

КОМПЬЮТЕРНОЕ ТЕСТИРОВАНИЕ КАК МЕХАНИЗМ ОЦЕНКИ КАЧЕСТВА ОБРАЗОВАНИЯ

А.В.Андриенко

// Тестирование в сфере образования: проблемы и перспективы развития: материалы II Всероссийской научно-практической конференции 19-20 мая 2009 г., Красноярск: СибГТУ, 2009. – С.267-277.

Повышение качества обучения, качества подготовки специалистов, было и является одной из основных задач любой образовательной системы. О значимости проблемы качества образования свидетельствуют решения 30-й сессии Генеральной конференции ЮНЕСКО, в которых представлены основные принципы формирования образовательной стратегии – доступность, качество, мобильность. Им соответствует Концепция модернизации Российского образования на период до 2010 года, выдвигающая приоритетным направлением – обеспечение качества образования.

Однако в существующей философской и психолого-педагогической литературе нет общепринятого определения понятию «качество образования», «качество обучения», «качество образовательного процесса», «качество подготовки специалистов». Тем не менее, многие преподаватели вузов справедливо считают, что повышение качества обучения и качества выпускаемых вузами специалистов возможно за счет активизации познавательной деятельности учащихся, их активной самостоятельной работы, а также посредством совершенствования методики преподавания различных дисциплин, привлечения школьников и студентов к исследовательской работе и т.д. При этом обычно, качество обучения и подготовки специалистов оценивается по приобретенным ими в процессе обучения знаниям, сформированным умениям и навыкам.

Большая часть вузов России придерживается установки соответствия знаний выпускника вуза образовательным стандартам. В этой позиции два недостатка. Первый – ложное представление, что качества можно достичь путем инспекции. Неизбежны столкновения интересов между теми, кто обучает студентов, и теми, кто инспектирует качество подготовки. Второй – пренебрежение потребностями образовательного рынка (стандарты не успевают за изменениями требований покупателей).

Малая часть вузов ориентируется на концепцию «соответствие применению», т.е. на удовлетворение нужд потребителя кадров, а не только на выполнение требований стандарта. Трудности этих вузов достаточно очевидны – наш рынок еще не осознает четко свои потребности. Поэтому эти вузы постепенно стали обращать внимание на концепцию «соответствие стоимости», стараясь совмещать высокое качество и низкую стоимость подготовки.

Лишь незначительная часть вузов уверенно ориентируется на концепцию «соответствие скрытым потребностям», т.е. пытается одновременно снижать затраты на подготовку специалистов и удовлетворять нужды потребителя кадров прежде, чем он их осознает. Для образовательного рынка России эта концепция очень перспективна, так как многие потребители кадров выдвигают слишком примитивные требования к качеству специалистов и им надо «открывать глаза» на новые грани качества высшего образования.

Если же в целом анализировать систему высшего образования России, то можно утверждать, что данная система только начинает ориентироваться на переход к более современным концепциям качества. Стремление вузов к удовлетворению Государственных образовательных стандартов путем «выбраковки» приводит к завышенным затратам на образование (посмотрите разницу между планом приема и планом выпуска студентов) [8].

Педагогическое тестирование как один из методов контроля усвоения обучаемыми знаний, умений и навыков обладает важными преимуществами перед традиционными методами контроля знаний. Вот некоторые из них:

Более высокая, чем в традиционных методах, объективность контроля. Обычно на оценку, получаемую студентами, влияют, помимо уровня его учебных достижений, многие другие факторы: личность преподавателя и самого студента, их взаимоотношения, строгость или, наоборот, либеральность преподавателя и т.п. В оценке, выставляемой на основе традиционных методов контроля, оказывается существенным субъективный компонент. Это подтверждается тем известным фактом, что за один и тот же ответ разные преподаватели могут поставить разные отметки. В тесте такое влияние субъективных факторов исключается. Если тест достаточно качественный, то получаемая оценка может рассматриваться как объективная.

Оценка, получаемая с помощью теста, более дифференцирована. В традиционных методах контроля пользуются четырехбалльной шкалой («отлично», «хорошо», «удовлетворительно», «неудовлетворительно»). Результаты тестирования, благодаря особой организации тестов, могут быть представлены, если необходимо, в более дифференцированных шкалах, содержащих больше градаций оценки. При этом обеспечивается более высокая точность измерений учебных достижений.

