

Г. С. ПРЫГИН

**ОСНОВЫ
ПСИХОДИАГНОСТИКИ**

Принципы и методы. История развития.
Основы психометрики

*Рекомендовано Советом по психологии УМО
по классическому университетскому образованию
в качестве учебного пособия для студентов высших
учебных заведений, обучающихся по направлению и
специальности "Психология"*

Москва

УМК "Психология"

2003 г.

УДК 159.95(075.8)

ББК 88я73

П-855

*Рекомендовано Ученым советом Института управления
(г. Набережные Челны)*

Рецензенты:

А. Н. Гусев, доктор психологических наук, профессор;
Л. М. Попов, доктор психологических наук, профессор.

Прыгин Г.С. Введение в психодиагностику: Принципы и методы. История развития. Основы психометрики: Учебное пособие. М.: УМК «Психология», 2003. - 213 с.

Учебное пособие является составной частью общего курса «Психодиагностики». В пособии освещены вопросы организации психодиагностических процедур и требований, предъявляемых к психодиагностическим методикам. Анализ психодиагностики в историческом контексте позволяет лучше понять современное состояние этой науки и дальнейшие пути ее развития.

Особо важное значение для специалиста психолога (вне зависимости от сферы его деятельности) имеет свободное владение психометрическими понятиями и законами конструирования и адаптации тестов. Этой цели посвящен такой раздел пособия, как «Основы психометрики», при изложении которого автор ставил перед собой задачу - сохранив полноту изложения материала, дать его, по возможности, в наиболее доступном виде. Именно поэтому текст пособия достаточно хорошо структурирован (выделены главные, существенные определения и положения).

Для студентов факультетов психологии высших учебных заведений, а также для слушателей курсов получающих второе высшее или дополнительное психологическое образование. Может быть использовано и более широким кругом лиц, интересующихся вопросами применения, конструирования и адаптации тестов.

ISBN 5-93692-052-6

@ Прыгин Г. С., 2000, 2003

@ УМК «Психология», 2000, 2003

Оглавление

№ §	Название	Стр.
	Тема 1: "Психодиагностика: синтез науки и практики "	4
1	Введение	9
2	Содержание и предмет психодиагностики	15
3	Системы классификации диагностических процедур	36
4	Условия тестирования и психодиагностические задачи	39
5	Профессиональные и этические принципы психологического тестирования	
	Тема 2: "История развития психодиагностики как науки"	51
1	Предыстория психодиагностики	54
2	Развитие тестов интеллекта	61
3	Развитие тестов специальных способностей и достижений	67
4	Развитие методов психодиагностики личности	72
5	Психологическая диагностика в России	
	Тема 3: "Основы конструирования психодиагностических методик"	83
1	Статистическая природа тестовых шкал	90
2	Взаимосвязь свойств пунктов теста и видов распределений тестовых баллов	98
3	Понятие нормы. Проблемы стандартизации показателей	119
4	Устойчивость распределения	121
5	Репрезентативность критериальных тестов	126
6	Надежность теста. Теория надежности	154
7	Дискриминативность теста	158
8	Валидность тестов	174
9	Достоверность самоотчета	183
10	Технология создания и адаптации методик	200
	Список литературы	204
	Приложение	

Тема 1: "Психодиагностика: синтез науки и практики "

§1. Введение.

Взаимосвязь психологии и практической деятельности людей очевидна. Одно из многих ее проявлений - непосредственная психологическая помощь различным категориям населения. Например, психолог-специалист, способен оказать эффективную консультативную психологическую помощь родителям в воспитании детей, супругам в ситуации семейного кризиса, детям, у которых нарушен ход нормального развития личности, юношам и девушкам при профессиональном самоопределении, руководителям в формировании производственных групп, стиля руководства, приемов общения.

Для психолога, применяющего психодиагностику, важно не *только* знание предмета диагностики (личностные особенности и состояния; семейные и производственные, внутри и межличностные конфликты и пр.), но и знание закономерностей самого процесса диагностирования, его стратегии и тактики.

