

Сутягина Ольга Владимировна

старший преподаватель

Бугрова Анастасия Александровна

студентка

ГБОУ ВО «Нижегородский государственный

инженерно-экономический университет»

г. Княгинино, Нижегородская область

СТАТИСТИЧЕСКИЙ АНАЛИЗ ОЦЕНКИ КАЧЕСТВА ТЕСТОВЫХ ЗАДАНИЙ

***Аннотация:** в статье проводится оценка качества тестовых заданий, размещенных в электронной информационно-образовательной среде Moodle, на надежность и валидность. Авторы резюмируют, что тестовые задания, используемые для оценки результатов успеваемости студентов, требуют обязательной проверки на надежность и валидность с целью дальнейшей их корректировки и повышения достоверности полученных результатов.*

***Ключевые слова:** валидность, дистанционное обучение, надежность, тестовые задания.*

В период пандемии, вызванной новой коронавирусной инфекцией, учебные заведения нашей страны были переведены на дистанционный формат обучения и традиционные способы оценивания результатов обучающихся стали недоступны. В этой связи в Нижегородском государственном инженерно-экономическом университете большинство преподавателей осуществляли срез знаний студентов преимущественно с помощью тестовых заданий, размещенных в электронной информационно-образовательной среде вуза на платформе Moodle [3]. Однако в сложившейся эпидемиологической ситуации у многих преподавателей практически отсутствовала возможность заранее подготовиться к контролю и осуществить пробное тестирование ввиду огромной нагрузки, свалившейся на их плечи. Исходя из этого большинство тестовых материалов, используемых для оценивания результатов обучающихся, можно поставить под

сомнение. Значимую роль в истинной оценке качества обучения играют роль сами тестовые задания, которые в обязательном порядке требуют проверку на адекватность.

Цель данной исследовательской работы заключается в оценке качества тестовых заданий, размещенных в ЭИОС вуза, на примере материалов по дисциплине «Теория вероятностей и математическая статистика».

В рамках изучения дисциплины «Теория вероятностей и математическая статистика» были проведено тестирование, содержащее восемнадцать заданий случайного типа, среди 38 студентов, обучающихся по направлениям подготовки 09.03.02 Информационные системы и технологии и 11.03.02 Инфокоммуникационные технологии и системы связи.

Основными критериями оценки качества тестовых заданий на адекватность считаются их надежность и валидность.

Надежность теста – это критерий, который отражает точность тестовых заданий, а также проверяет устойчивость результатов тестирования обучающихся к внешним факторам воздействия и, соответственно, позволяет судить насколько результаты тестирования соответствуют знаниям студентов [1, с. 177]. Надежность можно принять за внутреннюю согласованность каждого тестового задания с результатом за весь тест.

Надежность тестовых заданий возможно оценить мерой корреляции между результатами двукратного тестирования группы студентов, например, через некоторый временной промежуток – ретестовый метод, или использование двух его параллельных форм. Нами был использован метод расщепления, который не предусматривает повторного тестирования, что, в подавляющем большинстве случаев, доступно и удобно. Результаты проведенного тестирования были разбиты на две параллельные части – индивидуальные баллы за четные и нечетные задания, и произведен расчет коэффициента корреляции Пирсона

$$r_{\text{расщ.}} = 0,59275,$$

который требует обязательной корректировки, ввиду укороченности теста ровно в 2 раза, по формуле Спирмена-Брауна

$$r = \frac{2r_{\text{расщ.}}}{1 + r_{\text{расщ.}}} = 0,74431.$$

Желательно, чтобы величина коэффициента корреляции лежала в пределах от 0,8 до 1, что является хорошим показателем надежности теста. В нашем случае коэффициент корреляции принимает значение чуть большее, чем предельно допустимое значение – 0,7, демонстрирующее минимально возможную внутреннюю согласованность элементов теста, и требует исправление, исключение или замену части заданий.

Валидность – это комплексная характеристика теста, которая показывает, насколько пригоден данный тест для измерений знаний студентов по данному разделу, показывает эффективность и практическую полезность теста. Для вычисления валидности сопоставляются полученные результаты тестирования с независимыми данными тех же студентов (в данном случае – результатами текущего контроля). К тому же тест считается критериально валидным, если значение степени сходства результатов тестирования свойственно большей части тестируемых, а сами оценки распределены по нормальному закону.

