



**Современный
Гуманитарный
Университет**

Дистанционное образование

Рабочий учебник

Фамилия, имя, отчество _____

Факультет _____

Номер контракта _____

ПСИХОЛОГИЧЕСКИЙ ПРАКТИКУМ

ЮНИТА 8

**ПСИХОДИАГНОСТИКА: ВИДЫ НАДЕЖНОСТИ И
ВАЛИДНОСТИ ПРОЦЕДУР**

МОСКВА 1999

Разработано А.В. Мощенко, д-ром психол. наук

Рекомендовано Министерством общего и профессионального образования Российской Федерации в качестве учебного пособия для студентов высших учебных заведений

КУРС: ПСИХОЛОГИЧЕСКИЙ ПРАКТИКУМ

Юнита 1. Общая характеристика психологического исследования.
Юнита 2. Процедуры получения и описания эмпирических данных.
Юнита 3. Метод наблюдения в психологической практике.
Юнита 4. Стандартные способы обработки и представления данных.
Юнита 5. Психометрия: методы одномерного и многомерного шкалирования.
Юнита 6. Корреляционные исследования и квазиэкспериментальные планы.
Юнита 7. Эксперимент: внутренняя и внешняя валидность.
Юнита 8. Психодиагностика: виды надежности и валидности процедур.

ЮНИТА 8

Рассмотрены проблемы современной психологической диагностики, методические аспекты психодиагностики, принципы конструирования психодиагностических тестов.

Для студентов Современного Гуманитарного Университета

Юнита соответствует образовательной профессиональной программе №1

ОГЛАВЛЕНИЕ

ДИДАКТИЧЕСКИЙ ПЛАН	4
ЛИТЕРАТУРА	5
ПЕРЕЧЕНЬ УМЕНИЙ	6
НАУЧНЫЙ ОБЗОР	7
1. Проблемы современной психологической диагностики	7
1.1. Понятие психологической диагностики	7
1.2. Особенности современной психологической диагностики	10
1.3. Проблема влияния на человека общественной среды	14
1.4. Проблема выявления наследственно обусловленных особенностей психики	17
2. Методические аспекты психологической диагностики	18
2.1. Зарождение тестирования	18
2.2. Шкала Бине и ее модификации	25
2.3. Типология тестов по субъекту измерения	27
2.4. Типология тестов по форме предъявления	32
2.5. Формализация психодиагностических испытаний	36
3. Принципы конструирования психодиагностических тестов	38
3.1. Надежность психодиагностических тестов	38
3.2. Валидность психодиагностических тестов	44
3.3. Стандартизация психодиагностических тестов	52
3.4. Методика конструирования психодиагностических тестов	56
ЗАДАНИЯ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ	61
ТРЕНИНГ УМЕНИЙ	64
ГЛОССАРИЙ*	

* Глоссарий расположен в середине учебного пособия и предназначен для самостоятельного заучивания новых понятий.

ДИДАКТИЧЕСКИЙ ПЛАН

Проблемы современной психологической диагностики. Понятие психологической диагностики. Особенности современной психологической диагностики. Проблема влияния на человека общественной среды. Проблема выявления наследственно обусловленных особенностей психики.

Методические аспекты психологической диагностики. Зарождение тестирования. Шкала Бине и ее модификации. Типология тестов по субъекту измерения. Тесты специальных способностей. Интеллектуальные тесты. Личностные опросники. Типология тестов по форме предъявления. Бланковые тесты. Групповые тесты. Аппаратурные тесты. Формализация психодиагностических испытаний.

Принципы конструирования психодиагностических тестов. Надежность психодиагностических тестов. Валидность психодиагностических тестов. Стандартизация психодиагностических тестов. Методика конструирования психодиагностических тестов.

ЛИТЕРАТУРА

Базовая

- * 1. Дружинин В.Н. Экспериментальная психология: Учебное пособие. М.: Инфра-М, 1997. 256 с.
- * 2. Кэмпбелл Д. Модели экспериментов в социальной психологии и прикладных исследованиях. СПб.: Социально-психологический центр, 1996. 392 с.

Дополнительная

- * 3. Корнилова Т.В. Введение в психологический эксперимент: Учебник. М.: МГУ, 1997. 256 с.
- 4. Крылов В.Ю. и др. Математические методы в современной психологии. М.: ИП РАН, 1995. 342 с.
- 5. Логвиненко А.Д. Измерения в психологии: математические основы. М., 1993.
- 6. Паповян С.С. Математические методы в социальной психологии. М.: Наука, 1983.
- * 7. Гласс Дж., Стэнли Дж. Статистические методы в педагогике и психологии: Пер. с англ. М.: Прогресс, 1976. 495 с.
- * 8. Суходольский Г.В. Основы математической статистики для психологов. Л.: Изд-во ЛГУ, 1972. 429 с.
- * 9. Бурлачук Л.Ф., Морозов С.М. Словарь-справочник по психологической диагностике / Отв. ред. С.Б.Крымский. Киев: Наукова думка, 1989. 200 с.
- 10. Еремеев Б.А. Статистические процедуры при психологическом изучении текста: Учеб. пособие. СПб.: Образование, 1996. 54 с.
- * 11. Немов Р.С. Психология: В 3-х кн. Кн.3. Психодиагностика. Введение в научное психологическое исследование с элементами математической статистики. М.: ВЛАДОС, 1998. 632 с.
- 12. Татарова Г.Г. Методология анализа данных в социологии (введение): Учебное пособие для вузов. М.: Изд. дом "Стратегия", 1998. 224 с.
- 13. Клайн П. Справочное руководство по конструированию тестов. Киев, 1994.
- 14. Лаак Ян тер, Брюгман Г. Оценивание развития и процесс возрастных изменений // Мир психологии. 1999. №3.

Примечание. Знаком (*) отмечены работы, использованные при составлении тематического обзора.

Современный Гуманитарный Университет

ПЕРЕЧЕНЬ УМЕНИЙ

№ п/п	Умения	Алгоритмы
1	Расчет надежности теста методом расщепления	1. Расщепление теста на две эквивалентные части. 2. Расчет коэффициента корреляции двух частей теста $R_{1,2}$. 3. Вычисление индекса надежности двух частей теста $R(X,T)$.

1. ПРОБЛЕМЫ СОВРЕМЕННОЙ ПСИХОЛОГИЧЕСКОЙ ДИАГНОСТИКИ

1.1. Понятие психологической диагностики

Психологическая диагностика - область психологической науки, разрабатывающая методы распознавания и измерения индивидуально-психологических особенностей личности, методы классификации и ранжирования людей по психологическим и психофизиологическим признакам. Такое толкование позволяет рассматривать в ее рамках всевозможные психологические и психофизиологические методики, претендующие на то, чтобы отличать одного человека от остальных или одну группу людей от всей совокупности. Поскольку речь идет о методах классификации и ранжирования людей, то в число этих методов вполне правомерно войдут и адекватные статистические приемы, обеспечивающие объективность и доказательность использования диагностических методик. Такое понимание диагностики отнюдь не исключает в тех случаях, когда это требует исследовательская задача, и фиксирования неблагоприятия. Но нельзя только на нем сосредоточить всю психологическую диагностику.

Психологическая диагностика осуществляется с помощью большого и всевозрастающего набора средств, инструментов, которые можно объединить в понятие **диагностические методики**. Конструирование их, уточнение требований, которым они должны удовлетворять, правила их применения, границы выводов, на которые могут претендовать получаемые результаты, — все это образует специфическую для психологии область знаний. Диагностика представляет собой один из видов психологического экспериментирования и наблюдения. Ее массовость предполагает получение сравнимых результатов — без сравнения диагностика не выполнит своих задач. Благодаря сравнению результатов устанавливаются относительные точки отсчета. Поэтому вариабельность всех условий и факторов при диагностировании, охватывающем в разных местах и одновременно разные группы людей, должна быть минимальной. Тогда можно будет с достаточным основанием сделать вывод о действительных различиях испытуемых по диагностируемым признакам. Поэтому в диагностике придается столь большое значение формализации, и в частности статистике.

Используется ли диагностика в теоретическом или в практическом исследовании, цель ее заключается в том, чтобы дать материал для психологического заключения о группе лиц или отдельном лице, показав, чем это лицо отличается от других или чем эта группа отличается от других групп той же совокупности. И там и тут предполагается сравнение. Статистические приемы привлекаются для того, чтобы доказать, что сравниваемый признак (переменная) свидетельствует об отличии диагностируемых — лица или группы — от тех, с кем сравнивают, и что это отличие значимо. Статистика усиливает доказательства, получаемые при сравнении переменных, обосновывает их и показывает уровень их вероятности.

Переменная — любая реальность, наблюдаемые изменения которой по конкретным показателям методики могут быть зафиксированы и измерены в какой-либо шкале. **Латентная переменная** — гипотетическая переменная, которая не поддается измерению в исследовании, но в модели связи между переменными характеризует сумму всех неучтенных влияний на измеряемую переменную.

* Жирным шрифтом выделены новые понятия, которые необходимо усвоить. Знание этих понятий будет проверяться при тестировании.

Всегда ли психолог обязан прибегать к статистическим доказательствам? В некоторых случаях именно этот признак, и только он, отличает данное диагностируемое лицо от других и применения статистики не требуется. Достаточно описать признак и убедиться в его представленности. Центр тяжести доказательства перемещается в подобных ситуациях в другую плоскость: нужно установить не уровень выраженности признака, не степень его отличия от уровня, наблюдающегося у других людей, а прежде всего подметить, обнаружить данный признак как таковой. Если удастся это сделать, то никаких других показателей для диагноза не требуется, каждый специалист понимает, что наличие данного признака выделяет человека или группу из среды других людей, которые данным признаком не обладают. Диагностика, целью которой является только обнаружение определенного признака (констатация: есть — нет), отличается от диагностики, направленной на то, чтобы установить пункт нахождения отдельного испытуемого на оси выраженности признака. По центральной тенденции (средней арифметической, медиане, моде) определяют также пункт нахождения группы испытуемых.

Первый вид диагностики обычен для клинического типа исследования, где устанавливается и уточняется отклонение от нормы (в данном контексте имеется в виду не статистическое понятие нормы, а норма как отсутствие патологических симптомов в психике). Результаты этого вида диагностики помогают планировать лечебные мероприятия, углубить понимание отклонений и т.д. **Второй вид диагностики** применяется при изучении индивидуально-психологических различий для классификации испытуемых, разделения их на группы для последующих соотношений с каким-то “внешним критерием” — психологическими проявлениями испытуемых в лабораторных опытах, в их учебной и производственной деятельности.

Приведем пример диагностического экспериментально-психологического исследования психического больного (по Б.В.Зейгарник). Предъявлялась карточка лото, где изображен барабан. Описание больной: “Горшок с кисточкой”. Экспериментатор: “Может быть, что-нибудь другое?” Больная: “Булка, которую кладут в кастрюлю, а это кренделек (на палочку). Похоже и на шапку.” Важно отметить, что предшествующее экспериментально-психологическое исследование не выявило каких-либо грубых изменений психической деятельности больной. Больная правильно усваивала инструкцию, хорошо передавала содержание, подтекст прочитанной книги, понимала условный смысл пословиц, метафор. Но узнавание предметов грубо нарушено. Патологические изменения в направленном диагностическом исследовании выступили с такой резкостью и выпуклостью, что сравнивать результат с результатами нормы, прибегая к статистике, излишне.

Для каждого из видов диагностики приходится применять такой материал, который бы в наибольшей степени отвечал задачам исследования. И даже в случаях, когда используется тот же самый материал, он выступает в разных диагностических функциях: найти признак у испытуемого — одна функция, найти место испытуемого в континууме — другая. Так, в одном исследовании в качестве инструмента диагностики использовался тест “Простые аналогии”. Цель состояла в том, чтобы выяснить, отражает ли работа над тестом возрастные изменения в мыслительной деятельности у нормально развивающихся детей и дифференцирует ли она этих детей от умственно отсталых школьников. Исследование направлено не только на определение (по числу решенных заданий теста) места каждого испытуемого в континууме решений теста и на установление по тому же континууму мест группы нормально развивающихся и умственно отсталых детей, но и на выявление некоторых психологических особенностей работы умственно отсталых, т.е. выявление таких признаков, которые отличают их от нормальных детей.

Весьма возможно, что каждый из видов диагностики формирует у применяющих его психологов свое исследовательское мышление. Среди психологов-диагностов есть сознательные и стихийные сторонники только одного вида диагноза: того, который исключительно или наиболее часто по характеру их деятельности с успехом ими применялся, выступал как одно из звеньев их исследовательской работы.

Однако каждый вид диагностики хорош на своем месте, каждый имеет свое назначение, свои функции. Оба вида диагностики — особые формы сравнения. Если они противопоставляются один другому, то не по существу, а только в зависимости от задач, которые решаются исследованием.

Итак, диагноз вида “нахождение точки на оси континуума” — это установление места каждого ранжированного испытуемого (или группы), когда ранг выставляется по количественной оценке процесса или результатов, а диагноз “наличие или отсутствие признака” предполагает объединение испытуемых (или выделение одного испытуемого) по обнаруженному признаку. В том и другом виде диагностирования основой служит сравнение. В первом виде есть прямой резон опереться на статистическое доказательство, во втором это не является необходимостью, здесь нужны только убедительные доводы, свидетельствующие о том, что признак налицо, при этом не исключается возможность зарегистрировать степень его выраженности (едва представлен, заметен, очень резко выражен и т.п.). По терминологии С.С. Стивенса первый вид диагностирования позволяет в полной мере использовать при обработке статистику шкалы порядка, в некоторых случаях — шкалы интервалов; второй вид — статистику шкалы наименований, иногда шкалы порядка.

Как уже говорилось в начале главы, методики, объединяемые психологической диагностикой, предназначены для классификации (т.е. разделения на группы) и ранжирования людей по психологическим и психофизиологическим признакам. Поэтому методики, не способствующие решению этих задач, для диагностики непригодны.

Методики “нахождения точки на оси континуума”, т.е. ранжирования, имеют дело с возрастающими и убывающими признаками, и допустимо, что с какой-то точки на оси континуума характеристика приобретает определенное качество, например с определенного значения, полученного по методике “выраженность закона силы”, испытуемые будут квалифицироваться как представители группы сильных по преобладанию процесса возбуждения. Методики “наличия или отсутствия признака” диагностируют качество; они по большей части применяются в клинической психологии и нейропсихологии для отграничения нормы от патологии.

Иногда при определении психологической диагностики преимущественное внимание обращается на то, что с ее помощью можно получить целостную характеристику отдельной личности и что именно в этом состоит ее главное назначение. Трудно спорить с тем, что получение целостной характеристики личности — одна из заманчивых перспектив диагностики. Но диагностика может трактоваться так только в случае, если, во-первых, каждая составляющая этой характеристики возникает как продукт научного сравнения данных конкретной личности с данными какой-то достаточно большой совокупности людей, во-вторых, если будет найден научный принцип синтеза, соединения отдельных данных. Без этих двух условий получение целостной характеристики останется делом искусства, а не науки.

Неубедительным представляется такое определение диагностики, согласно которому, “определяя предмет психодиагностики, следует исходить из того, что психодиагностика предполагает постановку психологического диагноза, а диагноз, где бы он ни ставился — в медицине, в технике, в управлении, в прикладной психологии, — это всегда поиск, выявление скрытой причины обнаружившегося

неблагополучия, чаще всего в условиях множественности причинно-следственных связей” (В.И. Войтко. Ю.З. Гильбух, 1976, с. 18). Авторы, придерживающиеся этой точки зрения, пишут: “Психодиагностика не просто имеет дело с отдельным индивидом — она занимается лишь теми людьми, о поведении, деятельности, психическом состоянии которых заранее известно, что они характеризуются определенными отклонениями, недостатками и т.п.”. Они поясняют, что речь идет в данном случае о любом действии, поступке, состоянии, свойстве, качестве и т.п., которое рассматривается в том или ином отношении как неудовлетворительное, нежелательное. В такой постановке произвольно сужается область диагностики, ей предоставляется констатировать и исследовать только отклонения, только нежелательное. Но диагноз отсутствия отклонений и диагноз благополучия тоже имеет научную и практическую значимость, достаточно сослаться на исследования по дифференциальной психофизиологии. В некоторых случаях задача вообще не может трактоваться с позиций оценки “желательное — нежелательное” или “наличие и отсутствие отклонений”. Подобные оценки более чем относительны.

Высказывая в исследованиях свою точку зрения, эти авторы должны были бы решить методический вопрос — как выявить нежелательное, если нет научных и эмпирических данных о том, что такое желательное, и чем первое в диагностических показателях отличается от второго. Приходится вновь повторить, что без сравнения диагностика существовать не может.

Психологическое тестирование, понимаемое как объективное стандартизованное измерение индивидуальных различий, подразумевает *этапы* непосредственного конструирования методик, проверки их надежности и валидности, проведения тестовых испытаний, измерения и оценки индивидуальных результатов, завершаемых их целевой интерпретацией.

Психологическая диагностика находит, а в дальнейшем еще в большей степени найдет применение в школе и педагогических исследованиях; она занимает важное место в психологии изучения профессионального труда. Психологическая диагностика все шире внедряется и в область патопсихологии. Нет смысла заниматься перечислением тех областей теоретической и практической работы, где в диагностических исследованиях ощущается большая потребность, — их очень много.

Но, пожалуй, прежде всего школа испытывает нужду в применении диагностических методик. Так, с их помощью шире развернутся работы по выявлению детей с задержками психического развития. Перед психологами стоит задача создать специализированные методики для этого рода диагностики. Крайне необходимо включить диагностику в исследование уровней психического развития ребенка, контроля возрастной динамики. Несомненно, методики психологической диагностики займут место и в научно-методических психолого-педагогических исследованиях. Применение этих методик даст доказательную картину сравнительного значения средств и методов обучения и воспитания и для развития детей, и для проверки глубины усвоения ими пройденного материала. Должна занять свое место диагностика уровня знаний, так называемые тесты достижений, или успешности. Отличаясь по характеру решаемых задач от тестов психологической диагностики в собственном смысле, тесты успешности смогут выполнить свое назначение, если их конструирование, проверка и применение будут осуществляться с учетом уже разработанных психологами критериев.

1.2. Особенности современной психологической диагностики

Цель современной психологической диагностики в ее наиболее распространенных формах (тесты, проективные методики, психофизиологические методики, опросники) заключается в том, чтобы фиксировать и описывать в

упорядоченном виде психологические различия как между людьми, так и между группами людей, объединенных по каким-нибудь признакам. Так, психологическая диагностика устанавливает, что по данному признаку, например по уровню технического мышления, один ребенок 12 лет отличается от другого ребенка того же возраста и той же школьной подготовки. Или же методики зафиксируют, что V класс, состоящий из 30 учеников городской школы, отличается от V класса, состоящего также из 30 учеников, но уже сельской школы, по уровню обобщения. Во всех случаях диагностирования исследователю приходится упорядочивать полученный материал. Он может расположить данные в определенной последовательности, например по убыванию или возрастанию полученных результатов. Без последующей группировки результатов работа по психологической диагностике невозможна. После выполнения указанной операции группировки и классификации исследователь может при необходимости пойти дальше, найти смысловые и статистические критерии группирования и классификации, установить, насколько существенны различия между полученными данными, насколько сходны данные, полученные в одной и той же группе.

Если бы психологических различий не существовало, то отпала бы надобность в диагностике. В число диагностируемых признаков в зависимости от задач исследования могут входить психологические различия по возрасту, полу, образованию и культуре в самом широком понимании этих терминов, по психическим состояниям, психофизиологическим генетически детерминированным особенностям и др. Принимая во внимание огромное разнообразие признаков, подлежащих диагностированию, исследователи всегда подходят к конструированию методик в соответствии с представлениями о самом признаке. Прежде всего исследователь заботится о том, чтобы методика дифференцировала испытуемых, обнаруживала различия между ними по избранному признаку. Может произойти так, что методика будет стимулировать одинаковые ответы или другие реакции значительного большинства или всех испытуемых, например, когда уровень требований, содержащийся в методике, ниже уровня действительных различий испытуемых. Положим, методика требует выполнения простых арифметических действий, а испытуемые владеют хорошо отработанными умениями совершать такие действия. Естественно, что никакой дифференциации при выполнении подобных действий в данной выборке не будет. Однако то же получится, если уровень требований методики выше уровня действительных различий. Положим, методика требует владения умением интегрирования (имеется в виду математика), а испытуемые хорошо владеют лишь четырьмя действиями арифметики. В первой диагностической ситуации все испытуемые выполнят предложенное задание, а во второй задание не выполнит ни один человек, значит, никаких различий между испытуемыми в обеих ситуациях получить не удастся.

Различия, о которых шла речь, наблюдаются в каждом экспериментальном, теоретическом и прикладном исследовании. Это, конечно, не означает, что такие различия всегда интересуют исследователей, что они всегда склонны их выделять и интерпретировать. Любое исследование направлено на решение ограниченного числа научных задач. Иногда исследователь, руководствуясь принципами научной добросовестности, отмечает, что между испытуемыми обнаружены психологические различия, но далее этого не идет, т.е. не ранжирует испытуемых по различиям и не выдвигает по данному пункту никаких гипотез. Признавая концепцию целостности человеческой психики, следует все-таки согласиться с тем, что выступавшие в исследовании признаки, заметно и существенно отличающие одного человека (или одну группу людей) от других, не могут не отразиться и на результатах всего экспериментального исследования и его выводах. Признак, отличающий человека или группу людей от других, может, конечно, носить случайный характер, но и это нужно доказать. При отсутствии

такого доказательства остается подозрение, что наличие колебаний по выделенному признаку должно быть перенесено и на выводы. Обычно в выводах теоретического исследования констатируют, что после запланированных изменений в условиях эксперимента произошли какие-то изменения и с изучаемым признаком, что выразилось в различиях между экспериментальной и контрольной группами. Это и есть фиксация психологических различий между группами, а она входит в круг ведения психологической диагностики. В психологической диагностике обычно используются методики, предварительно прошедшие критериальную проверку, что позволяет повысить доказательность существования различий. Очень важно, чтобы использование методик, прошедших критериальную проверку, стало правилом, а не исключением в психологических исследованиях.

