

Актуальні питання навчання

УДК 004.89

О.Г. Молчанова, О.И. Морозова, И.В. Чумаченко

Национальный аэрокосмический университет им. Н.Е. Жуковского «ХАИ», Украина

АНАЛИЗ СИСТЕМ ТЕСТИРОВАНИЯ В КОМПЬЮТЕРНЫХ ОБУЧАЮЩИХ КОМПЛЕКСАХ

В статье рассмотрены существующие системы тестирования в компьютерных обучающих комплексах. Детально проанализированы методы тестирования и системы оценки результатов тестов в мировой практике. Рассмотрены оценки результатов в тестах SAT (Scholastic Assessment Test), ACT (American College Testing), GRE (Graduate Record Examinations) и ВНО (Внешнее независимое оценивание). Кроме этого, проведен обзор и анализ существующих компьютерных систем тестирования, выявлены их достоинства и недостатки.

Ключевые слова: системы тестирования, компьютерные обучающие комплексы, оценка результатов в тестировании, мониторинг качества образования, SAT (Scholastic Assessment Test), ACT (American College Testing), GRE (Graduate Record Examinations), ВНО (Внешнее независимое оценивание).

Введение

В настоящее время в мировой практике мониторинга качества образования нет однотипных моделей или установленных форм. Каждая страна при отборе абитуриентов руководствуется своими критериями и мотивами, образовательными и культурными традициями и законодательными нормами [1 – 3].

Модели поступления в университет фактически отображают следующие пять типов приема:

1. *На основании одного или нескольких школьных выпускных экзаменов*, проводимых на национальном или региональном уровнях правительственными учреждениями. Результаты экзаменов являются как единственным критерием отбора, так и в комплексе с академической успеваемостью в школе и изучением отдельных предметов (Франция, Австрия, Великобритания, Египет, Австралия).

2. *По результатам централизованных вступительных экзаменов* (Китай, Иран, Грузия) или на основании собственных экзаменов, которые проводят вузы (Аргентина, Парагвай), которые рассматриваются в комплексе с академической успеваемостью в школе (Турция, Испания, Сербия).

3. *По результатам стандартизированных тестов на способности* абитуриентов к обучению, в сочетании с другими факторами, характеризующими готовность выпускника к продолжению обучения (США, Швеция).

4. *На основании нескольких видов экзаменов* – национальных выпускных и вступительных, проводимых учебными учреждениями (Япония, Бразилия, Финляндия), вместе с дополнительными экзаменами, проводимыми правительством, учреждением образования, независимой организацией (Израиль, Индия).

5. *Базируясь на факторах, которые характеризуют абитуриента, без экзаменов:* академической

успеваемости в школе (Норвегия, Канада), личных достижений и академических способностей (отдельные университеты США).

Модели имеют преимущества и ограничения, связанные с факторами функционирования национальных систем образования, образовательными традициями и социальными мотивами [4, 5].

С 2005 года в Украине введено внешнее независимое оценивание (ВНО) знаний выпускников, которое с 2010 года выполняло функции государственной аттестации и отбора абитуриентов при поступлении в ВУЗ. Анализ опыта поступления в ВУЗы Украины показывает, что с введением ВНО произошли положительные сдвиги в обеспечении равного доступа граждан к высшему образованию на принципах объективности и справедливости [6, 7].

На сегодняшний день наиболее часто для проверки и оценивания уровня знаний используется тестирование. Тестирование отражает степень знания изученного материала и позволяет указать на пробелы в знаниях. Далее проанализируем существующие методы тестирования в мировой практике.

1. Методы тестирования и системы оценивания результатов тестов

1.1. Оценка результатов в тесте SAT

SAT (Scholastic Assessment Test) – стандартизованный тест для приема в ВУЗы США. SAT оценки основаны на процентильном отношении оценки студента к оценкам других студентов, прошедшими одинаковый тест.

