

**СРАВНЕНИЕ СУБЪЕКТИВНЫХ И ОБЪЕКТИВНЫХ ОЦЕНОК
КОМПЬЮТЕРНОГО ТЕСТИРОВАНИЯ**

О. Г. Альтишлер, О. М. Колесников, Т. Ю. Павлова

A COMPARISON OF SUBJECTIVE AND OBJECTIVE ASSESSMENT OF COMPUTER TESTING

O. G. Altshuler, O. M. Kolesnikov, T. Yu. Pavlova

В данной работе проанализированы результаты анкетирования студентов по вопросам применения компьютерного тестирования для итоговой оценки знаний и статистические характеристики проведенного тестирования. При общем положительном отношении к тестированию студенты считают, что результаты тестирования не должны быть единственным критерием итоговой оценки. В то же время перед применением созданного накопителя тестовых заданий для проведения нормативно-ориентированного теста, качество накопителя должно оцениваться с точки зрения выявления и удаления заданий с низкой степенью дискриминативности.

In this paper we analyzed the survey results of students in the use of computer-based testing for the final assessment of knowledge and the statistical properties of testing. With the overall positive attitude toward testing, students feel that the test results should not be the sole criterion of the final grade. At the same time, before using the test items base created for regulatory-based test, the quality of the base must be assessed in terms of identifying and removing the items with the low discriminative.

Ключевые слова: тестирование, компьютерное тестирование, сложность тестовых заданий.

Keywords: testing, computer testing, test items complexity.

О роли компьютерного тестирования в организации учебного процесса вуза

В настоящее время одной из популярных форм контроля знаний студента является компьютерное тестирование. Достоинства этого метода несомненны: оперативность и стандартизация процедуры, отсутствие

субъективизма со стороны преподавателя, быстрота обработки результатов тестирования.

Однако внедрение тестирования требует трудоемкой работы по созданию тестов, оценке качества разработанного накопителя тестовых заданий (НТЗ), степени трудности предъявляемых студентам заданий, валидности тестов. Нужно выявить и удалить из накопителя не-

корректные тестовые задания. Кроме того, существуют задания, на которые практически все обучающиеся отвечают правильно или все – неправильно. Согласно теории тестирования [1, с. 65 – 96], такие задания обладают низкой степенью дискриминативности, то есть не позволяют разделять испытуемых по уровню выполнения теста и должны быть исключены из НТЗ. Необходимо помнить также о том, что при тестировании по нескольким темам или по курсу в целом материал отдельных тем может иметь объективно неодинаковую сложность.

С другой стороны, нельзя забывать, что на результаты тестирования заметное влияние оказывает субъективное отношение обучающихся к процедуре тестирования. Целью данной работы было оценить созданный НТЗ с помощью объективных статистических характеристик и субъективных оценок обучающихся, выявленных с помощью анкетирования.

Авторами был сформирован банк из 223 тестовых заданий, на основе которого создан тест. Данный тест использовался при оценке знаний студентов химического факультета по разделу «Электричество и магнетизм» курса «Физика».

Оценка студентами процедуры и результатов компьютерного тестирования

Раздел «Электричество и магнетизм» является составной частью дисциплины «Физика» и изучается студентами 2 курса химического факультета в течение 3 семестра. Программа семестрового курса включает в себя 34 часа лекций, 17 часов практических и 34 часа лабораторных занятий.

Для контроля усвоения программного материала после изучения очередного блока проводилось промежуточное компьютерное тестирование с рейтинговой формой оценивания. Для оценки уровня освоения Государственного образовательного стандарта на экзамене проводилось итоговое компьютерное тестирование.

Если студент освоил данную дидактическую единицу (в качестве критерия берется правильное выполнение 60 % заданий из их общего числа), это значит, что он может получить по изучаемой дисциплине оценку «удовлетворительно». Для получения оценки «хорошо» и «отлично» необходимо показать более глубокое знание предмета, которое должно быть продемонстрировано при устном ответе на экзаменационные вопросы.

После сдачи экзамена 45 студентам было предложено ответить на вопросы анкеты, которая призвана оценить их отношение к тестированию, а также определить наиболее сложные, по их мнению, разделы курса. В обработку поступило 38 анкет.

Целью анкетирования было выяснение отношения студентов к оценке знаний с помощью компьютерного тестирования, методики подготовки их к тестированию, выполнения тестового задания. Вторая часть анкеты касалась содержания курса: какая из тем, приведенных вопросов является самой сложной.