Пока что такие методики чаще применяются в прикладных исследованиях, направленных на добытие психологических данных, например о различиях между людьми по практически значимому признаку. Далеко не всегда психологи сначала выдвигали теоретически обоснованную гипотезу о генезе и природе возможных различий, а затем, в соответствии с такой гипотезой, разрабатывали диагностические методики, чтобы проводить испытания. Исторически все происходило нередко в обратном порядке. Психологи подбирали такие методики (тесты, опросники и психофизиологические методики), с помощью которых устанавливались различия между людьми. Различия затем соотносили с успешностью учебной или производственной деятельности лиц, прошедших испытания. Теоретическому осмыслению психологического содержания диагностических средств не придавалось особенно большого значения, например, испытания интеллекта ведутся уже около века, а о том, что такое интеллект и каково его место в системе психологических понятий, тестологам известно не более, чем во времена А. Бине. Главное внимание обращается ими на нахождение связи результатов индивидуально-психологических испытаний с практикой.

Печать практической направленности в той или иной степени несут на себе самые известные батареи тестов — шкала А. Бине, тесты Д. Векслера, Р. Кеттэла. Сравнительно недавно в западных странах тестированием были охвачены большие контингенты людей — претенденты на рабочие места в промышленности и управлении, учащиеся и студенты, рекруты и военные. Распространение тестов вызвало к жизни особую отрасль издательского дела — печатание тестов, а издатели, всячески рекламируя выпускаемую продукцию, в свою очередь способствовали развитию тестирования.

В последние годы несколько меняется отношение к диагностическим испытаниям. Исследователей беспокоит теоретическая необоснованность некоторых методик. Одним из способов восполнения теоретических пробелов должны служить математические приемы стандартизации методик. Действительно, эти приемы сыграли некоторую роль как заслон против поделок дилетантов, изготовлявших психологически бессодержательные тесты и опросники. Но как бы ни были совершенны приемы установления надежности и валидности, они не могут создать психологический фундамент диагностики.

Статистические критерии выдвигают новые, собственно математические и методические проблемы, но несколько не продвигают разработку теории индивидуальных различий. Вместе с тем возникла особая линия развития диагностики, где содержательное психологическое толкование полученных данных заменялось подсчетами баллов, вычислениями их взаимосвязей, применением сложнейших приемов их формального анализа без обращения к самой психологии. Возникли удобные для работы и не требующие психологических знаний системы испытаний, итоги которых по заранее подготовленным таблицам с так называемыми нормами мог интерпретировать любой непсихолог. Тестам стала реально угрожать депсихологизация.

Новое отношение к диагностическим методикам, особенно ярко выразившееся в утрате доверия прежде всего к тестам, обозначилось в начале 60-х гг. В критическую переоценку ценностей включились видные психологи-тестологи. В последние 15—20 лет появляется все большее число публикаций, содержащих требования пересмотра отношения к тестам и другим диагностическим методикам.

Причина такой переоценки состоит в том, что произошла перестройка образования, оно превращается в институт развития способностей всего молодого поколения. Это привело к тому, что интеллектуальные тесты оказались непригодными для дифференцирования испытуемых. Большинство известных психологов считают, что тестами интеллекта испытывается уровень образования, а в более широком плане — степень приобщения к культуре. Ясно, что при таком толковании на первый план выступают новые, требующие специальной разработки методологические проблемы. Нуждается в более точном определении само понятие “культура”, возникают вопросы о тех ограничениях, которые накладывает культура на человека.

Из массы остро поставленных вопросов о влиянии культуры на психику человека выделяются те, которые касаются воздействий на индивидуальные особенности факторов наследственности и социального окружения. Психологическая диагностика не готова различать эти воздействия. Можно интерпретировать любой признак, выявляемый в диагностических испытаниях, как зависящий от неразделимого влияния факторов культуры и наследственности. При такой постановке как бы снимается и сама проблема изучения наследственности — наследственное всегда проявляется только в приобретенном.

Повседневная педагогическая практика настоятельно требует выделения динамических особенностей психики. Достаточно обратиться к так называемому **оптимальному индивидуальному темпу деятельности**, наследственная детерминированность которого изучалась в ряде отечественных работ. Допустим, учитель не знает о генетической природе темпа. Если сам он принадлежит к людям с высоким природным темпом, то чаще всего требует от воспитанников, чтобы они работали в таком темпе, который не то что неудобен, но и просто невозможен для одних и крайне труден для других. То, что одни воспитанники будут выполнять без особого напряжения, у других вызовет нарастающее напряжение; но даже и постоянное сохранение контроля за своими действиями, стремление соблюдать навязанный темп для этих воспитанников окажутся невозможными. Это лишь частный пример того, к чему может привести неумение или нежелание считаться с вполне учитываемым в его обособленности от психического содержания фактором, отражающим влияние наследственности.

Все рассмотренные аспекты свидетельствуют по крайней мере о том, что проблема наследственности в диагностике не может быть ни снята, ни отодвинута до лучших времен, что имеются научные перспективы ее разработки.

Влияние на психику человека материальной и общественной среды его существования, что может быть названо влиянием культуры, и влияние на психическую деятельность наследственных факторов составляют единство. Но это единство не освобождает психолога от научного исследования каждого из влияний. Вместе с тем в обычной процедуре тестовых испытаний решающую роль играет общее число правильно выполненных заданий. При этом в сущности не придается значения их психологическому содержанию. Предполагается как само собой разумеющееся, что о состоянии интересующей психолога стороны психики (или психической функции) можно безоговорочно судить по достигнутому каждым испытуемым количественному результату. Изучение влияния культуры применительно к этому уже ставшему общепринятым в тестологии положению заставило посмотреть на дело по-другому. Результат, учитываемый психологом,

завершает действия, выполняемые испытуемым: он воспринимает, например, текст и графические изображения, предложенные ему, старается применить к полученному заданию инструкцию, сообщенную ему перед выполнением теста, не упускает из виду и ограниченность времени, предоставленного на выполнение заданий. В анализе процесса выполнения теста неизбежно должны обнаружиться все те компоненты, которые в конечном счете сказались на результате, но в нем не различимы, не поддаются специальному и изолированному рассмотрению, ибо результат отражает их совокупное действие. Только процессуальный анализ психологического диагностического испытания позволяет выделить и собственно культурные, и генетически обусловленные факторы.

Разумеется, только перечисленными проблемами не ограничиваются задачи психологической диагностики. Так, в ее методическом арсенале еще не найдено окончательное место математико-статистическим приемам, хотя некоторые из них прочно вошли в практику диагностирования. Возникает много нерешенных вопросов и в связи со специфическими задачами применения методик психологической диагностики в разных практических ситуациях. Психологические методики возникали не из интереса к развитию теории, а как ответ на социальное требование. Поэтому их строение и содержание отражают социальную ситуацию, в которой их применяли. Диагностические методики чаще всего выступали как инструмент **психологического испытания**, т.е. их применяли в такого рода социальной ситуации, которая заранее предусматривала отсев, а в более широком плане — классификацию людей.

Современную психологическую диагностику характеризуют проблемы влияния на развитие человеческой психики материальной и общественной среды (культуры) и наследственных факторов. Можно надеяться, что при их рассмотрении откроются плодотворные пути для разработки других, не поставленных прямо проблем.

1.3. Проблема влияния на человека общественной среды

Одним из первых исследователей, обративших внимание на влияние культуры и подчеркнувших ее значение в диагностических исследованиях, был Б. Саймон (1958). Еще задолго до него подобные мысли высказывали и другие психологи: достаточно назвать Э. Меймана в Германии, Э. Торндайка в США. Б. Саймон особенно резко подчеркнул, что оценки испытуемых, которые получает исследователь, прежде всего отражают не их подлинные возможности, а те социальные условия, в которых они родились и выросли. Саймон приводит в качестве примера ряд вербальных тестов с использованием слов, значение которых должен знать ребенок, чтобы хорошо ответить на вопросы теста. Естественно, что используемые в тестах слова лучше известны одним детям, хуже другим, а иным они совсем неизвестны. И в результате дети, которые не имели возможности много читать или развивать разговорную речь, неизбежно оказываются в невыгодном положении. Трудно придумать более эффективный метод разделения детей в соответствии с социальной средой, чем стандартизованные вербальные тесты на определение умственной одаренности. **Вербальные тесты** - тип тестов, в которых материал тестовых заданий представлен в вербальной (словесной) форме.

Исследование Б. Саймона относится только к английским детям, т.е. детям, воспитывавшимся в одной национальной культуре (при всем ее разнообразии). Естественно, что эти свойства тестов выступают еще резче, когда объектами диагностики становятся представители разных этнических групп, разных национальных культур, а равно и лица иной общественной среды. В последние годы расширяются диагностические исследования, охватывающие детей и взрослых, сформировавшихся в условиях, отличающихся от тех, которые в целом

называют европейской культурой, например представителей некоторых африканских этнических групп. Становится очевидным, что мысленное оперирование любым психологическим материалом — предметами, словами, рисунками — при этом различно, поскольку все знания и умения приобретаются в жизненном опыте. Сюда включается школьное обучение, постоянное получение и усвоение значимой в данной социальной группе информации — нравственной, исторической, экономической, политической, эстетической, традиционные формы общения и усваиваемые ценностные ориентации и идеалы. Весь описанный круг явлений собственно и обозначают понятием “культура”. В таком значении этот термин и употребляется в настоящем рабочем учебнике.

Методики психологической диагностики, наиболее распространенные на Западе, проникшие в виде адаптированных переводов в нашу страну и вызвавшие массу подражаний, были ориентированы (быть может, без ясного осознания авторами и конструкторами этих методик) на определенную культуру. Конечно, эта культура не изолирована от других культур. В ней представлены элементы, имеющиеся и в других культурах. Но с наибольшей полнотой она отражала лишь социально-психологический опыт группы определенных западных стран, а среди них — США. Между тем даже в названных странах эта культура не может считаться единственной. Достаточный уровень владения ею присущ наиболее привилегированным социальным группам и тем социальным группам, которые к ним близки. Для других социальных групп многие элементы этой господствующей культуры чужды и овладение ими в естественном ходе развития молодых поколений имеет место лишь в очень ограниченных пределах.

Понадобилось немало экономических и политических потрясений, чтобы роль национальных культур стала осознаваться и учитываться психологами-диагностами (хотя какая-то часть психологов и теперь остается на старых позициях).

Особенности культуры проникают не только в психологическое содержание методик, культура столь же очевидно сказывается и на самой процедуре диагностирования. Так, инструкция (что и как нужно делать испытуемому) и порядок эксперимента в главных чертах воспроизводят отношение учителя к ученику в европейской и американской школах. Несомненно, что испытуемый, прошедший школьное обучение такого образца в сравнительно недавнее время, будет чувствовать себя в обстановке тестирования более привычно и свободно, чем тот, у кого подобного опыта нет или он незначителен. Испытуемому сообщают, что он должен выполнять задания с наибольшей доступной ему скоростью — это требуется почти во всех тестах и тоже совпадает с обычными порядками в европейской и американской школе. Но ведь в некоторых культурах воспитывается другая установка: неторопливое раздумье, сдержанность.

Историческая несомненность существования не единой и общей для всех наций, народов и социальных групп культуры, а множества своеобразных культур с их подразделениями — субкультурами — лишает какого бы то ни было научного значения положение о том, что представители разнообразных культур якобы имеют равные возможности, когда они подвергаются психологическому испытанию с помощью диагностических методик (тестов, опросников и т.п.), созданных на базе одной культуры.

Установление рангов по успешности выполнения психологических заданий всегда и неизбежно ставит представителей любой культуры, кроме той, на которую ориентировались составители методик, в неблагоприятные условия. В последние годы это обстоятельство стало предметом специального внимания. Оказалось, что в группах со смешанным национальным составом (что и совпадает в общем с принадлежностью к разным культурам) результаты тестирования дали невысокие совпадения с последующими успехами испытуемых, хотя применялись обычные в этом отношении тесты. В психологической терминологии появились обозначения — справедливые и несправедливые тесты. Последними называют тесты, по

результатам которых нельзя судить о будущей производственной эффективности, так как влияние культуры в них особенно сильно. Однако вряд ли есть надобность именовать какие-нибудь тесты несправедливыми. Все психологические методики отражают определенную культуру, и вопрос в том, насколько велик разрыв между культурой испытуемого и теми требованиями, которые заложены в психологической методике.

В последние полтора-два десятилетия западные психологи-диагносты все решительнее подчеркивают роль социально-экономических и культурных факторов в формировании индивидуально-психологических различий между людьми. Не исключается, впрочем, и роль наследственности. Выявленные диагностическими испытаниями особенности людей рассматриваются как продукт совместного действия среды и наследственности; вместе с тем обычно отмечается, что современное состояние наук о человеке (психология, нейрофизиология, генетика и др.) еще не позволяет найти долю того и другого фактора, особенно при анализе конкретных случаев. Это обстоятельство отчасти используется для объяснения или оправдания того, почему при психологических испытаниях принимался во внимание только результат диагностирования. Он принимался как определенная существующая данность, от чего бы он ни зависел. С этой данностью и производились все сопоставления жизненных показателей, “внешнего критерия”.

В последние годы получил широкое признание взгляд, согласно которому диагностическая психологическая методика раскрывает степень приобщения данного человека (или группы лиц) к культуре, служившей ориентиром при составлении данной методики. Это и может объяснить получение различных результатов при испытаниях людей, относящихся к разным культурам.

Однако вряд ли можно думать, что все люди, сформировавшиеся в какой-то определенной среде, будут совершенно тождественными по диагностическим результатам. И в самом деле, любые испытания даже культурно-гомогенных групп всегда дают разброс, дисперсию, а следовательно, возникает необходимость ранжирования и в таких группах. Не следует ли именно здесь включить и пустить в дело объяснительный принцип наследственности? Может быть, с его помощью удастся получить законченную картину: границы результатов зависят от культуры той группы, которая подвергается испытаниям, а колебания внутри этих границ — от наследственных особенностей тех лиц, которые вошли в группу?

В культурно-гомогенной группе имеются не относящиеся к наследственности влияния, как бы предопределяющие внутригрупповую дисперсию. Иными словами, группы, составленные по единому формальному признаку, нельзя считать подлинно гомогенными. И в них представлены лица, в разной степени владеющие той культурой, на которую ориентированы как диагностические методики, так и процедура испытания. Действительно, если тут и сказывается как-то фактор наследственности, то обнаружить его влияние вряд ли возможно. Даже не обращаясь к фактору наследственности, весь диапазон индивидуальных различий в результатах выполнения психологических методик можно, следовательно, объяснить факторами среды.

Понимание психики в отечественной психологии исключает возможность передачи по наследству того, что составляет содержание собственно психического (знаний, умений, ценностных ориентаций), и всего, что усваивается, приобретает человеком, что условно обозначают термином “культура”. Но именно содержание психического и раскрывают диагностические методики в суммарных результатах их выполнения. Сферу действия наследственных факторов исследуют в рамках различных подходов: объективного, субъективного и проективного.

Объективный подход - подход, при котором диагностика осуществляется на основе успешности и способа выполнения деятельности.

Субъективный подход - подход, при котором диагностика осуществляется на основе сведений, сообщаемых о себе, самоописания особенностей личности, поведения в тех или иных ситуациях.

Проективный подход - подход, при котором диагностика осуществляется на основе анализа особенностей взаимодействия с внешне нейтральным материалом, становящимся в силу его известной неопределенности объектом проекции.

В последние годы определился новый подход к индивидуальным различиям в интеллекте. Он заключается в преимущественном обращении к изучению процессуальных особенностей интеллектуальной деятельности и связанных с ним различий интеллектуального уровня. **Процессуальный подход** может рассматриваться как достаточно общий, пригодный для всех диагностических методик экспериментального характера (он труднее осуществим в методиках, основанных на наблюдении и самонаблюдении, т.е. опросниках). Среди процессуальных характеристик значительное организующее место принадлежит психофизиологическим динамическим характеристикам — накапливается все больше и больше научных оснований для того, чтобы соотносить эти характеристики с генотипом. В частности, речь идет о свойствах нервной системы, изучение которых начато было экспериментально на животных И.П. Павловым и его сотрудниками, продолжено на человеке Б.М. Тепловым, В.Д. Небылицыным и их сотрудниками.

1.4. Проблема выявления наследственно обусловленных особенностей психики

Систематическое изучение влияния наследственности начал Ф.Гальтон. Он полагал, что английская аристократия выдвинулась на роль руководителя общества лишь из-за благоприятных генетических факторов, передающихся от поколения к поколению. Эти идеи Гальтона потеряли к нашему времени какое бы то ни было научное значение. Серьезный недостаток, снижающий ценность его работ, обусловлен и тем состоянием биологических и социальных наук, которое исторически было предоставлено этому исследователю. Генетика только зарождалась. Отсутствовали отработанные статистические приемы доказательства значимости полученных данных. Тем не менее Гальтон сделал ряд важных методических открытий, забывая о которых не могут современные авторы. Он первым начал применять в исследовании близнецовый метод.

Применение близнецового метода значительно повлияло на выделение направления исследований генетически обусловленных особенностей психики человека, включило эти особенности в исследования психологии и психофизиологии. Если обратиться к наиболее изученным в настоящее время генетически обусловленным сторонам психики, к основным свойствам нервной системы, которыми, конечно, не ограничивается наследственный фактор психики, то нужно прежде всего вспомнить, что уже И.П. Павлов категорически высказался за их генотипическую природу. Эта мысль в несколько измененной формулировке была поддержана Б.М. Тепловым.

Опираясь на многосторонние и разнообразные исследования свойств нервной системы и их проявлений, можно с достаточным уровнем вероятности предполагать, как именно скажутся на процессуальной стороне диагностического испытания эти динамические особенности психики. Как уже говорилось, прежде всего надлежит отметить безусловно значительное воздействие темпа психической деятельности, признака индивидуального и генетически детерминированного. Оставляя в стороне некоторые неясности об отнесенности темпа к тому или другому свойству нервной системы (либо к лабильности, либо к подвижности, а может быть, в зависимости от конкретных условий его

проявления, то к тому, то к другому из упомянутых свойств), необходимо вспомнить, какое значение имеет темп в традиционной тестологии. При достаточной и равной отработке умений, обеспечивающих правильное исполнение теста, между испытуемыми останется только одно различие — по темпу.

Сходное воздействие на результаты тестирования оказывают и некоторые другие динамические особенности психики. Так обстоит дело с переустановкой, психическим свойством, о котором в самом нерасчлененном виде писал К. Марбе. Диффузное описание этого психического признака трудно включить в психологическую систему. Между тем, внося необходимые для определения ограничения, позволительно утверждать, что **переустановка** отражает реальную динамическую особенность психики и может быть истолкована как проявление подвижности нервных процессов. Проявление этого свойства наблюдается при переходе от одного действия к другому, с иными характеристиками. Известно также, какое значение имеет **работоспособность**, обусловленная силой нервных процессов в отношении возбуждения. В процессе исполнения тестов это свойство дает о себе знать в изменении кривой работы, в том, насколько проявится вместе со снижением работоспособности тщательность выполнения заданий.

Характерная черта перечисленных выше динамических особенностей психической деятельности состоит в известной широте диапазона их проявлений. Например, недостаточно сказать о подвижности только то, что для нее существенны моменты переходов от одного вида действий к другому. Высокая подвижность выражается и в склонности к такой смене видов действий. Неуловимые для психолога-тестолога причины успехов и неудач отдельных испытуемых, обнаруживающиеся в результатах исполнения диагностических психологических заданий, могут при углубленном анализе процесса исполнения найти высоковероятное объяснение в генетически обусловленной особенности индивидуальной психики.

Нельзя между тем утверждать, будто генетическая обусловленность психики может обнаружиться в значимых степенях при изучении разных привычек, стратегий, рабочих склонностей испытуемых. Имеется неопределенно большое число привычек, стратегий и т.п., одинаково успешно складывающихся на различной генетической почве. В развитии проблемы психологии индивидуальных различий большое значение имеют реальные достижения психологической науки. Они создают предпосылки, способствующие выяснению таких индивидуально-психологических особенностей, которые иногда прямо, иногда опосредованно обусловлены наследственными факторами.

2. МЕТОДИЧЕСКИЕ АСПЕКТЫ ПСИХОЛОГИЧЕСКОЙ ДИАГНОСТИКИ

2.1. Зарождение тестирования

Тест - стандартизированное, часто ограниченное во времени испытание, предназначенное для установления количественных и качественных индивидуально-психологических различий. Термин “тест” переводится как “испытание”, “проверка”, “проба”, “мерило”, “критерий”.

Тестология представляет в настоящее время систему знаний о методах измерения и оценки индивидуальных особенностей личности (в самом широком понимании этих слов), систему, используемую с дифференциальными целями в различных областях человеческой деятельности. Один из самых распространенных методов диагностики индивидуальных моделей психики человека — метод, составляющий ядро тестологии и получивший самое широкое распространение, — **психологическое тестирование**.

С помощью методик психологической диагностики, как известно, изучают и сравнивают между собой особенности психического развития отдельных людей

или групп, подобранных по какому-нибудь существенному признаку (возраст, образование, социальная принадлежность, отношение к профессии и т.п.). Эти методики адресованы различным особенностям психики, поэтому и требования к ним не совпадают. Для удобства изучения целесообразно разделить методики на классы, а в границах класса — на виды. Так, к одному классу можно отнести диагностические методики, выявляющие сформировавшиеся в прошлом опыте особенности психики. К другому классу относят диагностические методики, которые показывают не только психологические, но и физиологические особенности, как приобретенные, так и генетически детерминированные. В класс психологических методик попадают, таким образом, тесты, анкеты и опросники. В класс психофизиологических входят методики по изучению общих и частных свойств нервной системы и их жизненных проявлений.

Каковы характерные особенности тестирования как метода научной диагностики? Чтобы ответить на этот вопрос, необходимо хотя бы вкратце рассмотреть историю зарождения и развития психологического тестирования, прежде всего с точки зрения объективных причин, вызвавших это явление к жизни.

