

Лекция №4

Психометрические основы психодиагностики

Лектор: Сахарова В.Г.

- 1. Представление о психометрии;**
- 2. Требования к методикам;**
- 3. Требования к пользователям;**
- 4. Понятие стандартизации;**
- 5. Понятие надёжности;**
- 6. Понятие валидности**

Психометрия — область психологической диагностики, связанная с теорией и практикой измерений в психологии



*Это понятие было введено немецким исследователем **Х. Вольфом***

**Дифференциальная
психометрия**

Общая психометрия

- К общей психометрии относятся измерения **общепсихологического** (свойственного всем людям) характера, т. е. функциональные зависимости между свойствами стимула и свойствами субъективных реакций;

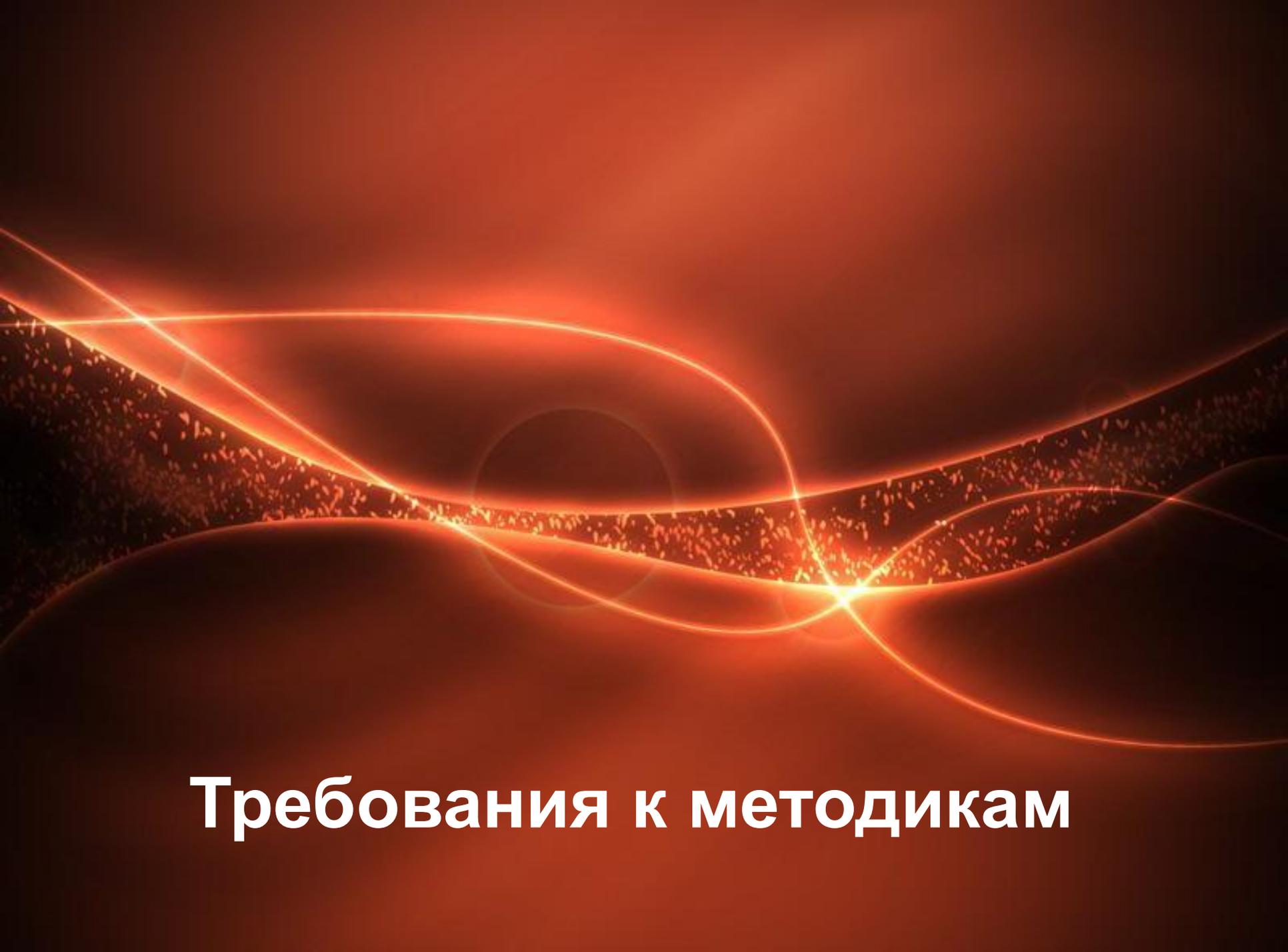
* Так, в психофизике определяется зависимость между физическими характеристиками стимулов и субъективными характеристиками ощущений (например, объективная прибавка веса и порог субъективного ощущения увеличения тяжести)