Оценки рассчитываются путем определения сырого балла по каждому из трех разделов, где правильный ответ на вопросы с множеством вариантов добавляет 1 пункт к сырому баллу, а неправильный выбор ответа вычитает 0,25 от сырого балла.

Сырой балл по разделу конвертируется во взвешенный балл, представленный в шкале 200-800 – для этого раздела, что позволяет экспертам скорректировать незначительные отклонения. Одинаковый уровень знаний приводит одинаковой взвешенной оценке, которая основана на кривой нормального распределения, где наибольшее количество баллов в диапазоне от 400 до 600.

1.2. Оценка результатов в тесте АСТ

АСТ (American College Testing) – стандартизированный тест для поступления в старшие классы или поступления в колледжи в США.

АСТ состоит из тестов с четырьмя вариантами ответа. Предметные тесты: английский, математика, чтение и научное мышление, где каждый оценивается по шкале от 1 до 36 баллов. Английский, математика и тесты на чтение имеют и дополнительные результаты от 1 до 18 (они никак не связаны с общей оценкой за предметный тест). «Общий результат» – это сумма всех четырёх тестов.

Дополнительно сдается письменный тест, оцениваемый по шкале от 2 до 12 баллов, объединённый результат по письменному тесту и тесту по английскому языку варьируется от 1 до 36. Возможно до четырёх комментариев к эссе от проверяющих. Эссе на общий результат не влияет. За верный ответ на «временный счет» начисляется 1 балл, за неверные ответы баллы не снимаются.

1.3. Оценка результатов в тесте GRE

GRE (Graduate Record Examinations) – тест для поступления в аспирантуру, магистратуру или иной последипломный курс в вузы США и ряда других стран.

Официальная оценка GRE состоит из трех частей: вербальная взвешенная оценка (по шкале от 200 до 800); количественная взвешенная оценка (по шкале от 200 до 800); оценка за аналитическое письмо (по шкале от 0 до 6). Первые два раздела состоят из тестовых вопросов, а третий из двух эссе.

В компьютерной версии GRE в общей сложности 58 ответов: 30 в словесном разделе и 28 в количественных разделах. Для вычисления взвешенной оценки используется алгоритм, учитывающий факторы: количество пройденных вопросов за установленное время с правильными ответами, статистические характеристики (уровень сложности). Компьютер предоставляет вопрос средней сложности по разделу и в случае правильного ответа, следующий вопрос будет сложнее, оценка корректируется вверх и наоборот. Взвешенный балл пересчитывается с учетом прогресса тестируемого. Вероятность расчета точной оценки повышается с увеличением количества вопросов теста.

1.4. Оценка результатов в тесте ВНО Украины

Определение результатов внешнего тестирования осуществляется в два этапа. После окончания

проверки бланков ответов определяется тестовый балл тестируемого как арифметическая сумма всех набранных баллов за каждое выполненное задание теста. Максимальное количество тестовых баллов за правильно выполненные задания, которые можно получить, приведено в спецификациях тестов [8, 9].

Далее на основе тестового балла с помощью компьютерных программных средств определяется оценка результатов участника внешнего тестирования по 100-бальной шкале (числа интервала от 100 до 200 баллов).

Оценка участника внешнего тестирования по шкале 100-200 баллов – рейтинговая. Участник тестирования, набравший максимальное количество тестовых баллов по предмету, будет иметь по этой шкале наивысшую оценку. Такая оценка наиболее приближена к эталонной шкале со средним значением результатов по каждому предмету. Она стандартизирована и распределение участников по этой шкале является приблизительно одинаковым, независимо от предмета тестирования.

Перевод тестового балла участника ЗНО в оценку по шкале 100-200 баллов осуществляется на основе эквипроцентильного превращения тестовых баллов в баллы по эталонной шкале со средним значением 150 баллов и среднеквадратическим отклонением 20 баллов.