Ранние исследования в области тестологии выполнены в основном западноевропейскими и американскими авторами. Работы отечественных ученых составляют лишь небольшую часть публикаций, посвященных тестологии, но они затрагивают самые принципиальные вопросы теории и практики психологического тестирования. Можно с уверенностью охарактеризовать исследования западных психологов как носящие явно выраженную прагматическую направленность, в противоположность отечественным исследованиям, обращенным в первую очередь к методологическому анализу и логике теоретических позиций. Это ведет к их более высокой практической результативности.

В современной зарубежной тестологии отсутствует какая-либо достоверная теоретическая концепция, в большом числе работ представлены противоречивые результаты. Аппарат достаточно надежного количественного сравнения и анализа полученных данных также еще не разработан. Основные исследования в тестологии были выполнены уже к моменту так называемого тестологического бума 20—30-х гг. XX столетия. Нельзя отрицать определенную ценность работ последних десятилетий, однако следует признать, что наиболее существенные исследования современного психологического тестирования, его принципы и характерные особенности были разработаны и сформулированы именно к первым десятилетиям нашего века.

Цель данного раздела — проследить связь между особенностями формирования тестирования как диагностической процедуры и объективными условиями, которые ее вызвали, поскольку вообще трудно представить появление той или иной методики диагноза, не преследующей научные или практические цели. Цели и условия создания диагностического метода неизбежно отражаются на его характерных особенностях.

Тесты можно классифицировать в зависимости от того, какой признак будет взят за основание деления. Так, различают тесты индивидуальные и групповые (коллективные), вербальные и невербальные, количественные и качественные, градуальные и альтернативные, общие и специальные и т.д. Наиболее значительными представляются следующие **три основания классификаций**: цель, содержание и формы психологического тестирования.

По цели различают тесты для отбора, распределения и классификации.

При отборе по отношению к каждому индивиду решается вопрос типа “да — нет”, “оставить или исключить”. Не следует забывать, однако, что отбор относится к финальным этапам тестового исследования, которым часто может предшествовать этап своеобразного “просеивания через решетку”, позволяющий значительно сократить и упростить основную процедуру.

При распределении большое значение придается результатам индивидуальных оценок по отдельным показателям и тестам, позволяющим установить наиболее оптимальное соответствие между оценками тестирования и особенностями требований внешних критериев. Использование психологических тестов при отборе и распределении непосредственно связано с решением практических задач.

Классифицирующий тест базируется не на одном, а на нескольких критериях, поэтому по отношению к нему постоянно определяется дифференциальная валидность, показывающая, с какой степенью точности может быть предсказан каждый критерий.

Характерные особенности создания психологического тестирования отражаются не только в целях, но и в содержании и форме каждого отдельного теста. Краткий экскурс в историю тестологии даст наглядное представление об этих особенностях.

2.1.1. Зарубежные работы по психологическому тестированию

Испытания, подобные тестированию, можно найти еще в древнем мире. Одно из первых упоминаний о системе гражданских испытаний более чем трехтысячелетней давности относится к расцвету Китайской империи. В Древней Греции испытания использовали для оценки физического совершенства и искусства мыслить и рассуждать. Известны предложенные Платоном серии испытаний военных способностей, проводимых при отборе воинов в описанном им идеальном государстве; перечень этих испытаний можно было бы считать первым систематическим описанием тестов способностей.

Однако историю возникновения собственно психологического тестирования, под которым понимается стандартизованное измерение индивидуальных различий, можно проследить лишь с конца XIX в. В качестве основной причины, обусловившей начальный этап развития психологического тестирования, принято считать выдвинутую врачебной практикой потребность в диагностике и лечении умственно отсталых и душевнобольных людей. Многочисленные институты Европы и Америки, созданные специально для решения этих проблем, испытывали острую необходимость в доступных стандартах и специальных классификациях, позволяющих решать задачу практической дифференцировки умственной отсталости от душевной болезни. Первые исследования, которые можно охарактеризовать как ранний этап зарождения психологического тестирования, были проведены на клиническом материале. В дальнейшем задачи тестирования изменились.

Одна из ранних публикаций, посвященных вопросам умственной отсталости, принадлежит французскому врачу Ж.Е.Д. Эскиролю, показавшему, что между крайними полюсами нормальности и идиотии существует континуум многочисленных степеней умственной отсталости. Эти данные были положены в основу первой в истории психологического тестирования классификации, позволяющей диагностировать различные степени и проявления умственной отсталости. Уже в этой работе отмечено, что индивидуальный характер употребления языка в значительной мере предопределяет интеллектуальный уровень человека. Большинство созданных впоследствии тестов в основном языковые, с представленным в них вербальным содержанием.

Особый вклад в развитие тестирования внес другой французский врач — Э. Сегэн, первым уделивший внимание такой важной проблеме врачебной практики, как обучение умственно отсталых детей, и основавший с этой целью в 1837 г. первую специальную школу. При помощи особых методик тренировки сенсорного различения и развития моторного контроля удалось опровергнуть господствующее мнение того времени о неизлечимости умственной отсталости.

Переезд Сегэна в Америку способствовал самому широкому распространению его идей в этой стране.

Первые психологи-экспериментаторы XIX в., заложившие основы психологического тестирования и диагностики индивидуальных различий, не признавали факта существования индивидуальных различий как таковых. Принципиальное мнение того времени требовало генерализованного описания поведения человека, индивидуальные различия игнорировались или исключались, и общее, таким образом, попадало в фокус внимания чаще индивидуального. Ярким примером здесь могут служить работы, вышедшие из Лейпцигской лаборатории, основанной Вундтом в 1879 г. Именно они повлияли на развитие психологического тестирования. Преимущественное исследование простых сенсорных и сенсомоторных феноменов нашло отражение в содержании первых тестов, конструируемых именно для объективного измерения этих феноменов. Кроме того, впервые четко выявилась необходимость стандартизации самой процедуры тестирования. Потребность в строгом контроле за условиями психологического тестирования вытекала из таких экспериментально установленных фактов, как влияние на ответ испытуемого (как в сторону увеличения, так и в сторону уменьшения) характера и содержания инструкции, а также особенностей и параметров предъявляемых раздражителей.

Первое систематическое изучение данных об индивидуальных различиях людей в отношении их конституциональных признаков связано с именем Гальтона. Систематизация потребовалась Гальтону для исследования проблемы наследственности, решение которой было в принципе невозможно без учета и измерения устойчивых и изменчивых индивидуальных параметров человека. Как известно, кроме антропометрических параметров в сферу научного исследования Гальтона входили также отдельные психологические показатели. Он полагал, что тестированием простейших сенсорных процессов можно получить данные об интеллектуальном уровне человека. Первое широкое использование опросников, употребление теста на свободные ассоциации, исследование воображения также связаны с именем Гальтона.

Особое место в истории психологического тестирования занимают работы Гальтона по применению математической статистики в экспериментальной практике. Развивая анализ индивидуальных различий с помощью статистических методов, он отобрал и адаптировал большое количество математических приемов, позволяющих количественно оценить результаты экспериментальной работы.

Первым исследователем, употребившим в психологической литературе термин “интеллектуальный тест”, был американский психолог Дж.М. Кеттэл (1890). Подготавливая в экспериментальной лаборатории Вундта докторскую диссертацию по проблеме индивидуальных различий времени реакции, читая затем лекции в Кембридже (1888) и непосредственно сотрудничая с Ф. Гальтоном, Кеттэл проявляет большой интерес к проблеме измерения индивидуальных различий. После возвращения в Америку Кеттэл создает серию тестов, измеряющих интеллектуальный уровень. Однако, находясь под влиянием Гальтона, Кеттэл пытается определить высшие интеллектуальные процессы, исходя из данных изучения простейших сенсорных процессов и времени реакции, которые, как показала школа Вундта, могут быть измерены достаточно точно и объективно. Оценка и измерение более сложных психических процессов для Кеттэла представлялись невыполнимой задачей. С помощью его тестов обследуются ученики школ, студенты колледжей и др.

Особенность исследований, проводимых в это же время в Европе, состоит в том, что, в отличие от работ Кеттэла, они направлены на изучение более сложных психических функций.

Представителем этого направления является Э. Крепелин, разработавший

на клиническом материале серию испытаний, позволяющих судить о таких процессах, как память, утомление, нарушение внимания, по показателям выполнения достаточно простых арифметических действий.

В работах Г. Эббингауза используются тесты на выполнение арифметических действий, непосредственного запоминания, оценки которых дают высокие корреляции с оценками школьной успеваемости.

Тогда же выполнен ряд работ под руководством Феррари. Подобно Крепелину, он на примере патологических случаев использует серии тестов, базирующихся на измерении совокупности физиологических процессов, и дополняет их данными моторных тестов.

Общая оценка состояния тестологии на рубеже двух столетий содержится в ранней публикации французских врачей — психологов и педагогов А. Бине и В. Анри. Они весьма критически относились к большинству действующих тестов, показывая их специфическую сенсорную направленность и выделение в них исключительно простейших, единичных и узковыраженных функций.

2.1.2. Отечественные работы по психологическому тестированию

Следует подчеркнуть особую роль, которую играют в развитии методических и теоретических аспектов тестологии работы отечественных исследователей. Первые работы по психологическому тестированию выполнялись в течение первой трети XX столетия в области психологии, педагогики и педологии. Подавляющее большинство методик были копиями западных психологических тестов. Отличия проявлялись в формах проведения тестовых испытаний, в обработке и интерпретации экспериментального материала. Однако самая важная отличительная черта, выделяющая отечественные работы, — осознанная потребность решить научно-методологические проблемы тестологии.

Вероятно, одна из первых отечественных работ по психологическому тестированию, представляющая законченное самостоятельное исследование, была выполнена Г.И. Россолимо (1910) в Московском университете. Решая задачу количественной оценки психических процессов применительно к норме и патологии, он разработал графическую систему представления измерений психологического профиля. Основу этой системы, описываемой с помощью оценок тестирования степени выраженности 11 параметров, составляли три группы психических процессов: внимание и воля, точность и прочность восприятия, ассоциации. Отличительная особенность **метода психологического профиля** — его независимость от возраста испытуемого. Форма профиля оказалась надежным критерием для диагностики умственной отсталости. Замечательно, что, публикуя результаты работы, предварительно доложенные еще на II Всероссийском съезде педагогической психологии в 1909 г., Г.И. Россолимо подчеркивал, что его исследование было вызвано не чем иным, как необходимостью разрешения общих и теоретических вопросов экспериментальной психологии.

Интересно мнение П.П. Блонского, высказанное по поводу методики определения психологического профиля: отмечая ряд недостатков, общих для тестологии в целом, он высоко оценивает эту методику. Работу Г.И. Россолимо он признает наиболее удачной среди всех отечественных работ, так как в ней отобраны “очень показательные для умственного развития тесты” (П.П.Блонский, 1934, с. 136).

Сопоставляя метод психологического профиля со шкалой Бине —Симона, М.Я. Басов отмечает его полную теоретическую самостоятельность. К несомненной удаче Г.И. Россолимо он относит определенность теоретического взгляда на структуру личности и интеллекта, поскольку уже “одно наличие его в основании метода составляет положительный факт” (М.Я. Басов, 1931, с. 637). М.Я. Басов отмечает и ряд недостатков профиля Россолимо, среди которых первостепенное

значение придает отсутствию возможности измерять динамику развития личности (что присуще шкале Бине—Симона). Психологический профиль Г.И. Россоломо не свободен и от одного из основных пороков тестологии, ибо в сущности метод Россоломо производит подмену качества количеством.

Определенный интерес с точки зрения развития новых форм тестирования представляет измерительная шкала ума А.П. Болтунова (1928), положившего в основу своей работы шкалу Бине—Симона, переведенную и адаптированную П.П. Соколовым для испытания умственного развития русских школьников. По сути дела шкала Болтунова, опубликованная автором в дискуссионном порядке, представляет собой самостоятельную разработку нового набора тестов. Несмотря на известную аналогию со шкалой Бине—Симона, шкала Болтунова имеет специфические особенности: в ней заменено большинство задач, введены совершенно новые задачи, предложена новая инструкция и форма ее использования, определены сроки решения тестовых задач, разработаны показательные возрастные ступени. Принципиальная особенность шкалы А.П.Болтунова состоит в возможности проводить групповые испытания. Это выделяет ее из общего круга работ по измерению интеллекта. И все-таки данная работа типична для психологического тестирования. В ней сильно сказывается утилитарный механистический подход к созданию диагностической методики.

Методы измерения интеллекта, среди которых тесты занимают ведущее положение, характеризует Ф.Р. Дунаевский: “Работа пары десятилетий, протекших со времени Бине, заключалась главным образом в том, чтобы привести тесты в большее согласие с основными принципами современной вариационной статистики” (Ф.Р.Дунаевский, 1925, с. 14). Иначе говоря, вместо изучения содержательной стороны психологического тестирования тестологи лишь тщательно отработывали вопросы формализации и обработки результатов. Поэтому создание новых тестов (анализ и отбор отдельных задач) определялось не какими-либо теоретическими соображениями, а более или менее случайным произволом конструкторской выдумки авторов тестов.

Особое место в отечественных психологических исследованиях занимают работы М.Ю. Сыркина (1929), специально изучавшего проблему сопряженности показателей тестов одаренности и признаков социального положения (факт, установленный еще в первых работах Бине). Связь между особенностями речевого развития и результатами тестирования к тому времени была доказана экспериментально (уже самые первые работы тестологов фиксировали эту зависимость). Однако с течением времени социальный аспект существования интеллектуальных различий для тестологии становится все более острым и значимым.

Трудно переоценить работы Сыркина, рассматривая их с методологической точки зрения, поскольку в отечественных исследованиях по психологическому тестированию ему первому удается наглядно доказать, сколь противоречивой системой является тестовая диагностика индивидуальных различий, допускающая прямо противоположную интерпретацию результатов исследования. Самостоятельные экспериментальные работы Сыркина показывают, что между тестовыми оценками и социальными признаками имеется линейная форма связи (достаточно тесная для некоторых факторов), обладающая высокой временной стабильностью.

Так, при изучении критериев точности тестовых испытаний М.Ю. Сыркину удалось разработать систему оценок основных параметров психологического теста. Положив в основу работы постулат, учитывающий тот факт, что тестовый балл является мерой умения данного лица решать задачи данного типа в момент испытания, М.Ю. Сыркин использует результаты ряда западных тестологов, дополняя их результатами собственных экспериментов. Ему удалось показать особенности учета многочисленных факторов, наиболее заметно влияющих на

результативность тестирования (объективность теста, его репрезентативность, рассеяние группы, точность тестового измерения, факторы стабильности и состояния, надежность и валидность, пригодность теста и др.).

Заметной методической работой следует признать исследование А.М.Мандрыки, в котором основное внимание уделено именно теоретическим аспектам прикладной статистики. Он достаточно объективно определяет общий методологический уровень работ по психологическому тестированию: “Мы имеем способы характеристики некоторых внешних связей, но скользим, как правило, по поверхности явлений. Более глубокое изучение в практических целях иногда даже отвергается принципиально, так как объявляется, что корреляционная связь и ее установление практически эквивалентны вскрытию причинной зависимости”. Научный анализ важнейших сфер прикладной математики (выделение показателей, интерпретация коэффициента корреляции, выбор статистических критериев, различные способы формализации данных в виде уравнений, вопросы эквивалентности и трудности шкал, критерии точности) позволяет А.М.Мандрыке сделать вывод о неправильном противопоставлении математических методов исследованию предмета по существу, тем более замены второго первыми.

А. М. Мандрыка доказывает, что статистические методы должны быть подчиненными составными частями общего исследования. Он считает недопустимым прямую и одностороннюю биологическую или психологическую трактовку и интерпретацию устанавливаемого исследованием математико-статистических связей и недоучет прочих факторов, и в частности фактора социального. Определенная трактовка результатов тестирования имеет целью обслуживание социального заказа, в результате чего из поля зрения исследователей выпадают связь и отношения личности со средой. Известные свойства личности, явно обусловленные социальными дефектами среды, притом дефектами, надлежащими устранению, объявляются неизменными, обусловленными самой природой, а потому и оправдывающими эти социальные дефекты.

Один из основных выводов исследования А. М. Мандрыки состоит в требовании растворить математические методы в методологии данного предмета, определяемой качественным своеобразием материала. Этот вывод можно считать общим для той части отечественных работ, в которых затрагиваются методические и методологические проблемы тестологии, что позволяет в конечном счете признать их более высокий уровень по сравнению с исследованиями западной тестологии.

В последнее время обнаруживается еще одна интересная тенденция в области применения математической статистики к психологическим исследованиям. К.М. Гуревич затрагивает методологические проблемы диагностики (анализ категорий надежности методик, проблему прогностической ценности экспериментальных данных, разработку новых подходов к анализу полученного материала), во многом пересматривает и изменяет типовые методологические требования к надежности и валидности методик и свидетельствует о появлении нового направления, где математическая статистика выступает в качестве логического аппарата доказательств в психологической диагностике.

В заключение обзора методических особенностей психологического тестирования правомерно привести высказывание Б.М. Теплова по поводу практического употребления психологических тестов. Он считал, что в принципе нельзя возражать против создания коротких стандартизованных методик, количественно измеряющих индивидуальные особенности личности. Возражать надо против того, что подавляющее большинство тестов не имеет под собой прочно обоснованной научной базы.

Поскольку тест не что иное, как “чисто эмпирически найденная проба”,

значимость которой должна быть подтверждена результатами массовых измерений, Б.М. Теплов считал необходимым отметить отношение к тестам: “Если нам не ясен физиологический или психологический смысл испытания, если это испытание является слепой пробой, то никакая статистическая обработка массового применения этого испытания, сколь бы сложной и остроумной она ни была, не даст научно ясных результатов”.

Перечисленные работы дают достаточно наглядное представление об одной из главных особенностей отечественных исследований — решении основных теоретических и методологических проблем психологического тестирования, без чего практическое употребление тестов признается ничем не оправданным действием.

Как бы ни назывался тест, какой бы уровень объективности и стандартизации ему ни приписывался, он в состоянии фиксировать и измерять только внешние, имеющиеся в момент испытания элементы индивидуальных моделей психической деятельности и поведения, и ничего больше. Что стоит за этим поведением, чем оно обусловлено и вызвано, что можно предположить по поводу будущего поведения индивида — все эти вопросы остаются нерешенными. Поэтому **диагностическая ценность** тестовых измерений должна тщательно анализироваться исследователем в каждой конкретной ситуации, хотя повышение гомогенности выборки, постоянная спецификация теста, создание очень узких локальных норм могут обеспечить высокий методический уровень теста и его надежность.

Таким образом, единственной положительной характеристикой психологического теста как метода диагностики индивидуальных различий, пожалуй, следует признать возможность осуществлять достаточно тонкую дифференцировку (методом ранжирования) обследуемой выборки по какому-либо признаку или их комбинации. Однако подобная процедура дает своего рода сиюминутный поперечный срез и в этом смысле позволяет получать информацию только о некоторой точке на кривой развития личности. И если о предыстории развития обследуемого человека еще можно составить в какой-то мере определенное суждение, то о его развитии в будущем можно сказать очень мало. Безусловно, диагностика и прогнозирование связаны между собой и отдельные моменты в истории развития психологического тестирования позволяют усматривать возможность решения проблем прогнозирования. Однако тесты этих проблем не решили. Следует признать, что в качестве орудия прогноза тесты себя не оправдали. Первоначальное мнение большинства тестологов о том, что тестовые испытания позволяют измерять некоторые стабильные и даже врожденные элементы поведения человека (естественно, повышающие прогностическую ценность тестов), не подтвердилось.

2.2. Шкала Бине и ее модификации

Успехи в разработке психологического тестирования связаны с именем А.Бине, посвятившего много лет продуктивным и изобретательным исследованиям интеллекта. Можно без преувеличения сказать, что почти все дальнейшие конструкции интеллектуальных тестов в той или иной мере связаны с работами Бине. Важно заметить, что именно после Бине в тестологии усиленно разрабатывается проблема измерения комплексных, или высших, психических процессов и функций.

Период до начала XX столетия с известными ограничениями можно охарактеризовать как период разработки допсихологических тестов. Он является подготовительным и одновременно переходным этапом на пути создания собственно психологического тестирования. Вводится новая классификация, за основание которой берется содержание тестирования. В этом плане психоло-

гические тесты делятся на четыре класса, или направления: тесты интеллекта, тесты способностей, тесты достижений и тесты личности.

Экспериментальные работы, проведенные под руководством Бине в школе, так же как и все предыдущие, возникли прежде всего под давлением требований практики. В данном случае было необходимо диагностировать степень умственного развития детей, поскольку практика проведения в жизнь закона об обязательном всеобщем обучении выявила значительные индивидуальные различия, осложняющие процесс обучения, а в некоторых случаях делающие его невозможным. Но задача диагностики степеней умственной отсталости с помощью специфических тестов того времени еще не была решена. Как признавал сам Бине, шкала 1905 г. давала возможность разделить испытуемых только на две группы по величине интеллектуального уровня. Одна из них достигала эмпирически выделенного уровня интеллектуальной нормы и превышала его, другая находилась ниже его. Результаты психодиагностического обследования школьников при помощи шкалы Бине давали основание отсеивать из школ детей, чей интеллектуальный уровень находился ниже нормы.

Психодиагностическое обследование - деятельность психолога, конечным результатом которой является психологический диагноз. **Этапы психодиагностического обследования** - сбор данных в соответствии с задачей исследования; обработка и интерпретация полученных данных; вынесение заключения (диагноз и прогноз). **Психологический диагноз** - конечный результат деятельности психолога, направленной на описание и выяснение сущности индивидуально-психологических особенностей личности с целью оценки их актуального состояния, прогноза дальнейшего развития.

А. Бине первым в тестологии стал ранжировать тестовые задачи по эмпирически определяемой степени трудности их решения. Шкала 1905 г. состояла из 30 задач, прошедших предварительную нормировку на смешанной контрольной группе. Возрастной диапазон группы колебался от 3 до 11 лет. Основные задачи теста были предназначены для испытаний рассудительности и понятливости (рассуждение и понимание), которые Бине считал основными компонентами интеллекта. Вторая редакция шкалы (1908) замечательна тем, что в ней помимо расширения возрастного диапазона до 13 лет, увеличения числа задач и устранения неудовлетворительных тестовых оценок вводится понятие **умственного возраста**. Вторая редакция шкалы Бине получила самое широкое распространение. При переводах и адаптации ее конструктивные и целевые особенности остаются постоянными: это методика, направленная прежде всего на диагностику умственно отсталых детей. Она и послужила основой проведенной в Стэнфордском университете (США) работы американского исследователя Л. М. Термена. Последняя (третья) редакция шкалы, опубликованная в год смерти Бине (1911), не вносит существенных изменений, за исключением дополнительных задач, позволяющих несколько увеличить возрастную норму.