- К дифференциальной психометрии относятся измерения, касающиеся **индивидуальных различий** между людьми по способностям, когнитивным функциям (внимание, память, мышление и т. д.), мотивам, эмоциональным особенностям и пр. Здесь числовые значения (баллы, ранги, шкальные значения) приписываются не стимулам, а индивидам



- **Дифференциальная психометрика** по отношению к психодиагностике и дифференциальной психологии выступает в качестве *технологическо-методической дисциплины*;

- Она обосновывает требования, которым должны удовлетворять психодиагностические методики, процедуру их разработки и применения. К числу таких требований относятся *стандартизация*, *надежность* и *валидность* методик

The background features a dark gradient with several glowing, wavy lines in shades of orange and red. A central horizontal band is filled with a dense stream of small, bright orange particles, resembling a comet tail or a data stream. The overall aesthetic is futuristic and dynamic.

Требования к методикам

а) должны быть однозначно сформулированы цели, предмет и область применения методики (подразумевается особая социальная среда или сфера общественной практики - производство, медицина, семейная жизнь и т. п., контингент испытуемых (пол, возраст, образование, профессиональный опыт, должностное положение). Должны быть конкретизированы цели использования результатов:

- **для прогноза успешности профессиональной деятельности;**
 - **для психологического вмешательства;**
- **для принятия правовых, административных решений;**
 - **для прогноза стабильности коллектива**

б) процедура проведения должна быть задана в виде однозначного алгоритма, пригодного для передачи лаборанту, не имеющему специальных психологических знаний, или ввода в компьютер, используемый для предъявления заданий и анализа ответов;

в) тестовые шкалы должны быть проверены на репрезентативность, надежность и валидность. Другие пользователи должны иметь возможность повторить стандартизационные исследования в своей области;

- г) процедуры, основанные на самоотчете, должны быть снабжены средствами контроля за достоверностью, позволяющими автоматически отсеивать недостоверные протоколы;**
- д) головная методическая организация определенного ведомства (области применения) должна вести банк данных, собранных по тесту, и производить периодическую коррекцию всех стандартов методики**

The background is a deep blue gradient. It features several glowing white, ethereal lines that curve and intersect across the frame. A central horizontal band is filled with a dense stream of small, bright white particles, resembling a data flow or a particle beam. The overall aesthetic is futuristic and digital.

Требования к пользователям

- а) психолог должен знать и применять на практике общие теоретико-методологические принципы психодиагностики, владеть основами дифференциальной психометрики, должен следить за текущей методической литературой по психодиагностике, самостоятельно вести картотеку и личную библиотечку методик, применяемых в заданной области;**
- б) отвечает за решения, принимаемые на основе тестирования, обеспечивая их соответствие репрезентативности и прогностической валидности методики;**

- в) пользуется преимущественным правом на проведение психодиагностики в заданной области, на использование протоколов в соответствии с профессионально-этическими принципами и интересами психологии;**
- г) в подборе методик в комплексную программу обследования не руководствуется субъективными предпочтениями и предубеждениями в оценке методик, а исходит из требования максимальной эффективности диагностики;**

- д) параллельно с использованием методик ведет научно-методическую работу, анализируя по собранным данным эффективность применения методики в заданной области;**
- е) обеспечивает тщательное соблюдение всех требований для проведения стандартных методик обследования. Подсчет баллов, интерпретация, прогноз делаются в строгом соответствии с методическими указаниями;**
- ж) обеспечивает конфиденциальность психодиагностической информации, полученной от испытуемого на основе «личного доверия»**

Стандартизация

- это единообразие процедуры проведения и оценки выполнения теста

Таким образом, стандартизация рассматривается в двух планах: как выработка единых требований к процедуре эксперимента и как определение единого критерия оценки результатов диагностических испытаний

К числу требований, которые необходимо соблюдать при проведении эксперимента, можно отнести такие:

- ❖ **инструкции следует сообщать испытуемым одинаковым образом, как правило, письменно; в случае устных указаний они даются в разных группах одними и теми же словами, понятными для всех, в одинаковой манере;**
- ❖ **ни одному испытуемому не следует давать никаких преимуществ перед другими;**
- ❖ **в процессе эксперимента не следует давать отдельным испытуемым дополнительные пояснения;**



- ❖ эксперимент с разными группами следует проводить в одинаковое по возможности время дня, в сходных условиях;
- ❖ временные ограничения в выполнении заданий для всех испытуемых должны быть одинаковыми и т. д.

