

Модель повышения надежности компьютерного теста

УДК 681.3(004):378.1©

Е.В. Шевчук, А.В. Шпак

Северо-Казахстанский государственный университет им. М. Козыбаева,
г. Петропавловск, Казахстан,
Лицей, г. Калачинск, Россия

Опубликовано в сб. материалов VI межд. науч. конф., посвящённом памяти
Б.А. Рогозина. Омск, 23 ноября 2018г. с.77-79

Современная тестология оставляет относительно открытым вопрос о выявлении тестируемых, которые, отвечая на тест, просто-напросто гадали. Подходы к определению качества тестового задания, при которых выравнивается доля очень легких и очень трудных вопросов, не меняет статус угаданных ответов.

Некоторые тестологи предлагают ввести перед выставлением оценки коррекцию баллов на угадывание. Но многие тестологи справедливо с этим не согласны, считая, что данная формула одинаково штрафует и тех, кто отвечал на тест, не прибегая к такому методу, как «угадывание», и тех, кто действительно отвечал на тестовые задания «наугад» [1].

В настоящей статье предлагается новый подход, при котором не происходит повального штрафования тестируемых путем коррекции баллов на угадывание. Существует категория тестовых заданий, на которые тестируемые по различным причинам преимущественно отвечают, не задумываясь, или гадая. Однако при традиционной статистической обработке теста эти задания зачастую не попадают в разряд мало отвеченных, или ненадежных. Без выявления таких «гадаемых» заданий перед определением уровня надежности и валидности теста традиционными методами, существует вероятность того, что надежность и валидность теста будет определена неправильно.

Предлагаемый в настоящей статье алгоритм выявления «гадаемых» тестовых заданий позволяет выявить группу тестируемых, которые, с большой долей вероятности, гадали, и группу ненадежных тестовых заданий, на которые большинство тестируемых отвечали по разным причинам (психологическим, методическим и др.) наугад.

Алгоритм определения «угадываемых» тестовых заданий (на примере закрытого компьютерного теста с одним вариантом правильного ответа из пяти).

Моделируется кластер тестируемых, которые отвечают на тест, не задумываясь, т. е. случайным образом выбирают ответы. Если стандартное тестовое задание состоит из вопроса и 5 вариантов ответов, то случайный тестируемый набирает наугад результат $20 \pm 4\%$ от максимума в сто процентов. При этом, используя методы математической статистики [2], можно сделать вывод, что количество одновременно выбравших правильные ответы на одно тестовое задание составит в среднем $20 \pm 6\%$ на каждом задании, при объеме заданий ~ 100 , объеме тестируемых ~ 300 .

77

Таким образом, при оценке качества тестов логично будет исключить из рассмотрения те тестовые задания, на которые ответили $20 \pm 6\%$ тестируемых от их общего числа, и логично предположить, что тестируемые, набравшие $20 \pm 4\%$ от общего числа баллов, скорее всего гадали.

Далее, на модели «гадающих» тестируемых вычисляется среднее число попарных перекрытий тестовых заданий, т. е. для каждой пары тестовых заданий вычисляется число одинаковых ответов (например, если студент одновременно отвечает или не отвечает на эти 2 вопроса). С надежностью 99 % это число – $67 \pm 1\%$ от общего числа тестируемых.

Из реальных «подозрительных» тестовых заданий, выявленных по схеме, изложенной в пункте 1 ($20 \pm 6\%$) выделяют те, число попарных перекрытий которых $67 \pm 1\%$. Ответы на остальные тестовые задания получились неслучайным образом (например, списыванием или действительно, не гаданием, а попыткой думать).

Совокупность оставшихся реальных подозрительных тестовых заданий (перечисление 3) проверяется на предмет схожести распределения по критерию Фишера с совокупностью смоделированных случайных тестовых заданий. В случае положительного результата с надежностью, принятой при вычислении статистики Фишера, можно принять решение о том, что большинство тестируемых отвечали на них случайным образом, гадая.

Вполне логично будет исключить из рассмотрения эти тестовые задания, как имеющие «шумовой эффект» перед процедурой традиционной статистической оценки теста на надежность и валидность. Исключив из рассмотрения ненадежные тестовые задания, можно подсчитывать результаты компьютерного тестирования. Причем нижней планкой для тестируемых в этом случае должен быть результат не ниже $20\% + 2 \times 4\% = 28\%$ от максимума (на конечном интервале разброс просуммированного фактора формируемого большим количеством независимых причин оценивается как 2σ [2]).

Также следует учитывать, что в интервал от 12 до 28 % попадут те тестируемые, которые «гадали», ниже 12 % – те, которые заблуждались сознательно. Описанная выше технология повышения надежности компьютерного тестирования запатентована авторами статьи и реализована в рамках информационно-образовательной среды вуза [3; 4].

Литература

- 1 Аванесов В.С. Научные проблемы тестового контроля знаний. М.: Иссл. Центр, 1994. - 135 с.
- 2 Кобзарь А.И. Прикладная математическая статистика. М.: Физматлит, 2006 816 с.
- 3 Мутанов Г.М., Шевчук Е.В. Экспертная система оценки знаний методом тестирования. Алматы: Казак университеті, 2012 152 с.
- 4 Шпак А.В., Шевчук Е.В. Информационно-образовательная среда. Опыт и перспективы. LAMBERT Academic, 2016, 99 с.