

Использование результатов тестирования для изучения уровня подготовленности обучающихся

Ядровская Марина Владимировна
к. ф.-м. н., доцент кафедры «Информационные технологии»
Донской государственный технический университет,
пл. Гагарина, 1, г. Ростов-на-Дону, (863)2340458
marinayadrovskaia@rambler.ru

Аннотация

В статье на основе анализа результатов тестирования предложена методика тестирования, осуществляющегося в рамках системы мониторинга и непрерывной оценки знаний «Эффективный контроль и мониторинг», разработанной и реализуемой с помощью информационно-коммуникационных технологий. Ее описание может быть использовано для демонстрации особенностей функционирования системы, поиска и устранения недостатков, определения направлений развития и повышения эффективности работы системы мониторинга.

In the article, based on the testing results analysis, a testing methodology was proposed that is implemented within the framework of the monitoring and continuous knowledge assessment system "Effective control and monitoring" developed and implemented using information and communication technologies. Its description can be used to demonstrate the features of the functioning of the systems, to find and fix the shortcomings, to determine the development trends and to improve the efficiency of the monitoring system.

Ключевые слова

контроль качества обучения, система мониторинга знаний, тестовый контроль, тестирование
quality control of training, knowledge monitoring system, test control

Информационные технологии – основа мониторинга знаний

Во всем мире отмечается повышенный интерес к проблемам качества образования. Современному обществу нужны специалисты с широким кругозором и высоким уровнем знаний. Как отмечает И.Г. Бартасевич, «в настоящее время остро стоит проблема повышения качества обучения студентов» [3]. Решению проблемы служит постоянный контроль знаний обучающихся как составляющая педагогической системы обучения.

Для контроля знаний широко применяют тестирование. Тестирования включает в себя тестовый метод, результат тестирования и интерпретацию результатов тестирования. Тестовый метод контроля качества обучения имеет ряд преимуществ перед другими педагогическими методами контроля. Одно из преимуществ состоит в возможности использования информационных технологий для реализации тестирования.

Сегодняшний уровень развития информационных технологий обеспечивает реальную основу для создания системы непрерывного тестового контроля знаний на разных ступенях обучения. Благодаря такой системе может осуществляться мониторинг знаний - процесс сбора, хранения, обработки и распространения

информации о результатах тестирования, а значит о деятельности педагогической системы, обеспечивающий непрерывную оценку ее состояния и прогнозирование её развития.

Такая система «Эффективный контроль и мониторинг» («ЭКиМ») создана в Донском государственном техническом университете (ДГТУ) и реализуется Управлением дистанционного обучения и повышения квалификации (УДО и ПК). Система использует портал СКИФ для проведения тестирования в автоматическом режиме и функционирует на основе инструментального средства Moodle. Разработаны и апробированы структуры базовых и специальных дисциплин, согласно которым тестовыми заданиями наполнены банки. Происходит непрерывное пополнение и редактирование заданий. В рамках этой системы осуществляется входной контроль знаний студентов по базовым дисциплинам всех форм обучения, текущий, рубежный и итоговый контроль, контроль знаний в рамках приемной комиссии для магистратуры. Результаты тестирования обрабатываются, проводится сравнительный анализ результатов, по которым регулярно готовятся отчеты и проводятся обсуждения.

Результаты тестирования для изучения уровня подготовленности обучающихся

Накопленные результаты тестирования позволяют проследить хоть и локальную, но важную для вуза динамику уровня подготовленности обучающихся. В.С. Аванесов предлагает называть общим термином «подготовленность испытуемого» определяемые в процессе педагогического измерения знания, умения, навыки, представления и компетенции. При этом, «в каждом конкретном случае педагогического измерения содержание понятия «подготовленность» определяется явно, в понятных терминах и на языке одной учебной дисциплины, если тест гомогенный, и нескольких дисциплин – если тест гетерогенный или интегративный». Уровень подготовленности будем оценивать по уровню знаний, который обычно представляется тестовым баллом обучающихся [1].

