

НАДЕЖНОСТЬ И ВАЛИДНОСТЬ МЕТОДОВ ИССЛЕДОВАНИЯ



Критерии и показатели в оценке методов исследования.

Надежность методов психолого-педагогического исследования во многом зависит от критериев и показателей, по которым происходит изучение выбранного для исследования образовательного феномена.



Критерий (от греч. *kriterion* – средство для суждения) – это признак, на основании которого производится оценка, определение или классификация чего-либо.

В диагностике критерием является переменная величина, принимающая разные значения в различных случаях или в разные моменты времени в рамках одного случая. Критерии дают возможность судить о состоянии объекта исследования.

Показатель (*indicator* – то, что доступно восприятию, то, что «показывает» наличие чего-либо) – это некоторая величина или качество переменной (критерия), которое может проявляться у конкретного объекта, т.е. это мера проявления критерия, его количественная или качественная характеристика, по которой судят о различных состояниях объекта; это внешне хорошо различимый признак измеряемого критерия. Можно сказать, что показатель выполняет роль эмпирического индикатора критерия.

Принято считать, что количество критериев должно составлять не менее трех, а по каждому критерию следует выделять как минимум три показателя. Только тогда можно говорить о полном отображении объекта и предмета диагностики и проявлении каждого соответствующего критерия



ОБЩИЕ ТРЕБОВАНИЯ К МЕТОДАМ ИССЛЕДОВАНИЯ

каждая методика должна иметь следующие *составляющие*:

- описание, обеспечивающее ее адекватное использование в точном соответствии со стандартами: предмет диагностики, сфера применения, контингент испытуемых, процедура применения;
- подробные сведения о процедуре разработки методики, полученных при этом данных о надежности и валидности;
- однозначное описание выборки стандартизации и характера диагностической ситуации в обследовании;
- процедура подсчета баллов и интерпретации должна быть описана с однозначной ясностью, позволяющей получить идентичные результаты при обработке одинаковых протоколов разными пользователями руководства.



Под операционализацией понимается требование, согласно которому при введении новых научных понятий необходимо четко указывать на конкретные процедуры, приемы и методы, с помощью которых можно практически удостовериться в том, что явление, описанное в понятии, действительно существует.

Операционализация предполагает указание на практические действия или операции, которые может выполнить любой диагност, чтобы убедиться в том, что определенное в понятии явление обладает именно теми свойствами, которые ему приписываются.

Требование верификации означает, что всякое новое понятие, вводимое в научный оборот и претендующее на получение статуса научного, обязательно должно пройти проверку на наличие методики экспериментальной диагностики описанного в нем явления. Качество результатов диагностики при этом принято оценивать по общепринятым критериям объективности, надежности, валидности и др.



Объективность характеризуется корреляцией (совпадением или согласованностью) между результатами, полученными двумя оценивающими лицами. Необходимо, чтобы коэффициенты корреляции в этом случае были близки к единице ($r = 1$).

Стандартизация – это единообразие процедуры проведения и оценки выполнения диагностического метода. Стандартизованность в диагностике – это неизменность заданных вопросов и задач, точность соблюдения испытуемыми инструкции, а диагностами – способов вычисления и интерпретации полученных показателей.



НАДЕЖНОСТЬ МЕТОДА ИССЛЕДОВАНИЯ

Надежность – один из критериев качества результата в диагностике, относящийся к степени точности и устойчивости показателей диагностируемого признака. Чем больше надежность методики, тем свободнее она от погрешностей измерения.

В самом широком смысле надежность – это характеристика того, в какой степени выявленные у испытуемых различия по результатам проведения методики являются отражением действительных различий в измеряемых свойствах и в какой мере они могут быть приписаны случайным ошибкам.



Приемы для оценки надежности диагностической методики

1. *Прием ретеста*, или повторной диагностики, позволяет обработать одни и те же задания, выполненные одними и теми же испытуемыми в разное время, и просчитать взаимосвязь результатов, выраженную в коэффициенте самокорреляции.
2. *Прием деления пополам* – подборка однажды выполненных заданий делится пополам (например, в первый полутест входят задания с нечетным порядковым номером, а во второй полутест – с четным), затем устанавливаются результаты каждого испытуемого по обоим полутестам и вычисляется коэффициент корреляции между полученными результатами.
3. *Прием параллельного теста* – для измерения одних и тех же знаний конструируются два различных набора заданий, которые по своему содержанию напоминают близнецов; оба параллельных набора заданий предлагаются непосредственно друг за другом или при удобном случае.

Во всех случаях при коэффициенте корреляции методик $r > 0,7$ методика считается надежной.



В тестовой методике принято учитывать три коэффициента надежности:

- *коэффициент стабильности*, или постоянства, – показатель корреляции между результатами первого и повторного испытаний одним тестом одной и той же выборки испытуемых;
- *коэффициент эквивалентности*, или коэффициент корреляции, результатов тестирования одного и того же контингента испытуемых с помощью вариантов одного и того же теста либо разными, но эквивалентными по форме и цели, тестами;
- *коэффициент внутреннего постоянства*, или внутренней однородности, который соответствует корреляции результатов частей теста, показанных одними и теми же испытуемыми.