Другим важным моментом в стандартизации методики является выбор критерия, по которому следует проводить сравнение результатов диагностических испытаний, поскольку диагностические методики не имеют заранее определенных стандартов успешности или неудачи в их выполнении

- Так, например, ребенок шести лет, выполняя тест умственного развития, получил балл, равный 117.

Хорошо это или плохо ? Часто ли такой показатель встречается у детей данного возраста?

- Количественный результат как таковой ничего не означает. Полученный дошкольником балл нельзя интерпретировать как показатель относительно высокого, среднего или низкого развития, так как это развитие выражено в мерах, присущих данной методике, и, таким образом, абсолютного значения полученные результаты иметь не могут. Очевидно, нужно располагать точкой отсчета и какими-то дозированными мерами, чтобы с их помощью оценивать полученные при диагностировании индивидуальные и групповые данные

Возникает вопрос, что за эту точку отсчета брать?

- В традиционном тестировании такая точка добывается статистическим путем — это так называемая ***статистическая норма***;
- В общих чертах стандартизация диагностической методики, ориентированной на норму, осуществляется путем ее проведения на большой репрезентативной выборке того типа, для которой данная методика предназначена

Относительно этой группы испытуемых, называемой ***выборкой стандартизации***, разрабатываются нормы, указывающие не только средний уровень выполнения, но и его относительную вариативность выше и ниже среднего уровня. В результате можно оценить разные степени успешности или неуспешности в выполнении диагностической пробы. Это позволяет определить положение конкретного испытуемого относительно нормативной выборки или выборки стандартизации

Выборка стандартизации

- При разработке и применении любой точки отсчета следует обращать особое внимание на выборку испытуемых, на которой проводится стандартизация диагностической методики;
- В математической статистике принято различать такие понятия, как **генеральная совокупность** (популяция) и **выборка**

- **Генеральная совокупность** - всякая большая совокупность людей, которую хотели бы исследовать или относительно которых собираются делать выводы;
 - **Выборка** — это часть или подмножество совокупности
- * Проводить исследование всей популяции не принято. Обычно из нее выделяют группу людей — выборку стандартизации, — которая реально подвергается тестированию, и с ее помощью оценивается генеральная совокупность

Чтобы оценки носили достоверный характер, выборка должна быть **репрезентативна**, **представительна** рассматриваемой популяции, т.е. ее вероятностные свойства должны совпадать или быть близкими к свойствам генеральной совокупности

А. Анастаси приводит пример формирования репрезентативной выборки

Выборка включала 1700 человек равным количеством мужчин и женщин. Испытуемые в возрасте от 16 до 64 лет были распределены по семи возрастным уровням. При формировании выборки исследователи опирались на данные последней переписи населения США. Учитывалось пропорциональное распределение населения по географическим районам, принадлежность к городскому и сельскому населению, принадлежность к белой или цветной расе, учитывались также уровень образования и профессия. На каждом возрастном уровне в выборку были введены один мужчина и одна женщина, находящиеся в учреждениях для умственно отсталых

По мнению А. Анастази, подавляющее большинство диагностических методик стандартизовано не для столь широких популяций, как многие полагают. Трудно рассчитывать, что по какому-либо тесту имеются адекватные нормы для таких обширных популяций, как, например, «взрослые американцы-мужчины» или «американские дети 14-летнего возраста»

- Так, если определить популяцию как «14-летние дети», а выборку стандартизации составить из 14-летних школьников, то ее нельзя рассматривать в качестве репрезентативной, поскольку не все 14-летние дети являются школьниками;
- В этом случае лучше сузить определение популяции (т. е. определить ее как «14-летние школьники»), чем переносить нормы, полученные на школьниках, на популяцию 14-летних детей

