

УДК 004:371.64: 681.3

К ТЕОРИИ И ПРАКТИКЕ ИСПОЛЬЗОВАНИЯ АДАПТИВНЫХ ТЕСТОВ**Сметанюк Л.В., Кравцов Г.М.
Херсонский государственный университет**

Представлены результаты моделирования тестов и процесса адаптивного тестирования в системе контроля знаний с использованием адаптивных тестов в системе дистанционного обучения «Херсонский виртуальный университет».

Ключевые слова: адаптивный тест, моделирование тестов, дистанционное обучение, мониторинг контроля знаний.

Введение

Одной из предпосылок вхождения Украины в единое Европейское образовательное пространство есть внедрение в систему высшего образования Европейской кредитно-трансферной и аккумулирующей системы (ECTS), которая функционирует на институциональном, региональном, национальном и Европейском уровнях, и является ключевым требованием Болонской декларации 1999 года.

Одним из важнейших стратегических задач на сегодняшнем этапе модернизации системы высшего образования Украины есть обеспечение качества подготовки специалистов на уровне международных требований. Отдельные высшие учебные заведения Украины имеют достаточно весомые наработки внедрения элементов ECTS, в частности модульно-рейтинговой системы оценивания знаний студентов. Улучшение существующей системы подготовки специалистов предусматривает решение нескольких задач, в том числе, повышение объективности оценивания знаний студентов, обоснование целесообразных подходов к системе оценивания знаний студентов в условиях внедрения кредитно-модульной системы [1].

В сфере образования идет поиск нового содержания и новых форм обучения, создаются новые образовательные технологии, расширяется использование личностно-ориентированных методов – таких, как дистанционное обучение, метод проектов и анализа ситуаций, адаптивное тестирование и рейтинговый контроль знаний. Все перечисленные формы обучения основаны на преимущественном использовании тестов.

В наше время практически все вузы оснащены компьютерной техникой, имеют свои локальные сети, доступ к сети Интернет. Это позволяет организовать как обучение, так и контроль усвоения знаний в виртуальном образовательном пространстве. Сеть Интернет содержит огромное множество различных учебных материалов, программ и систем, предназначенных для учебных целей, при этом порядка половины их числа составляют тестирующие программы. Проведение контроля усвоения учебного материала с использованием различных тестирующих программ является одной из ветвей информатизации образования и получило широкое распространение. Компьютерное тестирование студентов используется при проведении текущего, рубежного и итогового контроля знаний, при проверке остаточных знаний, при выставлении экзаменационных оценок по соответствующим дисциплинам [2].

Современная система дистанционного обучения (СДО) должна обеспечить создание, сохранение и использование учебных материалов в стандартизированном формате данных. В настоящее время принят стандарт, разработанный организацией IMS Global Learning Consortium, Inc. (IMS) [3]. Стандарт IMS содержит открытые спецификации поддержки учебной деятельности в рамках распределенного обучения, такие как создание, размещение и использование учебных материалов, спостереження за прогресом учня, складання звітів про успішність учня та обмін записами про учнів між адміністративними системами. Среди этих

требований важливое місце займає опрацювання єдиного стандарту (формату) збереження навчаючих інформаційних ресурсів.

СДО «Херсонский виртуальный университет» разработана с поддержкой стандарта IMS [4]. В частности, система тестирования поддерживает спецификации стандарта IMS QTI версии 2.1.

Адаптивное тестирование

Наряду с традиционными методами обучения и контроля знаний тестирование быстро становится необходимой частью учебного процесса. Это методическое направление в педагогике вновь возрождается в нашей стране. Существует множество видов и методов проведения тестирования. Особо выделим метод адаптивного тестирования.

Адаптивное тестирование – это широкий класс методик тестирования, предусматривающих изменение последовательности предъявления заданий в самом процессе тестирования с учетом ответов испытуемого на уже предъявленные задания.

