестве изменяемого параметра использовать уровень трудности следующего тестового задания.

Система использует базу тестовых заданий, ранжированных по уровням трудности. При создании базы тестовых заданий каждому заданию приписывается уровень трудности, определяемый при помощи системы экспертных оценок. По результатам исследований (на основе данных предварительного тестирования более чем 400 абитуриентов), проведённых в нашем университете, уровень трудности, полученный методом экспертных оценок, хорошо коррелирует с уровнем трудности, определённым методами классической теории тестов [3]. Полученный уровень трудности является начальным и уточняется в процессе функционирования системы с учётом накопленных статистических данных о результатах тестирования.

На этапе тестирования нами выбран следующий алгоритм адаптации. Пусть α_i – наибольший уровень трудности группы заданий, с которыми испытуемый справился успешно, а β_i – наименьший уровень трудности группы заданий, с которыми испытуемый не справился (i – номер этапа). Примем начальные значения α_0 и β_0 равными, соответственно, наименьшему и наибольшему значениям шкалы уровня трудности.

На начальном этапе испытуемому предлагается группа тестовых заданий уровня трудности $\gamma_1 = \left\lfloor \frac{\alpha_0 + \beta_0}{2} \right\rfloor$ ($\lfloor x \rfloor$ – целая часть x). В случае, если количество полученных правильных ответов превысит заданное число (критериальный балл), показывающее, что испытуемый достаточно успешно справляется с заданиями данного уровня трудности, то ему предлагается группа заданий

трудности β_0 . В противном случае трудность следующей группы предлагаемых заданий считается равной α_0 . На этом этапе можно чётко «отсечь» испытуемых, которые справляются с самыми трудными («отличники») и не справляющихся с самыми лёгкими тестовыми заданиями («двоечники»). Испытуемые, не попавшие в данные категории, тестируются дальше. При этом если испытуемый справился с заданиями уровня α_0 (но не справился с заданиями уровня γ_1), принимается $\alpha_1 = \alpha_0$, а $\beta_1 = \gamma_1$. В противном случае, когда испытуемый не справился с заданиями уровня β_0 (но справился с заданиями уровня γ_1), принимается $\alpha_1 = \gamma_1$, а $\beta_1 = \beta_0$. Далее полагая на каждом этапе α_i равным наибольшему уровню трудности группы заданий, с которыми испытуемый справился успешно, а β_i равным наименьшему уровню трудности группы заданий, с которыми испытуемый не справился, определяем уровень трудности следующей группы предлагаемых заданий по формуле

$$\gamma_{i+1} = \left\lfloor \frac{\alpha_i + \beta_i}{2} \right\rfloor.$$

Процесс повторяется до тех пор, пока разность $\beta_i - \alpha_i$ не станет равной единице (в выбранной шкале). Уровень знаний испытуемого соответствует последнему значению α_i и при желании может быть уточнён на основе информации о количестве правильных и неправильных ответов, данных испытуемым в процессе тестирования.

Следует отметить, что задания, предлагаемые испытуемому на каждом шаге, выбираются случайным образом среди заданий необходимого уровня трудности из базы тестовых заданий. Также при таком подходе, желательно, чтобы диапазон изменения уровня трудности был достаточно большим.

При апробации системы такие параметры, как количество заданий предлагаемых на каждом уровне трудности, и критериальный балл, соответствующий данному уровню трудности, задаются создателем теста, но в дальнейшем планируется проведение исследований для определения их оптимальных значений.

В результате функционирования системы каждый испытуемый получает вариант теста, соответствующий его уровню знаний, причём количество вопросов в отдельных вариантах различно.

Приведём сравнительный анализ эффективности тестов, сгенерированных в самонастраивающейся системе, и тестов, сгенерированных традиционной тестовой системой (к традиционным мы относим системы первых двух классов приведённой выше классификации). При тестировании с использованием самонастраивающейся системы после прохождения последнего задания уровень знаний R определён. При тестировании с использованием традиционной системы уровень знаний (в процентах) необходимо рассчитать по формуле: ,

$$R = \frac{100\%}{\max R} \sum_i R_i \cdot L_i$$
 , где $\max R$ – максимально возможная оценка за тест, R_i – оценка за i -й вопрос, L_i – весовой коэффициент, соответствующий трудности

i -го вопроса, суммирование ведётся по всем вопросам, на которые испытуемый ответил правильно.