Отметим, что без использования информационных технологий была бы затруднительна задача обработки и сравнения результатов тестирования в рамках функционирования системы мониторинга знаний. Проверка знаний требует постоянного и всеобщего контроля. Это увеличивает нагрузку на технические средства реализации мониторинга и приводит к большим объемам информации, которые необходимо оперативно обрабатывать, анализировать и предоставлять. Комплексное использование различных, ставших традиционными, средств и технологий обработки данных позволяет создать и применять инструмент мониторинга и целевого контроля качества подготовки студентов.

В статье предложен сравнительный анализ результатов тестирования по математике в 2013-2015 гг. студентов двух направлений подготовки. Это группы МКИС (мультимедиа коммуникации и информационные системы) кафедры «Медиакоммуникации и мультимедийные технологии» и группы АТК (Информационное обеспечение автоматизированных технологических комплексов) кафедры «Технология машиностроения». Рассматриваемые результаты тестирования относятся к первому рейтинговому контролю знаний обучающихся первого года обучения. Они были получены в результате тестирования знаний по базовым дисциплинам, проводимого в 2013, 2014, 2015 гг. В список testируемых групп три года подряд были включены группы кафедр «Медиакоммуникации и мультимедийные технологии» и «Технология машиностроения». Поэтому было интересно, во-первых, сравнить уровень подготовленности студентов, поступивших

и начавших обучение в разные годы. Во-вторых, необходимо было сравнить уровень подготовленности обучающихся разных кафедр. Структуры тестов были одинаковы. Использовалась единая база заданий, задания выбирались случайным образом из множества заданий каждой категории теста.

Представленное исследование может быть рассмотрено с позиций сравнения уровня подготовленности обучающихся разных кафедр по математике. Также, анализ результатов позволяет оценить методику тестирования и особенности реализации системы независимой оценки знаний обучающихся, а «анализ тестовых заданий посредством математических методов позволяет получить информацию об их скрытых дефектах, которые не удается выявить с помощью экспертных методов» [4, с. 76].

Сравнительный анализ результатов тестирования

Система позволяет отслеживать время, в течение которого каждый тестирующийся выполнял задания теста. Поэтому при анализе результатов мы учитывали время, которое каждый тестирующийся объективно потратил на тестирование. На рис.1 (а, б) представлено изменение среднего для группы тестового балла и среднего для группы времени тестирования для групп МКИС и АТК. Можно отметить, что за три рассматриваемых года значение среднего тестового балла для обеих групп возрастило. Для группы МКИС оно выше, чем для группы АТК, и составляет соответственно 55 % и 46 %. Можно предположить, что средний уровень знаний по математике у студентов групп МКИС выше, чем у студентов групп АТК. Незначительная разница может отражать действительное различие в уровне знаний и подготовленности обучающихся, а может быть связана с характеристиками тестирования, к которым отнесем время тестирования, трудность, надежность и валидность теста. Несмотря на одинаковую структуру теста, случайный выбор заданий теста из множества заданий категорий банка создает некоторые различия для обучающихся и групп в целом. С одной стороны, наличие множества заданий каждой категории банка заданий позволяет устранить проблему «списывания», а с другой – приводит к использованию индивидуальных тестов, тестов с различными характеристиками.

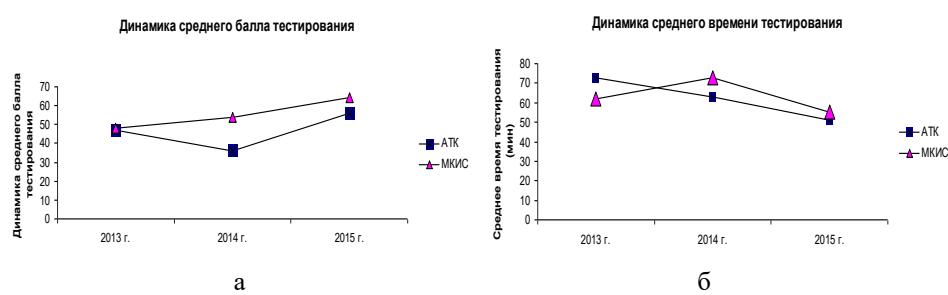


Рис. 1. Изменение среднего тестового балла и среднего времени тестирования для групп МКИС и АТК

