

условиях Стратегия развития образования: эффективность, инновации, качество: Материалы XIV науч.-метод. конф., посвященной 55-летию МГУТУ. Тематическое приложение к журналу «Открытое образование»: в 3 т. – М.: МГУТУ, 2008. – Т. 1. – С. 37-41

РАЗРАБОТКА АЛГОРИТМА ОЦЕНКИ КАЧЕСТВА ТЕСТОВЫХ ЗАДАНИЙ И ТЕСТА

Бозняков А.В., Муратова Е.А.

Томский политехнический университет, г. Томск

Научные руководители: Семенов М.Е., к.ф.-м. н., доцент кафедры ВММФ, Муратова Е.А., директор ЦОКО

Тестовые формы контроля знаний студентов на современном этапе являются одним из наиболее перспективных средств повышения эффективности процесса обучения. К безусловным достоинствам тестовых технологий контроля знаний относятся объективность и независимость, возможность оперативной оценки большего количества студентов при достаточно полном охвате материала изучаемого курса.

При работе с тестами часто возникают такие понятия, как качество теста и качество тестовых заданий, так как необходимо знать, насколько хорошо тест дифференцирует тестируемых по уровню подготовленности. Качество теста зависит от многих факторов, таких как: форма предъявления заданий, контролируемый материал, система оценивания ответов, работоспособность заданий.

В данной работе предложен алгоритм оценки качества тестовых заданий и теста, приведены результаты исследования разных систем оценивания ответов тестируемых.

Алгоритм оценки качества тестовых заданий и теста

Для анализа качества тестовых заданий и теста мы рассмотрим следующий алгоритм, состоящий из трех этапов.

1 этап. Вычисление трудности. Трудность отражает статистический уровень решаемости задания в протестированной группе. Обычно трудность это величина, обратная решаемости задания, и оценивается при сравнении числа тестируемых, правильно выполнивших задание N_p , к общему числу тестируемых N [1]:

$$U_i = (1 - N_p / N). (1)$$

Трудность изменяется $0 \leq U_i \leq 1$. Задания с нулевой или единичной трудностью рекомендуется исключать из теста.

2 этап. Вычисление дифференцирующей способности. Данная характеристика показывает, насколько эффективно тестовое задание различает тестируемых по уровню их подготовленности [2]:

$$I_i = 3 \left(\sum_{j=2N/3-1}^N A_{ij} - \sum_{j=1}^{N/3} A_{ij} \right) / N, (2)$$

где A_{ij} – суммарные значения правильных ответов в группах слабо и хорошо подготовленных студентов по матрице результатов. Дифференцирующая способность изменяется $-1 \leq I_i \leq 1$. Задания, со значениями $I_i < 0$ из теста исключаются, задания со значениями $0 \leq I_i \leq 0,2$ требуют корректировки.

3 этап. Вычисление надежности теста. Надежность теста вычисляется по формуле Кьюдера-Ричардсона [3]:

$$R = \frac{N}{N-1} \frac{1 - \sum_{i=1}^k P_i Q_i}{S^2}, (3)$$

где $P_i Q_i$ – вариация ответов тестового задания; S^2 – дисперсия по итоговому тестовому баллу, k – количество тестовых заданий. Коэффициент надежности изменяется $0 \leq R \leq 1$. Тест позволяет получить надежные результаты оценивания при $R \geq 0,7$ [3].

Исследование показателей качества тестовых заданий и теста

Рассмотренные показатели качества тестовых заданий и теста предложены для дихотомической шкалы оценивания: правильный ответ оценивается 1 баллом, неправильный – 0 баллов. Так как некоторые задания теста могут предполагать ответ, состоящий из выбора нескольких элементов (задания с выбором нескольких правильных ответов, задания на установление правильной последовательности, задания на соответствие, задания с выбором наиболее правильного ответа), то ответы могут оцениваться как частично верные. Необходимо исследовать как форма задания влияет на итоговый результат тестируемого и по результатам исследований, в случае необходимости, внести корректировку в предложенный алгоритм и расчетные формулы.

Применим алгоритм для анализа результатов тестирования по дисциплине "Химия". В тестировании приняло участие 420 студентов, на выполнение 24 заданий теста отведено 90 минут. Форма предъявления заданий: с выбором одного и нескольких правильных ответов, на упорядочивание элементов. Студент может получить за задание: 1 балл – правильный ответ; 0 баллов – не правильный ответ; 0,5 балла – частично верный ответ; «←» – не дан ответ.

