

ОЦЕНКА КАЧЕСТВЕННЫХ ХАРАКТЕРИСТИК ПЕДАГОГИЧЕСКОГО ТЕСТА**И.П. Овсянникова, О.Б. Ганак**Белорусский государственный университет пищевых и химических технологий,
policarpovna@mail.ru, ganak@mail.ru

Аннотация. В статье рассматривается тестовый контроль как один из способов оценки результатов знаний. Раскрываются понятия качественных характеристик теста, таких как надежность и валидность. Проведены расчеты коэффициентов надежности и валидности, стандартной ошибки измерения.

Ключевые слова: педагогический тест, надежность, валидность, коэффициент надежности, бинарная матрица, стандартная ошибка.

Одним из современных способов оценки достигнутых результатов студентов при обучении является педагогическое тестирование. Тестовый контроль знаний в вузовской системе обучения позволяет выявить степень овладения студентами учебным материалом, а также дать объективную оценку их знаниям. Инструментом для измерения по шкале достижений студента является правильно сконструированный качественный тест, который соответствует предмету обучения, его задачам и служит развитию системного подхода к изучению учебной дисциплины. Одними из важнейших качественных характеристик теста являются его надежность и валидность. Надежность – это показатель точности педагогического измерения и устойчивости результатов тестирования к воздействию посторонних или случайных факторов. Тест считается надежным, если он дает одни и те же (или очень близкие) показатели для каждого испытуемого при повторном тестировании. При этом необходимо, чтобы сами испытуемые не изменили свой уровень подготовки перед вторым тестированием, а их мотивация к получению наилучших результатов осталась прежней. Валидность характеризует пригодность результатов полученных после тестирования именно для той цели, ради которой и проводилось само тестирование. Чаще всего валидность и надежность оценивают с помощью математических методов и выражают в специальных показателях - коэффициентах валидности и надежности [1, с. 319].

Авторами была произведена оценка надежности тестов при проведении итогового контроля знаний студентов 1 курса по дисциплине «Информатика». Использовался метод расщепления, при котором оценка надежности строится на подсчете корреляции между двумя наборами результатов выполнения одного и того же теста или двух его параллельных форм. Чем выше корреляция, тем надежнее тест. Хорошим коэффициентом надежности теста считается тот, когда показатель колеблется в пределах $0,8 < r < 1$.

Метод расщепления теста удобен в практическом применении, так как ограничивается однократным тестированием. Он основан на допущении параллельности двух половин теста и предполагает деление результатов тестирования на две части: данные по нечетным заданиям теста (X) и по четным (Y). Коэффициент надежности вычисляется как коэффициент корреляции Пирсона по формуле 1 [1, с. 358]:

$$r = \frac{\sum_{i=1}^n x_i y_i - \frac{\sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n}}{\sqrt{\sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i\right)^2}{n}} \sqrt{\sum_{i=1}^n y_i^2 - \frac{\left(\sum_{i=1}^n y_i\right)^2}{n}}} \quad (1)$$

где x_i – индивидуальный балл i -го студента при ответе на тест по нечетным заданиям;
 y_i – индивидуальный балл i -го студента при ответе на тест по четным заданиям;
 n – число студентов.

Поскольку подсчет надежности ведется по расщепленному тесту, который в 2 раза короче, то оценка надежности корректируется по формуле Спирмена – Брауна:

$$r = \frac{2r_{pac}}{1 + r_{pac}} \quad (2)$$

В случае, когда выполнение задания оценивается дихотомически (1 балл – правильно; 0 баллов – неправильно), можно использовать способ оценки надежности с применением формулы Кьюдера – Ричардсона Коэффициент надежности находится по следующей формуле:

$$r = \frac{m}{m-1} \left(1 - \frac{\sum_{j=1}^m p_j q_j}{S_x^2} \right) \quad (3)$$

где m – число заданий в тесте;
 p_j – доля правильных ответов на j -е задание теста;
 q_j – доля неверных ответов на j -е задание, ($q_j=1-p_j$);
 S_x^2 – дисперсия индивидуальных баллов, рассчитываемая из соотношения

$$S_x^2 = \frac{\sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i\right)^2}{n}}{n-1} \quad (4)$$

где n – число студентов;
 x_i – индивидуальный балл i -го студента.

Для итогового контроля знаний 18 студентов был разработан тест, содержащий 20 тестовых заданий. Математико-статистическая обработка результатов тестирования осуществлялась с помощью пакета MS EXCEL, применялась дихотомическая шкала оценивания. По результатам тестирования была сформирована матрица размерностью 18×20, состоящая из 0 (неверный ответ) и 1 (верный ответ). Далее для расчета надежности теста потребовались следующие действия:

- 1) подсчитать индивидуальные баллы испытуемых (сумма баллов по строкам) X_i ;
- 2) найти количество правильных ответов на каждое задание (сумма баллов по столбцам) R_j ;
- 3) подсчитать количество неправильных ответов на каждое задание W_j ;
- 4) долю правильных ответов p_j и долю неправильных ответов на каждое задание q_j ;
- 5) рассчитать коэффициент надежности по формуле (3).