Тестирование обладает заметно более высокой эффективностью, чем традиционные методы контроля. Тесты можно одновременно проводить на больших группах студентов. Обработка результатов для получения окончательных оценок проводится легче, быстрее, чем, скажем, проверка контрольных работ. Особенно велика экономия времени, когда в форме тестирования проводятся экзамены.

Конечно, у тестирования как метода контроля есть и свои ограничения. Легче всего с помощью тестов проверять овладение просто организованным учебным материалом. Проверка глубинного понимания предмета, овладения стилем мышления, свойственным изучаемой дисциплине, с помощью тестов затруднена, хотя в принципе возможна. Отсутствие непосредственного контакта со студентом, с одной стороны, делает контроль более объективным, но, с другой стороны, повышает вероятность влияния на результат других случайных факторов. Например, невозможно проконтролировать случайные ошибки студента, вызванные невниманием или неправильным пониманием задания. Существуют довольно изощренные методы, позволяющие в значительной степени преодолеть эти недостатки.

Ограничения тестирования как метода контроля необходимо хорошо осознавать, чтобы правильно применять тесты. В ряде случаев наилучший эффект дает сочетание тестов с традиционными методами контроля. Достоинство тестов заключается еще в том, что можно оценить степень вероятности влияния на результат случайных факторов и, следовательно, знать, насколько точен полученный результат.

В нашу вузовскую жизнь на равных правах с системой менеджмента качества, Болонским процессом, стандартами третьего поколения, компьютерными тестовыми технологиями вошел Федеральный Интернет-экзамен в сфере профессионального образования (ФЭПО), разработанный Росаккредитацией. Аббревиатура ФЭПО знакома не только преподавателям вузов, но и преподавателям средне специальных учебных учреждений.

Европейские стандарты и руководящие принципы Болонского процесса (г. Берген, 19-20 мая 2005г.) устанавливают необходимость поставщикам высшего образования разрабатывать процессы, с помощью которых вузы смогли бы продемонстрировать свою отчетность, ...учебные заведения должны демонстрировать качество своей деятельности, как в стране, так и на международной арене.

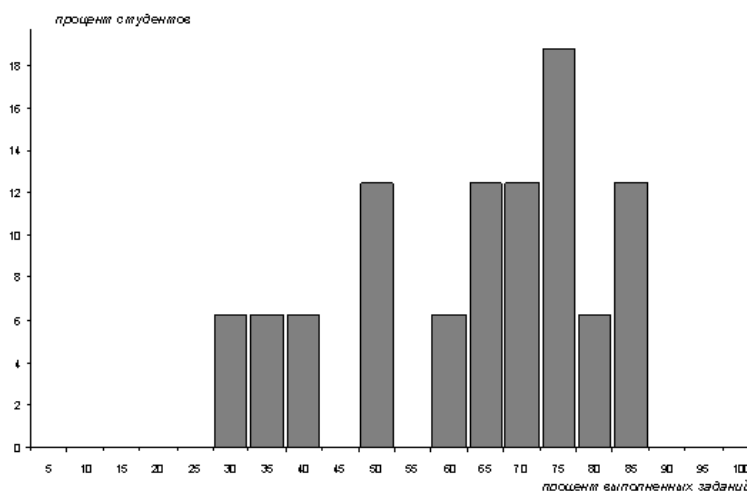
На заседании аккредитационной коллегии (протокол № 1-2008/АК от 07.02.2008) решили «Установить, что результаты Интернет-экзамена могут быть признаны в качестве оценки усвоения студентами программного материала при экспертизе соответствия содержания и качества подготовки обучающихся и выпускников требованиям государственного образовательного стандарта...».

ФЭПО позволяет проводить мониторинг оценки уровня подготовки студентов на соответствие требованиям ГОС в системе высшего и среднего профессионального образования по стране в целом в достаточно больших масштабах. С учетом приобретенного опыта Рособрнадзор может более четко определить условия учета результатов ФЭПО при государственной аккредитации вследствие существенно возрастающих объемов тестирования и выявления тенденций в полученных результатах. Проведение ФЭПО служит своеобразной экспериментальной площадкой для формирования общественного мнения о согласованном понимании требований ГОС, о содержании и уровне аттестационных педагогических измерительных материалов (АПИМ) для их оценки, для широкой апробации АПИМ. ФЭПО позволяет более обоснованно готовиться к процедуре государственной аккредитации всем подразделениям вуза (УМУ, кафедрам, компьютерной службе). ФЭПО может рассматриваться как одна из форм реализации Болонского соглашения в части механизма демонстрации вузами качества подготовки выпускников для всех заинтересованных слоев общественности: абитуриентов, родителей, работодателей.