Отбор испытуемых в выборку стандартизации осуществляется следующим образом:

- дается определение популяции с выделением в ее структуре переменных, значимых и малозначимых для изучаемого психического явления (возраст, образование, профессия и т. д.);
- популяция делится на части в соответствии со значимыми переменными;
- испытуемые отбираются в случайном порядке и пропорционально численности каждой значимой части совокупности

- **Случайный отбор может осуществляться по алфавиту, по таблице случайных чисел или другим способом. Важно, чтобы у всех представителей популяции были равные шансы попасть в выборку стандартизации. Это условие подразумевает, что каждый выбор не зависит от остальных;**
- **Объем выборки может варьироваться в широких пределах, но ее минимальный порог, необходимый для получения достоверных результатов, — порядка 200 человек**

Надёжность



- **Надежность** – это относительное постоянство, устойчивость, согласованность результатов теста при первичном и повторном его применении на одних и тех же испытуемых;
- Степень надежности методик зависит от многих причин. Поэтому важной проблемой практической диагностики является выяснение факторов, снижающих точность измерений и их классификация

Анализ показывает, что среди них наиболее часто называют следующие:

- *нестабильность диагностируемого свойства;*
- **несовершенство диагностических методик (небрежно составлена инструкция, задания по своему характеру разнородны, нечетко сформулированы указания, как методику предъявлять испытуемым, и т. д.);**
- **меняющаяся ситуация обследования (разное время дня, когда проводятся эксперименты, разная освещенность помещения, наличие или отсутствие посторонних шумов и т. д.);**

- различия в манере поведения экспериментатора (от опыта к опыту по-разному предъявляет инструкции, по-разному стимулирует выполнение заданий и т. д.);**
- колебания в функциональном состоянии испытуемого (в одном эксперименте отмечается хорошее самочувствие, в другом — утомление и т. д.);**
- элементы субъективности в способах оценки и интерпретации результатов (когда ведется протоколирование ответов испытуемых, оцениваются ответы по степени полноты, оригинальности и т. п.).**

- Одним из важнейших средств повышения надежности психодиагностической методики является ***единообразие процедуры обследования, его строгая регламентация:***

- одинаковые для обследуемой выборки испытуемых обстановка и условия работы,
- однотипный характер инструкций,
- одинаковые для всех временные ограничения, способы и особенности контакта с испытуемыми,
- порядок предъявления заданий

Основные показатели надёжности **К. М. Гуревич** предложил обозначать следующим образом:

- показатель, характеризующий измерительный инструмент, предлагается называть ***коэффициентом надёжности***;
- показатель, характеризующий стабильность измеряемого свойства, — ***коэффициентом стабильности***;
- показатель оценки влияния личности экспериментатора — ***коэффициентом константности***

- Для проверки надежности 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

- Можно делить задания и другим путем. Например, можно 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 **0,9** и более

Определить надежность самой методики - это не значит решить все вопросы, связанные с ее применением. Нужно еще установить, насколько устойчив, стабилен признак, который исследователь намерен измерять. Было бы методологической ошибкой рассчитывать на абсолютную стабильность психологических признаков

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

- **При определении стабильности признака большое значение имеет промежуток времени между первым и вторым обследованиями. Чем короче срок от первого до второго испытания, тем (при прочих равных условиях) больше шансов, что диагностируемый признак сохранит уровень первого испытания. С увеличением временного интервала стабильность признака имеет тенденцию снижаться, так как возрастает число посторонних факторов, влияющих на нее**

Следовательно, напрашивается вывод о том, что целесообразно проводить повторное тестирование через короткий срок после первого. Однако тут есть свои сложности: если срок между первым и вторым опытами небольшой, то некоторые испытуемые могут воспроизвести свои прежние ответы по памяти и таким образом отойдут от смысла выполнения заданий. В этом случае результаты двух предъявлений методики уже нельзя рассматривать как независимые

- **Трудно четко ответить на вопрос, какой срок можно считать оптимальным для повторного эксперимента. Только исследователь, исходя из психологической сущности методики, условий, в которых она проводится, особенностей выборки испытуемых, должен определить этот срок. При этом такой выбор должен быть научно обоснован.**
- **В тестологической литературе наиболее часто называются временные интервалы в несколько месяцев (но не более полугодя). При обследовании детей младшего возраста, когда возрастные изменения и развитие происходят очень быстро, эти интервалы могут быть порядка нескольких недель**