В.С. Ким отмечает, что «общее время тестирования определяется количеством и сложностью заданий», но «теоретически рассчитать это время невозможно, поэтому рекомендуется использовать эмпирические данные по результатам первичной апробации теста» [6]. Мы отмечали, что система фиксирует индивидуальное время выполнения теста каждым тестирующимся. Эту информацию можно использовать для того, чтобы понять, правильно ли в целом мы определяем

время для тестирования. Сначала изучим, как зависит средний тестовый балл от изменения времени, которое было потрачено обучающимися на тестирование. На рис.2 (а, б) графически представлена линейная зависимость среднего тестового балла и среднего времени, потраченного на тестирование, для групп МКИС и АТК. Коэффициенты корреляции такой связи соответственно для групп МКИС и АТК равны 0,51 и 0,5, т.е. практически одинаковы для обеих групп. Такое значение коэффициента корреляции находится на границе слабой и средней зависимости рассматриваемых величин. Можно предположить, что от времени, которое было потрачено на выполнение теста, результаты тестирования зависят слабо.

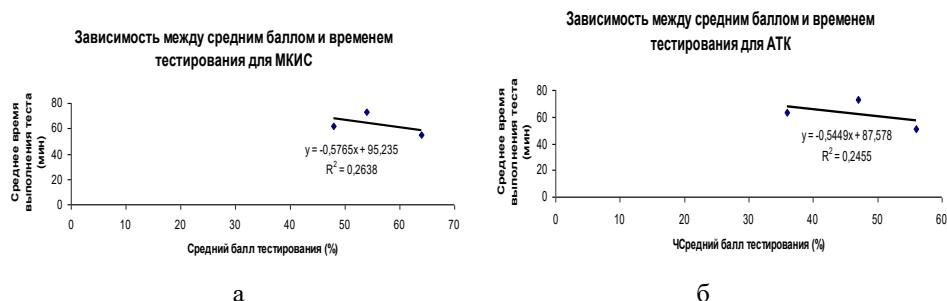


Рис. 2. Графическое представление зависимости среднего балла и среднего времени тестирования для групп МКИС и АТК

Из рис.3 (а, б) следует, что значение коэффициента корреляции для связи времени выполнения и трудности теста также находится на границе слабой и средней зависимости рассматриваемых факторов. Можно предположить, что отводимого времени достаточно для выполнения теста текущего контроля. Среднее за три рассматриваемых года значение среднего времени, потраченного на тестирование, для групп практически одинаково и равно соответственно 63 мин. для МКИС и 62 мин. для АТК. Таким образом, экспериментально мы получили, что для теста текущего контроля из 30 заданий время тестирования может составлять примерно 60 мин. Это согласуется с одним из положений рекомендаций по тестированию, которые дает В.С. Ким: «время тестирования определяется по расположению максимума дисперсии тестовых результатов и не должно превышать 60 минут» [6].

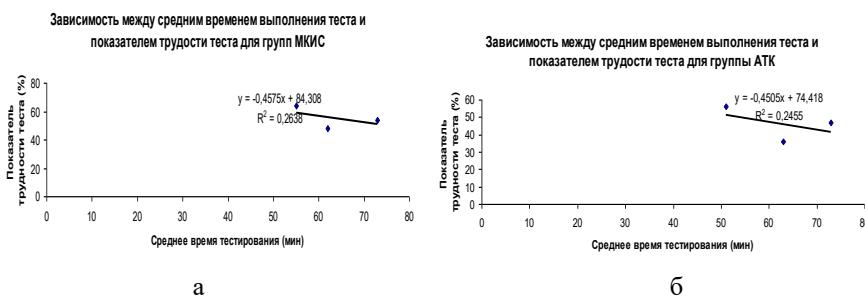


Рис. 3. Графическое представление зависимости времени выполнения и трудности теста для групп МКИС и АТК