Сегодня более 1300 высших учебных заведений добровольно принимают участие в процедурах ФЭПО. По результатам компьютерного тестирования учебному учреждению высылается аналитическая карта, в которой представлены статистически обработанные результаты анализа соответствия знаний студентов требованиям ГОС ВПО.

Анализ рейтинг-листов по результатам тестирования позволяет для каждого студента определить его индивидуальный рейтинг на основе верного или неверного выполнения каждого тестового задания, а также рейтинг-листы группы студентов по дисциплине, в которой можно выделить сложные дидактические единицы, с самым низким процентом освоения. По форме и положению гистограммы можно судить о характере распределения результатов для группы студентов. Гистограмма позволяет эффективно оценить расслоение результатов студентов по уровню подготовки.

**Гистограмма распределения результатов
педагогических измерений
Дисциплина: Информатика
ООП: 013100 «Экология»**



Процент выполненных заданий	Количество студентов	Процент студентов
от 80% до 100%	2	13%
от 60% до 80%	8	49%
от 40% до 60%	3	19%
до 40%	3	19%
Всего	16	100%

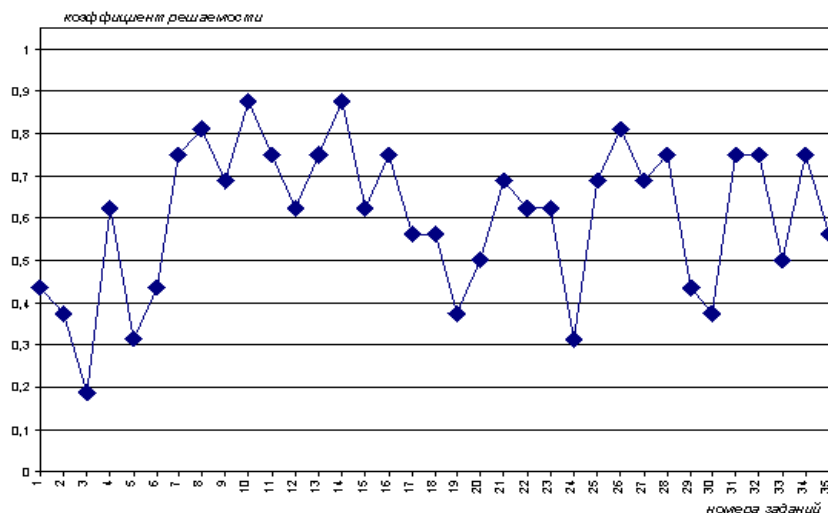
Рисунок 1 – Гистограмма распределения результатов педагогических измерений по дисциплине «Информатика»

Графическое представление результатов тестирования выступает карта коэффициентов решаемости заданий (рисунок 2) и карта коэффициентов освоения дидактических единиц (рисунок 3).

Карта коэффициентов решаемости заданий

Дисциплина: Информатика

ООП: 013100 «Экология»



Коэффициенты решаемости заданий	Количество заданий	Процент заданий
от 0,7 до 1	12	34%
от 0,4 до 0,7	17	49%
до 0,4	6	17%

Рисунок 2 – Распределение коэффициентов решаемости заданий по дисциплине «Информатика»

Карта коэффициентов решаемости позволяет оценить уровень подготовки испытуемых по отдельным темам измерительных материалов. По данному графику можно выявить темы, оказавшиеся трудными для студентов контролируемой группы (коэффициент решаемости менее 0,4)

Карта коэффициентов освоения ДЕ

Дисциплина: Информатика

ООП: 013100 «Экология»