Половина опрошенных (рис. 1) предпочитают компьютерное тестирование устному экзамену из-за сложности общения с преподавателем, треть думают о будущем и рассматривают его как тренинг ФЭПО, и только четверо из опрошенных отрицательно относятся к тестированию.

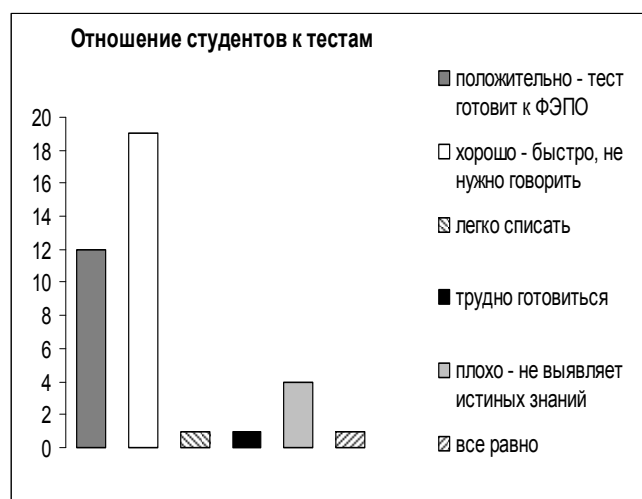


Рис. 1. Распределение ответов студентов на вопрос об их отношении к процедуре тестирования



Рис. 2. Распределение ответов студентов на вопрос по количеству попыток сдачи теста

Второй рисунок иллюстрирует печальный факт, что почти половина опрошенных с первой попытки тест не сдали. Ответы на следующий вопрос (рис. 3) проливают свет на причины данного факта. Почти половина тестируемых признаются в невнимательности. Треть признаются в слабой подготовленности, остальные винят себя в невнимании к пробному тестированию или незнании арифметики, однако находятся и такие, кто видит причину в излишней сложности тестовых заданий.

Рейтинг источников знаний для студентов (при их подготовке к экзамену) показывает рис. 4. Безусловными лидерами являются конспекты лекций и семинарских занятий, которые используются при повторном рассмотрении задач, решенных в течение семестра. На третьем месте находятся учебники и на четвертом – пробные тесты. В принципе результат предсказуемый. Данные рис. 5 с ним полностью согласуются.