По форме проведения в основном выделяют три вида адаптивного тестирования [5]:

пирамидальное – при отсутствии предварительных оценок всем испытуемым дается задание средней трудности и уже затем, в зависимости от ответа, каждому испытуемому дается задание легче или труднее;

flexilevel – испытуемый сам выбирает желаемый уровень трудности заданий [6];

stradaptive (от англ. stratified adaptive) – из банка заданий, разделенного по уровням, при правильном ответе следующее задание берется из верхнего уровня, при неправильном – из нижнего [7].

Можно выделить два подхода к созданию адаптивных тестов. При первом подходе принятие решения об изменении порядка предъявления тестовых заданий производится на каждом шаге тестирования (постоянная адаптация). Во втором подходе принятие решения об изменении порядка следования заданий осуществляется после анализа результатов отчетов испытуемого на специальный блок заданий (блочная адаптация).

С точки зрения порядка прохождения тестовых заданий существуют два подхода к созданию адаптивных тестов. Во-первых, существуют адаптивные тесты с *постоянной адаптацией* (детерминировано ветвящейся), когда принятие решения об изменении порядка следования тестовых вопросов производится на каждом шаге тестирования (рис. 1).



Рис. 1. Схема следования заданий при постоянной адаптации.
(О – ошибочный ответ, В – верный ответ).

Во-вторых, существуют адаптивные тесты с *блочной адаптацией* (варьирующей ветвящейся стратегией), когда принятие решения об изменении порядка следования заданий осуществляется после анализа результатов обработки некоторого специального блока заданий (рис. 2).

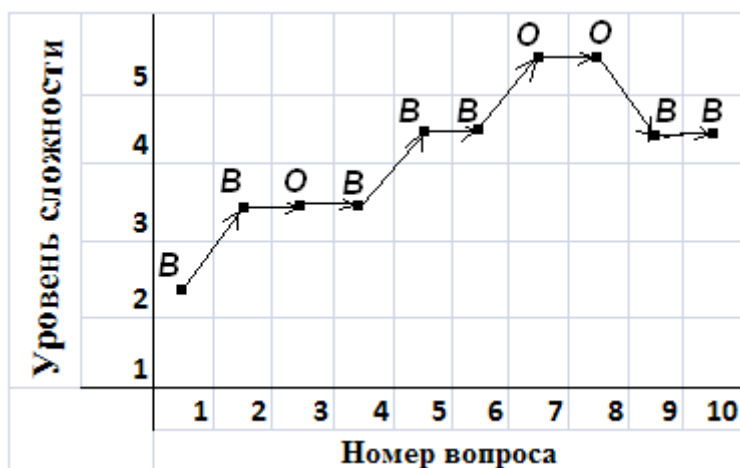


Рис. 2. Схема следования заданий при блочной адаптации (O – ошибочный ответ, B – верный ответ).

Время тестирования

Так же одним из ключевых вопросов при компьютерном адаптивном тестировании является вопрос о том, когда же можно прекращать тестирование?

При создании тестов учебных достижений время является одним из фундаментальных по значимости факторов, который определяет качество всего инструментария и качество получаемых в процессе тестирования результатов.

Время нередко называется В.С. Аванесовым в качестве системообразующего фактора при разработке и использовании тестов. Действительно, одно из соображений, положенных в основу создания тестов, – иметь инструмент быстрого и относительно точного оценивания больших контингентов испытуемых. Требование экономии времени становится естественным в массовых процессах, каковым и является образование.

Каждый тест должен иметь оптимальное время тестирования. Уменьшение или превышение оптимального времени снижает качественные показатели теста [8, 9].

Создатели тестов стремятся включить в тест как можно больше тестовых заданий. Такое положение диктуется двумя обстоятельствами – чем больше количество заданий, тем более надежным будет создаваемый тест и тем больший объем информации мы можем получить. С другой стороны, чем больше количество заданий, тем более продолжительной становится процедура тестирования.

