

МЕТОДЫ ОПРЕДЕЛЕНИЯ НАДЕЖНОСТИ ТЕСТОВ ДЛЯ УСТАНОВЛЕНИЯ ЭФФЕКТИВНОСТИ ТЕСТОВОГО КОНТРОЛЯ

Аннотация. В статье приводятся разные трактовки термина «надежность» тестовой методики контроля. Перечисляются факторы, влияющие на надежность теста и устойчивость тестовых показателей. Даются рекомендации, как можно повысить надежность психодиагностической методики и устранить влияние на результаты теста случайных факторов. Характеризуются три типа надежности теста (ретестовая надежность, надежность параллельных форм и надежность частей теста) и показываются методы определения каждого из них.

Ключевые слова: тестирование, надежность, ретестовая надежность, надежность параллельных форм, надежность частей теста, коэффициент корреляции.

Тестирование как инструмент контроля знаний, умений и навыков студентов-бакалавров стало неотъемлемой составляющей процесса обучения в вузах. Чтобы тест адекватно оценивал знания, умения и навыки, он должен обладать определенными характеристиками. Одной из таких характеристик и важным количественным показателем качества теста является надежность. Она отражает объективность и точность педагогического измерения, степень постоянства, стабильности результатов тестирования.

К. М. Гуревич определяет надежность как «признак психологического теста, который позволяет судить о том, насколько внушают доверие полученные при его применении данные, насколько обоснованно ожидание исследователя, что при сохранении известного минимума неизменных условий, испытуемые в выборке останутся и при повторных испытаниях примерно на тех же порядковых (ранговых) местах» [10, 163]. А. Анастаси под надежностью понимает «согласованность результатов теста, получаемых при повторном его применении к тем же испытуемым в различные моменты времени, с использованием разных наборов эквивалентных заданий или при изменении других условий обследования» [1, 97].

В словаре-справочнике Л. Ф. Бурлачука и С. М. Морозова [9] термин «надежность» рассматривается в широком и узком смысле. В широком смысле под надежностью понимается характеристика того, в какой степени выявленные у испытуемых различия по тестовым результатам являются отражением действительных различий в измеряемых свойствах и в какой мере они могут быть приписаны случайным ошибкам.

В узком смысле под надежностью подразумевается степень согласованности результатов теста, получаемых при первичном и повторном его применении, по отношению к тем же испытуемым в различные моменты времени, с использованием разных, но сопоставимых по характеру наборов тестовых заданий или при других изменениях условий обследования.

На устойчивость тестовых показателей и на надежность теста могут влиять следующие факторы:

- репрезентативная выборка объектов тестирования;
- степень однородности уровня подготовки испытуемых;
- нестабильность диагностируемого свойства;
- небольшое количество тестовых заданий, так как чем полнее охватывают вопросы теста измеряемые функции, тем надежнее тест;
- разнородность заданий, так как чем разнообразнее задания и чем больше их дифференцирующая сила, тем надежнее тест;
- трудность заданий, включение в тест слишком трудных или слишком легких заданий, то есть, тест должен иметь равномерное распределение заданий по трудности;
- небольшое время, отводимое на выполнение теста;
- непонимание испытуемыми инструкции или вопросов теста;
- случайность правильных ответов, умение выполнять тесты вообще;

- самочувствие испытуемых во время тестирования, степень усталости, интерес и мотивация;
- условия проведения теста (разное время дня, когда проводятся эксперименты, разная освещенность помещения, наличие или отсутствие посторонних шумов);
- различия в манере поведения экспериментатора (по-разному предъявляет инструкции и стимулирует выполнение задания);
- элементов субъективности в способах оценки и интерпретации результатов (когда ведется протоколирование ответов испытуемых, оцениваются ответы по степени полноты, оригинальности).

Надежность теста зависит от качества и числа включенных в него вопросов и от широты диапазона знаний выборки испытуемых. Чем больше вопросов в тесте при их одинаково высоком качестве, тем выше его надежность. Наличие в тесте плохих вопросов, которые не помогают отличить «знающих» от «незнающих», уменьшает общее число эффективных вопросов и тем самым показатель надежности. Слишком трудные для данного контингента вопросы ведут к увеличению числа ответов наугад, это снижает надежность теста. Слишком легкие вопросы, на которые могут правильно ответить все или почти все испытуемые, не помогают отличить «знающих» от «незнающих» и тоже снижают надежность. Чем больше «пестрота» знаний среди испытуемых, на которых определяется надежность, тем выше показатель надежности. При очень однородном контингенте этот показатель будет ниже.

