

В.С. Ноздренков,  
 ассистент,  
 И.Л. Лебединский,  
 к.т.н., доцент

Сумский государственный университет, г. Сумы

## ОЦЕНКА КАЧЕСТВА ТЕСТОВЫХ МЕТОДИК НА ОСНОВЕ КРИТЕРИЕВ ВАЛИДНОСТИ И НАДЕЖНОСТИ

Важным шагом в составлении теста является проверка качества предложенных задач. Если задания не достаточно качественны, то велика вероятность того, что с ними одинаково справится (или не справится) как отличник, так и двоечник. А это искажает истинную картину успеваемости. Поэтому необходимо произвести оценку теста на валидность и надежность.

Из литературы [1] известно, что проверка теста на валидность начинается с построения матрицы результатов. Матрица имеет размерность  $M \times N$ , где  $M$  – число заданий,  $N$  – число испытуемых. Обозначим через  $x_{ij} \in [0,1]$  числовую оценку успешности выполнения  $j$ -го задания, выполненного  $i$ -м испытуемым. Для каждого

испытуемого определяется суммарный балл  $X_i = \sum_{j=1}^M x_{ij}$ ,

определяющий успешность, а для каждого задания – суммарное

количество набранных баллов  $R_j = \sum_{i=1}^N x_{ij}$ . На основании

полученных данных строят упорядоченную матрицу, в которой испытуемые располагаются по их успешности, а задания – по их трудности.

После упорядочения матрицы результатов проводят ее обработку. Главной целью этого является расчет показателей связи между собой и с суммой баллов испытуемых. Основу этого составляет коэффициент корреляции Пирсона  $r_{yz}$ , который определяется по формуле

$$r_{yz} = \frac{SP_{yz}}{\sqrt{SS_y \cdot SS_z}}, \quad (1)$$

где  $y$  и  $z$  – параметры, связь между которыми рассчитывается;  $SP_{yz}$  – сумма произведений отклонений  $y$  и  $z$  от их средних значений;  $SS_y$  или  $SS_z$  – сумма квадратов отклонений по  $y$  или  $z$  от его предыдущего значения. Чем выше значения коэффициентов корреляции, тем выше качество теста, при этом нормой считается, если их значения не меньше 0,3 [1].

Для определения указанных выше величин используются формулы:

$$SS_y = SS_{x_{ik}} = \sum_{i=1}^N x_{ik}^2 - \frac{(\sum_{i=1}^N x_{ik})^2}{N} \quad (2)$$

Для связи между  $k$ -м заданием и суммой баллов

$$SS_z = SS_{X_i} = \sum_{i=1}^N X_i^2 - \frac{(\sum_{i=1}^N X_i)^2}{N}, \quad (3)$$

для связи между  $k$ -м и  $m$ -м заданиями

$$SS_z = SS_{x_{im}} = \sum_{i=1}^N x_{im}^2 - \frac{(\sum_{i=1}^N x_{im})^2}{N}. \quad (4)$$

Величина  $SP_{yz}$  вычисляется следующим образом:

$$SP_{yz} = SP_{x_{ik}z} = \sum_{i=1}^N x_{ik}z - \frac{\sum_{i=1}^N x_{im} \sum_{i=1}^N z}{N}. \quad (5)$$

На основании результатов расчетов строится корреляционная матрица, компонентами  $r_{ij}$  которой являются коэффициенты

связей между заданиями. Кроме того, рассчитываются коэффициенты  $r_{x_j X_i}$  ( $i = 1, \dots, N; j = 1, \dots, M$ ) связей всех заданий с суммой баллов для каждого испытуемого. Если для  $k$ -го задания  $r_{x_j X_i} > 0.3$ , то связь считается достаточной, и  $k$ -е задание остается в тесте. Если же будет  $r_{x_j X_i} < 0.3$ , то связь считается не достаточной,  $k$ -е задание необходимо удалить из теста.