Определение константности

Поскольку методика, разработанная для диагностических целей, не предназначена для того, чтобы вечно оставаться в руках своих создателей, крайне важно знать, в какой мере ее результаты поддаются влиянию личности экспериментатора. Хотя диагностическая методика всегда снабжается подробными инструкциями по ее применению, правилами и примерами, указывающими, как проводить эксперимент, однако регламентировать манеру поведения экспериментатора, скорость его речи, тон голоса, паузы, выражение лица очень трудно. Испытуемый в своем отношении к опыту всегда отразит то, как сам экспериментатор к этому опыту относится (допускает небрежность или действует точно в соответствии с требованиями процедуры, проявляет требовательность, настойчивость или бесконтрольность)

Коэффициент константности определяется путем корреляции результатов двух опытов, проведенных в относительно одинаковых условиях на одной и той же выборке испытуемых, но разными экспериментаторами. Коэффициент корреляции не должен быть **ниже 0,8**

Вопрос о валидности методики решается лишь после того, как установлена достаточная ее **надежность**, поскольку ненадежная методика не может быть валидной;

Но самая надежная методика без знания ее валидности является практически бесполезной



A. Анастazi

«**Валидность теста** —
понятие, указывающее
нам, что тест измеряет
и насколько хорошо он
это делает»

Валидность — 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, практическая полезность



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

*Проверка валидности
методики называется
валидизацией*



Валидизация



Теоретическая

исследователя
интересует само
свойство, измеряемое
методикой

Прагматическая

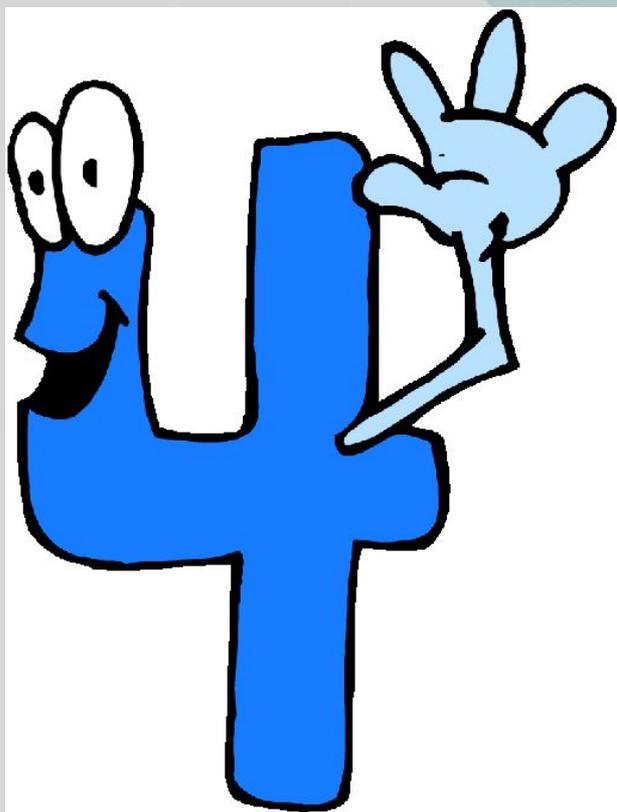
суть предмета измерения
оказывается вне поля
зрения;
Главный акцент сделан
на то, чтобы доказать,
что «нечто», измеряемое
методикой, имеет связь с
определенными
областями практики

**Провести теоретическую
валидизацию методики — это
доказать, что методика
измеряет именно то
свойство, качество, которое
она по замыслу
исследователя должна
измерять**

Прагматическая валидизация
подразумевает проверку методики с точки зрения ее практической эффективности, значимости, полезности, поскольку диагностической методикой имеет смысл пользоваться только тогда, когда доказано, что измеряемое свойство проявляется в определенных жизненных ситуациях, в определенных видах деятельности

Для проведения прагматической валидации методики, т. е. для оценки ее эффективности, действенности, практической значимости, обычно используется независимый **внешний критерий** — показатель проявления изучаемого свойства в повседневной жизни

Американские исследователи *Д. Тиффин* и *Е. Маккормик*, проведя анализ используемых для доказательства валидности внешних критериев, выделили четыре их типа



1) критерии исполнения (в их число могут входить такие, как количество выполненной работы, успеваемость, время, затраченное на обучение и т. п.);