Перевод тестового балла участника тестирования $г$ в оценку по шкале 100-200 баллов происходит с помощью таблицы $T(g)$, в основу которой положен метод эквипроцентилей. Таблица $T(g)$ строится по следующему алгоритму:

- тестовому баллу 0 должно отвечать 100 баллов по шкале 100-200 баллов, из дальнейшего рассмотрения исключаются работы участников с нулевыми тестовыми баллами: $T(0)=100$;

- каждому тестовому баллу $г$, отличному от 0, должна отвечать оценка s по шкале 100-200 баллов: $T(g)=s$;

- оценка s исчисляется как ближайшее число с одним знаком после десятичной запятой, которое не меньше числа x , для которого процентный ранг эталонного распределения равен процентному рангу значения $г$ тестового балла для распределения оценок участников внешнего тестирования.

Таблицы перевода тестовых баллов в оценки по шкале 100-200 баллов для каждого предмета утверждаются директором и размещаются на сайте Украинского центра оценивания качества образования (УЦОКО) через 3 дня после завершения проверки работ по каждому предмету.

Невозможность прямого переноса опыта проведения тестирования в Европе и США связано с традициями образования и интерпретации результатов тестирования. А именно потому, что в Украине тенденция критериального оценивания, а в Европе и США в основном рейтинговое оценивание.

2. Анализ существующих компьютерных систем тестирования

В форме компьютерных программ тесты впервые стали применяться в рамках первых систем автоматизированного обучения, разрабатываемых в соответствии с концепцией программируемого обучения. Тестовая форма оценки знаний учащихся позволяла сформировать обратную связь и значительно сократить затраты времени педагога на проведение теста и на первичный анализ его результатов [10, 11]. В самых первых тестирующих программах предполагался вывод вариантов ответов, ввод номеров ответов по выбору пользователя и накопление баллов.

Более универсальные системы позволяли хранение тестов, вопросов и вариантов ответов в отдельных файлах данных, позволяющих корректировать и дополнять тесты и данные.

Появилось первое разделение: на программы для создания тестов и программы для прохождения тестов. Важная особенность таких систем, по сравнению с примитивными тестирующими программами, – наличие более сложных и профессиональных интерфейсов, средств анализа ответов, что расширяет спектр используемых тестовых заданий, дает возможность проводить отложенное тестирование, автоматически создавать эквивалентные варианты тестов и пр.

Инструментальные средства баз данных позволяют использовать в заданиях различные виды невербального материала (не только текст) и имеют в своем составе два основных модуля: «модуль ученика» и «модуль учителя». Системы такого рода достаточно сложны и предусматривают сбор данных с большого количества рабочих мест, а к созданию сетевой версии систем тестирования перешли лишь немногие разработчики.

Технология компьютеризированного тестирования должна обладать следующими характеристиками: наличие интерактивной инструментальной среды; мультипредметное применение; полное отражение конструируемой модели предметной области; возможность выбора алгоритма тестирования; интегрируемость в различные образовательные технологии; профилируемость; масштабируемость; доступность; дружелюбность пользовательского интерфейса; ведение базы тестовых многоуровневых заданий; настраиваемое планирование и управление; достижение более высоких результатов и повышение мотивации.

В обучающих системах используют два принципа контроля знаний: оценка действий обучаемого и определение уровня знаний по знаниям о предметной области и правилам оценки действий обучаемого; стандартизированный контроль знаний по выборке специальных заданий и по ответам на них (тестовый).

В современном тестировании классические тесты интенсивно заменяются так называемыми «адаптивными тестами» [12].

Системы тестирования разрабатываются в форме серверных web-приложений, когда надо развернуть программную часть системы только на одном компьютере (web-сервере), а на рабочих местах потребуется только стандартный браузер (например, Internet Explorer из комплекта Windows). «Ученическая» и «учительская» части системы выполняются как замкнутые наборы web-страниц. Недостатком является то, что существуют некоторые ограничения на скорость работы и способ показа вопросов (ограничения формы). Достоинством является разработка путем использования языков создания приложений. Подразумеваются также меры, предупреждающие намеренное искажение результатов тестирования (большие банки заданий, ротация вопросов в тестах, исключение утечки информации о правильных ответах и пр.). Однако даже сложные системы тестирования не имеют в своем составе модулей статистической обработки результатов, за исключением примитивных суммирующих шкал, анализ выполняется внешними средствами, для таких расчетов существуют специально созданные прикладные пакеты.