Между результатами, которые может показать испытуемый и продолжительностью тестирования существует достаточно простая связь. Время тестирования не может быть чрезмерно большим. Простое увеличение времени тестирования приводит к утомлению испытуемых, что в свою очередь снижает результаты тестирования.

Таким образом, необходимо найти оптимальное время, которое отражало бы баланс между временем выполнения теста (количеством заданий) и утомлением испытуемых.

Существуют различные критерии для остановки прохождения теста [5]:

- исчерпаны все вопросы в банке тестовых заданий;
- достигнут конец теста;
- уровень знаний оценен с достаточной точностью;
- уровень знаний расположен на достаточно далеком расстоянии от значения критерия прохождения теста;
- испытуемый показывает свою несостоятельность при ответах на вопросы теста.

“Таким образом, адаптивный тест представляет собой вариант автоматизированной системы тестирования, в которой заранее известны параметры трудности и дифференцирующей способности каждого задания. Эта система создана в виде компьютерного банка заданий, упорядоченных в соответствии с интересующими характеристиками заданий. Самая главная характеристика заданий адаптивного теста – это

уровень их трудности, полученный опытным путем, что означает: прежде чем попасть в банк, каждое задание проходит эмпирическую апробацию на достаточно большом числе типичных учащихся интересующего контингента. Слова "интересующего контингента" призвано представлять здесь смысл известного в науке понятия более строгого понятия "генеральная совокупность" [9].

Об использовании адаптивного тестирования в ХГУ

Преподавателями кафедры информатики Херсонского государственного университета был проведен эксперимент. Из 344 студентов случайным образом были сформированы две группы по 60 человек и проведено исследование результатов тестирования единого потока студентов, проведенного по предмету «Новые информационные технологии». Тестирование проводилось в двух компьютерных аудиториях (по 12 компьютеров в классе).

Первой группе был предложен тест из вопросов сгенерированных случайным образом из 500 равносильных заданий по 40 вопросов в каждом с ограничением времени в 20 мин. Во второй проводилось адаптивный тест с блочной организацией.

Для создания тестовых заданий и проведения эксперимента использовалась система дистанционного обучения "Херсонский виртуальный университет", DB Version: 0,5 (01.11.2009), ASM Version: 1.0.1.24709.