К. М. Гуревич (1970) полагает, что именно с момента создания шкалы Бине произошло резкое смещение акцентов тестологических исследований. Первостепенное значение приобретает задача классификации нормальных детей и взрослых по степени умственного развития. С этой целью Термен использует коэффициент умственного развития (IQ), оценивающий отношение между умственным и хронологическим возрастными.

Следует заметить, что коэффициент IQ был предложен В. Штерном, вскрывшим существенный недостаток процедуры вычисления показателя умственного возраста по шкале Бине — Симона. Недостаток заключается в том, что одна и та же разность для различных возрастных ступеней имеет неодинаковое значение. Чтобы устранить этот недостаток, Штерн стал определять не абсолютную меру интеллекта — разность, а относительную, т.е. частное, получаемое при делении умственного возраста на хронологический. Этот показатель он и назвал

интеллектуальным коэффициентом. Таким образом, первая задача психологического тестирования — диагностика степени умственной отсталости — постепенно отходит на второй план.

Практика настоятельно требовала конкретного критерия, имеющего универсальный психологический характер и позволяющего решать любые задачи. Роль подобного критерия и по сей день, несмотря на внутреннюю противоречивость новой шкалы (колоссального количества степеней индивидуального разнообразия показателей нормальных людей, не находящих отражения в оценке интеллектуального уровня), выполняет коэффициент умственного развития.

Следующий этап развития психологического тестирования характеризуется изменением формы проведения тестового испытания. Все тесты, созданные к первому десятилетию XX столетия, индивидуальные и позволяют вести опыт только с одним человеком. Использовать их могли лишь специально подготовленные люди, имеющие достаточно высокую психологическую квалификацию.

Перечисленные особенности тестов ограничивали их распространение. Поэтому под давлением потребности практически распределять и отбирать большие массы людей (в промышленности и особенно в армии) в США в период первой мировой войны появилась новая форма тестовых испытаний — **групповое тестирование**. Необходимость как можно быстрее отобрать и распределить полуторамиллионную армию рекрутов по различного рода службам, школам и училищам заставила специально созданный комитет поручить А.С. Отису разработку новых тестов. Так появились две формы армейских тестов — “Альфа” и “Бета”. Первая из них предназначена для работы с людьми, знающими английский язык, вторая — для неграмотных и иностранцев. Вскоре после окончания войны эти тесты и их модификации получают самое широкое распространение.

Групповые (коллективные) тесты разработаны как инструмент массового тестирования, они не только делают реальными испытания больших групп, но наряду с этим значительно упрощают инструктирование, проведение процедуры и оценку результатов. Таким образом, последняя классификация — по форме проведения тестовой процедуры — включает два вида: индивидуальное и групповое тестирование.

Тестовый бум 20-х гг., основанный на неограниченном применении тестов, не привел к существенному прогрессу в развитии психологического тестирования. Наиболее показательна в этом смысле история интеллектуальных тестов. Большинство их измеряют прежде всего вербальные способности. Основная нагрузка тестов приходится на вербальный фактор, к которому добавляется способность к счету и некоторым абстрактным операциям с символами. Очень многие психологи признают, что термин **тест интеллекта** употребляется неправильно, поскольку в подобных случаях измеряются только его отдельные аспекты, в то время как общая оценка интеллекта представляется процедурой далекого будущего, тем более неясного, что и понятие интеллекта не определено.

Несомненно, тестами охватывается та часть способностей человека, которая имеет первостепенную значимость для включения в культурную жизнь общества. Фактически большинство интеллектуальных тестов связано с измерением тех или иных комбинаций способностей, требуемых для обучения в средней и высшей школе. Это отражается, например, в том, что некоторые тесты интеллекта были названы впоследствии тестами школьных способностей.

2.3. Типология тестов по субъекту измерения

2.3.1. Тесты специальных способностей

Ограниченные возможности тестов интеллекта потребовали дополнительной информации об особенностях психики человека. Так возникло новое направление

в тестологии — **тестирование специальных способностей**, призванное вначале дополнить оценки тестов интеллекта, а впоследствии выделившееся в самостоятельную область. Тесты специальных способностей были разработаны для отбора и классификации исследуемых в промышленности и военном деле. Таковы тесты механических, канцелярских, музыкальных и артистических способностей. Возникает еще одно специфическое направление в тестологии — статистическое **исследование черт личности**, основанное на обнаружении индивидуальных различий и определенной связи между отдельными показателями тестовых измерений.

Работы этого направления, начатые английским психологом Ч. Спирменом и продолженные затем американскими психологами Т.Л. Килли и Л.Л. Тэрстоном, получили название **факторного анализа**. Результаты тестовых испытаний рассматриваются при факторном анализе как индикаторы, свидетельствующие о наличии определенных факторов. Некоторые из них, наиболее репрезентативные, выделяются в определенных пропорциях после выполнения строго детерминированных математических процедур.

На основе факторных исследований были созданы многофакторные батареи тестов способностей, позволяющие измерять индивидуальный уровень каждой из способностей. **Тестовая батарея** - совокупность групп тестовых заданий (субтестов), объединенных в одну психодиагностическую методику и направленных на измерение различных сторон сложного психологического конструкта.

Батареи тестов множественных способностей широко распространяются уже после второй мировой войны и используются прежде всего в армии. В качестве примера можно упомянуть американские ВВС, в которых данный тип психологического тестирования проходят пилоты, штурманы, бомбардиры, радиооператоры и другие ведущие военные специалисты.

Психологическое тестирование интеллекта и способностей, применяемое в промышленности и военном деле для классификации и отбора, затронуло и обучение. Проведение экзаменов обогатилось новыми техническими приемами и формами, получившими общее название тестов достижения. **Тесты достижений** - тип психодиагностических методик, направленных на оценку достигнутого уровня развития способностей, навыков и знаний.

История развития стандартизованных тестов достижений может быть прослежена с момента смены в Бостонской школе (1845) устной формы экзаменов на письменную. Основными достоинствами новой формы экзаменационной процедуры считается наличие однотипной ситуации для всех экзаменуемых, позволяющей проверить более широкий объем знаний, исключить элементы случайности при опросе, а также устранить субъективизм в оценках.

Среди тестов достижений выделяются две группы процедур. Одна из них подготовлена Э.Л. Торндайком и представляет собой использование и стандартизацию принципов, разработанных в психологической лаборатории для тестирования особенностей правописания, арифметических вычислений и рассуждений. Другая представлена различными сериями тестов достижений, включая и новейшие формы тестирования с помощью машинной техники. Стандартизованные тесты достижений широко применяются при отборе сотрудников в промышленности и служащих в различного рода государственных учреждениях. В Америке стандартизованные задачи используются в некоторых государственных департаментах уже с 1872 г., а с 1883 г. их употребление становится регулярным и постоянным. Наиболее значительная разработка элементов техники конструирования тестов достижений выполнена в течение первой мировой войны и сразу после нее. Техническая сторона тестов достижений

находится в основном на уровне разработанности тестов интеллекта и способностей.

2.3.2. Интеллектуальные тесты

Тесты интеллекта (от лат. intellectus — понимание, познание) - тип психодиагностических методик, направленных на оценку общих способностей. Тесты интеллекта представляют собой совокупность методик, образованных в рамках **объективного диагностического подхода**. Предназначены для измерения уровня интеллектуального развития. Тесты интеллекта являются одними из наиболее распространенных в психодиагностике.

Проявления интеллекта многообразны, но им присуще то общее, что позволяет отличать их от других особенностей поведения. Этим общим является активизация в любом интеллектуальном акте мышления, памяти, воображения всех тех психических функций, которые обеспечивают познание окружающего мира. Соответственно под **интеллектом** как объектом измерения подразумеваются не любые проявления индивидуальности, а прежде всего те, которые имеют отношение к познавательным свойствам и особенностям. Это нашло свое отражение в многочисленных тестах для оценки различных интеллектуальных функций (тесты логического мышления, смысловой и ассоциативной памяти, арифметические, пространственной визуализации и т.д.). Этот эмпирически образовавшийся ряд тестов (“очищенных” от измеряющих мышечную силу, остроту слуха и зрения, время реакции и других переменных) достаточно четко ограничен от других методик измерения индивидуально-психологических особенностей — **тестов личности**, направленных на изучение интересов, эмоций, а также особенностей поведения в определенных социальных ситуациях.

Первые тесты интеллекта были созданы Ф. Гальтоном, стремившимся охватить “измерением и числом операции ума” (Ф. Гальтон, 1879). Ему же принадлежит заслуга разработки математико-статистических методов анализа данных по индивидуальным различиям. Заметный вклад в развитие идей Ф.Гальтона был внесен работами Дж. Кэттелла, положившими начало широкому использованию “умственных тестов”. Исследованиями Дж. Кэттелла, а за ним — многих европейских психологов и клиницистов были заложены основы современных тестов интеллекта, прообразом которых является широко известная шкала умственного развития Бине — Симона, разработанная в 1905 г. Мощным стимулом развития тестов интеллекта явились предпринятые уже в начале XX в. статистические работы по природе интеллекта, целью которых являлось установление взаимосвязей между показателями, полученными у различных обследуемых с помощью разнообразных тестов (Ч. Спирмен, 1904, 1927; Л. Тэрстоун, 1931, 1947). Родившийся в этих исследованиях **факторный анализ** способствовал появлению комплексных тестовых батарей, широко используемых в современной психодиагностике.

Полученные с помощью тестов интеллекта результаты выражаются количественно в виде **коэффициента интеллекта (IQ)**, на психологической интерпретации которого следует остановиться подробнее. Это необходимо потому, что за интерпретацией количественного показателя интеллекта стоят различные подходы к пониманию природы измеряемого явления.

Одной из доминирующих в западной психологии является концепция генетической обусловленности интеллекта, истоки которой находим в идеалистических философских системах. Значительная роль в распространении идеи о наследуемости интеллекта принадлежит Ф. Гальтону. В своей современной формулировке теория наследственной детерминации интеллекта утверждает, что примерно 80% вариаций IQ следует отнести за счет генетических различий между людьми (Г. Айзенк, 1979 и др.). Решающее значение при этом отводится

психометрическому изучению интеллекта монозиготных близнецов, воспитывавшихся порознь. Полученные данные, говорящие, на первый взгляд, в пользу наследуемости интеллекта, были подвергнуты впоследствии резкой критике, доказавшей их необоснованность, а в некоторых случаях — и фальсификацию (Л.Ф. Бурлачук, 1985). Под влиянием теорий генетического детерминизма предпринимались и не увенчавшиеся успехом попытки создания тестов, предназначенных для измерения интеллекта в “чистом” виде. Концепции генетической предопределенности интеллекта оказывают значительное воздействие на систему образования, являются “теоретической” основой расовой и социальной дискриминации.

Исследования, обращенные к изучению влияния социальных факторов, показывают их определяющее значение в формировании интеллекта. IQ не предопределен от рождения и существенно изменяется в зависимости от социальных и культурных условий. Во многих зарубежных работах установлены значимые корреляции между количественными показателями интеллекта и социоэкономическими факторами (доход семьи, профессия родителей и их образование и т.д.). Наиболее существенна зависимость уровня интеллекта, определяемого с помощью тестов, от образования, возможностей приобщения к культуре. Исследования отечественных психологов показывают возможность выделения в жизни индивида некоторых периодов, характеризующихся более высоким или относительно низким уровнем интеллекта, причем длительность этих периодов определяется не только уровнем образования, но и во многом связана с конкретными социальными требованиями, предъявляемыми к человеку в различные годы жизни.

Тесты интеллекта создавались и создаются не в вакууме, а во вполне определенных социальных условиях. В них всегда учитываются социальные нормы и ценности, стандарты поведения и мышления, присущие определенной культуре. Уже в тестах Бине — Симона нетрудно увидеть, что при их разработке авторы исходили из социальных требований, предъявляемых к учащимся школой того времени. О. Дункан (1972) остроумно замечал, что если бы первые тесты для определения коэффициента интеллекта были разработаны в обществе, в котором наиболее важное дело — охота, то под “общим интеллектом” подразумевали бы остроту зрения, скорость бега, а вовсе не словарный запас, умение им пользоваться, способность оперировать символами и т.п.

На результаты тестов интеллекта влияет не только принадлежность к определенной культуре, социальной группе, но и “субкультурные” различия. Даже в формально однородной выборке (обследуемые относятся к одному социальному слою, не различаются по полу и образованию) обнаруживаются существенные различия между отдельными лицами, связанные с конкретными условиями их микросреды.

Изначальная соотнесенность тестов интеллекта с культурой и опытом определенной социальной группы ограничивает сферу их применения. Тесты интеллекта оказываются неадекватными для обследования лиц, принадлежавших к иной культуре, нежели та, в которой они создавались. Сопоставление результатов тестирования интеллекта в различных этнических группах лишено смысла. Сравнивать количественные показатели интеллекта можно лишь тогда, когда его представляют в качестве универсального измерения. Игнорирование условий культуры, активизирующей, например, развитие не вербально-логического, а пространственно-образного типа переработки информации, может привести к постановке ошибочного диагноза умственной отсталости. Зависимость результатов тестов интеллекта от культуры рождает и известное разочарование в них. Некоторые специалисты в области психодиагностики за рубежом предлагают отказать от понятия интеллекта применительно к имеющимся тестам, предпочитая говорить об измерении “адаптивности” или индивидуально приобретенной

адаптации к культуре.

Решающее влияние социальных факторов на достигнутый уровень интеллектуального развития не означает, что в поиске причин индивидуальных различий в интеллекте необходимо сделать выбор между наследственностью и средой. Нельзя умалять роль природных данных человека, дифференциальных особенностей, не зависящих от опыта, степени приобщенности к культуре. Обнаруживаются ли эти природные особенности в тестах интеллекта?

Темп психической деятельности человека влияет на внутригрупповую дисперсию коэффициента интеллекта (разброс IQ). Однако природные особенности, как они фиксируются тестами интеллекта, даны в сплаве, единстве со множеством других факторов (пол, возраст, психическое состояние, образование и т.д.), выступают в “социализированном” обличье, и их невозможно выделить в чистом виде, так же, как невозможно измерить интеллект, свободный от прошлого опыта человека.

Наивными выглядят попытки некоторых современных исследователей (Г.Айзенк, 1979) “взвесить” долю природного и приобретенного в интеллекте, подсчитать процент того и другого. Любые рассуждения об относительной роли двух неперменных факторов – среды и наследственности – бессодержательны, ибо развитие человека определяется не их соотношением, а взаимодействием. Методологическая несостоятельность схем линейного детерминизма в изучении человеческого индивидуума раскрыта отечественными психологами (Б.Ф. Ломов, 1984 и др.). В отечественной психологии **проблема взаимодействия природного и приобретенного** в интеллекте получает свое решение в учении о задатках как предпосылках развития. Генетически детерминированные и врожденные анатомо-физиологические особенности мозга и нервной системы (задатки) являются лишь условиями формирования интеллекта, непосредственно его не определяя.

Интеллект, на измерение которого направлены тесты интеллекта, — относительно самостоятельная, динамическая структура познавательных свойств личности, возникающая на основе наследственно закрепленных (и врожденных) анатомо-физиологических особенностей мозга и нервной системы (задатков), формирующаяся во взаимосвязи с ними. Он проявляется в деятельности, обусловленной культурно-историческими условиями, и преимущественно обеспечивает адекватное взаимодействие с окружающей действительностью, ее направленное преобразование. Существующие тесты интеллекта позволяют как бы сделать срез с достигнутого уровня усвоения основных требований, предъявляемых условиями жизнедеятельности к познавательным особенностям личности. Конечно, достигнутый уровень не дает оснований для отдаленных прогнозов. Однако этот уровень — основа дальнейшего развития, и от него зависимы хотя бы ближайшие успехи в той или иной деятельности.

В отечественных исследованиях накоплен определенный опыт использования наиболее известных тестов интеллекта, свидетельствующий об их практической ценности в дифференциальной диагностике умственной отсталости, задержек психического развития детей, судебно-психологической экспертизе подростков. Обследование психически больных различной нозологической принадлежности позволило установить структурные особенности их интеллекта, учет которых позволяет “прицельно” планировать реабилитационные мероприятия. Диагностическое значение отдельных зарубежных тестов интеллекта (шкала измерения интеллекта Векслера, тест структуры интеллекта Амтхауэра, культурно-свободный тест интеллекта) подтверждено результатами исследований, направленных на углубленное изучение отдельных интеллектуальных функций в различные периоды онтогенеза (Б.Г. Ананьев, 1973 и др.). Осуществляется разработка отечественных тестов интеллекта, предназначенных для диагностики уровня интеллектуального развития учащихся средней школы (школьный тест умственного развития).

2.3.3. Личностные опросники

Особое направление в психологическом тестировании связано с разработкой различного рода тестов личности. Основу составляют тесты, созданные для измерения аффективных или неинтеллектуальных аспектов поведения. Единого толкования термина **тест личности** не существует; некоторые психологи употребляют его в широком смысле, включая в это понятие как неинтеллектуальные, так и собственно интеллектуальные черты. Чаще всего в психологическом тестировании этот термин охватывает сферу измерения характеристик, подобных эмоциональной устойчивости (регулированию), особенностям межличностных отношений, мотивации, интересов и т.д.

К ранним работам с использованием тестов личности следует отнести клинические исследования Э. Крепелина, который применял тест свободных ассоциаций, основанный на ответе субъекта первым пришедшим на ум словом на специально отобранные сигнальные слова-раздражители. Крепелин изучал такие феномены, как утомление, злость, наркомания.

Личностные тесты представлены в основном тремя группами: опросниками (самоотчет), ситуативными тестами и проективным тестированием.

Ситуативные тесты иначе называются тестами на выполнение работы. В большинстве тестов испытуемому предлагают выполнить некоторую деятельность, как правило, моделирующую элементы обычной жизни. Основное развитие ситуативных тестов приходится на второе и третье десятилетие XX в., когда была проведена их стандартизация (в основном на школьниках). С помощью ситуативных тестов диагностируются, в частности, такие аспекты личности, как склонность к мошенничеству, лжи, воровству, а также к упорству. Ситуативные тесты широко использовались в период второй мировой войны, например, в американской службе стратегического обеспечения. Как правило, ситуативные тесты носят комплексный характер, чаще всего они включают в себя изучение, наряду с социальными, и эмоциональных черт личности. Интерпретация ситуативных тестов весьма субъективна, чем отрицается возможность их применения вне чисто исследовательских целей в массовых обследованиях.

Особое направление в тестировании личности создает проективная техника, появившаяся и используемая в основном в клинической практике. Собственно говоря, **принцип свободных ассоциаций** следует признать одним из ранних типов тестов с применением проективной техники, которая предполагает, что в ответах на не полностью сконструированные задачи содержатся проецируемые субъектом характеристики типичных для него психических черт. Сильная зависимость от субъективности при интерпретации в сочетании с отсутствием нормативных данных осложняет применение этой техники.

2.4. Типология тестов по форме предъявления

2.4.1. Бланковые тесты

Бланковые тесты — тип психодиагностических методик, предусматривающих самостоятельную работу испытуемого. Испытуемому предлагаются специальный бланк обследования, брошюра, опросный лист и т.д., в которых содержатся инструкция и примеры решения, рабочие задания и графы для ответов испытуемого. Нередко испытуемый получает отдельно брошюру с тестовыми заданиями и инструкцией, а также бланк для регистрации ответов.

Бланковые тесты являются одними из наиболее распространенных и давно применяемых методик психологической диагностики. К их числу относятся **корректирующая проба, методика незаконченных предложений, исследования самооценки** и многие другие. Другим широко известным названием бланковых

тестов является “тесты карандаша и бумаги”. Здесь подчеркивается основная направленность такой процедуры — простота техники обследования, отсутствие необходимости в специальных аппаратных средствах.

Применение бланков распространено практически во всех видах тестовых методик, однако наиболее широко эта процедура используется в тестах достижений, тестах интересов, тестах общих и специальных способностей. Практически все личностные опросники, диагностические анкеты, большинство групповых тестов являются бланковыми методиками.

Оформление тестового материала (содержание, форма, расположение отдельных пунктов теста и граф для ответов, инструкция и т.д.) во всех бланковых тестах должно удовлетворять специфическим требованиям тест-объектов.

Бланк обследования — специальная форма фиксации результатов психодиагностического исследования. Предназначен как для учета результатов по отдельным методикам, так и для фиксации данных тематических исследований с помощью комплекса методов (карты обследования). Бланки обследования разделяются на рабочие листы для испытуемых, в которых обследуемый сам отмечает результаты решений задач теста или проставляет ответы на опросник, и формы для обработки и интерпретации данных обследования экспериментатором. Нередко оба вида бланков обследования объединены в одном блоке.

Бланки обследования обязательны для проведения большинства тестов интеллекта, проективных методик, личностных опросников а также всех групповых тестов. Конструктивное оформление бланков обследования обеспечивает удобство и быстроту работы испытуемого, сокращает непроизводительные затраты времени психодиагностического обследования, способствует правильному пониманию испытуемым инструкции, снижает количество случайных ошибок при работе обследуемого и оценивании ее результатов. Культура составления бланков обследования особенно важна при проведении массовых обследований. Имеется ряд правил, следование которым обязательно при разработке бланков обследования:

1. Бланки обследования следует составлять таким образом, чтобы свести к минимуму непродуктивную работу испытуемого по отыскиванию места (пропуски в тексте, графы, шкалы и т.д.) для проставления ответа.

2. Предпочтительно, по возможности, свести ответ к подчеркиванию, обведению кружком, обозначению простыми символами (например, “+”, “—” и т.д.). Это требование обязательно для задач закрытого типа.

3. Бланк не должен быть перегружен цифрами, делениями, строками и другими обозначениями.