Для грамотного проведения психодиагностики от психолога требуется:

- знание основных принципов разработки тестов;
- знание психологических особенностей оцениваемой области поведения;
- принятие на себя ответственности за социальные и этические последствия использования тестов;
- ясное представление о типах имеющихся тестов и методик и возможности их релевантного применения для решения конкретной практической задачи.

Очевидно, что психологическая помощь будет эффективнее тогда, когда субъект сам осознает переживаемое им неблагополучие. Это пере-

живание может быть острым и выражаться в недовольстве собой, окружающими, жизнью в целом, а иногда и в глубокой продолжительной депрессии. В подобных случаях требуется оказание не только консультативной, но и психотерапевтической помощи.

Подобная психотерапевтическая помощь по своей форме и цели может совершаться в виде беседы, дискуссии, игры (т.е. в процессе общения) и бывает направлена на избавление человека от страданий и ликвидацию причин, которые их вызывают.

Следует отметить, что психологическое вмешательство в жизнь человека отличается от медицинского вмешательства по двум существенным аспектам:

1. природа переживаемого неблагополучия кроется не в болезненных процессах, происходящих в организме человека, а в особенностях его личности, специфике жизненной ситуации и характере взаимоотношений с окружающими;

2. обратившийся за помощью объективно не является и субъективно не признает себя больным.

В какой бы форме ни осуществлялась психологическая помощь (в форме психологического консультирования или в форме немедицинской психотерапии), она обладает общей характеристикой - *направленностью на конкретного субъекта*.

Такая индивидуализация основывается на глубоком понимании личности обратившегося за помощью человека, его эмоциональной сферы, установок, субъективной картины мира, структуры взаимоотношений с окружающими. Для такого понимания часто бывает недостаточно одних лишь общепсихологических знаний и интуиции, требуется применение *специальных психодиагностических методов* [15].

Термин «диагностика» образован от известных греческих слов «диа»

и «гнозис» и буквально истолковывается как «различительное познание». В настоящее время он используется не только в психологии и педагогике, но и в медицине, в технике, в других областях науки и общественной практики. Согласно современному общенаучному представлению диагностикой называется распознавание состояния определенного объекта или системы путем быстрой регистрации его существенных параметров и последующего отнесения их к определенной диагностической категории с целью прогноза его поведения и принятия решения о возможностях воздействия на него в желательном направлении. Например, в технике любой агрегат проходит через техническую диагностику - испытывается на особых «тестовых стендах». Это позволяет выявить скрытые дефекты в изделии и устранить их еще до ввода устройства в эксплуатацию.

Соответственно, о психодиагностике мы говорим тогда, когда речь идет об особом рода объектах диагностического познания - о наделенных сознанием конкретных людях.

Например, если рассматривать систему обучения детей по аналогии с системой управления, то психодиагностику следует считать важнейшим средством обеспечения «обратной связи» любого педагогического воздействия.

Все мы в роли пациентов знакомы с диагностикой в медицине. По комплексу признаков-симптомов врач ставит больному диагноз, т.е. относит его заболевание к определенной диагностической категории, для которой, как правило, известно оптимальное лечение - перечень и порядок приема лекарств и других лечебных процедур.

Переход от поверхностных, наблюдаемых, симптомов и признаков к диагностическому выводу в научной психодиагностике требует применения особых методик и процедур - измерительных тестов и качественно-количественных экспертных шкал [16].

В психологическом словаре дается определение психодиагностики (от греч. *psyche* - душа и *diagnosis* - распознавание, определение) как науки и практики постановки психологического диагноза, т. е. выяснения наличия и степени выраженности у человека определенных психологических признаков.

В широком смысле слова, психодиагностика - это наука и практика *постановки психологического диагноза*.

В более узком понимании психодиагностика - это область психологической науки и одновременно важнейшая форма психологической практики, связанная с разработкой и использованием разнообразных методов распознавания индивидуальных психологических особенностей человека [16].