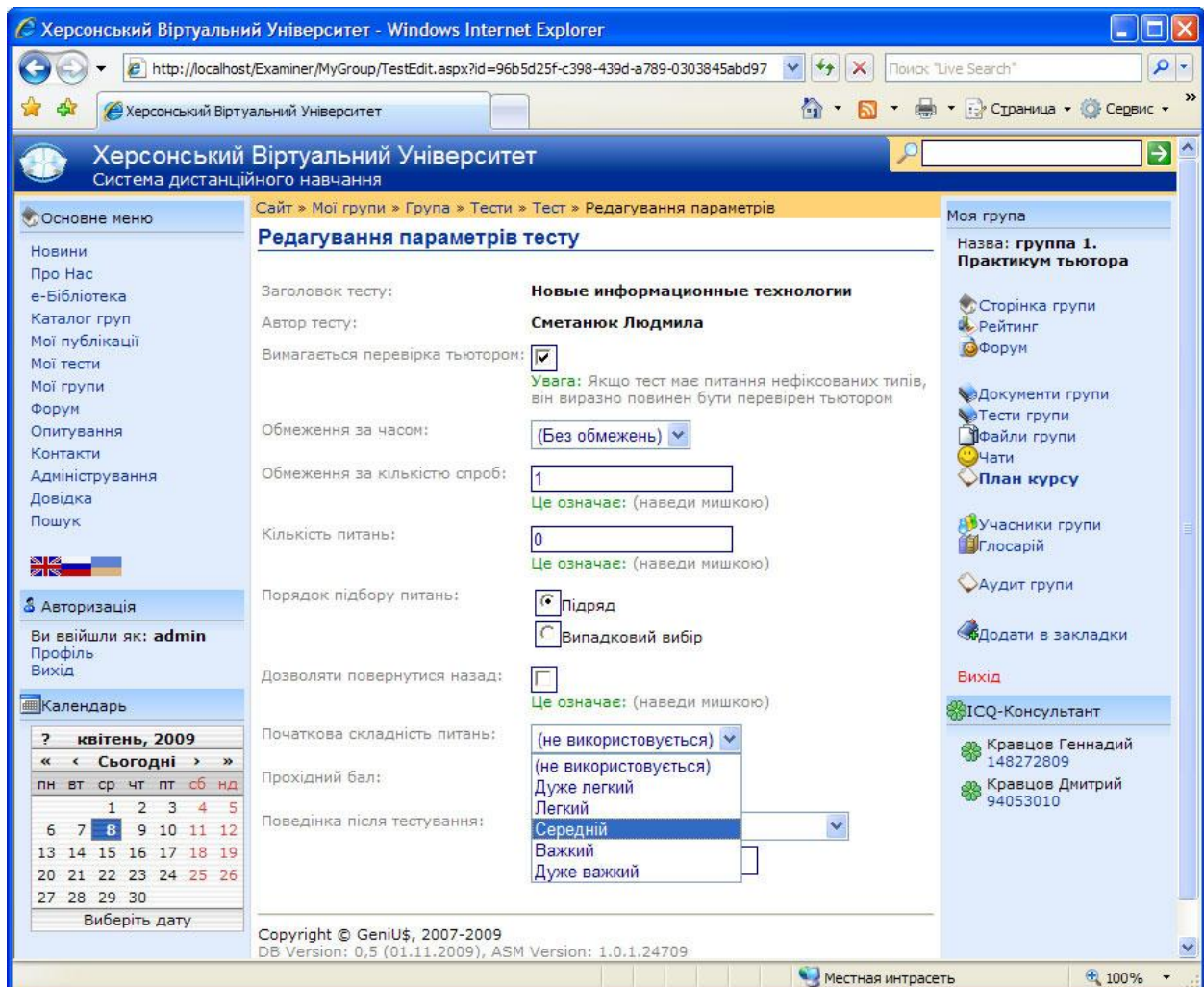


Рис. 3. Окно редактирования параметров теста.

Система тестирования данного сайта поддерживает несколько типов вопросов [10]:

- Выбор одного варианта из многих.
- Выбор некоторых из многих.
- Ввод текста.

- Ассоциативность.
- Упорядочивание.
- Сопоставление.
- Текст в контексте.
- Выбор в контексте.
- Множественный выбор в контексте.
- Ниспадающий список в контексте.
- Выбор "горячих точек" на изображении.
- Упорядочивание точек на изображении.
- Flash-объект

и позволяет задавать один из пяти начальных уровней сложности создаваемого вопроса: самый легкий, легкий, средний, сложный, самый сложный (рис. 3).

Ответ тестируемого обрабатывается в модуле «Обработка ответа» (Response Processing). Оценивание ответа на вопрос теста в модуле может происходить двумя различными путями: 1) дифференцированная оценка по всему вопросу и 2) накопление оценки по вариантам ответа.

В СДО «Херсонский виртуальный университет» реализованы 14 типов вопросов [11], которые относятся к группе так называемых «Простых элементов» (Simple Items) спецификации IMS QTI и полностью их покрывают. В спецификации QTI имеется описание типов вопросов, которые относятся к группе «Адаптивных элементов» (Adaptive Items) и «Объектных элементов» (Object Items). Согласно спецификации IMS особенности данных типов заключаются в многоэтапности прохождения тестового вопроса при выполнении задания. Имеет место обратная связь с тестируемым, которая определяет поведение тестовой системы в зависимости от их ответов на каждом этапе, и формирует таким образом вариативность ответа. В спецификациях адаптивных и объектных тестов могут быть дополнительные параметры, которые не описаны в стандарте IMS. Примером реализации модуля объектного типа может служить интерактивная Flash-анимация, в которой запрограммирована определенная функциональность:

- инициализация модуля с некоторыми входными параметрами,
- реализация интерактивной игровой ситуации, в которой принимает участие тестируемый,
- вывод выходных данных, как результат действия тестируемого.