ВАЛИДНОСТЬ МЕТОДА

Валидность метода исследования и диагностики показывает, в какой мере измеряется то качество (свойство, характеристика), для оценки которого он предназначен.

Валидность (адекватность) говорит о степени соответствия метода своему назначению. Чем ближе раскрывается в диагностике тот признак, для обнаружения и измерения которого предназначен метод, тем выше его валидность .

Понятие валидности относится не только к методике, но и к критерию оценки ее качества, *критерию валидности*.



Такими критериями могут стать следующие:

- поведенческие показатели – реакции, действия и поступки испытуемого в различных жизненных ситуациях;
- достижения испытуемого в различных видах деятельности – учебной, трудовой, творческой и др.;
- самоорганизация, данные, свидетельствующие о выполнении различных контрольных проб и заданий;
- данные, получаемые при помощи других методик, валидность или связь которых с проверяемой методикой считается достоверно установленной.

Чем выше коэффициент корреляции методики с критерием, тем выше валидность.



Виды валидности диагностических методик

1. *Теоретическая (концептуальная) валидность* определяется по соответствию показателей исследуемого качества, получаемых с помощью данной методики, показателям, получаемым посредством других методик (с показателями которых должна существовать теоретически обоснованная зависимость). Теоретическую валидность проверяют по корреляциям показателей одного и того же свойства, получаемым при помощи разных методик, связанных с одной и той же теорией.
2. *Эмпирическая (прагматическая) валидность* проверяется по соответствию диагностических показателей реальному жизненному поведению, наблюдаемым действиям и реакциям испытуемого. Если, например, с помощью некой методики мы оцениваем черты характера у данного испытуемого, то применяемая методика будет считаться практически или эмпирически валидной тогда, когда мы установим, что данный человек ведет себя в жизни именно так, как предсказывает методика, т.е. в соответствии с имеющейся у него чертой характера



3. Внутренняя валидность означает соответствие содержащихся в методике заданий, субтестов, суждений и т.п. общей цели и замыслу методики в целом. Она считается внутренне невалидной или недостаточно внутренне валидной тогда, когда все или часть включенных в нее вопросов, заданий или субтестов измеряют не то, что требуется от данной методики.

4. Внешняя валидность – это примерно то же самое, что и эмпирическая валидность, с той лишь разницей, что в данном случае речь идет о связи между показателями методики и наиболее важными, ключевыми внешними признаками, относящимися к поведению испытуемого.

5. Очевидная валидность описывает представление о методе, сложившееся у испытуемого, т.е. это валидность с точки зрения испытуемого. Методика должна восприниматься обследуемым как серьезный инструмент познания его личности, чем-то схожий с медицинским диагностическим инструментарием.

6. Конкурентная валидность оценивается по корреляции разработанной методики с другими, валидность которых относительно измеряемого параметра установлена.



7. *Прогностическая валидность* устанавливается с помощью корреляции между показателями методики и некоторым критерием, характеризующим измеряемое свойство, но в более позднее время.

8. *Инкрементная валидность* имеет ограниченное значение и относится к случаю, когда один тест из батареи тестов может иметь низкую корреляцию с критерием, но не перекрывается другими тестами из этой батареи. В этом случае данный тест обладает инкрементной валидностью. Это может быть полезно при проведении профотбора с помощью психологических тестов.

9. *Дифференциальная валидность* может быть проиллюстрирована на примере тестов интересов. Тесты интересов обычно коррелируют с академической успеваемостью, но по-разному для разных дисциплин. Значение дифференциальной валидности, так же как и инкрементной, ограничено.



10. Содержательная валидность определяется через подтверждение того, что задания методики отражают все аспекты изучаемой области поведения. Содержательную валидность часто называют «логической валидностью» или «валидностью по определению». Она означает, что методика валидна по мнению специалистов. Обычно она определяется у тестов достижений. На практике для определения содержательной валидности подбираются эксперты, которые указывают, какая область (области) поведения наиболее важна.

11. Конструктивная валидность демонстрируется полным, насколько это возможно, описанием переменной, для измерения которой предназначена методика. Конструктивная валидность включает в себя все подходы к определению валидности, которые были перечислены выше.

Существует прямая связь валидности с надежностью. Методика с низкой надежностью не может обладать высокой валидностью, поскольку неверен измерительный инструмент и нестабилен тот признак, который он измеряет.



ДОПОЛНИТЕЛЬНЫЕ ТРЕБОВАНИЯ К МЕТОДАМ ИССЛЕДОВАНИЯ

- *Точность* методики отражает ее способность тонко реагировать на малейшие изменения оцениваемого свойства, происходящие в ходе эксперимента.
- *Однозначность* методики характеризуется тем, в какой степени получаемые с ее помощью данные отражают изменения именно и только того свойства, для оценивания которого данная методика применяется.
- *Репрезентативность* означает, что свойства более широкого множества объектов представлены в свойствах подмножества.
- *Диагностическая ценность* методики определяется путем проведения предварительного опыта с так называемой нейтральной группой, результаты которого в дальнейшей диагностической работе не используются.