Если учитывать эти факторы и устранить условия, снижающие точность измерения, можно добиться приемлемого уровня надежности. Одним из важнейших средств повышения надежности психодиагностической методики и устранения влияния на результаты теста случайных факторов является единообразие процедуры обследования, его строгая регламентация, то есть одинаковые для всех испытуемых обстановка и условия работы, однотипный характер инструкций, одинаковые для всех временные ограничения, способы контакта с испытуемыми, порядок предъявления заданий и т. д.

Анализ литературы по проблеме надежности тестов [1–11] позволил выделить три типа надежности теста: ретестовую надежность, надежность параллельных форм и надежность частей теста.

Ретестовая надежность используется для определения стабильности изучаемого признака путем повторного обследования испытуемых с помощью одного и того же теста. Этот показатель вычисляется по соответствию результатов первого и второго обследований или по сохранению испытуемыми ранговых мест в выборке при ретесте.

Тестирование проводится дважды в одной и той же группе испытуемых. Период времени между первым и вторым тестированием не должен быть очень большим. Коэффициент корреляции, который называется коэффициентом стабильности, выражает величину стабильности результатов.

Он уменьшается с увеличением промежутка времени между проведением двух тестов, так как у испытуемых за этот промежуток времени может произойти сдвиг в развитии измеряемого качества. Но если этот отрезок времени будет слишком коротким, испытуемые могут в процессе повторного тестирования вспомнить свои ответы и механически воспроизвести их.

Для вычисления ретестовой надежности можно подсчитать коэффициент корреляции Пирсона для интервальных шкал или коэффициент ранговой корреляции Спирмена для шкал порядка.

Коэффициент Пирсона высчитывается по следующей формуле:

$$r = \frac{\sum x_{1i}x_{2i} - \frac{\sum x_{1i} \sum x_{2i}}{n}}{\sqrt{(\sum x_{1i}^2 - (\sum x_{1i})^2 / n)(\sum x_{2i}^2 - (\sum x_{2i})^2 / n)}},$$

где x_{1i} – тестовый балл i -того испытуемого при первом измерении; x_{2i} – тестовый балл того же испытуемого при повторном измерении; n – количество испытуемых.

Коэффициент Спирмена высчитывается по формуле:

$$p = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)},$$

где d_i – разность рангов i -того испытуемого в первом и втором ранговом ряду; n – количество испытуемых.

Использование метода ретестовой надежности вызывает у методистов ряд возражений. Во-первых, двукратное применение одного и того же теста уменьшает количество контрольных вопросов, с помощью которых проверяется усвоение учебного материала обучающимися; повторное тестирование не дает новой информации о степени усвоения обучающимися данного материала. Во-вторых, после проведения контрольной работы учащиеся обычно обсуждают ее, вновь рассматривают некоторые наиболее трудные вопросы, обращаются к учебникам и учебным пособиям, чтобы установить, правильно ли они ответили на те или иные вопросы, таким образом, они оказываются более подготовленными к повторному тестированию, чем к первому. В-третьих, при повторном тестировании почти не реализуется обучающая функция контроля; выполнение одного и того же задания снижает интерес обучающихся к выполнению теста.

Надежность параллельных (эквивалентных) форм определяется с помощью взаимозаменяемых форм теста. Дается аналогичный вариант уже проведенного теста, позволяющий проверить правильность отбора заданий, оптимальность структуры теста, возможность построения различных вариантов на основе типового теста. Для определения этого вида надежности подготавливаются два теста, примерно одинаковые по охвату материала и степени трудности. Оба теста проводятся в одной и той же группе испытуемых. Коэффициент корреляции между результатами испытуемых является коэффициентом эквивалентности форм теста. Он учитывает стабильность результатов, так как второй тест может проводиться сразу после

первого или через некоторый промежуток времени, а также постоянство ответов испытуемых, потому что вопросы с одинаковыми номерами обоих тестов содержат однотипные задания. Использование методики эквивалентных форм часто оказывается затруднительным, потому что составление двух эквивалентных форм теста требует большого времени и искусства составителей. Каждый вопрос первого теста должен иметь точный аналог во втором тесте, но не должен быть тождественным ему. Это относится не только к содержанию вопроса, но и к степени его трудности, форме, в которой излагается задание и набор ответов для выбора.