Результат выполнения задания объектного (адаптивного) типа может быть определен в модуле прохождения теста с учетом значения максимальной оценки за правильное прохождение теста и использован при автоматическом (программном) оценивании. Как альтернатива, оценка может быть определена тьютором при проверке.

В процессе прохождения адаптивного теста, тестируемые с высоким уровнем подготовки получали большее число сложных вопросов, с низким уровнем подготовки – легких. Доли правильных ответов у некоторых совпадали, но с учетом уровня сложности вопросов каждый набирал различное количество баллов, что увеличивает уровень достоверности результатов контроля знаний учащихся.

Для каждого испытуемого набор заданий являлся уникальным, в результате чего минимизируется возможность подсказок, выучивания правильных ответов и т.д. При повторном прохождении теста испытуемые решали новые задания, что уменьшило влияние эффекта тренированности. Поскольку банк заданий содержит 500 вопросов (по 100 вопросов на шкалу), относящихся к различным уровням сложности и к тому же при создании теста было использовано 8 типов вопросов, что значительно снизило влияние эффекта угадывания на результаты тестирования.

В результате эксперимента было установлено, что при адаптивном контроле знаний было сэкономлено 12,5% времени отводимого на тест по сравнению с первой группой, а также оценивание знаний стало более дифференцированным, что отображено на рис. 4.

