

Психометрические требования к построению и проверке методик Этапы разработки и стандартизации

На этапе разработки методики проводится стандартизация, которая подразумевает единообразие процедуры проведения и оценки выполнения теста. Если мы хотим, чтобы показатели, полученные разными людьми, были сравнимыми, условия тестирования должны быть одинаковыми для всех.

Процедура стандартизации включает три этапа:

- 1) единообразие процедуры проведения теста для получения сравнимых с нормой результатов;
- 2) единообразие оценки выполнения теста;
- 3) определение нормы выполнения теста для сопоставления с ними показателей, полученных в результате обработки данных тестирования.

Первый этап стандартизации психодиагностической методики состоит в создании единообразной процедуры тестирования. Она включает определение следующих моментов диагностической ситуации:

- условия тестирования (помещение, освещение и др. внешние факторы). Очевидно, что объем кратковременной памяти лучше измерять, когда нет внешних раздражителей, таких как посторонние звуки, голоса и т.д.;
- содержание инструкции и особенности ее предъявления (тон голоса, паузы, скорость речи и т.д.);
- наличие стандартного стимульного материала. Например, достоверность полученных результатов существенно зависит от того, предлагаются испытуемому оригинальный стимульный материал или искаженные копии;
- временные ограничения выполнения теста. Например, выполнение заданий тестов на интеллект обычно ограничено определенным промежутком времени;
- стандартный бланк для фиксации результатов, использование которого облегчает процедуру обработки;
- учет влияния ситуационных переменных на процесс и результат тестирования. Под переменными подразумевается состояние испытуемого (усталость, перенапряжение и т.д.), нестандартные условия тестирования (плохое освещение, отсутствие вентиляции и др.), прерывание тестирования;
- учет влияния поведения диагноста на процесс и результат тестирования. Например, одобительно-поощряющее поведение экспериментатора во время тестирования может восприниматься испытуемым как подсказка «правильного ответа» и др.;
- учет влияния опыта испытуемого в тестировании. Естественно, что испытуемый, который уже не в первый раз проходит процедуру тестирования, преодолел чувство неизвестности и выработал определенное отношение к тестовой ситуации. Например, если испытуемый уже выполнял тест Равена, то, скорее всего, не стоит предлагать ему его во второй раз.

Второй этап стандартизации психодиагностического теста состоит в создании единообразной оценки выполнения теста: предварительной обработки и интерпретации полученных результатов. Этот этап предполагает также сравнение полученных показателей с нормой выполнения этого теста для данного возраста (например, в тестах интеллекта), пола и т.д.

Третий этап стандартизации психодиагностической методики состоит в определении норм ее выполнения. Нормы, как правило, определяются разработчиками методики в результате исследования большой выборки испытуемых определенного

возраста и пола с последующим усреднением полученных оценок и дифференциацией по ряду релевантных показателей.

Пример: если нормальные 8-летние дети правильно решают 12 задач из 50 в тесте на типичное арифметическое рассуждение, значит, норма для 8-летнего ребенка по этому тесту соответствует 12.

Определение надежности и валидности теста

Любой научный метод измерения должен соответствовать определенным критериям качества. Существуют основные и второстепенные критерии качества. Основными критериями качества являются:

1. **Объективность:** Независимость теста от условий, в которых он применялся, от человека, проводящего испытание, анализирующего и интерпретирующего полученные результаты [4].

2. **Надёжность:** согласованность результатов, полученных при каждом повторном выполнении теста одним и тем же испытуемым, с результатами его первого тестирования [1].

3. **Валидность:** соответствие результатов теста той характеристике, для измерения которой он предназначен [4].

Второстепенными критериями качества являются:

- **Полезность:** Полезен ли метод и насколько удобен для изучения конкретной характеристики? Иногда это называют внешней валидностью.

- **Репрезентативность:** Могут ли проходить измерение люди с разным полом, возрастом, местом жительства и т. д.?

- **Экономичность:** Являются ли затраты на исследование адекватны полученным данным?

- **Прозрачность:** Является ли инструкция и порядок выполнения исследования понятным для испытуемого? Предусматривается ли процедура объяснения инструкции?

- **Фальсификация:** Устроен ли процесс исследования таким образом, что испытуемый имеет возможность умышленно исказить результат?

- **Рациональность:** Насколько рационально предусмотрено время исследования, учтены умственные способности испытуемых и окружающие физические условия?

- **Нормализация:** Существует ли нормальный показатель по используемому методу? Каким образом можно сравнить полученные от разных людей результаты?

Валидность является общим оценочным суждением об адекватности и степени пригодности полученных эмпирическим путём выводов.

Важно отметить, что валидность может измеряться только у объективного и надёжного теста. Другими словами, объективность, надёжность и валидность находятся в иерархических отношениях друг к другу.

Остановимся на более подробном рассмотрении этих показателей, характеризующих с разных сторон надёжность психодиагностической методики.

Определение надёжности измерительного инструмента. От того, как составлена методика, насколько правильно подобраны задания с точки зрения их взаимосогласованности, насколько она однородна, зависит точность, объективность любого психологического измерения. Внутренняя однородность методики показывает, что ее задания актуализируют одно и то же свойство, признак.

Для проверки надёжности измерительного инструмента, говорящего о его однородности (или гомогенности), используется так называемый метод "расщепления".