Проверка гипотезы о нормальном распределении проводилась по критерию χ^2 распределения Пирсона. По результатам тестирования была сформирована матрица в *MS Excel* размерностью 38×18 , где 0 – неверный ответ, 1 – верный ответ. Далее была сформирована выборка (количество правильных ответов каждого студента). По данным были составлены графики теоретического и эмпирического распределений выборки (рис. 1), а также подсчитаны критическое значение $\chi^2_{\text{кр.}} = 11,0705$ при уровне значимости 0,05 и наблюдаемое значение $\chi^2_{\text{набл.}} = 10,8763$.

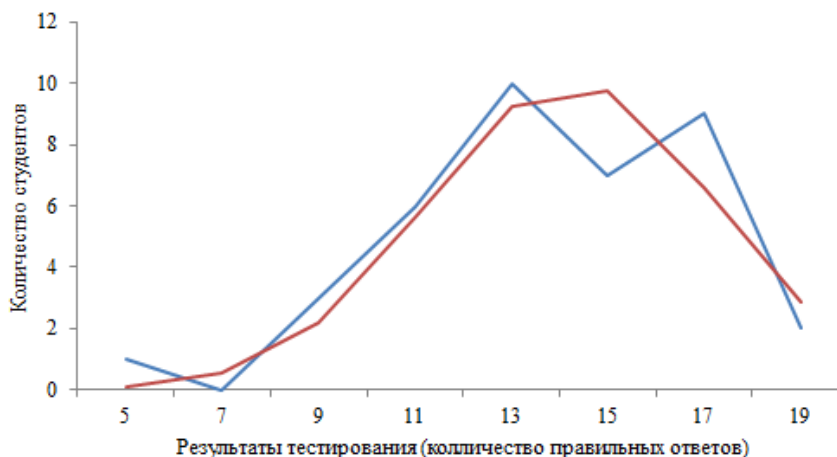


Рис. 1. Графики теоретического и эмпирического распределений количества правильных ответов студентов

Статистический анализ показал, что для рассматриваемой выборки можно принять гипотезу о том, что баллы, набранные студентами за выполнение тестовых заданий по дисциплине, есть случайная величина, распределенная по нормальному закону. Это видно, во-первых, из графика – фактическая частота почти совпадает с теоретической, а, во-вторых, критическое значение критерия больше наблюдаемого ($11,0705 > 10,8763$).

Несмотря на нормальность полученного распределения данная выборка плохо коррелирует с результатами текущей успеваемости студентов, осуществляемой по балльно-рейтинговой системе, включающей в себя оценку текущих контрольных работ, активности на практических и лабораторных занятиях, посещаемости (рис. 2). Коэффициент корреляции Пирсона составил величину равную всего 0,132124098.

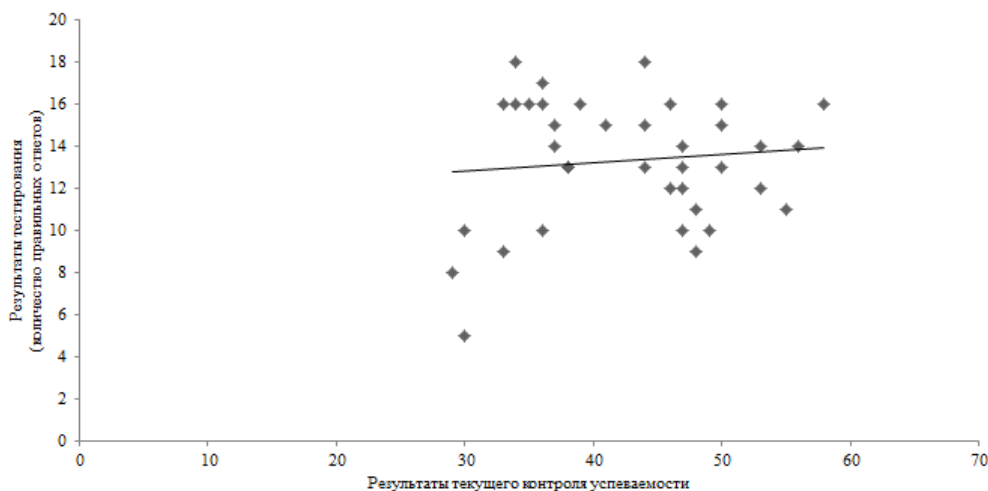


Рис. 2. Регрессионная модель зависимости результатов текущего контроля успеваемости и результатов тестирования

На полученный результат повлиял ряд факторов.