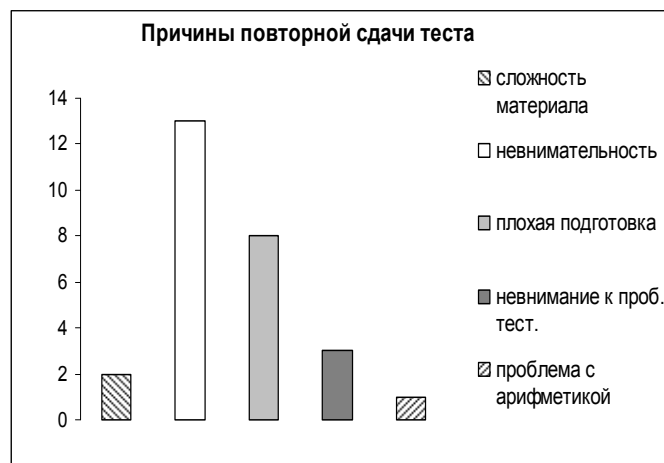


Рис. 3. Мнения студентов о причинах повторной сдачи тестов

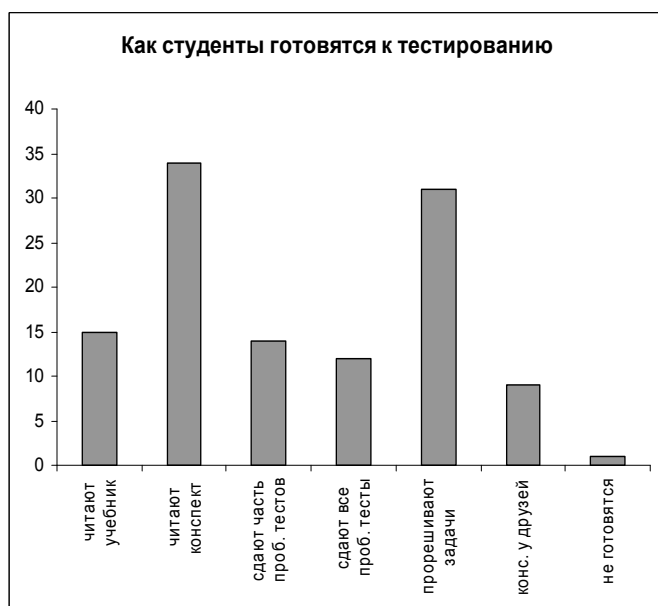


Рис. 4. Рейтинг способов подготовки к тестированию

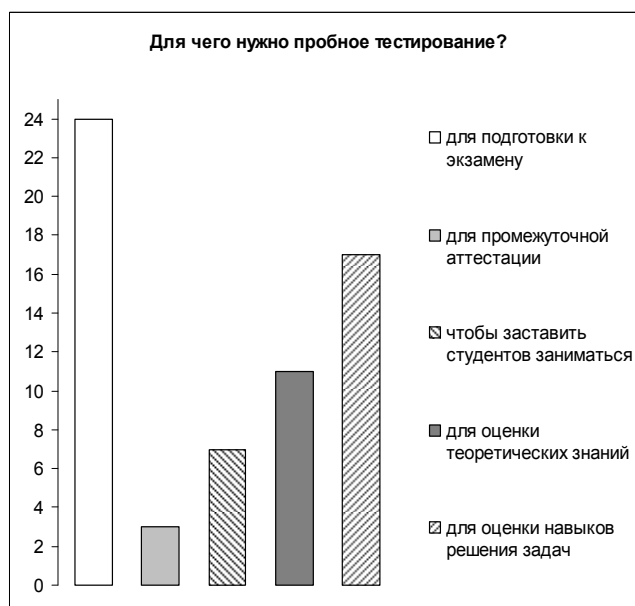


Рис. 5. Оценка студентами целей пробного тестирования

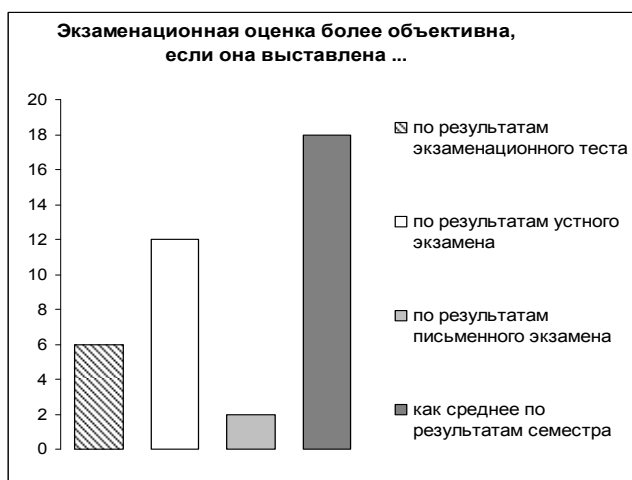


Рис. 6. Мнения студентов о процедуре получения итоговой оценки по учебному курсу



Рис. 7. Частотная диаграмма предъявления тестовых заданий

В основном студенты считают, что промежуточное тестирование предназначено для подготовки к экзамену и оценки навыков решения задач.

Результаты опроса, представленные на рис. 6, раскрывают психологическое восприятие тестовой формы оценки знаний студентами. Так, большинство опрошенных опасаются делать ставку только на результат тестирования, они не хотят, чтобы их семестровые усилия пропали даром; тем более, треть опрошенных считают, что они сильны в устном изложении своих знаний.

Статистика результатов компьютерного тестирования

Тестирование студентов проводилось с помощью накопителя, содержащего 223 тестовых заданий, разбитых на четыре темы. При настройке теста был выбран режим случайной выборки тестовых заданий (вопросов). Всего тестовые задания были предъявлены 2120 раз. Выбор такого режима тестирования приводит к тому, что задания предъявляются разное число раз. Частотная диаграмма выбора заданий приведена на рис. 7. Следует отметить, что 12 заданий не были включены в итоговый тест, так как при проведении предварительного тестирования более 70 % студентов ответили на них неправильно. В среднем одно задание предъявлялось 9,5 раз, несколько вопросов были использованы по 19 – 22 раза. В целом частотная диаграмма напоминает колокол нормального распределения, теоретически соответствующего случайной выборке, хотя имеющиеся отклонения свидетельствуют о не очень хорошем качестве генератора случайных чисел в системе АСТ.