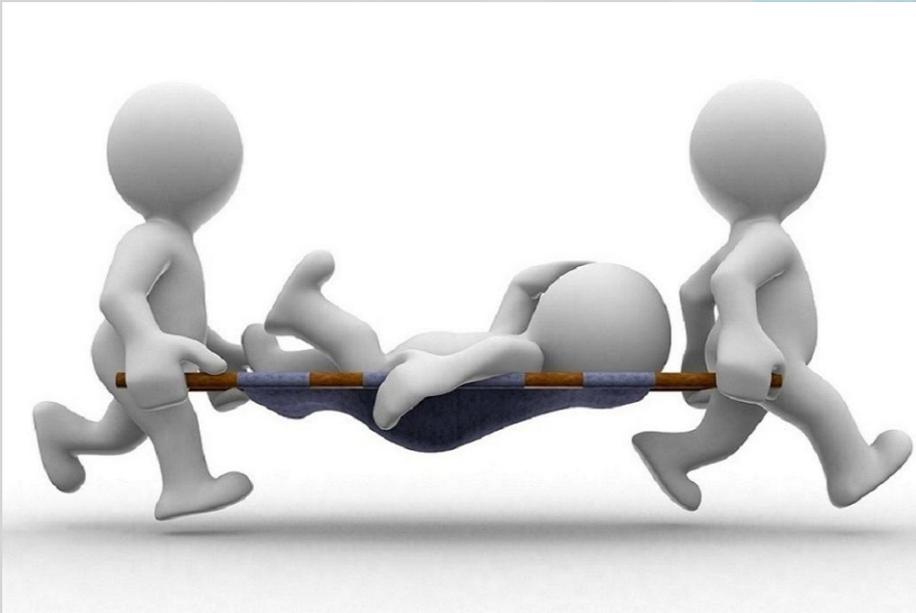
2) **субъективные критерии** (они включают различные виды ответов, которые отражают отношение человека к чему-либо или к кому-либо, его мнение, взгляды, предпочтения; обычно субъективные критерии получают с помощью интервью, опросников, анкет);



3) **физиологические критерии** (они используются при изучении влияния окружающей среды и других ситуационных переменных на организм и психику человека; измеряется частота пульса, давление крови, симптомы утомления и т. д.);



4) критерии случайностей
(применяются, когда цель исследования касается, например, проблемы отбора для работы таких лиц, которые менее подвержены несчастным случаям)



Существует несколько **ВИДОВ
валидности, обусловленных
особенностями диагностических
методик, а также временным статусом
внешнего критерия**

Валидность «по содержанию»:

Этот прием используется, например, 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 одновременности»

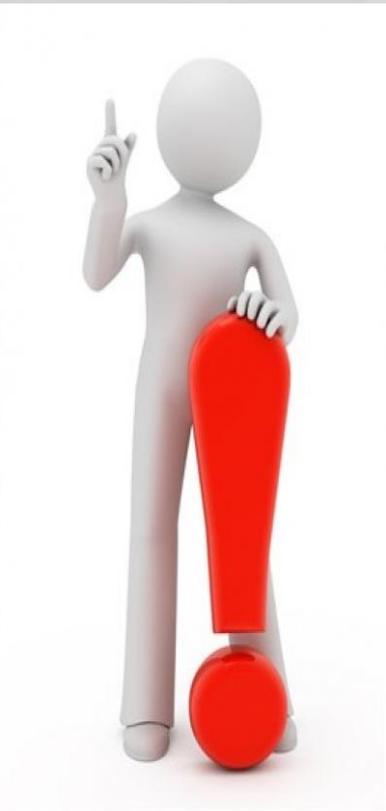
- Определяется с помощью внешнего критерия, по которому информация собирается одновременно с экспериментами по проверяемой методике;
- Другими словами, собираются данные, относящиеся к настоящему времени: успеваемость в период испытания, производительность в этот же период и т. д.
 - С ними сопоставляют результаты успешности по тесту

Прогностическая валидность

- **Определяется также по внешнему критерию, но информация по нему собирается некоторое время спустя, после испытания;**
 - **Внешним критерием обычно бывает выраженная в каких-либо оценках способность человека к тому виду деятельности, для которой он оценивался по результатам диагностических испытаний;**
 - **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**