Для проверки *надежности частей теста*, говорящей о его однородности, внутренней согласованности, задания делятся на четные и нечетные. Испытуемые выполняют один тест, но для подсчета результатов он разбивается на два субтеста и подсчет баллов осуществляется отдельно по каждому из них. Для определения коэффициента внутреннего постоянства используется формула Спирмена-Брауна:

$$r_t = \frac{2r_0}{1 + r_0},$$

где r_t – коэффициент надежности всего теста, а r_0 – начальный коэффициент надежности. Методика признается надежной, если полученный коэффициент не ниже 0,75-0,85.

Основная трудность при использовании метода определения внутренней согласованности теста состоит в том, чтобы разбить тест на две равноценные части. Для этого нельзя использовать первую и вторую половины теста, так как в большинстве тестов вопросы в конце обычно более трудные, чем в начале, не всем испытуемым удастся ответить на все вопросы теста до самого конца.

Индекс трудности задания можно вычислить по формуле:

$$P = \frac{R}{N} \times 100\%,$$

где R – количество правильных ответов, N – общее количество ответов.

Чем больше индекс трудности задания, тем оно легче. Индекс трудности должен лежать в интервале от 20 до 80.

Среднюю трудность теста можно вычислить по формуле:

$$P_0 = \frac{R}{N \times n} \times 100\%,$$

где R – количество правильных ответов, N – общее число испытуемых, n – количество заданий в тесте.

Таким образом, надежность является важной характеристикой, которую нужно учитывать при разработке тестов по разным дисциплинам. Качество диагностических материалов является ключевым моментом контроля знаний, умений и навыков студентов-бакалавров. Во избежание неадекватной оценки знаний, умений и навыков необходимо использовать только статистически проверенные и обоснованные тестовые материалы, обладающие достаточно высоким уровнем надежности.

ЛИТЕРАТУРА

1. *Аванесов, В. С.* Композиция тестовых заданий / В. С. Аванесов. – М. : Центр тестирования, 2002. – 240 с.
2. *Айсмонтас, Б. Б.* Общая психология : Схемы / Б. Б. Айсмонтас. – М. : Изд-во ВЛАДОС-ПРЕСС, 2003. – 288 с.
3. *Анастази, А.* Психологическое тестирование. В 2 кн. Книга 1. / А. Анастази ; под ред. К. М. Гуревича, В. И. Лубовского. – М. : Педагогика, 1982. – 320 с.
4. *Балыхина, Т. М.* Педагогический контроль и тестирование по русскому языку как иностранному : история, опыт, проблемы / Т. М. Балыхина, Н. М. Румянцева, М. Б. Челышкова // Тестирование в обучении русскому языку как иностранному : современное состояние и перспективы / под общ. ред. Т. М. Балыхиной. – М. : Московский государственный университет печати, 2003. – С. 18–33.
5. *Батаршев, А. В.* Тестирование : Основной инструментальный практического психолога / А. В. Батаршев. – М. : Дело, 2003. – 240 с.
6. *Бернштейн, М. С.* К методике составления и проверки тестов / М. С. Бернштейн // Вопросы психологии. – 1968. – № 1. – С. 51–66.
7. *Беспалько, В. П.* Основы теории педагогических систем. Проблемы и методы психолого-педагогического обеспечения технических обучающих систем / В. П. Беспалько. – Воронеж : Изд-во Воронежского университета, 1977. – 304 с.
8. *Блейхер, В. М.* Психологическая диагностика интеллекта и личности / В. М. Блейхер, Л. Ф. Бурлачук. – Киев : «Вища школа», 1978. – 144 с.
9. *Бурлачук, Л. Ф.* Словарь-справочник по психологической диагностике / Л. Ф. Бурлачук, С. М. Морозов ; отв. ред. Крымский С. Б. – Киев : Наук. думка, 1989. – 200 с.
10. *Гуревич, К. М.* Надежность психологических тестов (Опыт психологической интерпретации) / К. М. Гуревич // Педагогическая диагностика, ее проблемы и методы / под ред. К. М. Гуревича, В. И. Лубовского. – М., 1975. – С. 162–177.
11. *Рапопорт, И. А.* Тесты в обучении иностранным языкам в средней школе / И. А. Рапопорт, Р. Сельг, И. Соттер. – Таллин : «Валгус», 1987. – 352 с.