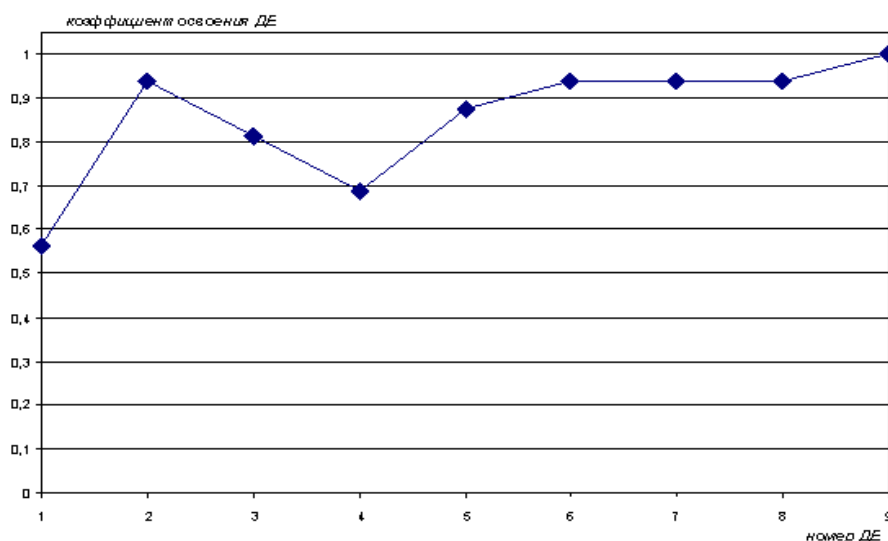


Рисунок 3 – Распределение коэффициентов освоения дидактических единиц по дисциплине «Информатика»

График позволяет оценить степень освоения тематических разделов дисциплины. Наглядно выявляются разделы, освоенные на недостаточном уровне (коэффициент освоения менее 0,7).

Основными областями использования результатов ФЭПО в вузе являются: рейтинговые технологии оценки учебных достижений студентов; внешняя независимая оценка в рамках системы менеджмента качества; самообследование, а также процедуры при прохождении государственной аккредитации.

Рейтинговые технологии оценки учебных достижений студентов позволяют оценить качество текущих знаний студентов, осуществлять контроль разделов или модулей дисциплины, а также засчитывать результаты тестирования в качестве экзамена.

Внешняя независимая оценка в рамках системы менеджмента в вузе позволяет анализировать результаты компьютерного тестирования в рамках оценки освоения требований ГОС на уровне:

- ректората (освоение материала по образовательным программам);
- деканата (освоение материала по дисциплинам и циклам дисциплин);
- кафедры (освоение материала по ДЕ и дисциплине);
- преподавателя (освоение материала по темам и ДЕ), а также проводить мониторинг результатов по дисциплинам и образовательным программам

При проведении самообследования важно помнить о двух важных задачах. Во-первых, компьютерное тестирование в рамках ФЭПО позволяет оценить освоение дисциплины на соответствие требованиям ГОС ВПО. Во-вторых, мониторинг результатов этой оценки для образовательных программ (не менее 3-х лет). Студенческие группы при самообследовании должны подбираться из положения о завершении изучения дисциплины не позднее, чем за 1 год до самообследования. В программу по самообследованию должны быть включены все аккредитуемые основные образовательные программы (ООП), все циклы дисциплин по каждой основной образовательной программе и все важнейшие дисциплины цикла: не менее 3-х по циклам ГСЭ и ЕН и не менее 5-и по циклам ОПД и СД. Число контролируемых дисциплин при уровне надежности 0,8 – 0,9 составит от 14 до 16. При этом, общее количество дисциплин в основной образовательной программе 30 – 40.

Результаты тестирования:

- позволяют делать разностороннюю диагностику содержания и уровня подготовки студентов;
- являются основой для сравнения результатов освоения требований ГОС студентами других вузов и ООП;
- позволяют проводить мониторинговые исследования базового уровня подготовки студентов;
- могут эффективно использоваться во внутривузовских системах менеджмента качества при внешней оценке базового уровня подготовки;
- отвечают всем требованиям, предъявляемым к результатам оценки освоения требований ГОС при самообследовании вуза.

Современный контроль уровня учебных достижений студентов основан на дидактических измерениях. В традиционном понимании дидактическое измерение есть ни что иное, как анализ и преобразование преподавателями ответов экзаменуемых на задачи или вопросы различной меры трудности. В смысловом отношении результат измерений содержит оценку (традиционно отметку) совокупности ответов студентов, которая отображается на шкалу отдельного преподавателя. Поскольку каждый преподаватель имеет собственную шкалу оценивания, измерения такого рода будут всегда субъективными. По мнению Васильева В.И., Тягуновой Т.Н. для устранения субъективизма дидактических измерений можно прибегать к групповому контролю, когда уровень учебных достижений студента оценивается сразу несколькими преподавателями, а отметка формируется экспертным путем. Это в определенной степени повышает объективность оценивания, но приводит к резкому возрастанию учебной нагрузки профессорско-преподавательского состава института. Более того, даже при экспертном подходе к проверкам достижений невозможно избежать ошибок, которые связаны с ограниченным числом вопросов, предъявляемых экзаменуемому. Программные средства и системы должны быть удобны и надежны в использовании, иметь современный дизайн. Существуют две тенденции в разработке программных продуктов: использование языков программирования высокого уровня и применение программно-инструментальных средств (оболочек). Готовые оболочки позволяют преподавателям создавать тестовые задания, не владея языками программирования.