При статистической обработке результатов тестирования анализировалась трудность заданий тестов. Трудность p_j тестового задания j равна доле испытуемых, правильно ответивших на это задание: $p_j = n_j/n$, где n_j — число правильных ответов на j -задание; n — общее число испытуемых, отвечавших на j -е задание. Трудность всего теста рассчитывалась аналогично как доля правильных ответов на задание теста (см.

табл. 1). Среднее за три рассматриваемых года значение показателя трудности для групп МКИС и АТК соответственно равно 55 % и 46 %. Средний тестовый балл непосредственно зависит от трудности теста.

Таблица 1

Сравнение показателей трудности тестов

Год тестирования	Показатель трудности тестов по группам	
	МКИС	АТК
2013 г	0,48	0,47
2014 г	0,54	0,36
2015 г	0,64	0,56

Проанализируем трудность тестов с помощью гистограмм распределения тестовых баллов, представленных на рис. 4 (а-в) и рис. 5 (а-в).

В хорошо сбалансированном по трудности тесте распределение баллов имеет вид нормальной кривой [6]. Как видим на рис.4 и рис.5 распределение баллов, в основном, бимодально. Бимодальная конфигурация указывает на то, что по результатам выполнения теста выборка testируемых разделилась на две группы. Одна группа справилась с большинством легких, а другая с большинством трудных заданий теста. Левосторонняя скошенность диаграмм распределения, характерная для МКИС в 2013 и 2015 гг., означает использование более легких тестов.



Рис. 4. Гистограммы распределения тестовых баллов для группы МКИС (2013-2015 гг.)

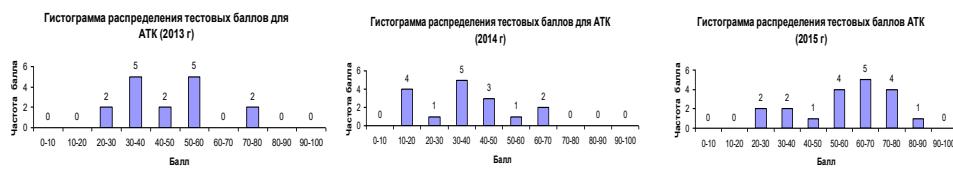


Рис. 5. Гистограммы распределения тестовых баллов для группы АТК (2013-2015 гг.)

В нашем случае более высокие баллы тестирования группы МКИС связаны с более высоким уровнем подготовленности по математике, так как использовался один банк заданий для тестирования, задания выбирались случайным образом, но показатели трудности тестов для групп МКИС были выше, чем для групп АТК.

Общей рекомендацией, например, для теста текущего контроля, является включение в тест 20% заданий малой трудности (показатель трудности > 70%), 70% – средней трудности (29-69% трудности) и 10 % заданий большой трудности (<25% трудности) [2]. В целом, рекомендуется стремиться к использованию в тесте заданий с трудностью 40-60 %. Рекомендация включать в тест больше заданий средней

трудности оправдана с точки зрения определения надежности измерения по формулам классической теории тестов. Проверка этого положения показала, что наиболее сбалансированы по трудности тесты 2013 г. для обеих групп (рис. 6).