Например, универсальный профессиональный статистический пакет SPSS 13 объединяет в себе функции ввода и управления данными, работу с графиками, таблицами, вывод информации в форме web-страниц и позволяет работать с данными, представленными в различных форматах, с метками значений, производить преобразование данных, расщепление и сортировку. В статистический аппарат включено несколько сотен процедур и удобные инструменты вывода результатов – трехмерные гистограммы и точечные диаграммы (плотности точек или «облака»). Есть модуль, позволяющий строить деревья классификаций, идентифицировать группы и модули для дисперсионного анализа и исследования регрессионной модели. Центр тестирования в России уже 20 лет исследует проведение централизованного тестирования на базе вузов. Результаты тестирования засчитываются как вступительные экзамены в вуз. Кроме этого, Центр тестирования занимается обработкой результатов и организационно-технологическим обеспечением проведения Единого государственного экзамена. Для шкалирования результатов используется современная теория тестов (IRT); для ЕГЭ используется модель Partial Credit, определяющая зависимость вероятности получения определенного балла за задание от уровня подготовленности ученика и параметров задания, она позволяет работать с дихотомическими и полиномическими заданиями. Модель Partial Credit превращается в однопараметрическую модель Раша, если в тесте только дихотомические задания. Особенность этой модели в

том, что участники тестирования, набравшие одинаковый первичный балл, получают в итоге одинаковый тестовый балл в 100-балльной шкале, т.е. первичные баллы являются достаточными статистиками.

Стандартизованная процедура проведения тестирования в Украине базируется на бланковой и компьютерной моделях, они несколько лет совершенствовались на больших выборках тестируемых. В вузах созданы программно-технические комплексы, где в форме бланкового тестирования проводятся экзамены, информация сканируется в компьютерную базу данных. Функции программных комплексов и автоматизированных систем: регистрация участников, сканирование и распознавание бланков, обработка результатов с заданным уровнем надеж-

ности и конфиденциальности, выставление оценок, распечатка итоговых ведомостей, проведение апелляции; статистический анализ; исследование качества тестовых заданий. Компьютерное тестирование значительно облегчает работу преподавателя при предъявлении и оценивании результатов выполнения тестов, но негативные последствия его не выявлены в полной мере. Выбор компьютерного формата экзамена должен быть обоснован, так как ставит обучаемых в неравные условия [13 – 15].

Компьютерное тестирование проводится в различных формах, различающихся по технологии объединения заданий в тест (рис. 1). Часть из них пока не получили специального названия в литературе по тестовой проблематике.



Рис. 1. Формы компьютерного тестирования

Первая форма, когда готовый тест, предназначенный для текущего контроля, вводится в специальную оболочку, функции которой могут различаться по степени полноты. При итоговом тестировании оболочка позволяет представлять задания на экране, оценивать результаты их выполнения, формировать матрицу результатов тестирования, обрабатывать ее и шкалировать первичные баллы испытуемых путем перевода в одну из стандартных шкал для выдачи каждому испытуемому тестового балла и протокола его оценок по заданиям теста.

Вторая форма компьютерного тестирования – автоматизированная генерация вариантов теста с помощью инструментальных средств. Варианты создаются перед или во время экзамена из банка калиброванных тестовых заданий с устойчивыми статистическими характеристиками. Калибровка достигается путем предварительной работы по формированию банка, параметры заданий которого получают на репрезентативной выборке учащихся, на протяжении 3 – 4 лет с помощью бланковых тестов. Содержательная валидность и параллельность вариантов обеспечиваются за счет определенного отбора заданий каждого варианта в соответствии со спецификацией теста.