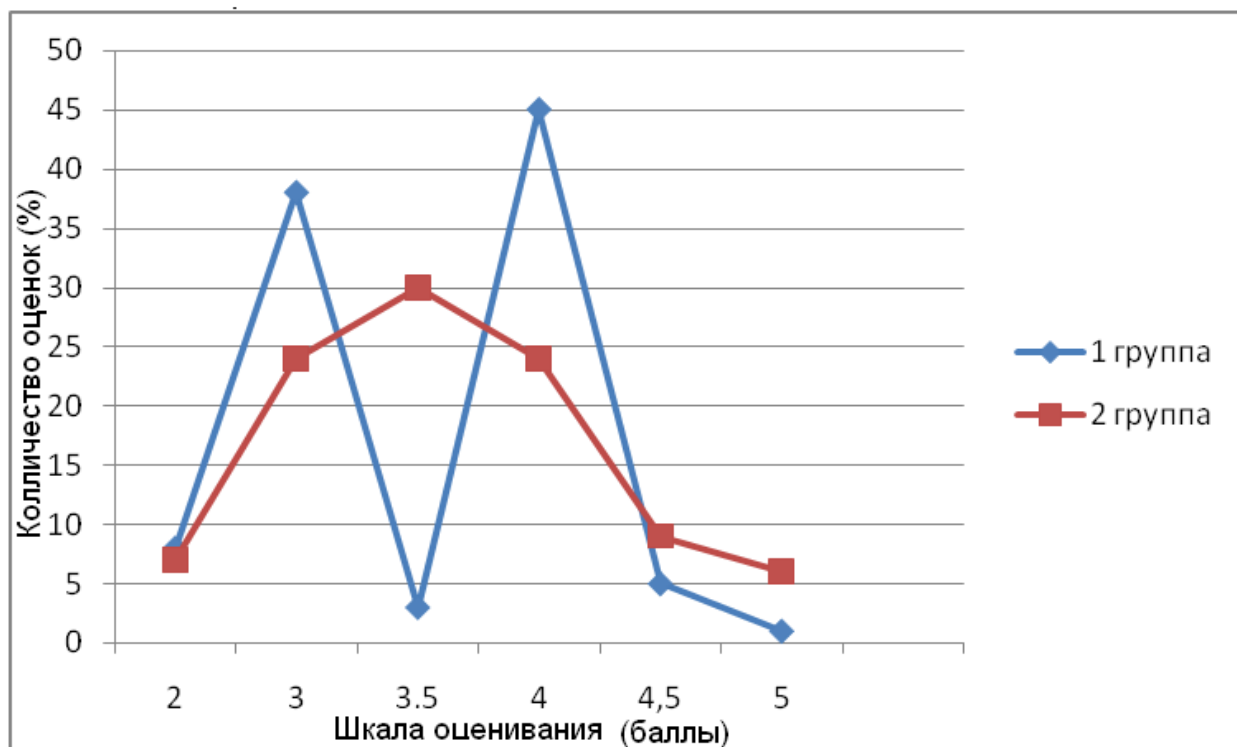


Рис. 4. График достоверности отображения знаний.

Статистика и анализ результатов прохождения теста

Несмотря на растущую популярность систем тестирования, существует несогласованность в работе по созданию тестовых материалов. Долгое время этому способствовало отсутствие теоретической основы (почти вся содержательная литература по тестированию была иностранной), которая позволяла бы не только написание авторских тестов, а построение тестов с единых методических позиций.

Уровень знаний обучаемых является латентным параметром (то есть недоступным для непосредственного измерения), для получения его значения необходимо использовать современные научные методы составления тестов и статистический анализ результатов тестирования. Существуют объективные и субъективные причины, по которым авторы тестов в полной мере не проводят статистическую обработку тестов (с помощью принятых в мире пакетов компьютерных программ). *Главным препятствием на этом пути является слабое владение авторами-предметниками методами и средствами анализа результатов тестирования.*

Современное понимание тестов и тестирования можно отнести к трем уровням [8]:

Первый – «бытовой» уровень. Здесь тест понимается как набор вопросов с вариантами ответов, который стоит в одном ряду с кроссвордами, головоломками и служит в большей степени для развлечения и удовлетворения познавательных интересов. Педагоги с таким пониманием тестирования считают тестирование очень ненадежным, ограниченным, а создание тестов простейшим делом.

Второй уровень понимания тестирования можно назвать «словарным». В этом понимании выделяются основные составляющие понятия тестирования. При этом не учитываются особенности процедуры создания, использования, анализа, специфичные для той или иной сферы применения. Для этого понимания характерны разночтения и противоречия в понятиях и определениях. Современное состояние развития тестологии находится именно на этом уровне. Многие понятия до конца не определены, многие авторы трактуют по-разному одни и те же понятия и, в свою очередь, одно явление может иметь несколько названий.

Третий уровень понимания может быть назван научным. Он наиболее точен, учитывает особенности тестов и отражает требования к тестам, которые появляются в процессе развития и научного обоснования тестирования.

Широкие возможности современной тестовой технологии в решении ряда важнейших для высшей школы задач в настоящее время почти не задействованы, а там, где их пытаются задействовать, делают это самостоятельно и часто неудовлетворительно. Отчасти это объясняется тем, что ввиду отсутствия требуемой литературы профессорско-преподавательский состав слабо информирован в вопросах теории и методики тестового контроля, а также из-за широко распространенных заблуждений о кажущейся легкости создания теста. Поэтому современная высшая школа Украины "засорена" самостоятельными кустарными псевдотестами и тестовыми заготовками, которые их авторы выдают за тесты.

Поэтому в помощь преподавателю-предметнику в системе тестирования Херсонского виртуального университета создан модуль статистики и анализа результатов прохождения теста. В нём обработка результатов тестирования производится посредством формирования матрицы тестовых результатов. Система автоматически рассчитывает качественные параметры теста – валидность и надежность результатов измерений, выдаются рекомендации по корректировке теста с предварительно установленным видом интерпретации тестовых результатов, проводится оценка итоговых результатов в рамках выбранной педагогической шкалы; результаты изображаются в удобной графической форме, отчет выводится на печать. Предусматриваются следующие виды интерпретации тестовых результатов: содержательно-ориентированная интерпретация, критериально-ориентированная интерпретация и нормативно-ориентированная интерпретация тестовых результатов. Каждый вид тестов сопровождается самостоятельной формой отчета, которая учитывает специфические особенности применяемого вида интерпретации.