Время T , затрачиваемое на прохождение теста, рассчитывается по формуле $T = \sum_{i=1}^n T_i$, где T_i – время, отводимое для ответа на i -е задание, n – число за-

даний в тесте. Очевидно, что чем меньше n , тем меньше T . Число заданий в тесте, в свою очередь, может быть рассчитано по формуле $n = \sum_i n_i$, где n_i –

количество предлагаемых заданий i -го уровня трудности. При генерации теста в традиционной системе, как правило, i принимает k различных значений ($i = 0, 1, \dots, k-1$), где k – количество уровней трудности. При генерации теста в самонастраивающейся системе количество различных значений i даже в самом неблагоприятном случае не превосходит $\left[\frac{k}{2}\right] + 1$. Это число является

верхней оценкой, которая при $k > 7$ оказывается больше истинного значения максимального количества различных значений i .

Сравнительные данные о количестве вопросов в тесте для $1 < k < 16$ приведены в таблице. Для удобства рассмотрен простейший случай, когда $n_i = 1$ по всем i , то есть тестируемому предлагается по одному заданию каждого уровня трудности.

Количество уровней трудности	Количество вопросов в тесте, сгенерированном традиционной системой	Максимальное количество вопросов в тесте, сгенерированном самонастраивающейся системой
2	2	2
3	3	2
4	4	3
5	5	3
6	6	4
7	7	4
8	8	4
9	9	4
10	10	5
11	11	5
12	12	5
13	13	5
14	14	5
15	15	5

В традиционной тестовой системе, согласно классической теории тестов, в генерируемый тест должны быть включены вопросы каждого уровня трудности, поэтому числовые данные первого и второго столбцов таблицы совпадают. Числовые данные третьего столбца получены эмпирически посредством построения траекторий процесса тестирования для самых «неблагоприятных» случаев (согласно описанному выше алгоритму).

Приведённые выкладки показывают, что использование самонастраивающейся системы с описанным выше алгоритмом адаптации позволяет существенно сократить время, затрачиваемое на прохождение теста. При этом упрощается алгоритм определения уровня знаний и сохраняется точность измерения.

В заключение хотелось бы отметить, что при создании компьютерной тестовой системы (либо при выборе готовой) необходимо чётко сформулировать цели её создания (использования) и определить границы и место её применения в учебном процессе. Использование компьютерной тестовой системы в учебном процессе должно быть целесообразным и экономически эффективным, а не являться данью моде. Тестовые технологии могут отрицательно влиять на учебный процесс в тех случаях, когда используются некачественные тесты, либо когда их использование не целесообразно. Также не стоит рассматривать тестирование, в частности реализованное на ЭВМ, как единственно приемлемый вид контроля.

Литература

1. Аванесов В. С. Основы научной организации педагогического контроля в высшей школе. М.: МИСиС, 1989.
2. Карпенко Д. С., Карпенко О. М., Шлихунова Е. Н. Система автоматического повышения качества тестовых заданий и мониторинг процесса усвоения знаний // Телекоммуникации и информатизация образования, 2001. № 1. С. 42–57.
3. Мутанов Г. М., Шевчук Е. В. Экспертная система оценки знаний методом тестирования. Омск, 2001.
4. Орлов М. М. Автоматизированная система тестирования, на базе «нечёткой логики». <http://www.omm2002.chat.ru>
5. Гусаков В. П., Куликов В. П., Панченко К. А. Универсальная оболочка нестандартного тестирования для дистанционного обучения // Материалы международной конференции «Наука и образование – ведущий фактор стратегии. Казахстан–2030». Караганда, 2000. С. 126–129.
6. Шмелёв А. Г., Бельцер А. И., Ларионов А. Г., Серебряков А. Г. Адаптивное тестирование знаний в системе «Телестинг» // Школьные технологии. 2000. № 1. С. 170–171.

О НЕКОТОРЫХ ЗАДАНИЯХ ТЕСТОВОГО ТИПА (НА МАТЕРИАЛЕ ТЕМЫ «МОРФЕМИКА»)

Т. А. Кальщикова

Нижнетагильский государственный педагогический институт

Тесты и задания тестового типа продолжают привлекать внимание преподавателей гуманитарных дисциплин. Достоинства заданий тестового типа неоспоримы: они позволяют проверить знания определенного раздела быстро и объективно. Кроме того, компьютер может представить результат в наглядной форме и даже дать рекомендации по устранению пробелов в знани-