Рассмотрим разные варианты вычисления характеристик заданий с использованием 1) оценок за частично верные ответы, 2) «строгое» оценивание, когда за частично верные ответы ставилось 0 баллов, 3) «лояльное» оценивание – за частично верный ответ выставляется 1 балл. На рис. 1 приведено распределение трудности заданий, вычисленное по формуле (1). На рис. 1 средняя линия показывает трудность тестовых заданий с дихотомической системой оценивания выполненных заданий теста. Верхняя кривая на рис. 1 показывает трудность при округлении вверх («лояльное» оценивание), нижняя кривая – при округлении вниз («строгое» оценивание). Отметим, что значения совпадают для заданий с выбором одного правильного ответа (например, 1 и 2 задания).

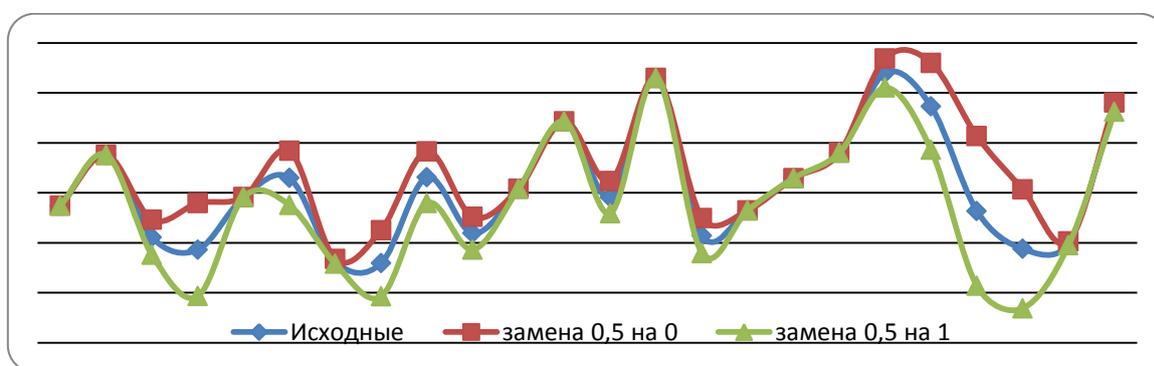


Рис. 1. График распределения трудности заданий

Для оценки влияния разных систем оценивания на итоговый результат, вычислим значения корреляции ответов тестируемых с использованием дихотомической шкалы и шкал «строгого» и «лояльного» оценивания с использованием коэффициента ранговой корреляции Спирмена. Наименьший коэффициент (0,93) получен для пары дихотомическая и «лояльная» система оценивания ответов. Наибольшее значение (0,95) получено для пары дихотомическая и «строгая» система оценивания ответов. В паре «строгая» и «лояльная» система оценивания ответов значение коэффициента корреляции составило 0,77. Это позволяет сделать вывод: при использовании ответов с частичным оцениванием предпочтительна замена частично правильного ответа (0,5 баллов) на неправильный ответ (0 баллов).

Для определения значений дифференцирующей способности заданий, все тестируемые были разделены на три равные группы в соответствии с итоговым тестовым баллом. Первая группа «хорошо подготовленных» (140 студентов) характеризуется высокими значениями тестового балла, вторая группа – средними значениями тестового балла, третья группа «слабо подготовленных» – низкими

значениями тестового балла. В полученных группах рассчитали доли правильных ответов студентов и определили дифференцирующие способности заданий теста с использованием разных систем оценивания правильных ответов (рис. 2).

Для оценки влияния разных систем оценивания на итоговый результат использованы коэффициенты ранговой корреляции Спирмена. В качестве исходных данных использованы значения дифференцирующей способности заданий, вычисленных по отметкам, полученным по разным системам оценивания. Наибольшее значение характерно (0,99) для дихотомической и «строгой» системы оценивания ответов студентов. Для дихотомической и «лояльной» системы оценивания коэффициент корреляции составил 0,94.