Внедрение процедуры тестирования требует трудоемкой работы по созданию тестов, однако даже высокий профессионализм преподавателя, разработавшего тестовые задания, не гарантирует высокого качества созданного НТЗ (накопителя тестовых заданий). Речь идет не просто о некорректных формулировках или о закраившихся ошибках в тестовых заданиях. В литературе [1, 2] существует общепринятое определение степени трудности задания, как отношения числа правильных ответов к числу предъявлений задания p_i (i – номер задания). В таком определении кроется некоторая про-

тиворечивость, ведь, чем больше p_i , тем больше тестируемых ответили правильно и тем легче задание, однако мы будем придерживаться этого классического определения.

На практике считается, что тестовые задания, степень трудности которых менее 0,2 или более 0,8 (т. е. на которые неправильно отвечают менее 20 % или более 80 % студентов), необходимо изъять из НТЗ или переработать. Однако степень трудности задания p_i должна быть рассчитана при предъявлении этого задания более 200 раз. В этом и заключается основная сложность объективной оценки трудности задания, так как в небольших вузах курсовой поток студентов 70 – 200 человек, что при случайной выборке заданий дает в среднем не более 30 предъявлений вопроса при тестировании потока студентов, а в нашем случае, как уже упоминалось, только 9,5. Более того, обычно за одну процедуру тестируется группа порядка 25 человек и многие задания вовсе не предъявляются.

Как же преподавателю оценить качество заданий теста? На рис. 8 представлены зависимости трудности задания p_i (выражена в %) от числа его предъявлений: а – для заданий средней трудности, б – для сложных заданий. На каждом графике представлено изменение величины p_i для трех заданий, p_i , соответствующая одному заданию на графике, представлена одним значком. График на рис. 8(а) соответствует трем заданиям, трудность которых при 18 – 22 предъявлениях стала порядка 0,6 (60 %). Будем считать такие задания заданиями средней сложности. Виден большой разброс величины p_i при малом числе предъявлений задания, так что оценить качество этого задания в таком случае невозможно. Совсем другое поведение демонстрирует p_i в случае сложного задания (рис. 8(б)). Даже при малом числе предъявлений она мала и разброс ее значений при малом количестве предъявлений невелик. Такое же поведение демонстрирует p_i для легких тестовых заданий. Следовательно, если при тестировании при 5 – 10 предъявлениях задания величина p_i меньше 0,2 или больше 0,8, задание обладает низкой степенью дискриминативности и его лучше переработать.

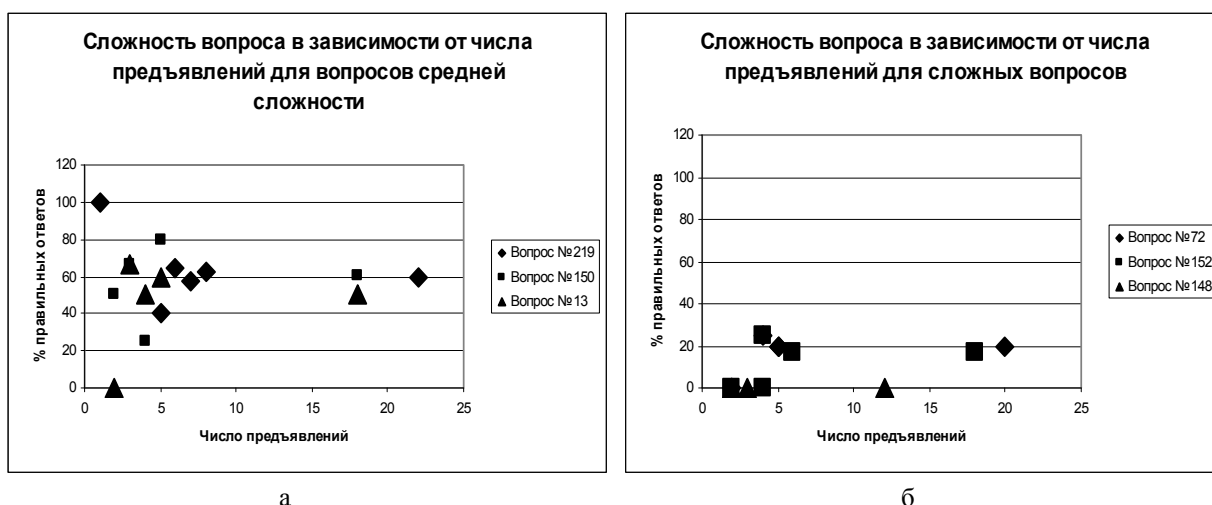


Рис. 8. Зависимость трудности задания p_i (выражена в %) от числа его предъявлений: а – для заданий средней трудности, б – для сложных заданий