Следует обратить внимание на циклический характер нескольких узлов схемы, среди которых центральным является база данных статистической информации о результатах тестирования. На этапе тестового эксперимента статистические данные в совокупности с предварительными, а затем последующими экспертными оценками позволяют "отбраковать" некачественные тестовые задания, неудачные тесты, определить сложность и другие качества тестовых заданий и тестов в целом. Многократный проход по циклическому пути с привлечением большого числа тестируемых (для увеличения статистики) и большого числа экспертов с их методиками и алгоритмами диагностики и измерения знаний позволяет совершенствовать измерительную шкалу и повышать точность диагностики знаний.

В дальнейшем, в рабочем режиме тестирования, система продолжает накапливать статистику по всем пользователям, совершенствовать и корректировать измерительную шкалу.

Выводы

С появлением адаптивных тестов теория тестирования вышла на новый уровень образования и получила более широкое применение. Был решён ряд вопросов по повышению качества контроля знаний учащихся, которые не могли быть решены при традиционном тестировании.

На примере использования системы дистанционного обучения «Херсонский виртуальный университет» показано, что использование адаптивного тестирования повышает эффективность тестовых измерений, точность оценок испытуемых, дает более объективную оценку знаний, умений и навыков обучаемых, а также позволяет экономить время и, соответственно, стоимость проведения тестирования.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Рішення Колегії МОНУ від 24 квітня 2003 року № 5/5-4. Про проведення педагогічного експерименту щодо запровадження кредитно-модульної системи організації навчального процесу у вищих навчальних закладах III-IV рівнів акредитації. –

2. http://www.kspu.kr.ua/index.php?option=com_content&task=blogcategory&id=11&Itemid=88888959.
3. Сметанюк Л.В. Методические и организационные проблемы использования программных средств учебного назначения в высших учебных заведениях при проведении контроля знаний студентов в виде тестирования. Інформаційні технології в освіті: Збірник наукових праць. Випуск 1. – Херсон: Видавництво ХДУ, 2008. – С.141 – 145.
4. Стандарт СДО IMS. – www.imsglobal.org.
5. Система дистанционного обучения “Херсонский виртуальный университет”. – <http://dls.kherson.ua/dls>.
6. Н.В. Белоус, С.А. Пархоменко. Компьютерное адаптивное тестирование. Проблемы высшей школы. Вестник ХГТУ №2(18), 2003. – С.421 – 423.
7. DeAyala, R. J.; Koch, William R. A Computerized Implementation of a Flexilevel Test and Its Comparison with a Bayesian Computerized Adaptive Test, 1986. – www.eric.ed.gov/ERICWebPortal/recordDetail?accno=ED269437.
8. Vale, C. David; Weiss, David J. A Study of Computer-Administered Stradaptive Ability Testing. Research Report 75-4, 1975. – www.eric.ed.gov/ERICWebPortal/recordDetail?accno=ED118602.
9. Майоров А. Н. Теория и практика создания тестов для системы образования. (Как выбирать, создавать и использовать тесты для целей образования). – М., «Интеллект-центр», 2001. – 296 с.
10. Аванесов В.С. Теория и методика педагогических измерений (Материалы публикаций в открытых источниках и Интернет) Подготовлено ЦТ и МКО УГТУ-УПИ, 2005. – <http://testolog.narod.ru/Theory41.html>.
11. Кравцов Г.М., Кравцов Д.Г., Козловский Е.О. Система дистанционного тестирования на основе стандарта IMS / «Information Technologies in Education for all». Киев. – 2006. – С.283 – 292.
12. Кравцов Г.М., Кравцов Д.Г. Адаптивные и объектные тесты в модели контроля знаний по стандарту IMS / УСиМ № 1, 2008. – С.42 – 48.