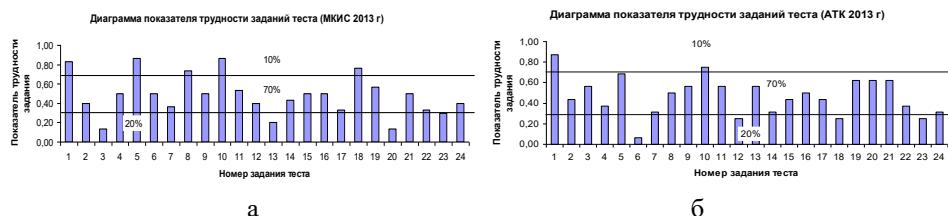


Рис. 6. Диаграммы показателей трудности заданий тестов для групп МКИС (а) и АТК (б) для 2013 г.

В teste рекомендуется задания [7, с.180] располагать в порядке возрастающей трудности, независимо от содержания тем, разделов и от учебных дисциплин. Проанализируем изменение трудности заданий в teste. Из графиков (рис.7, рис.8) видно, что тесты не имеют рекомендуемую структуру трудности заданий. Такая ситуация возникает из-за того, что задания категорий банка дисциплины не поделены на группы в зависимости от трудности. Поэтому нет возможности варьировать задания по показателю трудности. А также из-за того, что задания одной категории из множества выбираются случайным образом и имеют различный тип.

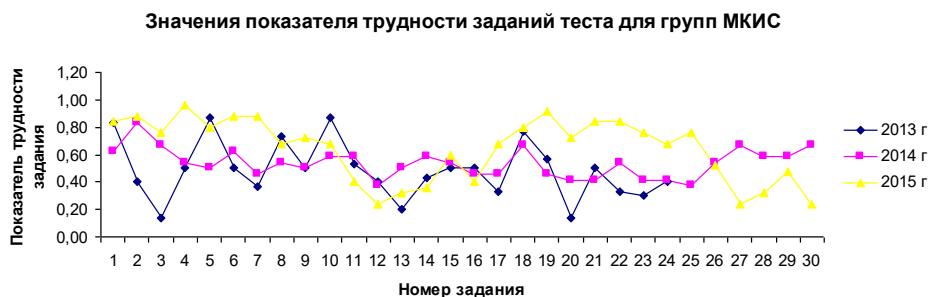


Рис. 7. Изменение показателя трудности заданий тестов для группы МКИС (2013-2015 гг.)

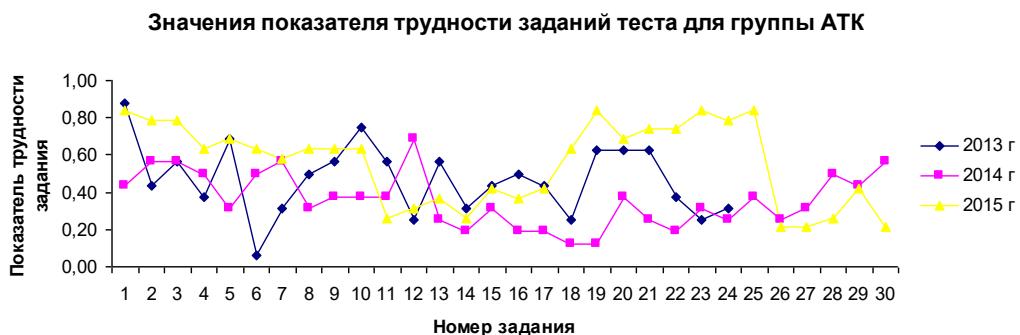


Рис. 8. Изменение показателя трудности заданий тестов для группы АТК (2013-2015 гг.)

Одна из важных характеристик теста – валидность. Валидный тест – это тест, который измеряет то, для измерения чего предназначен. В тестологии выделяют следующие виды валидности: очевидная, содержательная, конструктная, конкурентная, критериальная. Критериальная валидность – это степень соответствия теста определенному критерию, например, критерию оценки эффективности обучения. Согласно В.С. Ким «критериальная валидность (эмпирическая валидность) предполагает наличие внешнего критерия, корреляция с которым определяет валидность теста» [6]. Рассмотрим связь результатов тестирования с рейтинговой оценкой, выставленной преподавателем по результатам обучения за первый рейтинг.