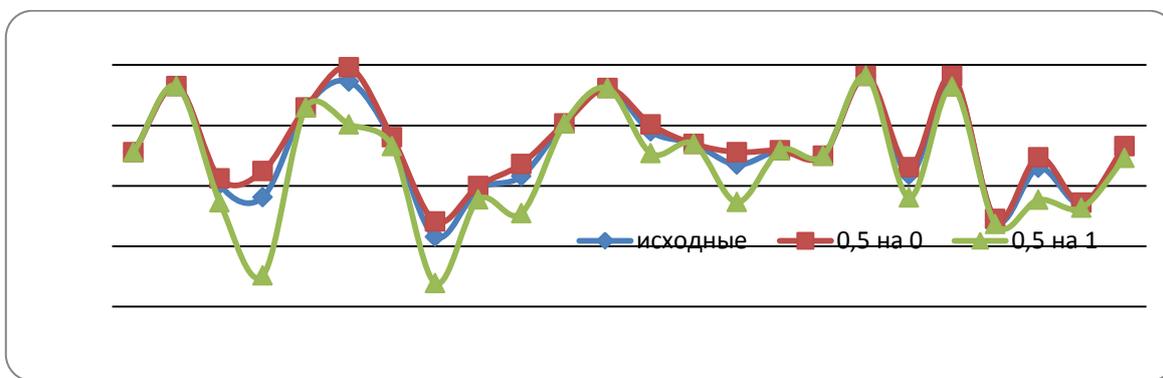


Рис.2. График коэффициентов различия заданий

На последнем этапе были вычислены коэффициенты надежности теста для всех используемых систем оценивания. В итоге получены следующие значения: дихотомическая шкала оценивания $R=0,68$, «строгое» оценивание $R=0,63$, «лояльное» оценивание $R=0,67$.

Выводы. В работе предложен алгоритм оценки качества тестовых заданий и теста. Предложенный алгоритм использован для исследования качества заданий теста и теста с использованием дихотомической, «лояльной» и «строгой» системы оценивания ответов тестируемых на задания с возможностью дать частично верный ответ. Приведен иллюстративный пример применения предложенного алгоритма. При включении в тест заданий с частично верными ответами будем считать студентов, как справившимися и учитывать их балл в расчетных формулах, как 0,5. При использовании данного метода оценивания предоставляется возможность не только более качественной дифференциации тестируемых, но и получить достоверный результат без искажения данных.

Список информационных источников

1. Чельшкова М.Б. Теория и практика конструирования педагогических тестов: Учебное пособие. – М.: Логос, 2002. – 432 с.
2. Аванесов В.С. Теория и методика педагогических измерений – [Электронный ресурс]. – режим доступа: <http://testolog.narod.ru/Theory21.html> – свободный.
3. Переверзев В.Ю. Технология разработки тестовых заданий: справочное руководство. – М.: Е-Медиа, 2005. – 265 с.

ОСОБЕННОСТИ ИССЛЕДОВАНИЯ МОТИВОВ И ИНТЕРЕСОВ ФИЗИКУЛЬТУРНОЙ ДЕЯТЕЛЬНОСТИ

Бондаренко Е.В., Черепанова Г.И.

Томский политехнический университет, г. Томск

На основе обзора проблем, связанных с состоянием здоровья и трудностей измерения оптимального мотивационного комплекса, изложен подход изучения, результаты диагностики интересов и степени заинтересованности в физической деятельности ТПУ.

Ускорение темпов перехода от одного технологического уклада к другому, делает жизнь человека более комфортной, стремительное развитие информационных технологий помимо рационализации и повышения эффективности труда вынуждает людей больше потреблять информации (как полезной – образовательной, необходимой в работе и т.п., так и развлекательной). В результате все больше людей страдает от гиподинамии.

Противоречие между комфортом, обеспечивающим экономию двигательных актов и физиологической потребностью в движении, стало нарушать генетически заложенную программу двигательной активности. Уже сегодня мы видим, что около 20–25 % первокурсников (каждый 4–5 студент) Томского политехнического университета (ТПУ) по результатам медицинского освидетельствования имеет низкие показатели здоровья.

Такие студенты направляются в специально-медицинскую группу или группу по лечебной физической культуре. Наибольшее число заболеваний первокурсников (65% от общего числа заболеваний) связано с заболеваниями опорно-двигательного аппарата (ОДА): плоскостопие III степени, сколиозы III степени, аномалии развития опорно-двигательного аппарата. По результатам функциональных тестов, определяющих развитие кардиореспираторной системы среди