Термин «психодиагностика», распространившийся в психиатрии вслед за появлением в 1921 г. работы Г. Роршаха «Психодиагностика», довольно быстро вышел за пределы медицины. В настоящее время термин «диагноз» понимается как:

- определение у субъекта (или группы субъектов) наличия или отсутствия некоторого признака, или же оценка степени его выраженности в соответствии с имеющимися критериями (нормами);

- распознавание любого отклонения от нормального функционирования или развития у конкретного лица (или группы лиц). Понятие психодиагностики распространилось и на профилактическое обследование индивидов и групп [3, 15, 18].

Можно выделить, по крайней мере, две существенные области использования психодиагностики - научные и практические исследования.

Научные исследования (в том числе с применением психодиагностики) ориентированы на поиск неизвестных закономерностей. В этих исследованиях используются «известные» (определенные по какому-

либо признаку) испытуемые; при этом, как правило, пренебрегают их индивидуальными отличиями.

Практические исследования (точнее говоря - обследования) обладают важной характеристикой, отличающей их от научных исследований. В этих исследованиях именно индивидуальные отличия являются главным объектом изучения; поэтому они ориентированы на поиск известных закономерностей у «неизвестных» обследуемых.

Существует несколько методов сбора информации для решения различных психодиагностических задач:

1. наблюдение за обследуемым в ходе оказания ему психологической помощи (психолог ведет наблюдение за клиентом только во время консультирования и/или психотерапии);

2. наблюдение за обследуемым в реальных условиях его жизни (например, наблюдение за поведением ребенка в школе).

Эти способы дают очень ценные сведения об обследуемом, однако они крайне затратные, не всегда доступны и не дают информацию до начала работы психолога, а только в ходе такой работы.

Поэтому в психодиагностике получил распространение третий метод сбора информации - применение специальных *психодиагностических методик*, которые используются не только в сфере консультирования и психотерапии, но и во всех тех случаях, когда необходимо получить оценку какой-либо психической особенности конкретного индивида или группы. Этот метод обладает следующими особенностями:

1. позволяет собрать диагностическую информацию в достаточно короткие сроки;

2. дает не общую информацию о человеке, а конкретно о тех или иных его особенностях (интеллекте, уровне тревожности, самоотношении, чувстве юмора, наиболее выпуклых личностных чертах и т.п.);

3. информация поступает в виде, позволяющем дать как качественное, так и количественное сравнение обследуемого индивида с другими людьми;

4. информация, получаемая с помощью психодиагностических методик, полезна с точки зрения выбора средств коррекции, прогноза ее эффективности, а также общего прогноза развития, общения, эффективности той или иной деятельности индивида [15].

Контрольные вопросы

§1. Введение.

1. Что понимается под психодиагностикой?
2. В чем разница между практической и научной диагностикой?
3. Чем обусловлено широкое применение в психодиагностике специальных психодиагностических методик?

Литература

- Анастаси А.* Психологическое тестирование: В 2 кн. М., 1982.
- Бурлачук Л.Ф., Морозов С.М.* Словарь-справочник по психодиагностике. СПб., 1999.
- Общая психодиагностика / Под ред. А.А.Бодалева, В.В.Столина. М., 1987.
- Основы психодиагностики / Под ред. А.Г.Шмелева. Ростов-на-Дону, 1996.

§2. Содержание и предмет психодиагностики.

Существует большое количество определений психодиагностики. Это связано с тем, что диагностика применяется во многих областях знаний и с разными целями. При широком понимании психодиагностики в ней можно выделить четыре компонента [22]:

1. Теорию тестов или психометрию.
2. Тесты и измерительные процедуры, созданные в соответствии с теоретически обоснованными конструктами, например, такими, как личностные черты, модели поведения и пр.
3. Теории, описывающие индивидуальные различия, особенности

среды, а также теоретические представления о развитии.

4. Процесс (технология) психодиагностического обследования.

Внутри каждого компонента можно выделить *три уровня представлений о диагностике*:

1. Первый уровень базируется на практических знаниях, т.е. житейских представлениях людей о человеке и его развитии.

2. Второй уровень базируется на связи основных теоретических положений психологической науки (теории индивидуальных различий и теории развития) с диагностикой.

3. Третий уровень - на математическом моделировании важных для психодиагностики психологических феноменов.

