

ДОСТОВЕРНОСТЬ ОЦЕНКИ ЗНАНИЙ ОБУЧАЕМЫХ ПО РЕЗУЛЬТАТАМ ТЕСТИРОВАНИЯ

С.Н. Нефедов, В.И. Пачинин

Институт информационных технологий Белорусского государственного университета информатики и радиоэлектроники, Минск, Беларусь, nefedov_sn@tut.by

Abstract. Questions the reliability of the test results when assessing the knowledge of the trainees. Analysed the major factors that have the greatest impact on the reliability of the tested. A comparison of the classical approach, and ITR when building tests and test results.

Весьма важным аспектом педагогической деятельности является оценивание преподавателем знаний обучаемых. В технологии дистанционного обучения при отсутствии непосредственного контакта обучаемого с преподавателем тестирование становится одним из основных средств контроля знаний, поэтому особенно остро встает проблема создания качественных тестов, которые могли бы быстро, объективно и адекватно измерять уровень знаний обучаемых.

В педагогической среде достаточно распространено негативное отношение к тестам. Однако, традиционная форма оценивания уровня знаний в форме опроса, экзамена, проводимого человеком, весьма субъективна. Столь неточный «измерительный прибор», каковым является человек, существенно снижает эффективность диагностики учебного процесса. По этой причине, в качестве контрольно-измерительного мероприятия предпочтительно выбирать тестирование даже при обычной форме обучения.

Сам процесс тестирования учебных достижений разбивается на три этапа:

- 1) разработка теста;
- 2) процедура тестирования;
- 3) обработка и интерпретация результатов тестирования.

Важность тестов в учебном процессе давно осознана за рубежом. Там теория и практика тестирования развиваются уже сотню лет. К сожалению, в 1936 году в СССР вышло постановление ЦК ВКП(б) «О педологических извращениях в системе Наркомпросов». Тестирование было признано противоречащим советской идеологии со всеми вытекающими последствиями. В послевоенные годы, работы в области тестирования начали возрождаться, а в 70-80-х годах прошлого столетия педагогическое тестирование стало усиленно развиваться в рамках технологии программированного обучения. В настоящее время данные работы получают все большее развитие на постсоветском пространстве. Так в РФ в 2002 г. Был создан "Институт педагогических измерений Министерства образования Российской Федерации", с 2004 года издается журнал «Педагогические измерения» и т.д.

Педагогический тест является измерительным инструментом и это должен быть качественный инструмент, позволяющий получать достоверные результаты. В создании качественных педагогических тестов чрезвычайно велика роль стандартов, которым должны соответствовать педагогические тестовые материалы.

Педагогический тест, в отличие от, например, контрольной работы, можно рассматривать как своеобразный измерительный инструмент определенной разрешающей силы и точности.

Измерение — операция для определения отношения одной (измеряемой) величины к другой однородной величине, которая берётся за единицу. Получившееся значение будет численным значением измеряемой величины.

Педагогическое измерение - это процесс установления соответствия между оцениваемыми характеристиками обучаемых и точками эмпирической шкалы, в

которой отношения между различными оценками характеристик выражены свойствами числового ряда.

Педагогический тест - это система тестовых заданий различной трудности, которая позволяет качественно и эффективно измерить уровень и структуру подготовленности испытуемых.

Достоверность результатов тестирования, которая характеризует степень соответствия результатов тестирования фактическим знаниям испытуемых, зависит от многих факторов. В [1] приведено 22 наиболее распространенных влияющих факторов. Эти факторы вносят различный вклад в достоверность результатов тестирования. К наиболее существенным факторам следует отнести:

1. качество теста (качество инструментария), которое определяется степенью соответствия количества тестовых заданий объему оцениваемых знаний, количеством тестовых заданий в тесте и порядком их выбора, формой тестовых заданий и их реализацией и др.;

2. характеристики испытуемого (его знания и способности, физическое эмоциональное состояние и здоровье, мотивация при проведении тестирования и др.);

3. условия проведения тестирования и время на выполнение тестовых заданий.

Данные факторы оказывают влияние на достоверность результатов тестирования, а также могут учитываться при составлении тестов.

Форма тестовых заданий – это способ организации, упорядочения и существования содержания теста. Соединившись с содержанием, форма придает заданию конкретный облик, или иначе, содержание принимает определенную форму. Форма может рассматриваться как инвариант. В тестах по разным учебным дисциплинам может использоваться одна и та же форма заданий.