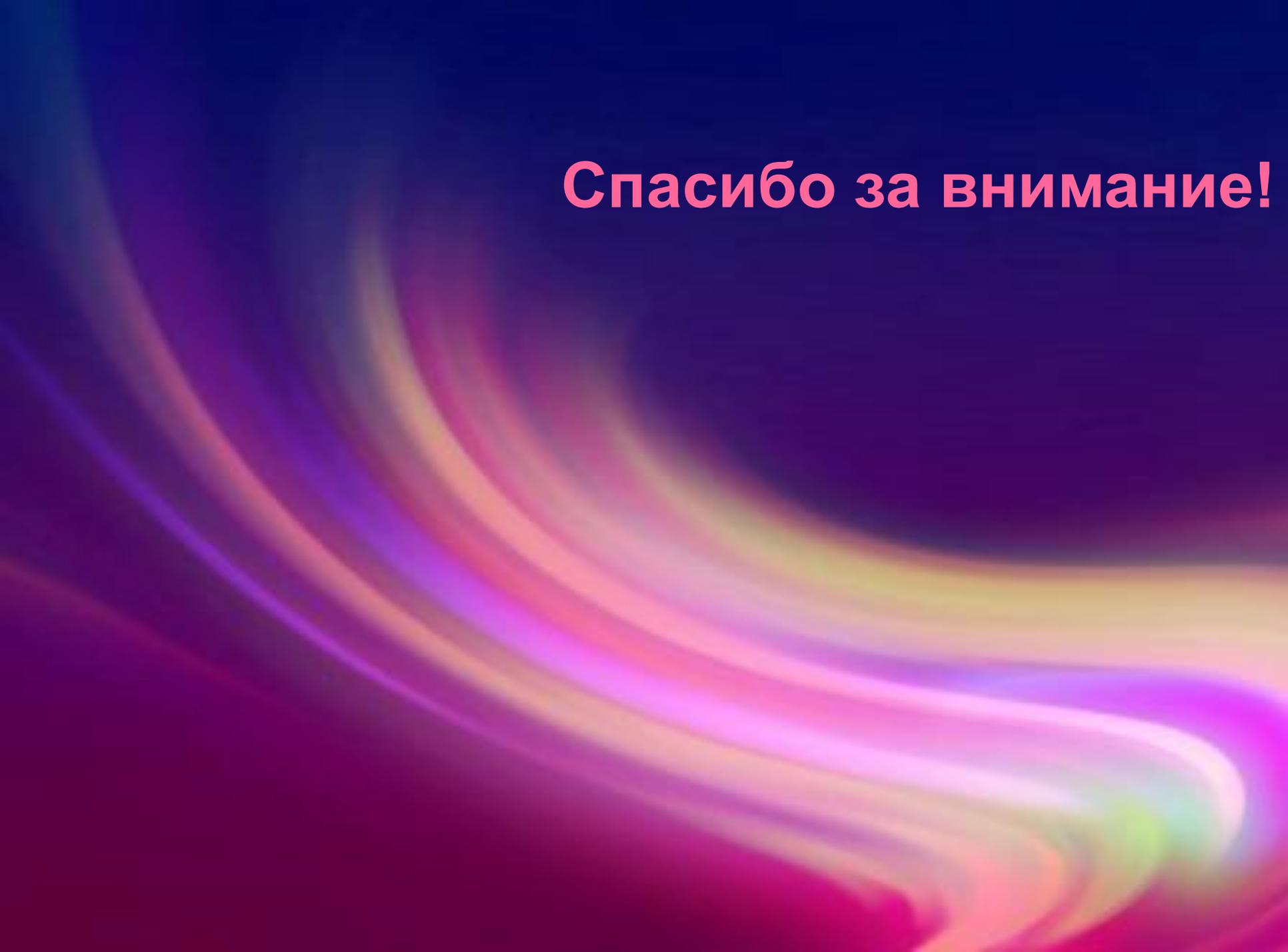
«Ретроспективная» валидность

- **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**



- ❖ 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, подобных тем, на которых он определялся

Спасибо за внимание!

The background of the slide is an abstract, artistic composition of flowing, wavy lines. The colors are vibrant and varied, including deep purples, magentas, pinks, and soft blues. The lines appear to be moving and blending together, creating a sense of motion and depth. The overall effect is a dynamic and visually appealing gradient that transitions from darker tones on the left to lighter, more saturated colors on the right.