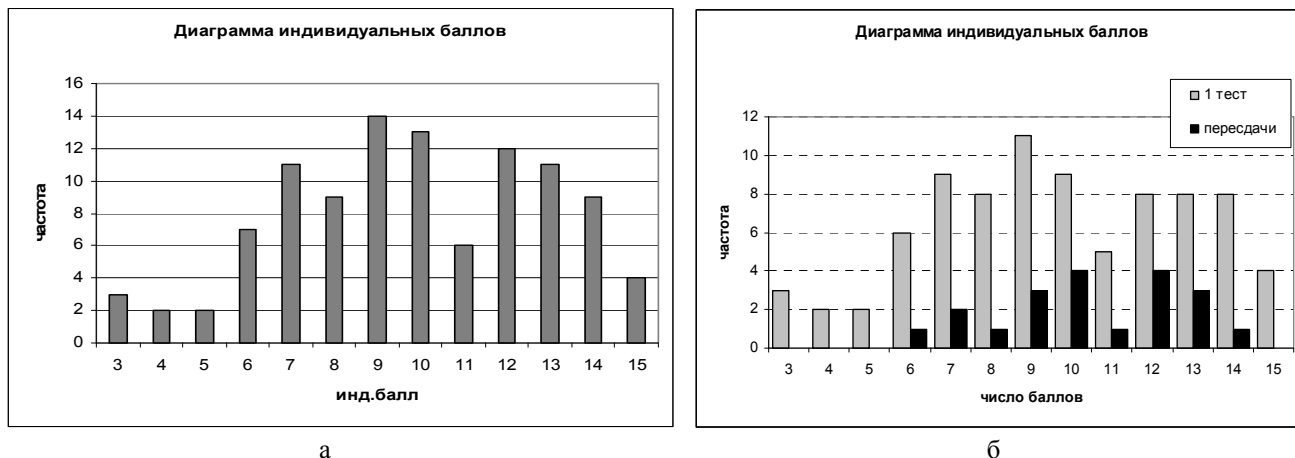


Рис. 9. Диаграмма распределения индивидуальных баллов студентов, набранных при тестировании: а – в целом по всем результатам тестирования, б – отдельно, по результатам первого тестирования и по результатам пересдач

Качественную оценку теста в целом можно произвести, анализируя баллы студентов, набранные во время тестирования. Индивидуальный балл каждого студента рассчитывается как число заданий, на которые он дал правильные ответы. Средний балл студентов по результатам всех попыток тестирования равен 9,8. На рис. 9 (а) представлена частотная диаграмма индивидуальных баллов (число студентов с данной величиной индивидуального балла). Серединой интервала разброса индивидуальных баллов (0 – 15) является 7,5. Из рисунка видно, что частотная кривая имеет вид нормального распределения (за исключением провала в области 11 баллов), среднее которого (9,8) сдвинуто относительно центра интервала вправо. В теории тестирования такой сдвиг частотной кривой расценивается как легкий тест.

С другой стороны, если построить частотную диаграмму индивидуальных баллов, набранных студентами во время первого тестирования, становится очевидным, что провал в районе 11 баллов на диаграмме 9 (а) не случаен. На рис. 9 (б) четко видно, что студенты делятся на группы: «сильных» (средний балл порядка 13), «слабых» (средний балл 8,5) и неготовых к тестированию (баллы до 6). Для группы «сильных», хорошо подготовленных студентов тест действительно был легким. При пересдаче, часть хорошо подготовившихся к тесту студентов присоединились к группе «сильных».

Преподаватели обычно практикуют нормативно-ориентированные тесты, позволяющие сравнивать учебные достижения отдельных испытуемых друг с другом. Однако те же результаты можно рассматривать в критериально-ориентированном плане, когда оценивается, в какой степени испытуемые овладели необходимым учебным материалом. Для этого сравним результаты тестирования, показанные студентами при выполнении заданий, соответствующих разным темам курса.

На рис. 10 представлены результаты тестирования по заданиям разных трех курса: «Электростатика», «Магнитостатика» и «Уравнения Максвелла». Объективными характеристиками, по которым можно сравнить степень освоения учебного материала, являются средняя трудность тестовых заданий, доли очень легких и очень трудных заданий. Очевидно, что наиболее трудной для студентов была тема «Уравнения Максвелла», т. к. средняя трудность заданий (доля правильных ответов) для

нее наименьшая. С этим результатом коррелируют и доли легких ($p_i > 0,8$) и сложных ($p_i < 0,2$) вопросов. Если рассматривать тест как критериально-ориентированный, то сложные для студентов вопросы либо отражают объективную сложность учебного материала, либо недостатки в методике его подачи, что для преподавателя является достаточно ценной информацией.