Умение правильно, ясно и лаконично оформлять задания есть необходимое, но недостаточное условие для создания теста. Это обусловлено тем, что даже хорошо сформулированное задание еще не является тестовым. Необходима еще достаточно трудоемкая процедура эмпирической проверки задания, статистическая обработка результатов его применения. Иногда возможно чисто визуально, на экспертном уровне, определить будет ли задание тестовым. Однако нередко бывает, что задание в тестовой форме не выдерживает статистической проверки, то есть не становится тестовым заданием.

Все задания разбиваются на две большие группы - задания в открытой форме и задания в закрытой форме. В основу классификации положено наличие или отсутствие ввода дополнительной информации испытуемым. Если дополнительная информация нужна, то это задание в открытой форме. Если не нужна, то это задание в закрытой форме.

Под надежностью, или релябильностью, измерения понимается степень надежности, или точности, с какой может быть измерен тот или иной конкретный признак. Надежность теста характеризует воспроизводимость его результатов. Надежность характеризуется коэффициентом надежности. Коэффициент надежности, это корреляционный коэффициент, показывающий степень совпадения результатов тестирования осуществленного в одинаковых условиях одним и тем же тестом.

Высокая надежность теста это необходимое, но недостаточное условие получения высококачественного теста. Тест еще должен быть валидным. Валидность – это важнейшая характеристика теста, без указания которой, его нельзя считать измерительным инструментом. Тест может иметь высокую надежность, но низкую валидность. Тест с высокой валидностью обязательно имеет высокую надежность. Если

тест имеет низкую валидность, то применять его нельзя, даже если он имеет высокую надежность.

Выделяют три вида валидности – содержательную, критериальную и конструктивную. При обработке результатов тестов используется либо классическая теория тестирования, либо IRT (Item Response Theory), позволяющая измерять уровень достижений испытуемого в специальных единицах измерения – логитах. В данных теориях используются различные способы расчета показателей надежности теста. Сравнение двух этих подходов приведено в таблице.

Таблица 1 - Сравнение классической теории тестов и IRT

Классическая теория	Item Response Theory
Основная задача теста: Получить значение истинного балла (Т) испытуемого исходя из наблюдаемого результата (Х), с учётом случайной ошибки измерения (Е), откуда вытекает основной постулат классической теории тестирования: $X_i = T_i + E_i$	Основная задача теста: Получить устойчивую объективную оценку латентного параметра уровня знаний испытуемого по исследуемому предмету, не зависящую от конкретного теста.
Возможности: Извлечь первичную информацию о тесте, на основании матрицы результатов апробационного тестирования, а именно: 1) Оценить статистическую сложность заданий 2) Взаимную корреляцию между заданиями теста и корреляцию баллов заданий и внешнего критерия (суммы баллов испытуемых), для определения валидности тестовых заданий 3) Получить оценку надёжности результатов тестирования посредством корреляционного анализа баллов испытуемых по тесту, либо по нескольким его вариантам. 4) Построить доверительный интервал, в пределах которого находится истинный балл испытуемого, либо получить точечную регрессионную оценку. 5) Интерпретировать результаты в терминах Z-шкалы, либо процентной шкалы, то есть определить место (рейтинг) испытуемого в выборке.	Возможности: 1) Установление связи между латентными параметрами испытуемых и наблюдаемыми результатами выполнения теста. Наблюдаемые результаты выполнения теста порождаются взаимодействием двух множеств латентных параметров теста: уровнем знаний испытуемых и трудности заданий. 2) Параметры уровня знаний испытуемого и трудности заданий теста отображаются в единую шкалу логитов, что позволяет реализовать идею адаптивного тестирования. 3) Существует возможность помимо стандартных критериев качества, ввести новый критерий- эффективность теста, путём введения особого класса информационных функций. 4) Реально только в рамках Item Response Theory реализована возможность применения не только рейтинговой, но и интервальной шкалы, а это значит, что уровень подготовки можно оценить количественно.

Литература

1. Майоров А.Н. Теория и практика создания тестов для системы образования. (Как выбирать, создавать и использовать тесты для целей образования). - М., «Интеллект-центр», 2001. - 296с.
2. Аванесов В.С. Методологические и теоретические основы тестового педагогического контроля. Дисс. докт. пед. наук. С-Пб. Госуниверситет, 1994.- 339с.