При сравнении коэффициентов корреляции результатов тестирования с рейтинговой оценкой (рис.9) можно сказать, что рейтинговые оценки, данные преподавателями групп АТК, были более близки к результатам тестирования (см. табл. 2).

Таблица 2
Сравнение коэффициентов корреляции результатов тестирования и рейтинговой оценки

Год тестирования	Коэффициент корреляции по группам	
	МКИС	АТК
2013 г	0,57	0,6
2014 г	0,77	0,93
2015 г	0,1	0,25

Для обеих групп результаты тестирования и рейтинговые оценки преподавателей наиболее разнятся в 2015 г. Наоборот, в 2014 г. результаты тестирования максимально были приближены к рейтинговой оценке преподавателей для обеих групп. Можно отметить, если речь идет о проверке базовых знаний, компьютерное тестирование является незаменимым средством проверки уровня подготовленности обучающихся, так как позволяет за небольшое время протестировать большое количество обучающихся по различным дисциплинам, обеспечивая правильную оценку существующего уровня подготовленности (см. табл. 2). Чтобы доверять результатам тестирования, необходимо иметь, на наш взгляд, качественные задания банка тестирования.

Динамика коэффициента корреляции результатов тестирования и рейтинга для разных групп

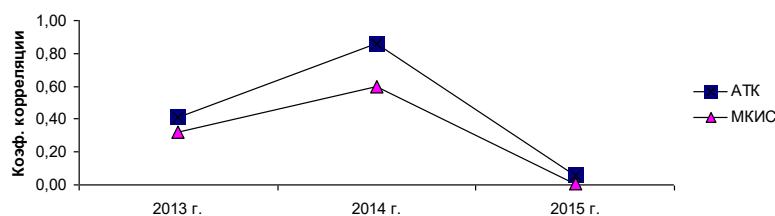


Рис. 9. Динамика коэффициента корреляции для связи факторов: результаты тестирования и результаты рейтинга

Как видим из диаграмм (рис. 10, рис. 11), большинство обучающихся по результатам тестирования получили удовлетворительную оценку знаний. Более высокая оценка знаний может быть выставлена преподавателем с использованием дополнительной проверки.

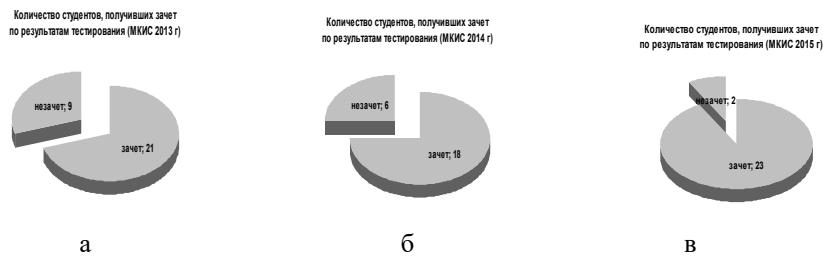


Рис. 10. Диаграммы результатов тестирования для группы МКИС (2013-2015 гг.)

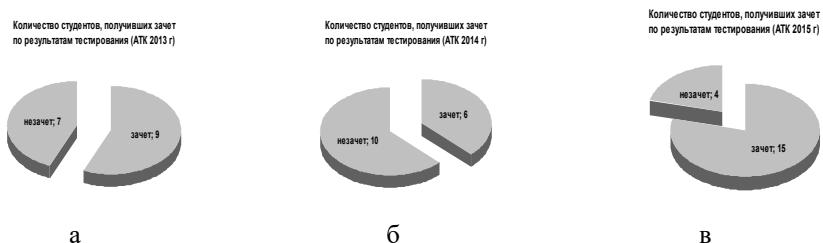


Рис. 11. Диаграммы результатов тестирования для группы АТК (2013-2015 гг.)