Третья форма – компьютерное адаптивное тестирование – базируется на специальных адаптивных тестах. В их основе лежат соображения о том, что учащемуся не надо давать задания теста, которые он выпол-

нит правильно без затруднений или гарантированно не справится в силу высокой трудности.

Компьютерное тестирование имеет преимущества по сравнению с бланковым тестированием, они особенно заметны при массовых проверках, например при проведении национальных экзаменов типа ЕГЭ (Единый государственный экзамен). Размещение вариантов теста на компьютере экономит средства на печать и доставку бланковых тестов, можно повысить информационную безопасность и предотвратить расклевывание теста за счет высокой скорости передачи информации и кодировки электронных файлов.

Преимущества компьютерного тестирования проявляются в текущем контроле, при самоконтроле и самоподготовке учащихся; также можно незамедлительно выдать тестовый балл и принять неотложные меры по коррекции усвоения нового материала на основе анализа протоколов по результатам выполнения корректирующих и диагностических тестов.

Возможности педагогического контроля при компьютерном тестировании увеличиваются за счет расширения спектра измеряемых умений в инновационных типах тестовых заданий, использующих возможности компьютера при включении аудио- и видеофайлов, интерактивности, динамической постановки проблем с помощью мультимедийных средств и др. Недостатками компьютерного тестирования являются: негативные психологические реакции испытуемых на

компьютерное представление тестов; воздействие на результаты тех, кто проходит тестирование предшествующего опыта работы на компьютере; влияние интерфейса на результаты тестирования, влияние ограничений при компьютерном предъявлении теста на надежность его результатов.

Таким образом, компьютерное тестирование повышает возможности процесса контроля, сбора дополнительных данных о динамике прохождения теста для осуществления дифференциации пропущенных и не достигнутых заданий.

Заключение

Проведя анализ существующих методов использования тестирования в национальных образовательных системах, выявлены их достоинства и недостатки и показано, что наиболее актуальной и сложной является задача создания информационной технологии для организации и проведения тестирования высокой ответственности. В результате детального рассмотрения существующих принципов построения систем тестирования обоснована необходимость создания нового инструментария для построения системы тестирования.

Список литературы

1. Тайджман А. Мониторинг стандартів освіти [Текст]; за ред. А. Тайджмана і Т. Невілла Постлвейта. – Львів: Літопис, 2003. – 328 с.
2. Васильева Г.І. Інформаційно-комп'ютерні технології в освіті: шляхи реалізації [Текст] / Г.І. Васильева, Г.П. Досенко // Інформаційні технології в освіті: зб. наук. праць. – Херсон: Видавництво ХДУ, 2008. – Вип. 2. – С. 99-105.
3. Інформаційна підтримка інноваційних підходів в освіті [Електронний ресурс]. – Режим доступу до ресурсу: http://www.nbuv.gov.ua/portal/soc_gum/Pipo/2010_28-29/inform1.pdf, вільний.
4. Андреев А.А. Педагогика высшей школы. [Электронный ресурс] / А.А. Андреев. – Режим доступа к ресурсу: <http://www.gdenet.ru/bibl/education/communications/5.1.html>. – Загл. с экрана.
5. Седов А.О. Стан та перспективи розвитку інноваційних освітніх технологій в Україні [Текст] / А.О. Седов // Інформаційні технології в освіті: Збірник

наукових праць. – Херсон: Видавництво ХДУ, 2008. – Вип. 1. – С. 122-125.