Предполагается, что *все эти три уровня взаимосвязаны*, но попытки свести их к какому-то одному, наиболее значимому, несостоятельны, поскольку каждый из них имеет собственный язык описания, собственную логику и способен внести самостоятельный вклад в понимание человеческого поведения.

Выделяя эти компоненты, Ян тер Лаак [22] исходит из следующих основных положений:

1. вышеназванные компоненты психодиагностики образуют единое функциональное целое;

2. противоречия, возникающие при сопоставлении между собой трех уровней анализа (не сводимых друг к другу), имеют важное значение и относятся к сфере содержания и методов психологии.

С нашей точки зрения, говоря о *предмете* психодиагностики, следует различать психодиагностику как самостоятельную научную область и психодиагностику как прикладную дисциплину.

В этом аспекте предметом общей психодиагностики как *самостоятельной научной дисциплины* являются принципы разработки

психодиагностических средств и их конкретное воплощение в диагностических методиках, включая их методологическое и теоретическое обоснование, проверку валидности и надежности и пр. [15].

Если рассматривать психодиагностику как *прикладную дисциплину*, то ее предметом являются умения, общие и специальные способности, особенности психических процессов, состояния, мотивы, потребности, интересы, черты личности и многое другое.

Существует и другая точка зрения [22], которая состоит в том, что психодиагностика не имеет *собственного, отдельного от других психологических дисциплин предмета исследования*. Более того, предполагается, что предмет психодиагностики определяется, исходя из общей методологии психологии. Действительно, «то, что должно быть диагностировано, определяется конкретным предметным содержанием психологических дисциплин, а то, *как это должно быть сделано* [выделено нами. - Г.П.], - общей методологией исследования» [22, с. 25].

Бесспорно, что психодиагностика как научная дисциплина опирается на общепсихологическое знание диагностируемых свойств. Однако, с другой стороны, собственной методологической основой психодиагностики может выступать *психометрика* - наука об измерении индивидуально-психологических различий. Именно психометрика разрабатывает технологию создания конкретных психодиагностических методик (тестов) и определяет методологию обеспечения следующих *научных требований* к ним:

- *надежности* - внутренней согласованности частей теста и воспроизводимости результатов при повторном тестировании;
- *валидности* – отражениям в результатах теста именно того свойства, для диагностики которого он предназначен;
- *дискриминативности* - способности отдельных пунктов (заданий)

теста дифференцировать испытуемых относительно «максимального» и «минимального» результата теста;

- *достоверности* - защищенности теста от влияния на результаты, стремления обследуемого изменить их в желательную для него сторону;

- *репрезентативности*, показывающей, что измеряемый признак распределен в выборке примерно так же, как и в генеральной совокупности (т.е. можно ли нормы, рассчитанные для генеральной совокупности, применить для конкретной выборки).

Следует отметить, что психометрические требования в разной степени приложимы к разным группам тестов: в наибольшей степени к объективным тестам и личностным опросникам; в наименьшей - к проективным техникам [19].

Итак, психодиагностику можно рассматривать и как теоретическую дисциплину, и как сферу практической деятельности психолога. Рассмотрим данное положение несколько подробнее.

Как *теоретическая дисциплина* общая психодиагностика рассматривает правила создания «диагностических умозаключений», а также закономерности вынесения валидных и надежных диагностических суждений, с помощью которых осуществляется переход от признаков определенного психического феномена к подтверждению их наличия (и выраженности) или отсутствия [15].

Любая выделяемая для психодиагностики переменная должна быть не только теоретически определена в соответствующей области психологической науки, но и иметь практическую значимость для решения той или иной научной или прикладной задачи.

Такая связь является обязательным условием успешности разработки диагностической процедуры (иначе диагностика носит «призрачный» характер, т.е. ищутся способы выявления того, чего на самом деле не

существует). С этой точки зрения психодиагностика является не только воплощением теоретических понятий соответствующих дисциплин в конкретных методиках, но и способом проверки истинности психологических концепций.