4. Ответы по отдельным сериям заданий (субтестам) должны выделяться на бланках обследования отдельно. Там, где задания не разбиваются на группы и их количество велико, графы для ответов должны разделяться на отдельные блоки, например, по 10—25 заданий.

5. Из бланка для испытуемого необходимо исключить название теста.

6. На бланках обследования должны быть предусмотрены места для проставления необходимых паспортных данных (если исследование не является анонимным), а также номера регистрации обследования и даты проведения.

7. Иногда целесообразно кратко продублировать инструкцию по обследованию на бланке для испытуемого. Это требование обязательно при проведении группового обследования.

8. При переходе к ответам на новую группу заданий необходимо поместить примеры заполнения граф.

9. При использовании шаблонов для получения первичных оценок необходимо соответствующее оформление граф для ответов (чаще всего матрица).

В методических материалах, как правило, предлагаются образцы бланков обследования или дается описание условий их составления. Для ряда психологических методик незначительное изменение внешнего вида бланка для испытуемого может существенно изменить результаты его работы. Так, показатели выполнения корректурной пробы будут значительно отличаться при применении бланков с различной величиной шрифта, формой букв, интервалом между строками, жирностью шрифта, оттенком бумаги и т.п.

2.4.2. Групповые тесты

Групповые тесты — тип методик, предназначенных для одновременного обследования группы испытуемых. Групповые тесты не следует смешивать со средствами диагностики социально-психологических особенностей групп, коллективов и т.д.

Основной целью разработки этого вида психодиагностических методик в свое время явилось стремление повысить эффективность практических служб психодиагностики за счет увеличения контингентов обследуемых. Одними из первых тестов для массовых обследований стали разработанные в США армейские тесты “Альфа” и “Бета”, направленные на оценку общих способностей при отборе новобранцев для прохождения службы. В настоящее время методики, допускающие одновременное обследование группы испытуемых, имеют хождение в тестах интеллекта, тестах достижений, особенно при исследованиях в школьной психодиагностике, методиках профессиональной психодиагностики. В качестве примеров наиболее распространенных методик такого рода можно привести САТ—Батарея Службы учебно-образовательных тестов США; GATB — Общих способностей батарея тестов; СЕЕВ — основной армейский профориентационный тест в США и др. Для группового обследования приспособлено большинство тестов интересов. Групповая форма обследования возможна и для практически всех личностных опросников. Групповые тесты с успехом применяются в целях отсеивания (скрининга), где с наибольшим успехом реализуются выгоды групповой формы обследования.

По характеру групповые тесты являются, как правило, типичными психометрическими тестами с весьма жесткой регламентацией процедуры обследования, обработки и интерпретации данных. Для группового применения могут быть приспособлены и отдельные аппаратные тесты, особенно при условии введения в диагностическую систему компьютерных средств обработки информации.

Большинство групповых тестов являются по форме бланковыми. Материал, необходимый для проведения теста (тестовая тетрадь, бланки для регистрации результатов и т.д.), отличается, однако, некоторыми специфическими особенностями. Он должен быть вполне приспособлен для самостоятельной работы испытуемого при минимуме пояснений и вмешательства со стороны экспериментатора. Инструкция к проведению обследования должна быть краткой, не требующей дополнительных разъяснений.

Количество одновременно тестируемых лиц ограничивается, как правило, возможностями контроля и наблюдения со стороны обследующего. Одним из основных требований успешного обследования с помощью групповых тестов является строгое соблюдение условий самостоятельности работы каждого испытуемого, исключение возможности влияния со стороны других тестируемых лиц (заимствование результатов, отвлекающие контакты, групповые эффекты и др.). Обычно максимально допустимое количество лиц в обследуемой группе — 20-25 человек.

Как правило, при групповой процедуре обследования показатели надежности и валидности теста изменяются (обычно оказываются ниже, чем при проведении

того же теста в индивидуальной форме). Иногда в руководствах к тестам даются различные нормы оценок для групповой и индивидуальной формы применения.

К числу недостатков группового обследования следует отнести ограничение возможности наблюдения, контакта в ходе обследования, качественного анализа психодиагностической информации, что ставит групповые тесты в ряд вспомогательных средств психологической диагностики.

Групповые методики наряду с возможностью значительного увеличения производительности диагностической службы имеют и другие преимущества. Среди них — простота инструментария и проведения, отсутствие специальных требований к профессиональной подготовке лиц, проводящих обследование и т.д.

В отдельных случаях традиционные индивидуальные тесты выгодно проводить в небольших группах из 3—5 испытуемых. Иногда это целесообразно для повышения мотивированности при обследовании детей за счет реализации эффекта соревнования. Такая форма обследования для детей является более привычной, так как напоминает естественные условия обучения и осуществления контроля знаний в классе.

2.4.3. Аппаратурные тесты

Аппаратурные тесты — тип методик, требующих применения специального оборудования для проведения исследования или регистрации полученных данных.

Аппаратурные методики распространены среди критериально-ориентированных тестов, где с помощью специальных установок моделируются условия критериальной деятельности. **Критериально-ориентированные тесты** — тип тестов, предназначенных для определения уровня индивидуальных достижений относительно некоторого критерия на основе логико-функционального анализа содержания заданий.

Аппаратурные тесты являются стандартизированной методикой оценки психофизиологических свойств. К числу аппаратурных тестов можно отнести и диагностическое применение широко известных приборов для исследования показателей времени реакции (реактометры, рефлексометры), типологических особенностей высшей нервной деятельности (например, прибор Хильченко и т.д.).

Большие перспективы распространения аппаратурных тестов открываются в связи с развитием **компьютерной психодиагностики**. Автоматизированные пульты для испытуемых объединяются в этом случае с информационными системами, позволяющими оперативно и всесторонне анализировать различную психодиагностическую информацию при обследовании широких контингентов испытуемых.

К достоинствам аппаратурных тестов следует отнести более высокую точность и объективность результатов обследования, возможность автоматизации сбора первичных данных. Испытуемый, работая с установкой, нередко склонен сообщать более достоверную информацию, касающуюся отдельных сторон его личности. Эмоционально он чувствует себя менее скованно. Применение аппаратурных тестов положительно сказывается на очевидной валидности методики. Благодаря более широкой возможности устранения фактора воздействия личности экспериментатора на результат в аппаратурных тестах достигается наиболее высокая степень стандартизации процедуры обследования, что положительно сказывается на надежности теста.

К недостаткам аппаратурного способа проведения тестовых исследований следует отнести высокую стоимость необходимого оборудования и сложность технического обеспечения психодиагностической лаборатории. Применение

аппаратурных тестов ограничивает возможности качественного анализа результатов работы испытуемого. Учитывая последнюю особенность и в то же время признавая необходимость широкого внедрения аппаратурных тестов, следует помнить, что стремление ввести в процедуру психодиагностического исследования максимум технических средств не должно являться самоцелью, в угоду которой приносится непосредственный контакт с испытуемым.

Аппаратурные тесты широко используются в области психодиагностики школьной и психодиагностики профессиональной, особенно при групповых обследованиях уровня подготовки, достижений, профессиональных навыков, знаний.

2.5. Формализация психодиагностических испытаний

Важнейшей методической задачей, определяемой целью и содержанием теста, является необходимость сравнивать между собой оценки, получаемые различными испытуемыми, при соблюдении постоянства самой процедуры тестирования.

Как правило, при конструировании теста подготавливают детальные указания и руководства для его проведения. Составление таких руководств — неотъемлемая часть работы, направленной на получение достаточно объективных результатов. Сюда относятся такие требования, как точность представления материалов, лимит времени, характеристики подачи инструкции, ее форма, даже интонация и жесты экспериментатора и т. д.

Максимально значим последний этап психологического тестирования — интерпретация данных. Выделим в этой связи понятие **стандартизации**, которая рассматривается в двух планах. Первый заключается в постоянстве основных экспериментальных процедур. Второй полностью относится к понятиям нормы. Без понятия нормы, позволяющего вести от него отсчет, тестовые оценки вообще не могут быть интерпретированы. Тесты в принципе не имеют определенных стандартов, поэтому индивидуальная оценка тестового испытания может быть получена только путем сравнения с оценками других испытуемых. Основная статистическая концепция, с помощью которой в большинстве диагностических методик решается проблема нормы, базируется на кривой нормального распределения Гаусса. Она получается в некоторых случаях в результате обследований больших популяций. Далее будет рассмотрена правомерность этой концепции.

Стандартизация теста проходит два этапа. Первый относится ко времени конструирования теста и непосредственно связан с характеристикой наполняющих тест задач. Исторически этот этап берет начало с работ А.Бине, впервые расположившего задачи в порядке возрастающей трудности. Процедура определения трудности задач, хотя и носила эмпирический характер, позволяла тем не менее повысить объективность испытаний. Такой принцип определения трудности задач в основном сохранил свою значимость и по настоящее время.

Классификация и выбор задач для включения их в тест всегда связаны с необходимостью соблюдать определенную пропорцию в степени трудности применяемых задач, так как трудность теста непосредственно зависит от трудности вопросов, которые в него входят. Поэтому окончательная проверка трудности самого теста обеспечивается испытанием его на репрезентативной выборке, которая, как правило, имеет достаточно большой количественный объем до 2—4 тыс. испытуемых. Эта процедура позволяет расположить оценки, получаемые отдельными индивидами, по их уровню так, чтобы имелась группировка индивидуальных оценок около центра шкалы распределения и постепенное снижение ее в обе стороны по мере приближения к крайним точкам. Практически кривая нормального распределения вполне удовлетворительно описывает результаты испытаний подобного рода. Однако это имеет место только в конце процесса

конструирования теста, так как на начальных этапах его формирования, как правило, наблюдаются различного рода отклонения. Кривая, например, может иметь асимметрию в том случае, если тест наполнен очень легкими или, наоборот, очень трудными задачами. Регулировка структуры и содержания теста продолжается до тех пор, пока распределение оценок испытуемых не приблизится к нормальному. Фактически это достигается подбором испытуемых с разным уровнем развития исследуемой функции. Тесты умышленно конструируются и специально модифицируются таким образом, чтобы получить среднюю оценку правильно выполненных заданий, равную примерно 50%. Только таким путем возможно достигнуть максимальной дифференцировки индивидуальных различий при применении того или иного теста. Пятидесятипроцентный уровень правильно решаемых задач является, таким образом, максимально удобным для получения нормального распределения.

Использование понятия нормы на этапе интерпретации результатов тестового обследования несколько отличается от этапа составления теста. Непосредственные результаты тестового испытания (сырые оценки) в этом случае используются редко. Материалом, дающим основание для заключений о результатах психологического тестирования, являются так называемые приведенные показатели. Преимущества приведенных показателей в следующем. Прежде всего они позволяют определить положение индивидуальной оценки на оси нормального распределения. Более того, приведенные показатели обеспечивают экспериментатора мерой сравнения результатов деятельности отдельных индивидов. На практике используются три основные группы приведенных показателей: возрастные, процентильные и стандартные.

Концепция умственного возраста, введенная в период реконструкции шкалы Бине—Симона, в стэнфордской редакции соотносит успешность выполнения тестовых задач с хронологическим возрастом испытуемого. Однако существенный недостаток понятия умственный возраст — его значительная изменчивость во времени. В этом и заключается основная трудность интерпретации. Кроме того, как показала практика, нормировка возраста после 15 лет практически невозможна из-за большого разброса параметров. Наконец, последнее: с течением времени на результатах тестовых испытаний начинают сказываться также различного рода неучитываемые дополнительные влияния. Основное употребление понятие умственный возраст и связанная с ним процедура возрастного приведения находят при определении интеллектуального уровня.

Следующая группа приведенных показателей — **процентили** (% испытуемых, выполнивших определенную часть теста). Как правило, процентили высчитывают при выполнении тестов способностей и достижений. Значительное преимущество процентилей состоит в легкости вычислительных процедур и отсутствии затруднений при интерпретации получаемых результатов, что в известной мере обеспечивает их универсальность. Однако недостаток процентилей — это существенное неравенство единиц отсчета в том случае, когда анализируют крайние точки распределения. При использовании процентилей определяется только относительное положение индивидуальной оценки, но не величина различия между отдельными показателями.

Последний вид приведенных показателей — **стандартные показатели**, среди которых различают линейные и нормализованные. **Линейные показатели** подразумевают обычную линейную трансформацию исходных показателей в показатели стандартного отклонения. **Нормализованные показатели** позволяют наиболее удовлетворительно получать приведенные показатели, когда исходные данные трансформируются с помощью таблиц в нормальное распределение. Единственным ограничением в употреблении нормализованных стандартных показателей и одновременно требованием к ним является необходимость выведения окончательных оценок на больших репрезентативных выборках.

Отклонения от нормы с высоким уровнем статистической значимости в этих случаях объясняются недостатками самого теста, а не индивидуальными особенностями ответов отдельного индивида.

Несколько слов о соотношении между приведенными показателями. Как правило, при соблюдении двух условий — тщательности процедуры приведения показателей тестовых измерений и разумной их интерпретации — становится возможной трансформация одного вида показателей в другой с помощью специальных шкал и таблиц.

Понятие нормы включает в себя элементы специфичности. Дело в том, что норма тесно связана с контрольной выборкой, являющейся совокупностью конкретных лиц и показателей, извлекаемой из исходной популяции. Следовательно, норма не есть абсолютный и универсальный фактор, поэтому в отдельных случаях возможны и даже необходимы специализированные нормы. Речь идет о стандартизации тестов на узкозаданных выборках, формирование которых преследует локальные цели. Тесты в этих случаях весьма специфичны и наиболее эффективны.

Приведенный краткий обзор истории развития и становления психологического тестирования свидетельствует о широком разнообразии тестов как по содержанию, так и по форме. Возникновение психологического тестирования, во-первых, вызвано одним общим фактором — непосредственным требованием практики, во-вторых, преследует одну общую цель — удовлетворение этих требований.

3. ПРИНЦИПЫ КОНСТРУИРОВАНИЯ ПСИХОДИАГНОСТИЧЕСКИХ ТЕСТОВ

3.1. Надежность психодиагностических тестов

Конструирование психодиагностических тестов представляет собой процесс создания инструментов измерения, удовлетворяющих основным психометрическим требованиям. Всякий тест должен быть надежным, валидным и стандартизированным.

Одно из решающих значений при конструировании тестов и интерпретации данных, полученных путем тестирования, имеет надежность. **Надежность** - характеристика методики, отражающая точность психодиагностических измерений, а также устойчивость результатов теста к действию посторонних случайных факторов.

Надежность — крайне сложное и многоплановое понятие, одна из основных функций которого — оценить постоянство показателей тестовых испытаний. В принципе можно сказать, что надежность должна обосновывать ошибку измерения — она должна показывать, какая часть изменчивости показателей ошибочна. Известно несколько основных факторов, определяющих уровень надежности. Так, надежность всегда будет иметь тенденцию к увеличению, если соблюдается постоянство условий проведения процедуры тестирования, так как это снижает ошибку вариабельности измеряемого параметра. В то же время многочисленность целей, сложность проблемы, изменчивость ситуации, как правило, увеличивают ошибку измерения, уменьшая тем самым надежность.

При изучении вопросов разработки и использования тестов важную роль играет анализ **ошибок измерения**. Результаты тестирования всегда содержат ошибки, которые могут быть существенными или ничтожно малыми. Ни один тест не может быть правильно использован без тщательного изучения этого важного аспекта измерения. Если ошибки тестирования игнорируются или оцениваются небрежно, то это может привести не только к бессмысленным затратам времени и труда, но, что особенно важно, когда измерения касаются

людей, использование ненадежных тестов может стать причиной педагогических или административных ошибок, последствия которых трудно предугадать.

При проведении тестирования, как при любом измерении, возникают три основных класса ошибок: промахи, систематические и случайные ошибки.

Промахи — следствие грубых нарушений процедуры тестирования. У опытных экспериментаторов промахи встречаются довольно редко. Они могут быть легко выявлены и устранены путем отбрасывания резко отклоняющихся значений.

Систематические ошибки измерения остаются постоянными или закономерно меняются от измерения к измерению. В силу этих своих особенностей они могут быть предсказаны заранее, а в некоторых случаях и устранены. К этой группе относятся ошибки, возникающие в связи с использованием различных методов сбора данных. Каждый метод сбора данных о личности ("L", "Q" и "T") обладает своими специфическими инструментальными искажениями (подробнее см. юниту 2).

Случайные ошибки имеют место, когда при последовательных измерениях постоянной характеристики получаются различные числовые оценки, т.е. при определении случайных ошибок предполагается, что измеряемая характеристика не изменяется во времени, а все отклонения обусловлены неточностью измерения.

В действительности возможна ситуация, когда измеряемая характеристика тоже флуктуирует. Например, если в течение некоторого фиксированного периода времени измеряется доминирующий фон настроения, то различия будут объясняться не только ошибочностью измерений, но и его колебаниями. В этом случае данный анализ по-прежнему справедлив, однако получаемые оценки будут другими.

Деление ошибок измерения на систематические и случайные является довольно содержательным для анализа процедуры тестирования. Вместе с тем не следует забывать, что граница между этими двумя типами ошибок достаточно условна. Например, склонность некоторых испытуемых к экстремальным, крайним ответам на вопросы анкеты является систематической ошибкой, когда рассматривается соответствующая группа испытуемых, и случайной, если дело идет о рандомизированной выборке.

Суммарная ошибка любого теста состоит из ошибок двух этих типов, но в различном соотношении. Относительный вес каждого типа ошибок зависит от качества теста и условий проведения обследования. В общем результаты тестирования, как бы тщательно оно не проводилось, всегда содержат ошибки. Однако, используя методы математической статистики, можно оценить величину суммарной ошибки тестирования и использовать ее для оценки надежности теста. Без статистической оценки надежности теста результаты измерения считаются недостоверными. Основным средством для оценки надежности служит коэффициент корреляции — цифровой показатель надежности.

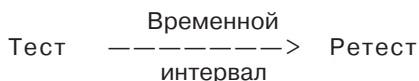
На практике используются три основных метода оценки надежности тестов: 1) повторное тестирование; 2) параллельное тестирование и 3) расщепление. Рассмотрим каждый из них в отдельности.

Метод повторного тестирования

Ошибка измерения может быть определена путем многократного тестирования. Ошибочная вариация при этом свидетельствует о случайных флуктуациях результатов одной и той же деятельности. Поэтому совершенно естественно, что повторное тестирование выборки испытуемых одним и тем же тестом через определенный интервал времени и при одинаковых условиях стало одним из основных методов измерения надежности. Повторное тестирование

обычно называют **ретестом** (retest), а надежность, измеренную таким способом, — ретестовой надежностью.

Ретестовая надежность - характеристика надежности психодиагностической методики путем повторного обследования испытуемых с помощью одного и того же теста. Схема оценки ретестовой надежности имеет вид:



В этом случае за **индекс надежности** принимается коэффициент корреляции между результатами двух тестирований. Вычисляется корреляция между оценками, полученными одними и теми же испытуемыми при первом и повторном проведении теста, причем тестирование одним и тем же тестом происходит в идентичных условиях.

Метод повторного тестирования обладает как достоинствами, так и недостатками. К числу достоинств относятся естественность и простота определения коэффициента надежности. Недостаток этого типа надежности — зависимость результатов от тренировки, имеющей место в ходе самого тестирования. К недостаткам следует отнести неопределенность в выборе интервала между двумя измерениями. Возникновение временной неопределенности связано с тем, что повторное тестирование отличается от первичного. Испытуемые уже знакомы с содержанием теста, помнят свои первоначальные ответы и ориентируются на них при повторном выполнении теста. Поэтому при повторном тестировании нередко наблюдается или “подгонка” под первоначальные результаты, или, как следствие негативизма, демонстрация “новых” результатов. На величину коэффициента надежности влияет и природа теста, так как иногда, уловив структуру теста, испытуемый при повторении легко справляется с предлагаемыми задачами.

В результате оказывается, что, проводя повторное измерение для оценки одних источников ошибок, мы вызываем к жизни другие. Это специфическая психометрическая проблема, так как человек является не только объектом, но и субъектом измерения, активно влияя на результаты тестирования. В качестве одного из эффективных средств устранения ошибок, связанных с повторным применением одного и того же теста, рекомендуется проводить повторное тестирование через достаточно большие интервалы времени, чтобы первые начальные ответы были забыты, а по возможности и сам факт первого измерения изгладился из памяти. Однако при больших интервалах времени между двумя обследованиями могут происходить изменения в самих измеряемых характеристиках личности. В этом случае низкие значения коэффициента надежности будут следствием изменчивости измеряемой характеристики, а не низкой надежности теста. Поэтому, приводя в руководстве к тесту его ретестовую надежность, следует указывать, какому интервалу времени она соответствует. В связи с тем, что ретестовая надежность уменьшается с ростом временного интервала, наиболее доверительными являются высокие коэффициенты надежности, полученные при явно больших интервалах между тестированиями. Недостаточно высокие коэффициенты надежности могут быть следствием неоптимального определения временных интервалов.

Метод параллельного тестирования

В этом случае одни и те же испытуемые выполняют несколько вариантов одного и того же теста. Многократность измерения организуется с помощью параллельных, или эквивалентных, тестов. Рассчитанный между двумя тестами

коэффициент корреляции называется **эквивалентной надежностью**.

Надежность параллельных форм - характеристика надежности психодиагностической методики с помощью взаимозаменяемых форм теста.

Этот тип надежности — **тестирование варьирующими тестами** — представляет собой альтернативную форму надежности. Как правило, практическое использование данного типа надежности связано со значительными затруднениями, поскольку крайне сложно построить несколько вариантов одного теста таким образом, чтобы испытуемый не мог обнаружить их психологическую однородность. Да и искажающее влияние тренировки в этом случае не полностью снято. Кроме того, возникает вопрос: являются ли альтернативные типы надежности характеристиками именно надежности теста, а не параметрами эквивалентности испытаний? Ведь если две формы испытания проводятся при одготипных константных условиях, то, вероятнее всего, исследуются показатели эквивалентности двух форм тестирования, а не показатели надежности самих тестов. Ошибка измерения в этом случае определяется флюктуациями выполнения теста, а не флюктуациями структуры теста.

