

Лещенко В.Г., Шеламова М.А., Мансуров В.А.
ОЦЕНКА НАДЕЖНОСТИ ТЕСТОВЫХ МЕТОДИК В ОБУЧЕНИИ
Белорусский государственный медицинский университет
Минск, Беларусь

Аннотация. Рассматриваются принципы построения контролирующих и обучающих тестов для оценки усвоения слушателями учебного материала. Наряду с качеством самих тестовых заданий, важным является и алгоритм определения качества и надежности теста в целом. Как слишком трудные, так и слишком легкие задания не позволяют корректно оценить уровень знаний учащихся. Для оценки сложности и надежности теста рекомендовано использовать среднее значение и медиану распределения всех полученных оценок.

Ключевые слова: качество, валидность, надежность тестов.

Leschenko V.G., Shelamova M.A., Mansurov V.A.
ASSESSMENT OF THE RELIABILITY OF TEST METHODS IN TRAINING
Belarusian State Medical University
Minsk, Belarus

Abstract. The principles of development of control and training tests for assessment of mastering of educational material by students are considered. Alongside with the quality of test tasks themselves, the algorithm for determining the quality and reliability of a test as a whole is also important. Both too difficult and too easy tasks do not allow to assess correctly the level of students' knowledge. To assess the complexity and reliability of a test, it is recommended to use the average value and the distribution median of all the obtained scores.

Keywords: quality, validity, reliability of tests.

Развиваемый в настоящее время компетентностный подход к образованию ставит вопрос о разработке соответствующих методов контроля успешности обучения учащихся. Одним из педагогических методов оценки знаний учащихся является тестирование, наряду с мониторингом и рейтингом. При этом основным формализованным средством контроля качества обучения является тестирование.

Этот метод основан на предъявлении испытуемым стандартизованных заданий, которые позволяют оценить их знания и умения, и имеет ряд преимуществ: высокая научная обоснованность заданий (их соответствие программе), технологичность, наличие одинаковых для всех испытуемых правил проведения и интерпретации результатов.

В настоящее время тестирование проводится в основном с использованием технических средств – прежде всего персональных компьютеров. Валидность тестов в значительной мере обеспечивается большим количеством (пулом) заданий разного уровня сложности, разработанных преподавателями заранее; компьютер по специальным программам выдает из этого пула заданий

индивидуальный набор тестов для каждого испытуемого. Задания выводятся на экран монитора, учащийся выбирает или вычисляет вариант ответа, а компьютер их анализирует и по введенным заранее программам выставляет оценку.

Принципы построения тестовых заданий хорошо известны [1]:

- коллегиальность подготовки тестов - позволяет уменьшить субъективность оценки знаний;
- подбор типов и трудности тестовых заданий;
- формирование наборов адекватных по сложности заданий, оценки их сложности и введения «веса» коэффициента;
- централизованное накопление заданий в базе данных системы тестирования.

Выработка алгоритма выставления результирующей оценки.

Важнейшим является принцип построения тестов по методике, обеспечивающей выполнение требований соответствующего образовательного стандарта.

В настоящее время можно отметить следующие виды компьютерных тест-программ:

- Контролирующие тесты- позволяют оценить степень усвоения учащимся учебного материала;
- Обучающие тестирующие программы - предполагают объединение процессов контроля знаний и одновременного обучения, часто с использованием гипертекстовых систем;
- Моделирующие – требуют от учащихся правильного воспроизведения алгоритма действий или рассуждений для решения поставленной задачи;
- Тренажеры – предназначены для отработки и закрепления технических навыков решения тех или иных задач;

Типы предъявляемых заданий:

- закрытая форма;
- открытая форма;
- установление соответствия;
- установление последовательности.

Методы оценки знаний по тестовому контролю:

- на основе количественных критериев – предполагает использование некой количественной шкалы, обычно суммы баллов, полученных за правильные ответы на тестовые задания;
- на основе вероятностных критериев – определение вероятности правильного ответа испытуемого в зависимости от уровня его подготовленности и сложности тестового задания;
- на основе классификационных критериев – предполагают отнесение испытуемого субъекта к определенному устойчивому классу с учетом совокупности некоторых признаков.

Для успешности теста очень важным является подбор заданий по индексу трудности. При слишком трудных заданиях резко снижаются валидность и надежность теста, а слишком простые задания ведут к малой вариативности результатов и не позволяют корректно оценить уровень знаний учащихся.

Наряду с качеством тестовых заданий, важным является и алгоритм определения результатов (итоговой оценки) тестирования, т.е. качества теста.

Часто в качестве критерия используют индекс трудности теста $\lambda = (1 - r)$, который определяется относительным числом учеников r , давшим правильный ответ на данный тест. Например, если на тест правильно ответили только 20% учащихся ($r = 0,2$), то $\lambda = 0,8$ и тест следует считать трудным, а если правильно ответили 80%, то $\lambda = 0,2$ и тест считается легким.

Существуют и более сложные метод расчета *коэффициента надежности теста*.

Тесты, используемые в нашей практике, обычно содержат 20–30 заданий, и оцениваются по 10-бальной шкале, поэтому для определения сложности и надежности теста мы предлагаем рассчитать статистические параметры распределения этих оценок, такие как среднее значение и медиана этого распределения. Эти параметры обычно близки (при нормальном, гауссовом распределении, они совпадают). На наш взгляд тест можно считать валидным и надежным, если эти значения лежат в интервале 7–8; меньшие значения этих параметров (4–6) указывают, что тест является сложным, а при значениях 8–9 - легким.

Такой подход к оценке сложности тестов, используемых на кафедре медицинской и биологической физики показал, например, что тест по «Биомеханике и биореологии» для англоязычных студентов оказался сравнительно легким, т.к. средняя оценка равна 8, медиана – 8,5, тогда как тест по «Медицинской электронике», напротив оказался для них сложным: средняя оценка – 5,9, медиана – 6. Очевидно, что оба теста требуют корректировки заданий.

Литература

1. Белоус В.В. Тестовый метод контроля качества обучения и критерии качества образовательных тестов. Обзор / Белоус В.В., Домников А.С., Карпенко А.П. // Наука и образование. – 2011. – №4.