6. Современная педагогическая интеграция, ее характеристики [Электронный ресурс]. – Режим доступа к ресурсу: http://jeducation.ru/1_2009/9.html, свободный. – Загл. с экрана.
7. Антюхов А.В. Проектирование системы управления качеством образования в вузе [Текст] / А.В. Антюхов // Инновации в образовании. – М.: Современ. гуманитар. ун-т, 2010. – № 9. – С. 34-43.
8. Вісник ТІМО. Тестування і моніторинг в освіті: журнал. – Х.: Факт, 2012. – № 1-2. – 97 с.
9. Вісник ТІМО. Тестування і моніторинг в освіті: журнал. – Х.: Факт, 2010. – № 4-5. – 96 с.
10. Информационные технологии обучения [Электронный ресурс]. – Режим доступа к ресурсу: <http://www.pedagogics-book.ru/articles/6-10.html>, свободный. – Загл. с экрана.
11. Типы обучения в процессе развития образовательной деятельности [Электронный ресурс]. – Режим доступа к ресурсу: <http://sprosp.com.ua/obuchenie/tipu-obucheniya-v-processe-razvitiya-obrazovatelnoj-deyatelnosti.html>, свободный. – Загл. с экрана.
12. Федорук П.І. Використання адаптивних тестів в інтелектуальних системах контролю знань [Текст] / П.І. Федорук // Штучний інтелект. – Донецьк: Наука і освіта, 2008. – № 3. – С. 380-387.
13. Сергеев В.В. Адаптивное тестирование в системах дистанционного обучения [Электронный ресурс] / В.В. Сергеев. – Режим доступа к ресурсу: <http://technomag.edu.ru/doc/65577.html>, свободный. – Загл. с экрана.
14. Информационные технологии в образовании [Электронный ресурс]. – Режим доступа к ресурсу: http://bigor.bmstu.ru/?cnt?doc=Default/050_iteduc.cou, свободный. – Загл. с экрана.
15. Зеневич С.Я. Математическое моделирование процесса обучения [Электронный ресурс] / А.М. Зеневич, С.Я. Жукович // Материалы конференции «Подготовка научных кадров высшей квалификации с целью обеспечения инновационного развития экономики». – Мн.: ГУ «БелИАС», 2006. – 146 с. – Режим доступа к ресурсу: <http://belisa.head.by/ru/izd/other/Kadr2006/kadr29.html>, свободный. – Загл. с экрана.

Поступила в редколлегию 31.08.2012

Рецензент: д-р техн. наук, проф. В.М. Вартамян, Национальный аэрокосмический университет им. Н.Е. Жуковского «ХАИ», Харьков.

АНАЛІЗ СИСТЕМ ТЕСТУВАННЯ В КОМП'ЮТЕРНИХ НАВЧАЛЬНИХ КОМПЛЕКСАХ

О.Г. Молчанова, О.І. Морозова, І.В. Чумаченко

У статті розглянуті існуючі системи тестування в комп'ютерних навчальних комплексах. Детально проаналізовано методи тестування і системи оцінки результатів тестів у світовій практиці. Розглянуто оцінки результатів у тестах SAT (Scholastic Assessment Test), ACT (American College Testing), GRE (Graduate Record Examinations) і ЗНО (Зовнішнє незалежне оцінювання). Крім цього, проведено огляд та аналіз існуючих комп'ютерних систем тестування, виявлено їх переваги та недоліки.

Ключові слова: системи тестування, комп'ютерні навчальні комплекси, оцінка результатів у тестуванні, моніторинг якості освіти, SAT (Scholastic Assessment Test), ACT (American College Testing), GRE (Graduate Record Examinations) ЗНО (Зовнішнє незалежне оцінювання).

ANALYSIS OF TESTING SYSTEMS IN COMPUTER TRAINING COMPLEXES

O.G. Molchanova, O.I. Morozova, I.V. Chumachenko

The existing systems of testing in computer training complexes are discussed in the article. Detailed analysis of test methods and evaluation system tests in the world are carried out. The evaluation of tests SAT (Scholastic Assessment Test), ACT (American College Testing), GRE (Graduate Record Examinations) and EIE (External independent evaluation) are considered. In addition, a review and analysis of existing computer testing systems are held and their strengths and weaknesses are identified.

Keywords: testing systems, computer training complexes, evaluation of the testing, monitoring the quality of education, SAT (Scholastic Assessment Test), ACT (American College Testing), GRE (Graduate Record Examinations), EIE (External independent evaluation).