Высокие значения коэффициента корреляции между параллельными формами указывают не только на надежность этих двух тестов, но и на эквивалентность их психологического содержания. Поэтому, когда эквивалентная надежность низкая, то это может быть связано либо с неэквивалентностью психологического содержания тестов, либо с низкой надежностью, либо с тем и другим вместе. Чтобы исключить некоторые альтернативы, следует получить оценки ретестовой надежности для параллельных тестов. При низкой ретестовой надежности низкую эквивалентную надежность логично считать следствием ненадежности тестов, в противном случае, когда ретестовая надежность высокая, приходится признать, что параллельные формы неэквивалентны по психологическому содержанию. Если имеются сомнения в эквивалентности психологического содержания тестов, то эти сомнения не снимаются никакими статистическими расчетами. Эквивалентность подтверждается психологическим анализом, экспертными суждениями специалистов и только в дополнение к этому — статистическими критериями.

Создание параллельных форм тестов связано с преодолением ряда трудностей. В мировой практике известно довольно мало тестов, имеющих параллельные формы. В качестве примера таких тестов можно назвать многофакторные личностные тесты, разработанные сотрудниками Иллинойского университета под руководством Р.Б. Кэттелла, — CPQ, CSPQ, HSPQ, 16PF, каждый из которых имеет две или более параллельных формы.

Метод расщепления

Надежность теста при однократном тестировании можно определить с помощью его деления на две половины. Это одна из наиболее простых проверок теста, когда вычисляется коэффициент корреляции между его половинами. Метод расщепления является развитием метода параллельного тестирования и базируется на допущении о параллельности не только отдельных форм теста, но и отдельных заданий внутри одного теста.

Надежность частей теста - характеристика надежности психодиагностической методики путем анализа устойчивости результатов отдельных совокупностей тестовых задач или единичных заданий теста. Для вычисления коэффициента надежности методом расщепления тест разбивается на отдельные задания или группы заданий. Каким же образом поделить тест на две половины, чтобы иметь возможность выровнять обе половины по тому или иному конкретному основанию? Наиболее распространена процедура "расщеплений" теста на две части: в одну часть собираются результаты четных заданий,

а в другую — нечетных. Главное достоинство этого типа надежности состоит в независимости результатов тестирования от таких элементов деятельности, как тренировка, практика, утомление и т.д.

При разделении теста на две части **индекс надежности** вычисляется по формуле Спирмена — Брауна:

$$R(X,T) = \frac{2R_{1,2}}{1 + R_{1,2}},$$

где $R(X,T)$ — индекс надежности, $R_{1,2}$ — коэффициент корреляции двух половин теста.

Разделить тест на две равные части можно разными способами, и каждый способ даст новую численную оценку надежности. Для преодоления этого недостатка метода расщепления разрабатываются способы вычисления коэффициентов надежности, которые исходят из предположения деления теста не на две части, а на большее число частей, в пределе равное числу заданий. В психометрической литературе ведется широкое обсуждение этой проблемы, результатом которой являются различные предложения: от простых вычислительных процедур до сложных математических выкладок. Но никто еще не дал исчерпывающего теоретического обоснования этой проблемы: на практике большинство критериев опираются на тот факт, что матрица корреляций отдельных заданий надежного теста состоит из коэффициентов, величина которых близка к единице. Поэтому наиболее естественно рассматривать в качестве индекса надежности средний модуль коэффициента корреляции всех заданий теста или средний коэффициент детерминации.

Вообще говоря, большинство методов измерения надежности путем расщепления по существу являются способами приближенной оценки ранга корреляционной матрицы. Ранг матрицы - это количество линейно независимых столбцов и строк. Надежному тесту соответствует корреляционная матрица с рангом 1. При ранге 2 задания теста образуют уже две разные группы. Чем выше ранг корреляционной матрицы, тем разнородней состав заданий в тесте. Поэтому коэффициент надежности, полученный методом расщепления, называют **коэффициентом внутренней согласованности**, или **гомогенности, теста** (homogeneity). Гомогенность теста свидетельствует о том, что все задания теста устойчиво измеряют одну и ту же психологическую характеристику. Это означает, что если испытуемые имеют одинаковые тестовые оценки, то и степень выраженности у них измеряемого качества одинакова. Таким образом, гомогенность теста — обязательное условие того, чтобы результаты измерения имели однозначную трактовку: в случае негомогенного теста одни и те же оценки у разных испытуемых допускают различное объяснение.

Итак, мы рассмотрели три эмпирических метода оценки надежности тестов: повторное тестирование одним и тем же тестом, повторное тестирование параллельной формой теста и расщепление теста. Какой из этих методов дает истинную оценку надежности теста? Каким из методов следует пользоваться? Ответ на этот вопрос зависит от личной симпатии и целей исследования.

При использовании метода повторного тестирования получаем оценку степени устойчивости результатов во времени и в зависимости от условий тестирования. Поэтому **ретестовый коэффициент** надежности называют также **коэффициентом устойчивости**, или **стабильности, теста**.

При использовании метода параллельных форм и метода расщепления оценивается степень взаимной согласованности частей теста. Поэтому коэффициенты надежности, полученные этими двумя методами, интерпретируются как показатели **гомогенности, однородности тестов**.

Помимо показателей устойчивости и гомогенности Р.Б. Кэттелл считает необходимым рассматривать показатель **переносимости** (transferability). Он представляет собой оценку способности теста сохранять точность измерения в разных выборках, субкультурах и популяциях. Совместно устойчивость, гомогенность и переносимость образуют комплексную характеристику надежности, которую Кэттелл называет **постоянством** (consistency) и определяет как степень, в которой тест продолжает предсказывать то, что он однажды предсказал, несмотря на изменения (в определенных пределах):

- а) объема, в котором тест применялся;
- б) условий, в которых он применялся;
- в) состава выборки, в которой он применялся.

Выделение этих компонентов надежности Р.Б. Кэттелл проводит, опираясь на анализ **источников вариаций** результатов тестовых измерений. Эти источники изображены на рис.3.1 как векторы пятимерного пространства.

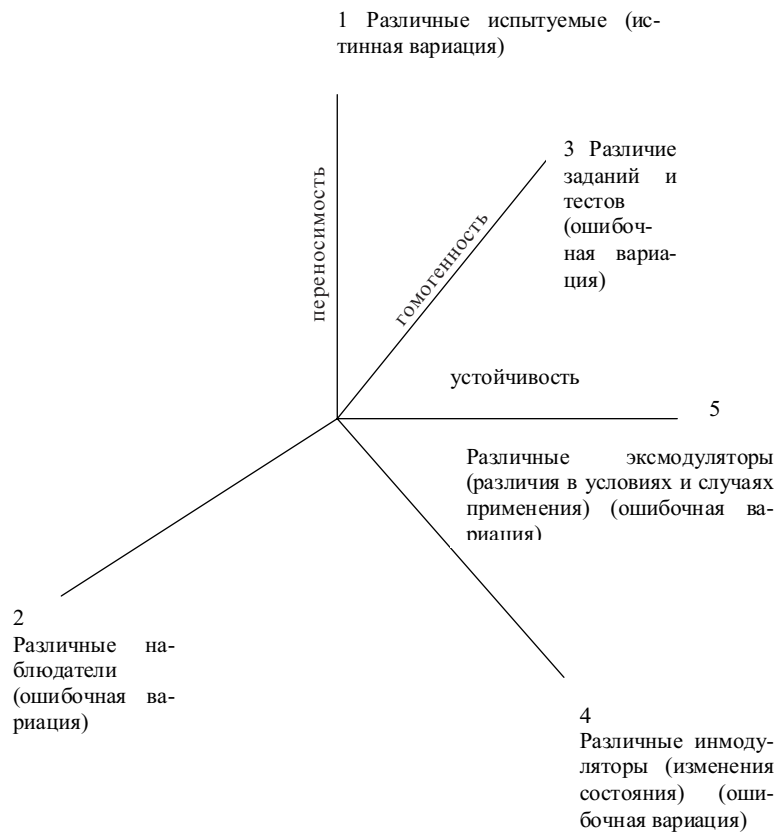


Рис.3.1

При исследованиях устойчивых характеристик личности измерению подлежит только истинная вариация, показанная осью 1. Оси 2—5 показывают вариации, которые формируют суммарную ошибку тестирования. Она может возникнуть из-за различий между экспериментаторами, осуществляющими тестирование, различий в самих тестах (разнородность, гетерогенность заданий), изменения состояний испытуемых во времени (инмодуляторы), а также из-за неконтролируемости внешних условий эксперимента (эксмодуляторы). Контроль источников ошибочных вариаций является самым эффективным средством повышения надежности тестирования.

Надежность результатов тестирования зависит не только от надежности самого теста и процедуры его проведения. Важным фактором, влияющим на результаты интерпретации данных, является специфика конкретной выборки. Наиболее существенными характеристиками выборки, с этой точки зрения, следует признать социально-психологическую однородность по разным параметрам; учитывается также возраст и пол.

Существует тип надежности, непосредственно адресованный к надежности самого лица, проводящего тестовое испытание. Оценка **надежности лица, проводящего тест**, получается с помощью независимого моделирования испытания двумя различными экспериментаторами.

Особо следует отметить еще одну характеристику тестов, выражаемую двумя подходами к их конструкции и интерпретации. Большинство тестов — это испытания на скорость. Индивидуальные оценки, получаемые в эксперименте, зависят от скорости деятельности. Корректность получения этих оценок обеспечивается такими факторами, как единообразие действия и невысокая сложность предлагаемых решений. Лимит времени настолько мал, что ни один испытуемый не в состоянии выполнить все предлагаемые задачи. При таких условиях каждая индивидуальная оценка зависит только от скорости, с которой выполняется работа.

Другой вид тестов преследует испытание того, что именуется мощностью (power). В этом случае отсутствует жесткий лимит времени — его достаточно, чтобы попытаться решить все предложенные задачи теста. Однако трудность задач постепенно нарастает, и тест иногда завершают практически заведомо нерешаемые задачи — настолько высок уровень их трудности (разумеется, в данной специфической выборке).

Объективная информация о соотношении скорости выполнения теста с показателем его мощности позволяет повысить общую его надежность при интерпретации результатов.

Рассмотренные понятия тестологии, безусловно, ее важнейшие атрибуты. Однако ни стандартизация, ни нормирование, ни высокие показатели надежности сами по себе не определяют практическую ценность теста. Ведущий фактор, который позволяет измерить целевые результаты психологического тестирования, — валидность.

3.2. Валидность психодиагностических тестов

Валидность (validity) - комплексная характеристика методики (теста), включающая сведения об области исследуемых явлений и репрезентативности диагностической процедуры по отношению к ним.

Валидность теста информирует о том, что должен измерять тест и насколько хорошо он это делает. В принципе эта информация может содержаться уже в наименовании, которое в краткой форме должно отражать цель конкретного тестирования. К сожалению, это требование и по сей день еще не стало реальностью. Выражение валидности в общих терминах наименования принципиально невозможно, так как ни один тест нельзя охарактеризовать высоким

или низким уровнем валидности абстрактно, не учитывая его конкретной целенаправленности. На деле между наименованием теста и его практическим назначением нередко существует несоответствие. Валидность теста должна определяться только относительно его специального использования.

Главная проблема теории тестов заключается в том, как по измеренным результатам определить истинные результаты. Существуют два основных определения истинных результатов тестирования, которые имеют значение для психометрии: специфические и обобщенные (генерализованные).

Под **специфическим истинным результатом измерения** понимается точный результат, полученный с помощью конкретного теста. Это понятие близко к понятиям действительных точных (или “выровненных”) оценок, используемых при измерении физических величин. Здесь истинность результатов понимается с точки зрения точности метода измерения. Для характеристики специфической истинности результатов тестирования и разработано понятие надежности. Оно характеризует соотношение наблюдаемого результата и его действительной величины.

Генерализованный (обобщенный) истинный результат — это термин, используемый для характеристики действительного уровня скрытых, ненаблюдаемых качеств или свойств. Скрытое от внешнего наблюдения свойство не может быть измерено только с помощью одного какого-либо теста. На практике для этой цели используется батарея с конечным числом разных тестов, направленных на измерение одного скрытого свойства, уровень которого требуется определить. Каждый из используемых тестов лишь частично покрывает исследуемое свойство. Это особый случай отношений между сущностью и явлением.

В отечественной психодиагностике, в отличие от диагностического позитивизма западной психометрии, исходят из предположения, что эти отношения носят закономерный и реальный характер, а не являются только результатом конструктивных построений.

Для характеристики способности теста измерять действительный уровень психического свойства или качества используется понятие валидности. Валидность теста показывает, в какой мере он измеряет то качество (свойство, способность, характеристику и т.п.), для оценки которого он предназначен. Инвалидные, т.е. не обладающие валидностью, тесты не пригодны для практического использования.

Валидность и надежность — связанные понятия. Их соотношение может быть проиллюстрировано следующим примером. Допустим, имеются два стрелка А и В. Стрелок А выбивает 90 очков из 100, а стрелок В — только 70. Соответственно, надежность стрелка А равна 0,90, а стрелка В — 0,70. Однако стрелок А всегда стреляет по чужим мишеням, поэтому на соревнованиях его результаты не засчитываются. Второй стрелок всегда правильно выбирает мишени. Поэтому валидность стрелка А нулевая, а стрелка В — 0,70, т.е. численно равна надежности. Если стрелок А начнет правильно выбирать мишени, его валидность тоже будет равна его надежности. Если же он будет продолжать иногда путать мишени, то часть результатов не будет зачтена и валидность стрелка А будет ниже надежности. В нашем примере аналогом надежности является точность, меткость стрелка, а аналогом валидности — тоже точность стрельбы, но не по любой, а по строго определенной, “своей” мишени. Поэтому надежный стрелок может стать валидным, не только сменив мишень, но и сменив команду. В истории известны случаи, когда тесты, признанные инвалидными для измерения одних свойств, оказывались валидными для других. Значит, надежность является необходимым условием валидности. Ненадежный тест не может быть валидным, и, наоборот, валидный тест всегда надежен. Надежность теста не может быть меньше его валидности; в свою очередь, валидность не может превышать надежности.

В современной психометрии выделяются три основных вида валидности: 1) содержательная (логическая); 2) эмпирическая и 3) конструктивная (концептуальная).

Содержательная валидность (content validity) - один из основных типов валидности методики, характеризующий степень репрезентативности содержания заданий теста измеряемой области психических свойств. Ее часто называют “логической валидностью” или “валидностью по определению”. Содержательная валидность означает, что, по мнению специалистов, тест является валидным.

Валидность по содержанию, определяемая путем систематического контроля содержания, должна показывать, насколько полно охватывает тест представительную выборку по совокупности измеряемых параметров. Сфера тестологического исследования должна постоянно изучаться и анализироваться с тем, чтобы тест был наполнен адекватными задачами. В основном данный тип валидности применяется в тестах достижения. Для тестов способностей и тестов личности этот тип валидности малоприменим и неадекватен.

Содержательную валидность следует отличать от очевидной, лицевой, внешней валидности. **Очевидная валидность** (face validity) - представление о тесте, сфере его применения, результативности и прогностической ценности, которое возникает у испытуемого, не располагающего специальными сведениями о характере использования и целях методики.

Очевидная валидность играет важную роль в процессе тестирования, поскольку именно она в первую очередь определяет отношение испытуемых к обследованию. В.А. Аванесов приводит пример отказа от выполнения теста испытуемыми из-за отсутствия у теста очевидной валидности. Во время одного социологического опроса моряков рыболовного флота он столкнулся со случаем отказа заполнять анкету под предлогом того, что анкета составлена “не для моряков”, в ней нет вопросов, относящихся к морякам, “на морские темы”. Если бы опросник содержал “морские термины”, то он бы казался “своим” и вызывал больше доверия у испытуемых. Именно в силу этого обстоятельства очевидную валидность иногда называют также доверительной валидностью.

Очевидная внешняя валидность может находиться в разных отношениях с содержательной валидностью. В одних случаях содержательная и внешняя валидность совпадают, в других внешняя валидность используется для маскировки содержательной валидности. С этой целью содержанию заданий и инструкции придается специальная форма, маскирующая истинную цель исследования. Тем самым предупреждается возникновение установочного поведения при тестировании. Это один из основных тактических приемов, используемых при создании объективных тестов.

В отличие от очевидной валидности содержательная валидность всегда связана с оценкой отношения содержания тестового материала к истинной, а не мнимой цели измерения. Для определения содержательной валидности используются **экспертные методы**. Как и во всякой экспертизе, необходимо решить следующие вопросы: отобрать компетентных экспертов, сформировать экспертную группу, организовать процедуру экспертизы. Специфичным является **объект экспертизы** — содержание теста. Эксперты должны оценить содержание заданий теста по их соответствию психическому свойству, объявляемому в качестве содержания валидизируемого теста. С этой целью экспертам предъявляются спецификация к тесту и список заданий. Если конкретное задание полностью соответствует спецификации, то эксперт обозначает его как соответствующее содержанию теста. В противном случае он бракует задание и письменно обосновывает свою оценку. Если несколько экспертов бракуют одни и те же задания, они признаются несоответствующими спецификации к тесту. Если между экспертами существуют расхождения в оценке заданий, их анонимно знакомят с письменными объяснениями экспертов, забраковавших задания, после чего экспертиза повторяется.

Конечно, следует предусмотреть возможность контроля компетентности и добросовестности экспертов. Экспертам всегда легче признать задание соответствующим спецификации, чем писать развернутое объяснение. Одним из простых способов такого контроля является введение заданий, не соответствующих спецификации к тесту. Оценкам тех экспертов, которые оказались несостоятельными в выделении контрольных заданий, не следует особенно доверять.

С целью получения итоговой количественной оценки содержательной валидности теста производится обобщение экспертных оценок. В простейшем случае это может быть процент заданий, признанных группой экспертов соответствующими спецификации к тесту.

Эмпирическая валидность (empirical validity) - совокупность характеристик валидности теста, полученных сравнительным статистическим способом оценивания. Идея эмпирической валидности заключается в определении способности теста служить индикатором или предсказателем строго определенной психической особенности или формы поведения человека. Для измерения эмпирической валидности теста вычисляется **коэффициент корреляции** тестового результата с внешним критерием — $R(X,T)$.

Критерий валидации - непосредственная и независимая от валидируемого теста мера психического свойства, на исследование которого направлена психодиагностическая методика.

В качестве критерия может выступать любой показатель, независимо и бесспорно измеряющий ту же психологическую характеристику, что и валидируемый тест. Вопрос о выборе критерия является в этом виде валидности основным. От выбора критерия зависят качественная и количественная оценки валидности. Условно можно выделить три группы внешних критериев: экспертные, экспериментальные и "жизненные".

Экспертный критерий предполагает использование экспертных оценок для независимого измерения свойств личности. Хотя экспертному методу измерения особенностей личности в последнее время уделяется значительное внимание, тем не менее применение экспертных оценок для валидации тестов не имеет достаточно разработанных унифицированных процедур, позволяющих получить достоверную и устойчивую информацию.

При валидации тестов, предназначенных для школьников, в качестве экспертов обычно используются учителя. Однако, как показали специальные исследования, экспертные оценки учителей имеют низкую валидность и надежность. Немногие педагоги могут выступить в качестве надежных экспертов. Это связано с тем, что их оценки в сильной степени подвержены субъективным искажениям (симпатиям и антипатиям, переносу отношений с успеваемости на качество личности и т.п.).

Другой причиной ошибок является недостаточное понимание содержания оцениваемого качества. Следует добиваться соответствия понятий по объему для теста и экспертной оценки. В противном случае нельзя рассчитывать на высокую корреляцию.

Если просто предложить учителю оценить такую-то особенность личности, то это будет означать постановку перед неспециалистом задачи, затруднительной даже для большинства психологов. Вначале следует объяснить учителю содержание психического свойства, для измерения которого предназначен валидируемый тест, ознакомить его со спецификацией и составом заданий. Только после этого можно просить учителя выставить экспертные оценки ученикам. Оценку следует выставлять по 7- или 9-балльной шкалам, что, во-первых, делает оценки более точными, а во-вторых, затрудняет механический перенос оценок успеваемости на измеряемое качество.

В связи с низкой надежностью и трудностями организации экспертизы

экспертные критерии валидности используются редко, преимущественно в случаях, когда невозможно или чрезвычайно затруднительно использовать в качестве внешнего критерия валидности обычные количественные методы измерения. На практике значительно больше распространены **экспериментальные критерии**. В этих случаях в качестве критерия валидности используются результаты одновременного и независимого тестирования испытуемых другим тестом, предположительно измеряющим то же свойство личности, что и валидизируемый тест. Коэффициент корреляции между результатами двух измерений называется **эмпирической взаимной валидностью**. Его величина зависит как от степени совпадения содержания измерения, так и от надежности тестов. Поэтому максимальные коэффициенты эмпирической взаимной валидности имеют параллельные тесты. Коэффициент взаимной валидности параллельных тестов равен их коэффициенту надежности.

Это вытекает из определения параллельных тестов как тестов, измеряющих одно и то же свойство с одинаковой ошибкой. Если бы параллельные тесты были абсолютно надежны, их взаимная валидность была бы равна единице, так как по психологическому содержанию они идентичны.

Эмпирическая валидность теста — это всегда валидность по отношению к конкретному критерию. Поэтому процесс конструирования теста часто сопровождается поиском подходящего критерия. Если к моменту проверки валидности такого экспериментального критерия не оказывается, в качестве него используются характеристики реального поведения, которые на основе существующих психологических теорий предполагаются связанными с измеряемым свойством. Например, в качестве **жизненных критериев** для тестов интеллекта используются показатели успешности обучения, для экстравертированности — успешность административной деятельности, для тревожности — частота заболеваний неврозом и т.п. Однако успешность обучения, поведения и деятельности редко зависит только от одного-единственного свойства психики. Реальное поведение человека множественно детерминировано, т.е. зависит не от одного, а от целого комплекса психических свойств. Поэтому использование жизненных критериев для валидации тестов имеет смысл только в отношении многомерных тестов, таких, как MMPI, 16PF, HSPQ и др., представляющих многопрофильные батареи тестов. Для оценки валидности батареи тестов или многомерного теста в целом используются методы множественного регрессионного анализа.