Тест представляет собой систему заданий. При некорректности и плохом качестве тестовых заданий, велика вероятность, что слабые обучающиеся, не зная ответа, будут пытаться его угадать. Сильные испытуемые, зная верный ответ, но не находя его среди предложенных, также будут вынуждены случайным образом выбирать любой из ответов. В итоге, индивидуальные баллы будут представлять собой случайные последовательности, не повторяющиеся в разных сессиях тестирования. Вследствие этого воспроизводимость тестовых баллов будет полностью отсутствовать и надежность теста будет близка к нулю.

Для тестов рассчитывался показатель надежности – ретестовая надежность (устойчивость результатов теста) (см. табл. 3), который демонстрирует возможность получения одинаковых результатов у испытуемых в различных случаях. Для вычисления надежности теста нужны результаты двух испытаний, которые организуются следующими способами: тестирование с помощью двух параллельных тестов (parallel-form reliability); повторное тестирование с помощью одного и того же теста (test-retest reliability); расщепление теста (split-half method) [6]. В работе использовался метод расщепления. Полученные результаты делились на две группы: в первую вошли результаты по четным заданиям, во вторую – результаты по нечетным заданиям. Затем вычислялся коэффициент корреляции между этими группами по формуле:

$$r_{xy} = \frac{N \sum_{i=1}^N X_i Y_i - \left(\sum_{i=1}^N X_i \right) \cdot \left(\sum_{i=1}^N Y_i \right)}{\sqrt{\left(N \sum_{i=1}^N X_i^2 - \left(\sum_{i=1}^N X_i \right)^2 \right) \cdot \left(N \sum_{i=1}^N Y_i^2 - \left(\sum_{i=1}^N Y_i \right)^2 \right)}}$$

Поскольку для определения надежности используется лишь половина теста, то полученное значение r_t является заниженным. Для коррекции значения r_t используется формула Спирмена-Брауна:

$$r_t' = \frac{2r_t}{1+r_t}$$

где r_t' – исправленный коэффициент надежности; r_t – коэффициент надежности по половинкам расщепленного теста [6]. Согласно показателю, различают надежность отличную, хорошую, удовлетворительную, неудовлетворительную.

Таблица 3
Сравнение показателей надежности тестов

Год тестирования	Показатель надежности теста по группам	
	МКИС	АТК
2013 г	0,81(хорошая)	0,71 (удовлетворительная)
2014 г	0,8 (хорошая)	0,22 (неудовлетворительная)
2015 г	0,1 (неудовлетворительная)	0,11 (неудовлетворительная)

Как видно из рис.12, на котором изображены показатели надежности тестов для разных тестирований по группам, три теста имеют удовлетворительную и хорошие показатели надежности.

Динамика показателя надежности тестов по группам

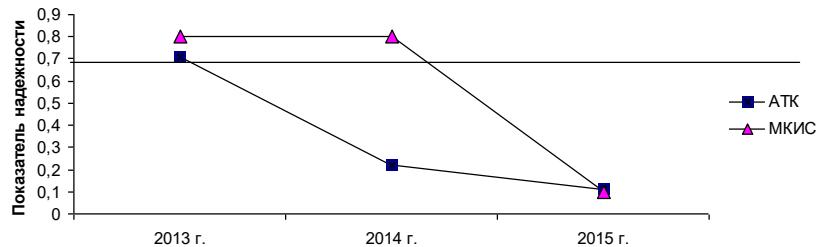


Рис. 12. Динамика показателя надежности тестов для групп МКИС и АТК (2013-2015 гг.)

Для них показатель надежности выше 0,69 (см. табл. 2). В 2013 г. началось создание и наполнение банка тестовыми заданиями с одновременной проверкой качества заданий, поэтому надежность тестов оказалась приемлемой. В последующем происходило слияние банка с ранее подготовленными заданиями, проверка качества была низкая. В настоящее время происходит непрерывное пополнение банка заданиями и осуществляется проверка качества заданий с планированием сертификации банка заданий.

