

Лекция №9

Тема: Оценивание надежности и валидности педагогических тестов

Цель: Раскрыть значение понятия надежности и валидности педагогических тестов. Знакомство с видами оценки надежности и валидности педагогических тестов.

Вопросы для рассмотрения:

1. Общие замечания о надежности и методах ее оценивания
2. Надежность теста
3. Валидность теста

Краткое содержание лекции: Общие представления о надежности и валидности были введены ранее. Оценка надежности нормативно-ориентированных тестов проводится различными методами, которые по способу осуществления можно условно разделить на две группы. Первая группа методов базируется на двукратном тестировании, проводимом с помощью одного и того же теста либо с помощью двух параллельных форм теста. Вторая группа предполагает однократное тестирование при оценке надежности теста. На практике стараются использовать вторую группу методов, поскольку организация повторного тестирования, как и разработка параллельных форм, всегда сопряжена с определенными трудностями и дополнительными затратами со стороны создателей тестов. Обычно вне зависимости от метода оценка надежности строится на подсчете корреляции между двумя наборами данных. Логика рассуждений при этом довольно проста: чем выше корреляция, тем надежнее тест.

В ходе проверки надежности теста проводится оценка постоянства полученных результатов при повторном проведении испытания. Расхождения данных должны отсутствовать или же быть незначительными. В противном случае невозможно относиться к результатам теста с доверием. Надежность теста — это критерий, который свидетельствует о точности измерений. Существенными считаются следующие свойства тестов:

- воспроизводимость результатов, полученных по итогам исследования;
- степень точности методики измерения или соответствующих приборов;
- устойчивость результатов на протяжении определенного периода времени.

Способы оценки надежности теста. При определении надежности теста могут быть использованы следующие методики. Метод повторного тестирования является одним из самых распространенных. Он позволяет установить степень корреляции между результатами исследований, а также временем, в которое они были проведены. Данная методика отличается простотой и эффективностью. Тем не менее у испытуемых, как правило, повторные исследования вызывают раздражение и негативные реакции.

Метод проверки внутренней согласованности не берет во внимание постоянство получаемых при повторном исследовании результатов. Он устанавливает взаимосвязь ответов, которые были даны в рамках одного эксперимента. Вопросы теста делятся на два перечня (по определенному принципу), после чего рассчитывается коэффициент корреляции между результатами.

Метод эквивалентных форм заключается в использовании двух или более тестов с разными формулировками заданий, но с одинаковой сутью, формой и степенью сложности выполнения. О надежности теста свидетельствуют одинаковые или приближенные результаты, которые были получены с использованием одного и того же измерительного прибора или вычислительной формулы. Если же итоги сильно расходятся, то, скорее всего, они были искажены намеренно или же испытуемый не очень ответственно подошел к процессу опроса.

Важнейшей характеристикой теста является его надежность, определяющая воспроизводимость результатов тестирования, их точность. Допустим, у нас есть гипотетическая группа испытуемых, которые немедленно забывают содержание теста по

его завершении. Тогда, в случае надежного теста, повторяя тестирование многократно, мы должны получать одни и те же индивидуальные баллы. Для малонадежного теста результаты будут меняться каждый раз.

Тест представляет собой систему заданий. Качество заданий определяет надежность теста в целом. Рассмотрим пример. Допустим, тест состоит из заданий в закрытой форме, в которых по ошибке не указаны правильные ответы. Слабые испытуемые, не зная ответа, будут пытаться его угадать. Сильные испытуемые, зная верный ответ, но не находя его среди предложенных, так же вынуждены будут случайным образом выбирать любой из ответов. В итоге, индивидуальные баллы будут представлять собой случайные последовательности, не повторяющиеся в разных сеансах тестирования. Воспроизводимость тестовых баллов будет полностью отсутствовать и надежность теста будет близка к нулю. Низкая надежность теста обусловлена низким качеством тестовых заданий.

Для определения надежности реальных тестов можно использовать коэффициент корреляции Пирсона для индивидуальных баллов разных сеансов тестирования. Для организации разных сеансов тестирования можно использовать либо параллельные тесты, либо повторное тестирование через определенный промежуток времени. Можно также использовать результаты одного сеанса тестирования. При этом выполняют расщепление теста, например, на четные и нечетные задания и, затем, находят корреляцию между этими двумя половинами.

Надежность теста определяется разными методами. Рассмотрим их.

Из классической теории теста следует, что надежность теста есть

$$r_t = 1 - \frac{S_E^2}{S_X^2}$$

где S_E^2 - дисперсия ошибочного вклада тестовый балл, S_X^2 - дисперсия наблюдаемого тестового балла.

Валидность теста — это критерий, который определяет достоверность измерения. Можно сказать, что это пригодность того или иного инструментария к оценке определенной психологической характеристики. Стоит отметить, что валидность, надежность теста — это взаимодополняемые критерии, по отдельности они являются несущественными.

Валидность может рассматриваться с точки зрения теоретического и прагматического аспекта. В первом случае речь идет о методе оценки или же измерительном инструменте. Что касается второго понимания валидности, то оно касается цели проведения исследовательских мероприятий. Стоит отметить, что данный критерий может существенно отличаться для одного и того же теста, в зависимости от круга испытуемых. Наивысшая оценка может колебаться в районе 80%.

Валидность психологического теста может оцениваться в соответствии с количественными или качественными показателями. В первом случае речь идет о проведении математических расчетов. Качественная оценка производится описательно, на основании логических выводов.

Вопросы и задания для самоконтроля:

1. Что такое надежность теста?
2. Как проверить надежность и валидность теста?
3. Что такое валидность?
4. Какие методы используются для оценки надежности тестов:
5. Источники повышения валидности теста.

Учебная литература:

1. Засобина, Г.А. Педагогика : учебное пособие / Г.А. Засобина, И.И. Корягина, Л.В. Куклина. - Москва ; Берлин : Директ-Медиа, 2015. - 250 с [<http://biblioclub.ru/index.php?page=book&id=272316>].
2. Зудина, Е.В. Курс по выбору «Импровизация в профессионально-педагогической деятельности» : учебное пособие / Е.В. Зудина ; Министерство образования и науки Российской Федерации, Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Волгоградский государственный социально-педагогический университет» ; науч. ред. Н.К. Сергеев. - Волгоград : Издательство ВГСПУ "Перемена", 2012. - 117 с [<http://biblioclub.ru/index.php?page=book&id=429359>].
3. Курзаева, Л.В. Управление качеством образования и современные средства оценивания результатов обучения: учеб. пособие [Электронный ресурс] : учеб. пособие / Л.В. Курзаева, И.Г. Овчинникова. — Электрон. дан. — Москва : ФЛИНТА, 2015. — 100с [<https://e.lanbook.com/book/70446>].
4. Беленчук, Л.Н. История отечественной педагогики : учебное пособие / Л.Н. Беленчук. - Москва : Институт эффективных технологий, 2013. - 120 с [<http://biblioclub.ru/index.php?page=book&id=232198>].
5. Марусева, И.В. Современная педагогика (с элементами педагогической психологии) : учебное пособие для вузов / И.В. Марусева. - Москва ; Берлин : Директ-Медиа, 2015. - 624 с [<http://biblioclub.ru/index.php?page=book&id=279291>].
6. Столяренко, А.М. Общая педагогика : учебное пособие / А.М. Столяренко. - Москва : ЮнитиДана, 2015. - 479 с [<http://biblioclub.ru/index.php?page=book&id=436823>]