Если тесты сконструированы специальным образом и состоят из довольно большого числа заданий, детально отображающих определенную область содержания, то появляется возможность выявления причин затруднений обучаемых. Это может быть неумение мыслить или выполнять что-либо по известной схеме в стандартных задачах, а может быть и неумение переносить знания в незнакомую ситуацию, искать необходимую информацию для выполнения творческих задач. Естественно, сделать тесты для целей диагностики непросто, однако необходимо, поскольку их отсутствие наносит заметный урон качеству подготовки специалистов. По нашему мнению, создание на базе ведущих вузов центров тестирования, в которых бы разрабатывалась не только методология тестирования, но и апробировались различные аттестационные педагогические измерительные материалы, помогло бы разрешить проблемы педагогического контроля в современных образовательных условиях.

Как отмечает М.Б. Челышкова «в последние годы в практике образования складывается ситуация, под влиянием которой традиционное тестирование, осуществляемое с помощью стандартизированных тестов фиксированной длины, перерастает в современные эффективные формы адаптивного тестирования, базирующегося на отличных от традиционных теоретико-методологических основах и иных технологиях конструирования и предъявления тестов.

На протяжении ряда лет вопросы адаптивного тестирования неоднократно привлекали внимание зарубежных ученых. Об этом свидетельствуют многочисленные исследования фундаментального характера и публикации таких авторов как J.A. Arter, R.K. Hambleton, J.L. Horn, C.D. Jensema, G.G. Kingsbury, F.M. Lord, J. Millman, L. Nauels, R.J. Owen, K.J. Patience, M.D. Reckase, J. Spray, H. Swaminathan, V.W. Urry, M. Waters, D.J. Weiss, R.W. Wood, A.R.

Zara и многих других. С начала 90-х годов компьютерное адаптивное тестирование получило за рубежом широкое признание в сфере практики. В значительной мере этому способствовали прикладные теоретические исследования таких ученых, как C.V. Bunderson, D.K. Inouye, G.G. Kingsbury, J.B. Olsen, H. Wainer, D.J. Weiss, работы которых были не только поддержаны педагогической общественностью, но и нашли широкое применение в практике профессионального отбора в ряде структур промышленного и военного комплексов» [9, с. 4-6].

Так что же представляет собой адаптивное тестирование и по какой причине оно вызывает интерес специалистов, занимающихся психологическим тестированием или тестированием знаний? Основная идея, движущая исследователей в сфере адаптивного тестирования, заключается в том, что тестовые задания необходимо адаптировать (подогнать) по трудности к уровню подготовленности испытуемых в тестируемой группе. При этом исследователи исходят из тех соображений, что слабым испытуемым бесполезно давать трудные задания, так как с большой вероятностью они не сумеют их выполнить верно. Отдельные исключения из сформулированного выше утверждения, конечно, бывают, но они крайне нежелательны, поскольку приводят к снижению валидности профилей ответов слабых испытуемых, способствуя тем самым росту ошибочного компонента в их тестовых оценках.

Столь же бесполезными выглядят легкие задания при тестировании сильных испытуемых. Очевидно, что использование слишком легких заданий может привести к тому, что все или почти все получают, примерно одинаковые высокие баллы и, следовательно, измерение не состоится по причине несоответствия уровня трудности заданий подготовленности тестируемых учеников или студентов.

Не менее важным является вопрос секретности заданий: для каждого испытуемого набор заданий является уникальным, в результате чего минимизируется возможность подсказок, выучивания правильных ответов и т.д. При повторном прохождении теста испытуемый также вынужден решать новые задания, что уменьшает влияние эффекта тренированности. Банк заданий содержит 450 вопросов (по 75 вопросов на шкалу), относящихся к различным уровням сложности. Количество вариантов ответов для большинства вопросов расширено до 5 (традиционно их было 4), что уменьшает влияние эффекта угадывания на результаты тестирования [12].