Например, если предполагается, что не существует психически здорового человека, который не имел бы какой-либо мотивации, то не может существовать валидных методов, которые бы диагностировали отсутствие всякой мотивации у конкретного лица, не являющегося психически больным. Если была бы обнаружена группа психически здоровых людей, у которых отсутствовала всякая мотивация, то это означало бы существенную недоработку самого понятия мотива.

Общая психодиагностика преимущественно связана с общей, социальной и дифференциальной психологией, *частная психодиагностика* - с медицинской, возрастной, консультативной, юридической, военной и другими областями психологии.

Итак, первую из основных частей психодиагностики составляет *теоретическая психология соответствующей предметной области*.

Второй базовой дисциплиной, составляющей фундамент общей психодиагностики (и ее основной частью), является *психометрика* - наука, обосновывающая и разрабатывающая измерительные диагностические методы.

Третье основание психодиагностики - *практические сферы* применения психологического знания; именно в них ставятся психодиагностические задачи и обосновывается выделение комплексных, интегральных переменных, выступающих как предметы психодиагностики.

Например, существуют профессии, в которых чрезвычайно важна стрессоустойчивость - способность сохранять контроль за собственным состоянием и работоспособность в опасных ситуациях. Значимость этой переменной выделена практикой: если бы не существовало про-

фессий, связанных со стрессом, не было бы необходимости ее диагностировать. Однако практика не только показывает важность того или иного диагностируемого качества, свойства, но и позволяет его выделить.

Таким образом, общая психодиагностика как наука образуется пересечением трех областей психологического знания: *предметной области психологии*, изучающей данные явления, *психометрики* - науки об измерении индивидуальных различий в диагностируемых переменных и *практики* использования психологического знания.

Можно сказать, что практическая психодиагностика относится к теоретической так же, как инженерная эксплуатация техники к ее разработке и конструированию. Как и всякая эксплуатация достаточно сложных устройств в реальных «полевых» условиях, практическая психодиагностика предполагает наличие навыков и интуиции, богатого клинического и житейского опыта.

Помимо этого, в практической психодиагностике используются правила применения психодиагностических инструментов, основанные на знании свойств измеряемых переменных и измеряющих инструментов, а также этических и профессиональных норм психодиагностической работы.

Например, практик-психодиагност должен уметь квалифицировать условия проведения обследования и учитывать их при сопоставлении индивидуальных данных с нормативами.

Практическая психодиагностика предполагает также:

- умение поддержать мотивацию клиента на обследование;
- умение оценить состояние обследуемого в целом;
- умение сообщить обследуемому информацию о нем так, чтобы произвольно не нанести ему вреда, и многое другое [15].

В американской научной литературе существует еще один подход к психодиагностике - она рассматривается как *процесс оказания помощи*

людям в решении их проблем.

В этом процессе выделяют следующие составляющие: 1) сбор информации; 2) интерпретация информации; 3) обобщение информации; 4) попытка решения проблемы.

На наш взгляд, не существует принципиального расхождения в определениях психодиагностики, принятых в отечественной и американской психологии, если рассматривать ее именно как *процесс оказания помощи* людям в решении их проблем, как составную часть *практики использования психологического знания*.

Контрольные вопросы для самопроверки

§2. Содержание и предмет психодиагностики

1. Что понимается под термином "диагностика" вообще и "психодиагностика", в частности?
2. Имеет ли психодиагностика, как наука свой предмет изучения (какие точки зрения есть на эту проблему)?
3. Что является методологической основой психодиагностики?
4. Что определяет структуру психодиагностики, как науки?
5. Какие научные требования предъявляются к психодиагностическим методикам?
6. Как рассматривается психодиагностика в американской психологической литературе?

Литература к теме

1. Анастаси А. Психологическое тестирование. В 2-х кн. М., 1982.
2. Бурлачук Л.Ф., Морозов С.М. Словарь-справочник по психодиагностике. - СПб.: Изд-во "Питер", 1999. - 528 с.
3. Общая психодиагностика / Под ред. А. А. Бодалева, В. В. Столина.- М., 1987.
4. Основы психодиагностики / Под ред. А. Г. Шмелева. Ростов-на-Дону., 1996.
5. Ян тер Лаак. Психодиагностика: проблемы содержания и методов. - М.: Изд-во "Институт практической психологии", Воронеж, 1996. - 384с.