Значение информационных технологий для реализации тестового контроля

Постоянный анализ результатов тестирования и функционирования системы «Эффективный контроль и мониторинг» на предмет соответствия целям и задачам педагогической системы, в рамках которой они реализуются, способствует устранению недочетов, развитию системы и повышению эффективности ее работы. При этом становятся понятными как ее новые возможности, способы и функции их

реализации, так и новые информационные технологии, которые могут при этом использоваться.

Можно отметить, что для системы ЭКиМ остаются неизменными задачи формирования банков базовых и специальных дисциплин, накопления заданий и проведения экспертного и экспериментального оценивания их качества. Представленное исследование подтвердило актуальность задачи структуризации заданий банков по показателю трудности.

Сопоставление результатов тестирования и оценок критериальной валидности тестов позволяет доверять результатам тестирования, особенно, если речь идет о проверке базового уровня знаний.

Важной задачей системы мониторинга является задача расширения электронных технологий, используемых для тестирования. Уже имеющийся опыт позволяет особое внимание обратить на применение технологий мобильного тестирования.

Также актуальна и важна задача обучения преподавателей по программам повышения квалификации, связанным с компьютерным тестированием. Такие программы реализуются в УДО и ПК ДГТУ и связаны с: отработкой процедур компьютерного и мобильного тестирования; оперативной обработкой результатов тестирования и получением качественных выводов для использования в учебном процессе.

Комплексное использование различных, ставших традиционными, средств и технологий обработки данных позволяет создать и применять инструмент мониторинга и целевого контроля качества подготовки студентов. Задачи такого контроля состоят в: проверке соответствия знаний, умений, навыков обучающихся целям обучения на определённом этапе формирования компетенций; удовлетворении запросов обучающихся на объективную и независимую оценку знаний; получении объективной информации о результатах образовательной деятельности вуза.

Литература

1. Аванесов В.С. Знания как предмет педагогических измерений // Педагогические измерения, №3, 2005. Новая редакция 2008 г. URL: <http://testolog.narod.ru/EdMeasmt5.html> (дата обращения: 1.04.2017).
2. Антифеева Е.Л., Петрова Д.Г. Промежуточный контроль знаний студентов по курсу «Прикладная механика» / методика преподавания естественный и точных наук. URL: [https://lib.herzen.spb.ru/media/magazines/contents/19\(48\)/antifeeva_9_48_131_146.pdf](https://lib.herzen.spb.ru/media/magazines/contents/19(48)/antifeeva_9_48_131_146.pdf), (дата обращения: 4.04.2017)].
3. Бартасевич И.Г. Измерение уровня знаний – основного показателя качества обучения. Вестник Астраханского государственного технического университета. Серия: Управление, вычислительная техника и информатика. №1, 2010. URL: <http://cyberleninka.ru/article/n/izmerenie-urovnya-znaniy-osnovnogo-pokazatelya-kachestva-obucheniya>, (дата обращения: 5.04.2017)
4. Бузун, Д.Н. Компьютерные дидактические тесты: оценка качества / Д.Н. Бузун // Информационное обеспечение исторического образования: Сб. ст. / Под. ред. В. Н. Сидорцова, А. Н. Нечухрина, Е. Н. Балыкиной. – Минск: БГУ; Гродно: ГрГУ, 2003. – С. 76–86. (Педагогические аспекты исторической информатики; Вып. 3).].
5. Захарова О.А., Ядровская М.В. Анализ результатов внедрения в ДГТУ системы независимой оценки знаний студентов // Международный научно-методический симпозиум «Современные проблемы многоуровневого образования». Ростов н/Д: ДГТУ. 2014. – С. 50-77.

6. Ким В.С. Тестирование учебных достижений. Монография. Уссурийск, 2007.
URL: http://uss.dvfu.ru/static/kim_testing_monograph/index.html (дата обращения: 4.04.2017)
7. Пасховер И.Л. Педагогический тест как инструмент системы оценки качества образования // Народное образование. Педагогика. №4, 2011.