Аналогично рассчитывается коэффициент связи между двумя заданиями, например,  $k$ -м и  $m$ -м, т.е. элементы корреляционной матрицы. Необходимые формулы имеют следующий вид:

$$r_{x_{ik}x_{im}} = \frac{SP_{x_{ik}x_{im}}}{\sqrt{SS_{x_{ik}} \cdot SS_{x_{im}}}}; \quad (6)$$

$$SS_{x_{ik}} = \sum_{i=1}^N x_{ik}^2 - \frac{(\sum_{i=1}^N x_{ik})^2}{N}; \quad (7)$$

$$SS_{x_{im}} = \sum_{i=1}^N x_{im}^2 - \frac{(\sum_{i=1}^N x_{im})^2}{N}; \quad (8)$$

$$SP_{x_{ik}x_{im}} = \sum_{i=1}^N x_{ik} \cdot x_{im} - \frac{\sum_{i=1}^N x_{ik} \cdot \sum_{i=1}^N x_{im}}{N}. \quad (9)$$

Рассчитанная матрица коэффициентов связи между заданиями имеет размерность  $M \times M$ , поскольку число столбцов и строк равно числу заданий, и симметрична, так как  $r_{ij} = r_{ji}$ , т.е. эти коэффициенты выражают связь между одними и теми же заданиями. Диагональные компоненты матрицы равны единице, так как выражают связь задания с самим собой.

Одной из важнейших интегральных характеристик теста является средний коэффициент корреляции задания со всеми заданиями

$$\bar{r}_j = \frac{1}{M} \sum_{i=1}^M r_{ij}.$$

Нормой считаются значения

коэффициентов корреляции не ниже 0,3 [1]. В этом случае тест валиден по способности дифференцировать испытуемых.

Важной характеристикой теста является его надежность. О высокой надежности метода говорят в том случае, когда метод точно измеряет то свойство, для измерения которого он предназначен. В качестве критериев точности можно отметить следующие:

- при повторном применении метода к тем же самым испытуемым в одних и тех же условиях через определенный интервал времени результаты обоих тестирований существенно не различаются между собой;

- действия случайных посторонних факторов не оказывают существенного влияния на результаты тестирования.

В качестве посторонних факторов можно назвать следующие: эмоциональное состояние и утомление, если они не входят в круг исследуемых характеристик, температура, освещенность помещения и др. Такие посторонние факторы еще называют факторами нестабильности измерительной процедуры.

При повторном применении метода к тем же самым испытуемым через определенный интервал времени в измененных условиях результаты обоих тестирований существенно не различаются между собой.

Существуют различные методы оценки надежности.

Ретестовый метод — повторное тестирование выборки испытуемых одним и тем же тестом через определенный интервал времени при одних и тех же условиях. За индекс надежности принимается коэффициент корреляции между результатами двух тестирований.

Надежность взаимозаменяемых форм — повторное тестирование выборки испытуемых параллельной формой теста через минимальный интервал времени при одних и тех же условиях. За индекс надежности принимается коэффициент

корреляции между результатами тестирования двумя параллельными формами теста. Высокий коэффициент корреляции и большой интервал между двумя испытаниями свидетельствует о высокой надежности теста [2].

Метод расщепления состоит в том, что тест разбивают на две сопоставимые части. Респондент выполняет задания этих двух частей в течение одного сеанса с получением двух результатов (например, тест Равена содержит две части: четные и нечетные задания). За индекс надежности принимается коэффициент корреляции Спирмана-Брауна между результатами тестирования двумя этими частями. Он называется коэффициентом внутренней согласованности теста. Для его расчета используется формула:

$$r_i = \frac{2r_{1/2}}{1 + r_{1/2}}, \quad (10)$$

где  $r_{1/2}$  — коэффициент корреляции Пирсона, характеризующий в данном случае связь между двумя половинками теста — четных и нечетных заданий. Он рассчитывается по формуле (1).

Предложенные методики оценки валидности и надежности тестов позволяют выполнить оценку качества тестовых заданий.

### Список литературы

1. Аванесов В.С. Научные основы тестового контроля знаний. М.: Исследовательский центр, 1994. — 135 с.
2. Аванесов В.С. Композиция тестовых заданий. Учебная книга для преподавателей вузов, учителей школ, аспирантов и студентов педвузов. М.: Адепт, 1998. — 217 с.