Регрессионный анализ предназначен для изучения влияния независимых переменных на одну или несколько зависимых переменных. В общем случае использование регрессионного анализа для измерения валидности является по существу построением стохастической модели поведения, представляющей значения жизненного критерия в виде линейной комбинации оценок по шкалам валидизируемой батареи тестов.

Коэффициент валидности - статистический показатель эмпирической валидности теста. В качестве коэффициента валидности батареи тестов относительно жизненного критерия используется коэффициент множественной корреляции или совокупный коэффициент корреляции (R). Величина совокупного коэффициента корреляции показывает, какая часть дисперсии жизненного критерия может быть объяснена с помощью линейной комбинации тестовых оценок. Поэтому чем больше R, тем большая доля дисперсий элиминируется уравнением регрессии, тем ближе вычисленные значения жизненного критерия к его реальным значениям и тем, соответственно, выше валидность теста по отношению к данному жизненному критерию. Валидность теста по отношению к жизненному критерию называется иногда **практической валидностью**. Наиболее распространенные многомерные личностные тесты и батареи тестов обладают высокой практической валидностью по отношению к большому числу жизненных

ситуаций.

Критериальная валидность - комплекс характеристик, включающий текущую и прогностическую валидность методики и отражающий соответствие диагноза и прогноза определенному кругу критериев измеряемого качества.

Критериальная валидность определяется по связи с критериями с единственной, крайне важной для практики целью: оценить индивидуальную прогностичность теста. Для этого результаты выполнения теста сопоставляются с непосредственными и независимыми оценками того, что должен предсказывать тест. Это один из принципиальных вопросов использования тестов. Речь идет о сопоставлении между собой двух форм валидности: текущей (диагностической) и прогностической.

Текущая валидность - характеристика теста, отражающая его способность различать испытуемых на основании диагностического признака, являющегося объектом исследования в данной методике.

Критерии текущей (диагностической) валидности очевидны: они определяются самим содержанием и процедурой теста, и можно с определенной вероятностью ответить на вопрос о том, что за функция определяется в подобной ситуации. Прогнозирование же уровня выполнения в будущем какого-либо вида деятельности (к тому же непосредственно самой тестовой процедурой не исследуемого) в известном смысле напоминает гадание и имеет низкую вероятность.

Прогностическая валидность - информация о методике (тесте), характеризующая степень точности и обоснованности суждения о диагностируемом психологическом качестве по его результату спустя определенное время после измерения.

Для увеличения прогностической валидности употребляется процедура, позволяющая достигнуть достаточной чистоты теста. При этом добиваются его независимости от различного рода влияний, таких как содержание задач, оценочных критериев влияния знаний, способностей испытуемого. Для валидности данного типа используют как общие, так и специфические критерии, однако, несмотря на широкое применение общих, предпочтение в последнее время отдается специфическим критериям. Они используются в частных случаях, но обеспечивают более высокий уровень валидности, так как основаны на тестировании конкретными задачами, имеющими конкретные цели.

Возможны критерии валидности, являющиеся объединением обоих видов критериев. **Синтетическая валидность**, таким образом, представляет собой комплексную характеристику, включающую в себя в определенной пропорции оба отмеченных выше аспекта.

В настоящее время не возникает сомнений в практической ценности тестов.

Инкрементная валидность - один из компонентов критериальной и прогностической валидности теста, отражающий практическую ценность методики при проведении отбора. Однако использование результатов тестирования, т.е. диагностической информации, для предсказания результатов нетестовой деятельности в будущем более чем проблематично.

Количественно валидность чаще всего оценивается с помощью величины коэффициента корреляции между тестовой оценкой и критерием деятельности (внешним критерием). Известны основные факторы, влияющие на величину коэффициента валидности. К ним относятся: характер группы испытуемых, ее специфичность; однородность выборки (однородность); преселекция (предварительный отбор); различного рода изменения во времени.

Для интерпретации результатов тестирования большое значение имеет также форма отношения между тестом и внешним критерием. Как правило, между ними предполагается прямолинейная связь, которую оценивают с помощью **коэффициента корреляции Пирсона**. Однако этот коэффициент корректен лишь

до тех пор, пока исследуемая функция изменяется линейно и монотонно. На практике данное условие выполняется, к сожалению, далеко не всегда, поскольку комплексный характер причинно-следственных отношений между явлениями, значительное количество действующих факторов, учет которых возможен также не во всех случаях, сильно отклоняют измеряемые зависимости от прямолинейной формы. Как следствие этого, величина коэффициента валидности оказывается невысокой. Величину валидности пытаются выражать при помощи стандартной ошибки измерения, по прямой аналогии с ошибкой измерения надежности теста.

Популяционная валидность — это оценка того, насколько правомерно распространять обобщения, полученные на основании использования в исследовании ограниченного числа испытуемых, на более широкие выборки испытуемых.

С учетом того, что в различных выборках и подгруппах испытуемых валидность сильно зависит от личностных характеристик, в тестологии вводится понятие **уравнивающих переменных**. Установлено, что предсказуемая ошибка теста имеет гораздо меньшую величину, чем ошибка индивидуальной оценки. Поэтому при использовании предсказуемой валидности, полученной на нескольких конкретных выборках в случае объединения их даже в общую группу, коэффициент валидности может иметь очень малую величину. Однако перекомбинация индивидуальных особенностей в некоторое число идентифицированных характеристик (например, половые и возрастные параметры) позволяет повысить валидность теста по одним и понизить по другим, не имеющим высокой ценности, показателям.

Для предсказания практического критерия, как правило, используется не один тест, а целый их набор, батарея тестов. Одна из основных проблем, возникающих в этом случае, — проблема объединения оценок различных тестов в некую обобщенную характеристику испытуемого. Решение этой проблемы идет тремя путями: составлением уравнения множественной регрессии, определением множественных показателей границ и использованием клинических оценок.

Уравнение регрессии определяет величину предсказуемого критерия для каждого индивида на основании его оценок, полученных при выполнении заданий всей батареи тестов. Коэффициенты при членах уравнения оценивают вклад отдельной индивидуальной особенности, измеряемой конкретным тестом. Тесты, имеющие высокую корреляцию с критерием, будут наиболее весомыми. Валидность батареи тестов определяется вычислением корреляции между критерием и батареей. Она будет обладать максимальной предсказательной ценностью в том случае, если каждый отдельный тест имеет оптимальный вес для предсказания критерия в своих баллах. Эта оптимальность отдельных весов определяется из уравнения регрессии.

Процедура определения **многофакторных показателей границ** в основном направлена только на фиксацию минимальной оценки для каждого теста. В соответствии с этим каждый испытуемый, получивший хотя бы по одному тесту оценку ниже минимальной, исключается из выборки.

Ни одна статистическая процедура и формула не в состоянии полностью отразить многообразие индивидуальных оценок. Поэтому часто в практике тестирования используются **клинические оценки**, представляющие собой не что иное, как качественное описание сущности изучаемого феномена. В данном случае речь идет об использовании приемов, не опирающихся на статистическую обработку.

Конструктивная валидность (construct validity) - один из основных типов валидности, отражающий степень репрезентации исследуемого психологического конструкта в результатах теста.

Она устанавливается путем доказательства правильности теоретических

концепций, положенных в основу теста. Обращение к конструктивной (концептуальной) валидности необходимо в тех случаях, когда результаты тестовых измерений используются не просто для предсказания поведения, а как основа для выводов о том, в какой степени испытуемые обладают некоторой определенной психологической характеристикой. Причем измеряемая психологическая характеристика не может быть отождествлена с какой-либо наблюдаемой особенностью поведения, а представляет собой теоретическую концепцию. Примерами таких концепций являются интеллект, черты личности, мотивы, установки и т.д. Конструктивная (концептуальная) валидность имеет значение также при разработке принципиально новых тестов, для которых не определены внешние критерии валидности.

По существу проверка конструктивной (концептуальной) валидности теста проходит через три основных этапа:

1) определяется некоторая теоретическая концепция, которая предположительно объясняет выполнение валидизируемого теста;

2) из теоретической концепции выводятся одна или несколько гипотез, связанных с валидизируемым тестом;

3) выдвинутые гипотезы подвергаются эмпирической проверке.

Если эмпирические данные подтверждают гипотезу, то тем самым подтверждается психологическая концепция, положенная в основу теста, и способность теста служить инструментом измерения этой концепции. Чем убедительнее эмпирическое подтверждение, тем определеннее можно говорить о валидности теста по отношению к психологической концепции, положенной в его основу. Если эмпирические данные не подтверждают выдвинутых гипотез, то ошибка могла возникнуть из-за:

а) неправильности теоретической концепции, положенной в основу теста;

б) отсутствия соответствия между тестом и теоретической концепцией;

в) ошибок при выдвижении гипотез, относящихся к квалифицируемому тесту.

В отличие от эмпирической валидности, для которой обычно имеется единственный внешний критерий, при оценке конструктивной (концептуальной) валидности проверка подвергаются по возможности все гипотезы, вытекающие из теоретической концепции, лежащей в основе валидизируемого теста.

Конструктивная валидность дает информацию о степени измерения тестом теоретически (конструктивно) выделяемой черты. Наиболее часто данный тип валидности характеризует тесты интеллекта, вербальные, скоростные, нейротизма, тревожности. Конструктивная валидность требует постепенного накопления информации об изменчивости оценок. Фиксируются и анализируются данные, проливающие свет на отдельные черты и природу изучаемого явления, а также условия, воздействующие на его развитие и проявление. Особое место в процедуре определения конструктивной валидности занимает факторный анализ, когда статистические процедуры применяются для идентификации результатов психологических исследований.

Специфика определения валидности по конструкции зависит от нескольких факторов, и в первую очередь таких, как возрастные различия и степень связи с другими, ранее использованными тестовыми испытаниями. **Валидность по возрастной дифференциации** - один из компонентов конструктивной валидности, связанный с возрастной динамикой изменений исследуемого качества.

Практика психологического тестирования требует фиксировать еще одну разновидность валидности, выражаемую двумя противоположными характеристиками. Важное значение имеют данные не только о высокой корреляции теста с некоторыми другими переменными, с которыми он теоретически должен коррелировать. Существенно также, чтобы тест не был тесно связан с теми параметрами, с которыми он не должен коррелировать. В

соответствии с этим для определения концептуальной валидности широко используется предложенная Д.Т. Кэмпбеллом и Д.В. Фиске процедура оценки **конвергентной и дискриминантной валидности**.

Эта процедура предполагает использование наряду с валидируемым тестом специальной многомерной батареи контрольных тестов, подобранной таким образом, чтобы в нее входили как тесты, предположительно связанные с валидируемым тестом, так и не связанные с ним. Экспериментатор должен заранее предсказать, какие тесты будут высоко коррелировать с валидируемым тестом, а корреляция с какими тестами будет низкой. Тесты, которые, по предположению, высоко коррелируют с валидируемым тестом, называются **конвергирующими**, а не коррелирующие — **дискриминантными**. Концептуальная валидность может считаться удовлетворительной, если коэффициенты корреляций валидируемого теста с группой конвергирующих тестов статистически значимо выше коэффициентов корреляций с группой дискриминантных тестов.

Когда батарея контрольных тестов относительно малочисленна, для сравнительного сопоставления коэффициентов корреляций в группах конвергирующих и дискриминантных тестов достаточно визуального анализа матрицы корреляций. При анализе связей большого числа тестов, а именно десять и более, для оценки концептуальной валидности следует использовать процедуры группировки признаков: метод корреляционных плеяд, факторный анализ, компонентный анализ и т.д.

3.3. Стандартизация психодиагностических тестов

Стандартизация психодиагностических тестов представляет собой линейное или нелинейное преобразование тестовых оценок. Смысл преобразований исходных тестовых оценок заключается в изменении характера их распределения с тем, чтобы облегчить понимание и интерпретацию тестовых результатов. Чаще всего используются три основных вида преобразований: 1) приведение к нормальному виду; 2) приведение к стандартной форме и 3) квантильная стандартизация.

Преобразование распределения тестовых оценок к нормальному виду.

Стандартизация психодиагностических тестов основана на так называемой **аксиоме нормальности**, т.е. опирается на предположение, что все психические характеристики распределены в популяции по нормальному закону Гаусса.

Предположение о нормальности распределения тестовых результатов является некоторой идеализацией. На практике многие тесты дают результаты, распределение которых отличается от нормального. Поэтому часто возникает вспомогательная задача нахождения способа преобразования данных к нормальному виду. В самом начале поиска способа преобразования большую помощь может оказать построение гистограммы и полигона распределения. Они позволяют легко выявить левостороннюю или правостороннюю асимметрию, двугорбость и другие отклонения от нормальности. В психологических исследованиях часто встречаются логарифмические нормальные распределения, особенностью которых является крутая левая ветвь полигона и пологая правая (т.е. частоты резко падают с ростом тестовых оценок). При логарифмировании исходных тестовых данных левая ветвь кривой распределения растягивается и распределение принимает приближенно нормальный вид.

Для нормализации распределений с правосторонней асимметрией используются тригонометрические и степенные преобразования данных. Таким образом удастся преобразовать тестовые оценки, не подчиняющиеся закону нормального распределения, чтобы распределение новых, преобразованных оценок стало нормальным.

Преобразование тестовых оценок в стандартную форму. Нормальное распределение имеет вид симметричной колоколообразной кривой, которая растянута до бесконечности в положительном и отрицательном направлениях.

Нормальное распределение полностью определяется значениями двух параметров: генеральной средней (m) и стандартного отклонения (b). Положение кривой распределения в системе координат задается значением m , а степень ее крутизны или растянутости — величиной b . На рис.3.2 представлено графическое изображение функций $P(X)$ с различными значениями m и b . По оси абсцисс отложены значения тестовых оценок, а по оси ординат — плотность вероятности этих значений.

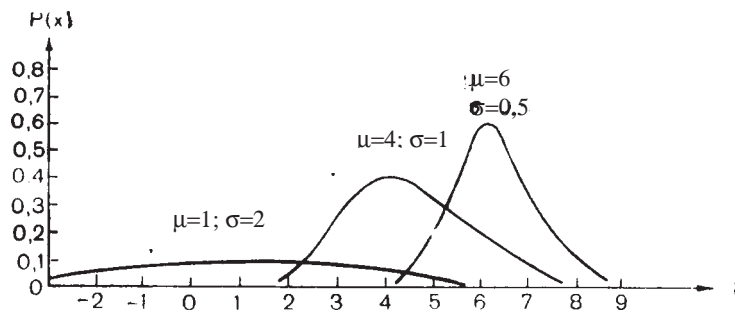


Рис.3.2

Интерпретация тестовых оценок невозможна без знания того, к какой кривой распределения они принадлежат, т.е. для того, чтобы оценить величину тестовой оценки и частоту ее реализации, необходимо соотнести их с генеральной средней m и стандартным отклонением b . Без этого исходные тестовые оценки ничего не скажут нам о степени выраженности исследуемой характеристики и о вероятности появления такого ее значения у других лиц. Вместе с тем именно эта информация особенно интересует экспериментатора, поскольку чаще всего тестовые обследования проводятся для сравнения испытуемых по исследуемой психологической характеристике.

По исходным оценкам мы можем судить только о том, что чем выше оценка, тем больше выражена соответствующая характеристика. О том же, какова она по отношению к среднему значению этого свойства в популяции, мы ничего сказать не можем.

Следующим существенным недостатком исходных тестовых оценок является невозможность сопоставления результатов, полученных с помощью разных тестов. Как правило, разные тесты имеют различные средние и стандартные отклонения, поэтому их результаты имеют различную размерность. Чтобы сделать возможным сопоставление результатов и устранить различия в размерности, необходимо тестовые оценки нормировать, введя единый для всех оценок масштаб.

Способ приведения тестовых оценок к виду, удобному для практического использования, предложен Р.Б. Кэттеллом. Он представляет собой перевод исходных тестовых оценок в 10-балльную равноинтервальную шкалу. Это достигается путем разбиения оси значений тестовых оценок на 10 интервалов, соответствующих долям стандартного отклонения (рис.3.3). При этом среднее арифметическое по группе принимается за среднюю точку и ей приписывается значение, равное 5,5 балла по стандартной десятибалльной системе. Всякая оценка в интервале $(x + 0,25b)$ переводится в 6 баллов, а оценка $(x - 0,25b)$ дает

стандартный балл, равный 5,0. Любое дальнейшее увеличение или уменьшение тестовой оценки на $0,5b$ увеличивает или уменьшает стандартную оценку на 1 балл.

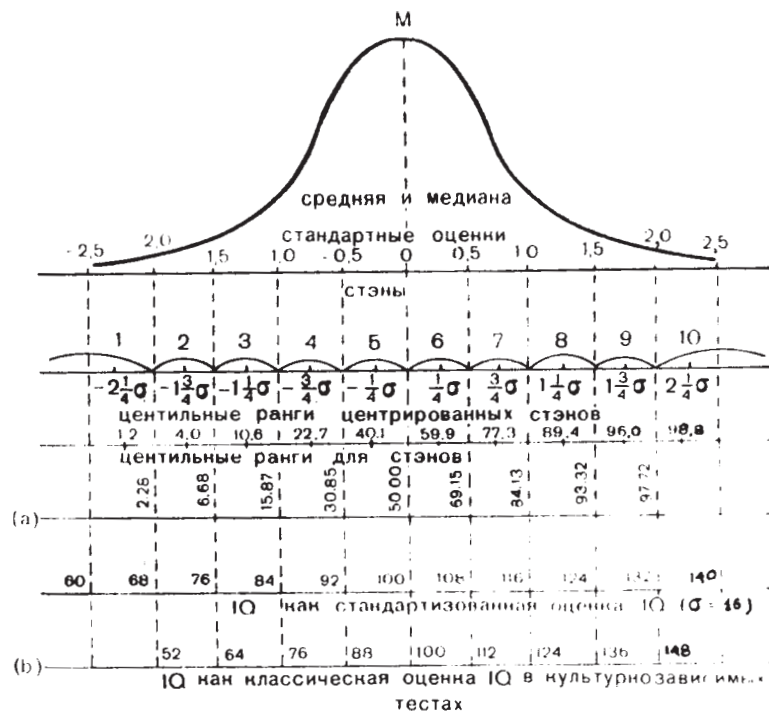


Рис.3.3

При такой системе стандартизации диапазон, который принято называть средним или нормой (диапазон в 1 стандартное отклонение), характеризуется стандартными оценками от 4 до 7 баллов. Только при получении стандартных оценок в 3 или 8 баллов следует думать о значительных индивидуальных отклонениях, выходящих за границы средней нормы.

Оценки 2 и 9 баллов получаются при отклонении индивидуальных оценок на $1,75b$ выше или ниже среднегруппового значения. Максимальная оценка в 10 баллов по десятибалльной системе достигается при отклонении индивидуального тестового результата на $2,0b$ вверх от средней нормы. Однако, чтобы включить в анализ 0,6 % выборки с отклонениями выше $2,0b$, оценка в 10 баллов распространяется и на все остальные оценки, отклоняющиеся от средней более чем на два стандартных отклонения. Аналогичным образом оценка в 1 балл ставится за все отклонения от средних значений ниже двух стандартных отклонений.

Метод стандартизации, предложенный Р.Б. Кэттеллом, — это метод огрубленного интервального представления данных, поэтому его разумно применять в случаях, когда не требуется высокой точности измерения. По этой же причине он может быть использован для согласования оценок по тестам,

стандартизованным разными способами.

Квантильная стандартизация. В некоторых случаях знания степени отклонения индивидуального результата от среднегруппового бывают недостаточны. Экспериментатору необходимо оценить место, которое занимает испытуемый в популяции по исследуемому показателю, т. е. узнать, какой процент испытуемых выполняет тест хуже или лучше обследованного лица, имеет более высокие или более низкие оценки и т.п. Ответ на эти вопросы может быть получен на основе распределения накопленных частот.

На рис.3.4 по оси абсцисс отложены нормированные значения тестовых оценок, а по оси ординат — накопленная частота. Кривая накопленной частоты позволяет легко определить место испытуемого в нормативной выборке, т. е. перейти от одной шкалы оценок к другой. Исходные оценки выражают результаты тестирования через задания теста, а преобразованные — через популяцию. Как видно из рис.3.4, обе эти шкалы связаны нелинейным образом.

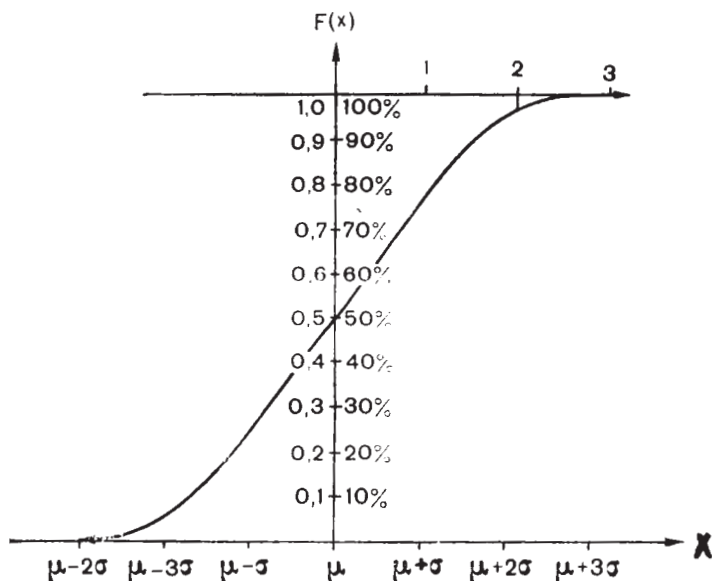


Рис.3.4

На практике используются не точные, а интервальные оценки места испытуемого в популяции. С этой целью ось накопленной частоты разбивается на фиксированное число равных интервалов. Точка на оси накопленной частоты, делящая ось в установленной пропорции, называется **квантилем**, поэтому этот вид стандартизации называется **квантильной стандартизацией**.