Обычно задания делятся на четные и нечетные, отдельно обрабатываются, а затем результаты двух полученных рядов коррелируются между собой. Для применения этого способа нужно поставить испытуемых в такие условия, чтобы они смогли успеть решить (или попытаться решить) все задания. Если методика однородна, то большой разницы в успешности решения по таким половинкам не будет, и, следовательно, коэффициент корреляции будет достаточно высоким.

Можно делить задания и другим путем, например, сопоставить первую половину теста со второй, первую и третью четверть — со второй и четвертой и т. п. Однако "расщепление" на четные и нечетные задания представляется наиболее целесообразным, поскольку именно этот способ наиболее независим от влияния таких факторов, как вработываемость, тренировка, утомление и пр.

Методика признается надежной, когда полученный коэффициент не ниже 0,75–0,85. Лучшие по надежности тесты дают коэффициенты порядка 0,90 и более.

Но на начальном этапе разработки диагностической методики можно получить невысокие коэффициенты надежности, например, порядка 0,46—0,50. Это означает, что в разрабатываемой методике присутствует некоторое число заданий, которые в силу своей специфичности ведут к снижению коэффициента корреляции. Такие задания необходимо специально проанализировать и либо переделать их, либо вообще изъять.

Для проверки стабильности диагностируемого признака, свойства используется прием, известный как тест-ретест. Он заключается в повторном обследовании испытуемых с помощью той же методики. О стабильности признака судят по коэффициенту корреляции между результатами первого и повторного обследования. Он будет свидетельствовать о сохранении или несохранении каждым испытуемым своего порядкового номера в выборке.

Другим после надежности ключевым критерием оценки качества методик является валидность. Вопрос о валидности методик решается лишь после того, как установлена достаточная ее надежность, поскольку ненадежная методика без знания ее валидности является практически бесполезной.

Валидность по своей сути — это комплексная характеристика, включающая, с одной стороны, сведения о том, пригодна ли методика для измерения того, для чего она была создана, а с другой стороны, какова ее действенность, эффективность. По этой причине не существует какого-то единого универсального подхода к определению валидности. В зависимости от того, какую сторону валидности хочет рассмотреть исследователь, используются и разные способы доказательства. Другими словами, понятие валидности включает в себя разные ее виды, имеющие свой особый смысл. Проверка валидности методики называется валидизацией.

Существует несколько видов валидности, обусловленных особенностями диагностических методик, а также временным статусом внешнего критерия:

1. Валидность "по содержанию". Этот прием используется в основном в тестах достижений. Обычно в тесты достижений включается не весь материал, который прошли учащиеся, а какая-то его небольшая часть (3—4 вопроса). Можно ли быть уверенным в том, что правильные ответы на эти немногие вопросы свидетельствуют об усвоении всего материала. На это и должна ответить проверка валидности по содержанию. Для этого проводится сопоставление успешности по тесту с экспертными оценками учителей (по данному материалу). Валидность "по содержанию" также подходит к

критериально-ориентированным тестам. Иногда этот прием называют логической валидностью.

2. Валидность "по одновременности", или текущая валидность, определяется с помощью внешнего критерия, по которому информация собирается одновременно с экспериментами по проверяемой методике. Другими словами, собираются данные, относящиеся к настоящему времени успеваемость в период испытания, производительность в этот же период и т. д. С ним коррелируют результаты успешности по тесту.

3. "Предсказывающая" валидность ("прогностическая" валидность). Определяется также по достаточно надежному внешнему критерию, но информация по нему собирается некоторое время спустя после испытания. Внешним критерием обычно бывает выраженная в каких-нибудь оценках способность человека к тому виду деятельности, для которой он отбирался по результатам диагностических испытаний. Хотя этот прием наиболее соответствует задаче диагностических методик – предсказанию будущей успешности, применять его очень трудно. Точность прогноза находится в обратной зависимости от времени, заданного для такого прогнозирования. Чем больше проходит времени после измерения, тем большее количество факторов требуется учитывать при оценке прогностической значимости методики. Однако учесть все факторы, влияющие на предсказание, практически невозможно.

4. "Ретроспективная" валидность. Определяется на основе критерия, отражающего события или состояние качества в прошлом. Может быть использована для быстрого получения сведений о предсказательных возможностях методики. Так, для проверки того, в какой мере хорошие результаты теста способностей соответствуют быстрому обучению, можно сопоставить прошлые оценки успеваемости, прошлые экспертные заключения и т. д. у лиц с высокими и низкими на данный момент диагностическими показателями.

При приведении данных о валидности разработанной методики важно точно указать, какой вид валидности имеется в виду (по содержанию, по одновременности и т. д.).

Желательно также сообщать сведения о численности и особенностях индивидов, на которых проводилась валидизация. Такая информация позволяет пользующемуся методикой исследователю решить, насколько валиден этот прием для той группы, к которой он собирается его применять. Как и в случае с надежностью, необходимо помнить, что в одной выборке методика может обладать высокой валидностью, а в другой – низкой. Поэтому если исследователь планирует использовать методику на выборке испытуемых, существенно отличающейся от той, на которой проводилась проверка валидности, ему необходимо заново провести такую проверку. Приводимый в руководстве коэффициент валидности применим только к группам испытуемых, подобным тем, на которых он определялся.