Сравнение субъективных и объективных результатов тестирования

В результате анкетирования выявлено, что два раздела «Магнитостатика» и «Уравнения Максвелла» студенты считают сложными, а раздел «Электростатика» – простым для изучения. Однако среднее число правильных ответов при тестировании по первой и третьей темам примерно одинаково: 52 % и 49 % соответственно, так что оценки студентов субъективны. «Уравнения Максвелла» действительно вызывают затруднения, среднее число правильных ответов на вопросы этой темы – всего 35 %.

Удивительно противоречиво студенты оценивают трудность задания. В анкете студентов просили указать, какой из трех приведенных вопросов является самым трудным, а какой – самым легким. В результате, первый вопрос признали самым сложным (самым легким) 42 % (34 %), второй – 29 % (24 %), третий – 24 % (39 %). Процент правильных ответов при тестировании составил 75 %, 42 % и 100 % соответственно.

Заключение

Анкетирование студентов наряду со статистической обработкой результатов тестирования по разделу «Электричество и магнетизм» курса «Физика» для студентов химического факультета, проведенные в данной работе, показывает, что при общем положительном отношении к тестированию студенты не могут адекватно оценить свои знания и предсказать результат тестирования. При разработке накопителя тестовых заданий для нормативно-ориентированных тестов необходимо провести предварительное тестирование для оценки качества НТЗ и лишь затем, использовать данный накопитель для оценки знаний студентов. Сравнение результатов тестирования студентов по разным темам курса позволяет оценить объективную сложность материала и выявить недостатки в методике его подачи.

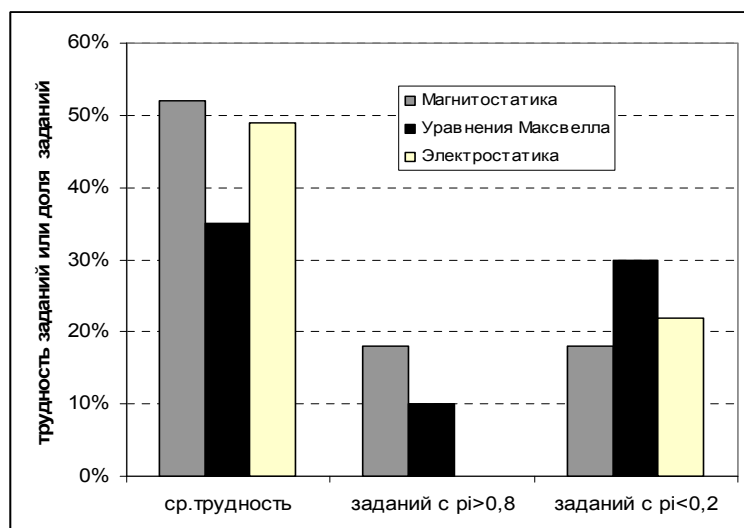


Рис. 10. Сравнение результатов тестирования по заданиям разных тем курса: средней трудности заданий, доли очень легких заданий и доли очень трудных заданий

Литература

1. Потеев, М. И. Основы теории тестирования. Основы аналитической дидактики: уч. пособие / М. И. Потеев. – СПб., 1992.
2. Чельшкова, М. Б. Теория и практика конструирования педагогических тестов / М. Б. Чельшкова. – М.: Логос, 2002.

Информация об авторах:

Альтшулер Ольга Генриховна – кандидат химических наук, доцент кафедры информационных технологий в образовании, КемГУ, 8(3842) 58-31-95, alt_og@bk.

Altshuler Olga Heinrichovna – Candidate of Chemistry., Associate Professor at the Department of IT in Education of KemSU.

Колесников Олег Михайлович – кандидат физико-математических наук, доцент кафедры информационных технологий в образовании, КемГУ, 8(3842) 58-31-95, colo@list.ru.

Kolesnikov Oleg Mikhailovich – Candidate of Physics and Mathematics, Associate Professor at the Department of IT in Education of KemSU.

Павлова Татьяна Юрьевна – кандидат физико-математических наук, доцент кафедры информационных технологий в образовании, КемГУ, 8(3842) 58-31-95, pavlova@kemsu.ru.

Pavlova Tatiana Jurievna – Candidate of Physics and Mathematics, Associate Professor at the Department of IT in Education of KemSU.