Квантиль — это общее понятие, а квартили, квинтили, децили и процентили — его наиболее частные реализации. Имеются, например, три квартиля (Q1, Q2, Q3), которые делят выборку на четыре равные части (кварти) таким образом, что 25% испытуемых располагаются ниже Q1, 50% — ниже Q2 и 75% — ниже Q3. Четыре квинтиля (K1, K2, K3, K4) делят выборку аналогичным образом на пять равных частей, девять децилей (D1, ..., D9) — на десять равных частей, а 99 процентилей (P1, ..., P99) — на 100 равных частей.

Номер квантиля используется в качестве новой преобразованной тестовой

оценки. Он показывает относительное положение испытуемого в нормативной выборке. Например, квартильная оценка 3 и процентильная оценка 75 указывают, что более высокую тестовую оценку могут иметь только 25% испытуемых.

Квантильная шкала является равноинтервальной только относительно накопленной частоты. В этом смысле квантильная стандартизация является методом стандартизации, как бы противоположным методу, предложенному Р.Б.Кэттеллом. В методе стандартизации Р.Б. Кэттелла равноинтервальная шкала строится на оси абсцисс, а в методе квантильной стандартизации — на оси ординат, т.е. метод Р.Б. Кэттелла группирует тестовые оценки, а квантильный метод группирует испытуемых.

Таким образом, в процессе подготовки тестов к практическому использованию тестовые результаты претерпевают три вида преобразований: приведение к нормальному виду, приведение к стандартной форме и квантильная группировка. Эти три вида преобразований следует рассматривать не как самостоятельные и независимые процедуры, а как последовательность шагов представления результатов тестирования в виде, удобном для осмысления и интерпретации.

3.4. Методика конструирования психодиагностических тестов

При конструировании теста психолог должен стремиться сделать тест удовлетворяющим основным психометрическим требованиям: валидности, надежности, стандартизованности. Специализированная методика конструирования психодиагностических тестов включает следующие этапы:

- 1) формирование информационной базы исследования;
- 2) конструирование интегральных показателей;
- 3) классификация испытуемых на однородные группы.

Формирование информационной базы исследования

Создать информационную базу для конструирования теста — значит отобрать систему признаков, построить на их основе “черновой” вариант теста и с его помощью провести выборочное обследование.

Наиболее сложной и трудно формализуемой частью работы является формирование исходной системы признаков. Это связано с тем, что, когда перед психологом впервые встает задача разработки теста, предмет измерения определяется в довольно общем виде, например: разработать тест для измерения скорости обучения, переноса навыков, социальной адаптации, психической устойчивости, умственного развития и т.д. Поэтому первым шагом должна стать **разработка содержания понятий**. Представляется, что здесь не может быть общих рецептов; необходимо прежде всего обсуждение и сопоставление различных точек зрения и выработка единой концепции. Лучшим основанием для содержательной разработки предмета тестирования является наличие хорошо развитой психологической теории, однако в социальных науках это явление довольно редкое. Поэтому конечным результатом содержательного анализа должно быть четкое вербальное определение анализируемого понятия и расчленение его на основные части. Например, успешность учебной деятельности определяется успеваемостью, дисциплинированностью, участием в общественной жизни школы, степенью социального признания детским коллективом и т.п.

Соревновательная надежность спортсмена может быть расчленена на устойчивость к сбивающим факторам, умение мобилизовать себя на борьбу с равным или более сильным противником, уровень технико-тактической готовности и т.п.

Естественно, что подобные описания не являются полными, но они должны

удовлетворять первичным требованиям к определению предмета измерения, уточняя и конкретизируя содержание измеряемого свойства.

После определения и разделения измеряемого понятия на составные части следует перейти к поиску системы признаков, адекватно отражающих основное содержание анализируемого понятия.

Разрабатываемая система признаков должна удовлетворять следующим общим требованиям:

1) **полнота описания**, т.е. система признаков должна охватывать все выделенные аспекты измеряемого понятия;

2) **экономность описания**. Это требование находится в диалектическом противоречии с первым. Система признаков должна быть не только разносторонней, но и экономной. Требование экономности диктуется временными ограничениями, трудностями сбора, обработки и анализа чрезмерно больших по объему массивов данных;

3) **четкая структурированность системы признаков**. Признаки должны группироваться, относительно равномерно описывая все ранее выделенные стороны измеряемого явления;

4) **количественная определенность отбираемых признаков**. В связи с последующей математической обработкой все признаки должны иметь количественное выражение в виде чисел или рангов.

Исходя из вышеназванных требований должен быть составлен список заданий и сконструирован начальный вариант теста. Соблюдение этих требований обеспечивает его содержательную валидность. Иными словами, содержательная валидность должна рассматриваться лишь как начальный этап в конструировании теста.

Построенный на первом этапе черновой вариант теста позволяет осуществить сбор численных данных для продолжения исследований.

Конструирование интегральных показателей (шкал)

Важным этапом в разработке теста является формирование обобщающих показателей. На этом этапе должны быть решены две взаимосвязанные подзадачи:

во-первых, необходимо оценить, насколько удачно исходная система признаков описывает содержание измеряемого свойства. Признаки, ошибочно включенные в тест, должны быть исключены;

во-вторых, на базе отобранных признаков следует построить обобщающий показатель.

Понятие обобщающего или интегрального показателя базируется на представлении о том, что все многообразие оценок по отдельным заданиям теста может быть сведено в единое понятие, которое можно охарактеризовать одномерным континуумом, т.е. осью или шкалой. Будучи по своей природе одномерной величиной, интегральный показатель должен в то же время наиболее полно отражать разнообразие испытуемых по измеряемому свойству.

Эти две подзадачи решаются по-разному при наличии или отсутствии внешнего критерия.

При наличии в экспериментальных данных внешнего критерия, независимо и достоверно измеряющего исследуемое свойство, интегральный показатель может быть построен с помощью метода наименьших квадратов. Для этой цели применяется множественный регрессионный анализ.

Мерой валидности, совпадения оценок теста с оценками по внешнему критерию служит коэффициент множественной корреляции (R). Используя распределение Фишера, можно проверить значимость коэффициента множественной корреляции. Если гипотеза о значимости коэффициента множественной корреляции отклоняется, то разрабатываемый тест не может

считаться валидным относительно выбранного внешнего критерия.

Квадрат коэффициента множественной корреляции (R^2), или коэффициент множественной детерминации, показывает, какая часть дисперсии внешнего критерия может быть объяснена с помощью показателей теста, а какая часть остается необъясненной. Причем каждый тестовый признак вносит свой вклад в объясняемую часть дисперсии внешнего критерия. Без большой ошибки из теста могут быть исключены все признаки, вносящие незначительный вклад в объясняемую часть дисперсии. В регрессионном анализе существуют специфические статистические методы отбора переменных, используя которые можно построить уравнение для вычисления интегрального показателя, включающего только наиболее значимые для оценки измеряемого свойства признаки.

В практике разработки тестов более типичной является ситуация, когда необходимо построить интегральный показатель без опоры на внешний критерий. Особенно важно это на начальных этапах измерения какого-либо психологического свойства, когда независимые эталонные способы измерения полностью отсутствуют.

Для построения интегрального показателя при отсутствии внешнего критерия можно воспользоваться методами факторного анализа. В этом случае обобщенный показатель ищется в виде линейной функции от рассматриваемых признаков теста.

Набор коэффициентов называется **весами** и подбирается таким образом, чтобы дисперсия значений интегрального показателя по всем испытуемым была максимальной. Полученный в этом случае интегральный показатель будет представлять собой так называемую **первую главную компоненту**, которая дает максимальный вклад в общую дисперсию.

Между двумя описанными методами построения интегральных показателей существуют довольно серьезные различия.

При использовании множественного регрессионного анализа обычно стремятся получить оптимальную оценку известного внешнего критерия. В противоположность этому при использовании факторного анализа пытаются построить новую, фиктивную переменную, влиянием которой объясняются основные изменения признаков, входящих в тест. Поэтому регрессионная модель интегрального показателя представляет собой линейную комбинацию признаков теста, максимально коррелирующих с внешним критерием и минимально друг с другом, а факторная модель, наоборот, является линейной комбинацией признаков теста, максимально коррелирующих между собой. В связи с этим интегральный показатель, построенный методами регрессионного анализа, будет обладать высокой эмпирической валидностью и низкой гомогенностью, а построенный методами факторного анализа — высокой гомогенностью и конструктивной валидностью.

Для повышения гомогенности факторного интегрального показателя его расчет ведется по следующей схеме. Вначале на основе собранного эмпирического материала методами факторного анализа строится интегральный показатель, обобщающий информацию, содержащуюся во всех заданиях теста. Если оказывается, что все признаки вошли в формулу для вычисления значения интегрального показателя с большими весами, то на этом работа заканчивается. Однако так бывает редко; в большинстве случаев среди признаков теста присутствуют признаки с малыми весами, которые без ущерба для существования дела могут быть отброшены. После сокращения исходного списка признаков делается повторный расчет для получения нового интегрального показателя. При этом может оказаться, что часть признаков опять будет иметь низкие нагрузки, так что потребуются дальнейшее сокращение и дополнительные расчеты.

Итогом такой работы будут список информативных признаков и формула

для вычисления по этим признакам значения интегрального показателя (фактора). Эта формула позволяет произвести математически обоснованную замену множества отдельных оценок одной интегральной оценкой. Причем интегральные оценки получаются в центронормированном виде, когда среднее арифметическое значение равно нулю, а дисперсия — единице. Это означает, что нулевое значение фактора соответствует среднегрупповому уровню, а отличие от нуля показывает отклонение от среднего в долях дисперсии. Другими словами, положительные значения интегрального показателя у испытуемого показывают относительный уровень превышения значения соответствующих качеств от среднего уровня в экспериментальной выборке, а отрицательные — относительный уровень снижения от среднего значения в выборке.

Интегральные оценки дают возможность выделить из однородной группы испытуемых лиц, отличающихся экстремальными или близкими к экстремальным значениям уровнями психических свойств. Эти испытуемые должны привлечь к себе внимание психологов, педагогов и администраторов.

Классификация испытуемых на однородные группы

Третий этап работы заключается в построении одномерных классификаций испытуемых. Эта задача возникает в связи с необходимостью сравнения испытуемых по уровню выраженности психических свойств и качеств.

Как определить, какие различия можно считать существенными, а какими можно пренебречь? Что считать высоким, средним или низким уровнем развития анализируемого качества? Ответить на эти вопросы можно, если разделить испытуемых на группы, существенно различающиеся по степени выраженности соответствующего интегрального показателя. Такая группировка испытуемых может быть осуществлена **методами автоматической классификации**, предназначенными для разбиения объектов на фиксированное число классов.

С помощью алгоритмов автоматической классификации ось каждого интегрального показателя разрезается на требуемое число упорядоченных интервалов. Алгоритмы автоматической классификации предназначены для выделения фиксированного числа классов, которое задается исследователем исходя из содержательных соображений и вида распределения значений интегрального показателя. Ось каждого интегрального показателя разбивается на 10 классов. Число классов (10) выбрано произвольно и оно при необходимости может быть изменено. При выборе такого числа классов мы руководствовались следующими двумя соображениями:

1) 10 классов вводят достаточно высокую дифференциацию между испытуемыми. Эмпирические результаты показывают, что при выделении 10 классов обычно 98—99% различий между испытуемыми приходится на различия между классами и только 1—2% - на различия испытуемых внутри классов;

2) в результате обучения десятичная система представляется для всех наиболее удобной и привычной.

После классификации каждому испытуемому может быть приписано число в соответствии с номером класса, в который он был отнесен по индивидуальному значению интегрального показателя. Так как классы упорядочены по оси интегрального показателя, номера классов могут быть приняты за ранговые оценки соответствующей ранговой шкалы. При таком ранжировании классы 8-10 включают лиц с высокими, а классы 1-3 – лиц с низкими значениями интегрального показателя.

От одномерных классификаций легко перейти к многомерной классификации испытуемых. С этой целью все выделенные интегральные показатели объединяются на одном графике в виде специфических “профилей”. Такой способ представления результатов психологического обследования значительно

облегчает восприятие и интерпретацию данных.

Из приведенного обзора видно, что уровень разработанности статистических аспектов тестологии достаточно высок. В большей степени это относится к этапам конструирования и апробации тестовых испытаний. Техника нормировки, стандартизации и определения надежности с формально-математической точки зрения безупречна. Исключение в этом отношении составляет только проблема определения валидности теста (особенно двух ее видов: по связи с внешними критериями деятельности и прогностическая валидность). Эта проблема появилась с момента зарождения тестологии и имеет общеметодологический научно-теоретический характер. Именно этот недостаток психологического тестирования, не устраненный и по настоящее время, часто сводит на нет результативность исследований, опирающихся на метод тестов.

ЗАДАНИЯ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ

- 1. Составьте логическую схему базы знаний по теме курса.**

2. Раскройте один из видов надежности или валидности теста (при выполнении этого задания студент имеет право выбора).

3. Выявите ошибки в предложениях:

а) Объективный подход - подход, при котором диагностика осуществляется на основе сведений, сообщаемых о себе, самоописания особенностей личности, поведения в тех или иных ситуациях.

б) Надежность - комплексная характеристика методики (теста), включающая сведения об области исследуемых явлений и репрезентативности диагностической процедуры по отношению к ним.

в) Валидность - характеристика методики, отражающая точность психодиагностических измерений, а также устойчивость результатов теста к действию посторонних случайных факторов.

4. Стрелок А выбивает 75 очков из 100, а стрелок В – только 67. Какова надежность стрелка А и стрелка В?

5. Поставьте в соответствие понятия и определения.

Понятия	Определения
Тесты интеллекта	— любая реальность, наблюдаемые изменения которой по конкретным показателям методики могут быть зафиксированы и измерены в какой-либо шкале.
Психологическая диагностика	— тип психодиагностических методик, направленных на оценку достигнутого уровня развития способностей, навыков и знаний.
Переменная	— тип психодиагностических методик, направленных на оценку общих способностей.
Тесты достижений	— область психологической науки, разрабатывающая методы измерения индивидуально-психологических особенностей личности, методы классификации и ранжирования людей по психологическим и психофизиологическим признакам.

6. Постройте классификацию видов надежности по методам их оценки.

- а)
- б)
- в)

ТРЕНИНГ УМЕНИЙ

1. Пример выполнения упражнений тренинга на умение № 1

Задание

В исследовании, моделирующем деятельность диспетчера морского порта, группа испытуемых мореходного училища проходила обследование с применением теста достижений. Испытуемые должны были решать задачи по выбору оптимального режима погрузочно-разгрузочных работ. В результате получены следующие данные:

		1-я часть теста	2-я часть теста
1	Т.А.	25	25
2	П.А.	23	28
3	Ч.И.	27	29
4	Ц.А.	19	26
5	С.А.	16	27
6	К.Е.	26	21
7	К.А.	18	23
8	Б.Л.	24	19
9	И.А.	22	26
10	Ф.В.	18	20

Проведите расчет индекса надежности частей теста $R(X,T)$.

Решение

Предварительно заполните таблицу, подобрав к каждому алгоритму конкретное соответствие из данной ситуации.

№ п/п	Алгоритмы	Конкретное соответствие данной ситуации предложенному алгоритму
1	Расщепление теста на две эквивалентные части	Отбор в одну часть теста результатов четных заданий, а в другую — нечетных.
2	Расчет коэффициента корреляции двух частей теста $R_{1,2}$	Вычисление коэффициента ранговой корреляции Спирмена (см. тренинг умений в юните 6): $R_{1,2} = 0,52$
3	Вычисление индекса надежности частей теста $R(X,T)$	Расчет индекса надежности частей теста по формуле Спирмена-Брауна: $R(X,T) = \frac{2R_{1,2}}{1 + R_{1,2}} = 0,68$

Решите самостоятельно следующие ситуации:

Ситуация 1.1

В исследовании, моделирующем деятельность диспетчера автовокзала, группа испытуемых проходила обследование с применением теста достижений.

Испытуемые должны были решать задачи по выбору оптимального расписания движения автобусов. В результате вычисления коэффициента корреляции двух частей теста $R_{1,2} = 0,27$. Проведите расчет индекса надежности частей теста $R(X,T)$.

Ситуация 1.2

В исследовании, моделирующем деятельность диспетчера железнодорожной станции, группа испытуемых проходила обследование с применением теста достижений. Испытуемые должны были решать задачи по выбору оптимального перемещения товарных и пассажирских составов. В результате вычисления коэффициента корреляции двух частей теста $R_{1,2} = 0,39$. Проведите расчет индекса надежности частей теста $R(X,T)$.

Ситуация 1.3

В исследовании, моделирующем деятельность диспетчера электростанции, группа испытуемых проходила обследование с применением теста достижений. Испытуемые должны были решать задачи по выбору оптимального режима снабжения электричеством различных объектов. В результате вычисления коэффициента корреляции двух частей теста $R_{1,2} = 0,74$. Проведите расчет индекса надежности частей теста $R(X,T)$.

ПСИХОЛОГИЧЕСКИЙ ПРАКТИКУМ

ЮНИТА 8

ПСИХОДИАГНОСТИКА: ВИДЫ НАДЕЖНОСТИ И ВАЛИДНОСТИ ПРОЦЕДУР

Редактор Н.М.Пилипенко
Оператор компьютерной верстки А.Б. Кондратьева

Изд. лиц. ЛР № 071765 от 07.12.1998

Сдано в печать

НОУ "Современный Гуманитарный Институт"

Тираж

Заказ

Современный Гуманитарный Университет

ГЛОССАРИЙ

№ п/п	Новое понятие	Определение
1	2	3
1	Психологическая диагностика	область психологической науки, разрабатывающая методы распознавания и измерения индивидуально-психологических особенностей личности, методы ранжирования и классификации людей по психологическим и психофизиологическим признакам
2	Переменная	любая реальность, наблюдаемые изменения которой по конкретным показателям методики могут быть зафиксированы и измерены в какой-либо шкале
3	Латентная переменная	гипотетическая переменная, которая не поддается измерению в исследовании, но в модели связи между переменными характеризует сумму всех неучтенных влияний на измеряемую переменную
4	Вербальные тесты	тип тестов, в которых материал тестовых заданий представлен в вербальной (словесной) форме
5	Объективный подход	подход, при котором диагностика осуществляется на основе успешности и способа выполнения деятельности
6	Субъективный подход	подход, при котором диагностика осуществляется на основе сведений, сообщаемых о себе, самоописания особенностей личности, поведения в тех или иных ситуациях
7	Проективный подход	подход, при котором диагностика осуществляется на основе анализа особенностей взаимодействия с внешне нейтральным материалом, становящимся в силу его известной неопределенности объектом проекции
8	Тест	стандартизированное, часто ограниченное во времени испытание, предназначенное для установления количественных и качественных индивидуально-психологических различий

9	Психодиагностическое обследование	деятельность психолога, конечным результатом которой является психологический диагноз
10	Этапы психодиагностического обследования	сбор данных в соответствии с задачей исследования; обработка и интерпретация полученных данных; вынесение заключения (диагноз и прогноз)
11	Психологический диагноз	конечный результат деятельности психолога, направленной на описание и выяснение сущности индивидуально-психологических особенностей личности с целью оценки их актуального состояния, прогноза дальнейшего развития
12	Тестовая батарея	совокупность групп тестовых заданий (субтестов), объединенных в одну психодиагностическую методику и направленных на измерение различных сторон сложного психологического конструкта
13	Тесты достижений	тип психодиагностических методик, направленных на оценку достигнутого уровня развития способностей, навыков и знаний
14	Тесты интеллекта	тип психодиагностических методик, направленных на оценку общих способностей
15	Бланковые тесты	тип психодиагностических методик, предусматривающих самостоятельную работу испытуемого
16	Групповые тесты	тип методик, предназначенных для одновременного обследования группы испытуемых
17	Аппаратурные тесты	тип методик, требующих применения специального оборудования для проведения исследования или регистрации полученных данных
18	Критериально-ориентированные тесты	тип тестов, предназначенных для определения уровня индивидуальных достижений относительно некоторого критерия на основе логико-функционального анализа содержания заданий
19	Надежность	характеристика методики, отражающая точность психодиагностических измерений, а также устойчивость результатов теста к действию посторонних случайных факторов

1	2	3
20	Ретестовая надежность	характеристика надежности психодиагностической методики путем повторного обследования испытуемых с помощью одного и того же теста
21	Надежность параллельных форм	характеристика надежности психодиагностической методики с помощью взаимозаменяемых форм теста
22	Надежность частей теста	характеристика надежности психодиагностической методики путем анализа устойчивости результатов отдельных совокупностей тестовых задач или единичных заданий теста
23	Валидность	комплексная характеристика методики (теста), включающая сведения об области исследуемых явлений и репрезентативности диагностической процедуры по отношению к ним
24	Содержательная валидность	один из основных типов валидности методики, характеризующий степень репрезентативности содержания заданий теста измеряемой области психических свойств
25	Очевидная валидность	представление о тесте, сфере его применения, результативности и прогностической ценности, которое возникает у испытуемого, не располагающего специальными сведениями о характере использования и целях методики
26	Эмпирическая валидность	совокупность характеристик валидности теста, полученных сравнительным статистическим способом оценивания
27	Критерий валидации	непосредственная и независимая от валидируемого теста мера психического свойства, на исследование которого направлена психодиагностическая методика
28	Коэффициент валидности	статистический показатель эмпирической валидности теста
29	Критериальная валидность	комплекс характеристик, включающий текущую и прогностическую валидность методики и отражающий соответствие диагноза и прогноза определенному кругу критериев измеряемого качества

30	Текущая валидность	характеристика теста, отражающая его способность различать испытуемых на основании диагностического признака, являющегося объектом исследования в данной методике
31	Прогностическая валидность	информация о методике (тесте), характеризующая степень точности и обоснованности суждения о диагностируемом психологическом качестве по его результату спустя определенное время после измерения
32	Инкрементная валидность	один из компонентов критериальной и прогностической валидности теста, отражающий практическую ценность методики при проведении отбора
33	Популяционная валидность	оценка того, насколько правомерно распространять обобщения, полученные на основании использования в исследовании ограниченного числа испытуемых, на более широкие выборки испытуемых
34	Конструктивная валидность	один из основных типов валидности, отражающий степень репрезентации исследуемого психологического конструкта в результатах теста
35	Валидность по возрастной дифференциации	один из компонентов конструктивной валидности, связанный с возрастной динамикой изменений исследуемого качества