Очевидно, что в данном случае алгоритм адаптивного тестирования является более «жестким» в отношении возможного эффекта экзаменационного (тестового) стресса: после выполнения первого задания программа осуществляет «резкий скачок» к значительно более трудному заданию, но зато в последующих пробах колебание кривой трудности довольно быстро затухает, так что уже к 20 заданию достигается стабилизация в рамках определенного доверительного интервала, допустимого с точки зрения ошибки измерения, типичной для тестов интеллекта. Важно подчеркнуть, что такой подход с психологической точки зрения оправдан тем, что «затухание кривой трудности» производится отдельно и независимо для каждой субшкалы теста - отдельно для «числового интеллекта», «вербального», «визуального», «эрудиции», «речевой логики» и т.п. Тем самым испытуемый, поднимаясь по шкале трудности довольно быстро в рамках одного тематического блока (субшкалы), остается на довольно щадящих по трудности заданиях в рамках того блока, который ему удастся менее всего [11].

Внедрение тестового контроля следует начинать с обучения профессорско-преподавательского состава методике разработки заданий в тестовой форме. При этом самое главное - донести до сознания преподавателей, что тест это не просто привычная проверка, испытание, пробы или оценка знаний с помощью традиционных вопросов и экзаменационных билетов, а система заданий в соответствующей форме. Тестовые задания должны быть краткими, так чтобы на выполнение каждого не тратилось много времени. Далеко не всякое задание включается в тест. Оно должно быть строго определенным по форме, содержанию и специальным параметрам, оцениваемым в ходе математического анализа. Свойства заданий

обуславливают свойства всего теста, поэтому говорят не просто о наборе, а о системе заданий. Повсеместное использование современных технологий в обучении приводит к необходимости освоения преподавателями новых средств обучения и контроля знаний студентов. Одним из средств контроля является компьютерное тестирование. В настоящее время существуют требования к составлению тестов и тестовых заданий. Требования высоки не только к содержанию тестовых заданий, но и к программному обеспечению.

Библиографический список

1. Аванесов, В.С. Форма тестовых заданий [Текст] В.С.Аванесов. - М.: Центр тестирования, 2005. – 153 с.
2. Анастаси, А. Психологическое тестирование [Текст] А.Анастаси. - М.: Педагогика, 1982. – 532 с.
3. Андриенко А.В. Система гарантий качества подготовки специалистов в ГОУ ВПО «Сибирский государственный технологический университет» // Материалы 5-ой научно-методической конференции «Инновационные методы и средства оценки качества образования» (25-26 апреля 2007 г.). – М., МГУП, 2007. – С. 112-117.
4. Андриенко А.В., Суворова Н.В. Об опыте организации компьютерного тестирования в ГОУ ВПО «Сибирский государственный технологический университет» // Материалы 5-ой научно-методической конференции «Инновационные методы и средства оценки качества образования» (25-26 апреля 2007 г.). – М., МГУП, 2007. – С. 3-13.
5. Васильев В.И., Тягунова Т.Н. Основы культуры адаптивного тестирования [Текст]. В.И. Васильев, Т.Н. Тягунова. – М.: Национальный институт бизнеса, 2003. – 370 с.
6. Майоров, А.Н. Теория и практика создания тестов для системы образования. (Как выбирать, создавать и использовать тесты для целей образования) [Текст] А. Н. Майоров. – М., «Интеллект-центр», 2001. – 296 с.
7. Наводнов, В., Масленников, А. Об Интернет-экзамене в сфере профессионального образования [Текст] В. Наводнов, А. Масленников // Высшее профессиональное образование в России. Научно-педагогический журнал МОН РФ. – 2006. - №4. – С.15-19.
8. Соколова, И.Ю. Психологическое обеспечение качества образовательного процесса: учебное пособие / И.Ю.Соколова. – Томск: Изд-во ТГПУ, 2005. – 230 с.
9. Челышкова, М.Б. Теория и практика конструирования педагогических тестов [Текст] М. Б. Челышкова. Учебное пособие. – М.: Логос, 2002. – 432 с.
10. Чернилевский, Д.В. Дидактические технологии в высшей школе [Текст] Д. В. Чернилевский: Учебное пособие для вузов. – М.: ЮНИТИ-ДАНА, 2002. – 437 с.
11. Шмелев, А.Г. Адаптивное тестирование знаний в системе «Телетестинг» [Текст] А.Г. Шмелев, А.Г. Ларионов // IX международная конференция-выставка «Информационные технологии в образовании»: Сборник трудов участников конференции. Часть II. - М.: МИФИ, 1999. С.405.
12. Шмелев А.Г. Компьютеризация экзаменов: проблема защиты от фальсификаций - Тезисы международной конференции «Информационные технологии в образовании» [Текст] А.Г. Шмелев. - Москва: ИПИ РАН, 2001.