Во-первых, форма проведения тестирования – дистанционная. Вопреки использованию процедуры наблюдения и контроля за тестируемыми, осуществляемой преподавателем с помощью платформы видеосвязи ZOOM (подключение камеры и микрофона, демонстрация рабочего стола), возможностей воспользоваться сторонней помощью у студентов было достаточно. Большинство средств «списывания» основаны на том, что преподаватель один, а студентов много. Дистанционный формат упрощает проведение процедуры оценивания обучающихся, но значительно ухудшает достоверность полученных результатов.

Во-вторых, адекватность тестовых заданий. Размещение заданий на платформе Moodle обладает еще одним достоинством, это возможностью провести анализ не только результатов обучающихся, но и самих тестовых заданий. Для этой цели достаточно просмотреть анализ структуры теста (Название теста → Результаты → Статистика). Проанализируем задания, предлагаемые в рамках промежуточной аттестации по курсу «Теория вероятностей и математическая статистика», по критериям, представленным в среде Moodle:

а) индекс легкости (ИЛ) – демонстрирует процент тестируемых верно ответивших на данное задание;

б) стандартное отклонение (СО) – характеризует рассеяние значений оценок, набранных за данное задание;

в) балл случайного угадывания (ИСУ) – оценка, которую смог бы получить тестируемый при случайном ответе на задание;

г) назначенный вес (НВ) – номинальный балл за правильный ответ за данное задание;

д) эффективный вес (ЭВ) – фактический балл, который желателен за правильный ответ на данное задание.

е) индекс дискриминации (ИД) – показатель взаимосвязи ответа на данное задание с ответами на остальные задания теста.

Из восемнадцати тестовых заданий ИЛ равный 100% не получило ни одно задание, но ИЛ первого задания по теме «Алгебра событий» был очень близок к этой величине, исходя из чего, ввиду его легкости, данное задание было выведено из тестовой базы. Помимо этого подлежат исключению еще два задания: второе – «Виды событий» и шестое – «Теоремы сложения и умножения вероятностей», имеющие отрицательный индекс дискриминации и показывающие, что правильность ответов на данные задания не согласуется с результатом за тест целиком, на них «сильные студенты отвечают хуже чем «слабые». Значительные результаты были получены в результате сравнения назначенного и эффективного весов, что позволило скорректировать оценки, заработанные за верный ответ на данные задания, и увеличить не только ценность теста, но его достоверность и адекватность.

Резюмируя вышесказанное, можно заключить, что тестовые задания, используемые для оценки результатов успеваемости студентов, требуют обязательной проверки на надежность и валидность с целью дальнейшей их корректировки и повышения достоверности полученных результатов. В среде Moodle данные возможности представлены достаточно широко, так внутреннюю согласованность теста можно считать условием его высокой надежности, что является предпосылкой оптимальной валидности теста.

Список литературы

1. Мороз Л.С. Методы определения надежности и валидности тестов для контроля знаний / Л.С. Мороз // Труды БГТУ. Серия VI. Физико-математические науки и информатика. – 2010. – Вып. XVIII. – С. 176–179.
2. Нестеров С.А. Оценка качества тестовых заданий средствами среды дистанционного обучения Moodle / С.А. Нестеров, М.В. Сметанина // Научно-технические ведомости СПбГПУ. – 2013. – №5 (181). – С. 87–92.
3. Электронная информационно-образовательная среда Нижегородского государственного инженерно-экономического университета [Электронный ресурс]. – Режим доступа: <https://ngiei.mcdir.ru/>