Были получены следующие результаты: бинарная матрица результатов тестирования итогового контроля знаний студентов; коэффициент надежности теста по формуле Кьюдера-Ричардсона, равный 0,8353 и коэффициент надежности по формуле Спирмена-Брауна, равный 0,8148 [2, с.164].

Если значение коэффициента надежности ниже 0,7, то надежность теста считается неудовлетворительной, так как возникает большая погрешность измерений. Анализ результатов свидетельствует о том, что надежность разработанного теста является допустимой. Для увеличения коэффициента надежности некоторые тестовые задания необходимо заменить.

Один из аспектов применения коэффициента надежности связан с определением стандартной ошибки измерения. Ошибка измерения определяется как статистическая величина, отражающая степень отклонения наблюдаемого балла от истинного балла испытуемого. Исходя из проведенных расчетов, доверительный интервал для истинного балла студента со значением 7 будет (6,14; 7,86). Следовательно, истинный балл испытуемого студента находится в промежутке от 6 до 8 баллов [3, с.99].

Оценку валидности всегда получают путем соотнесения характеристик результатов измерения с внешними критериями. В качестве таких критериев могут выступать оценки экспертов при анализе содержания теста и его адекватности целям измерения (содержательная валидность), результатов по другим тестам (конструктивная валидность), успешности дальнейшего обучения или работы (прогностическая валидность) [1, с.342].

Для бинарной матрицы [2, с.165] был вычислен коэффициент валидности по формуле [4, с.110]:

$$V = \frac{\frac{1}{n} \left(\sum_{i=1}^n Y_i \cdot y_i \right) - \bar{Y} \cdot \bar{y}}{S_Y \cdot S_y} \cdot \frac{n}{n-1} \quad (5)$$

где y_i - количество правильных ответов;

\bar{y} - средняя арифметическая тестовых баллов тестируемых 5;

Y_i - экспертная оценка;

\bar{Y} - среднее арифметическое экспертных оценок;

S_Y - стандартное отклонение количества правильных оценок ;

S_y - стандартное отклонение экспертных оценок.

С учетом статистического анализа бинарной матрицы результатов тестирования итогового контроля знаний студентов 1 курса по дисциплине «Информатика», получаем коэффициент валидности

$$V = \frac{\frac{1}{6} \cdot 155 - 4,83 \cdot 5}{1,36 \cdot 1,89} \cdot \frac{6}{6-1} = 0,7713 \quad (6)$$

Если V принимает значение в диапазоне от 0,6 до 1 – валидность считается высокой; если V принимает значение в диапазоне от 0,3 до 0,6 – валидность считается средней; если V принимает значение меньше 0,3 – валидность считается низкой. На основании расчетов можно сделать вывод о том, что валидность рассмотренного теста по дисциплине «Информатика», высокая.

Исходя из полученных расчетов показателей надежности и валидности, можно сделать вывод, что тест достигает своих результатов, т.е. разработчики теста справились со своим заданием – разработкой качественного теста по данной тематики.

Список использованных источников

1. Чельшкова, М.Б. Теория и практика конструирования педагогических тестов: учебное пособие. М.: Логос, 2002 - 432с
2. Овсянникова, И. П. Анализ надежности теста для оценки учебных компетенций студентов / И. П. Овсянникова, О. Б. Ганак // Качество подготовки специалистов в техническом университете: проблемы, перспективы, инновационные подходы : материалы V Международной научно-методической конференции, Могилев, 19-20 ноября 2020 г. / Учреждение образования «Могилевский государственный университет продовольствия»; редкол.: А. С. Носиков (отв. ред.) [и др.] - Могилев: МГУП, 2020. - С. 164 – 166.
3. Ганак, О.Б. Надежность педагогического теста и стандартная ошибка измерения / О. Б. Ганак, И. П. Овсянникова // Качество подготовки специалистов в техническом университете: проблемы, перспективы, инновационные подходы: материалы VI Международной научно-методической конференции 18 ноября 2022 г. / Учреждение образования «Белорусский государственный университет пищевых и химических технологий»; редкол.: А. С. Носиков (отв. ред.) [и др.]. – Могилев: БГУТ, 2022. - С. 99 – 100.
4. Овсянникова, И. П. Анализ валидности теста для оценки учебных компетенций студентов/ И. П. Овсянникова, О. Б. Ганак //Техника и технология пищевых производств: материалы XV Юбилейной Международ. науч.-техн. конф., 19–20 апреля 2023 г., в 2 т., Могилев / Учреждение образования «Белорусский государственный университет пищевых и химических технологий»; редкол.: А. В. Акулич (отв. ред.) [и др.]. – Могилев: БГУТ, 2023. -Т.2 – С. 110-111.