§3. Системы классификации диагностических процедур.

Классификация методов психодиагностики призвана облегчить

практическому работнику выбор методики, максимально соответствующей задаче исследования. Поэтому классификация должна отражать связь методов как с диагностируемыми психическими свойствами, так и с практическими задачами, для решения которых эти методы разрабатываются. Именно такие критерии чаще всего определяют выбор диагностируемых свойств, а на их основе и методик, с помощью которых эти свойства можно выявить.

Методом в психодиагностике, как правило, называется широкий класс методик. Если частная методика привязана к решению узкого класса практических задач и направлена на диагностику определенных свойств, то метод определяется общим родством технологических приемов и процедур в проведении всех методик данного типа. Например, проективному методу соответствуют такие частные методики, как: «Тест цветовых предпочтений» (Люшер), «Тест Сонди», «Чернильные пятна» (Роршах) и другие.

Как правило, психодиагностика проводится не только для получения информации о психических свойствах как таковой, а в основном для составления прогноза определенного социально значимого поведения.

Можно выделить один из видов такого поведения - *критериальное поведение*, которое соответствует определенному *социально значимому критериальному* показателю.

Что такое «критериальное поведение»? Поясним это на конкретном примере.

Например, в учебной деятельности критериальное поведение - это успешное обучение, а критериальный показатель - элементарная успеваемость учащегося. Мы предпринимаем диагностику способностей учащихся для того, чтобы спрогнозировать их успеваемость. Когда нас интересует моральная воспитанность школьника или наличие определенных противоправных тенденций в развитии личности, то в качестве критериального поведения выступает соблюдение моральных и дисциплинарных норм в школе, дома, во дворе, а в юридическом смысле - элементарное законопослушное поведение. На производстве критериальный показатель - про-

изводительность труда. В сфере здравоохранения - общий уровень физического здоровья (работоспособность) [16].

Успешное прогнозирование возможно в том случае, *если измеренное психическое свойство находится в причинных отношениях с критериальным поведением*, т.е. буквально служит причиной его появления. Тогда на основе информации об этом свойстве можно предупреждать нежелательное поведение и стараться изменить условия общения и развития субъекта так, чтобы, воздействуя на психическое свойство в желательном направлении, вызвать желательное поведение [16].

В настоящее время не существует единой общепринятой классификации психодиагностических методик, также как и не существует взаимоднозначного соответствия между задачами и методами.

В то же время существует множество методик, которые применяются для решения совершенно определенного вида задач. Классификацию методики с этой точки зрения можно условно назвать *прагматической*.

Однако имеется большой спектр методик, обладающих универсальностью, которые могут применяться для решения различных задач. При использовании таких методик требуется соблюдать определенные технологические требования. С учетом этих требований методики подобного типа можно классифицировать по *операционально-технологическим* признакам.

Кроме вышеуказанных, существует много других оснований для классификации психодиагностических методик, в частности можно различать:

- диагностические методы, основанные на заданиях, которые *предполагают правильный ответ*, либо на заданиях, на которые *правильных ответов не существует*. К первому типу относятся многие тесты интеллекта, тесты специальных способностей, некоторых личностных черт (например, тесты Равена, диагностическая процедура определения «поле-

зависимости-полнезависимости» Уиткина, и др.). Другие диагностические методики состоят из заданий, характеризующихся лишь *частотой и направленностью того или иного ответа*, но не его правильностью. Это большинство личностных опросников (например, тест 16-ти факторов Р. Кеттелла);

- *вербальные и невербальные психодиагностические методики*. Вербальные так или иначе опосредованы речевой активностью обследуемых; задания, содержащиеся в них, обращены к памяти, воображению, системе убеждений в их опосредованной языком форме. Невербальные включают речевую способность испытуемых только в плане понимания инструкций, само же выполнение задания опирается на перцептивные и моторные способности;

- методики также по методическому принципу, на основании которого они разработаны [15]. По этому основанию обычно различают:

1. *Объективные тесты*. Это, прежде всего тесты интеллекта и достижений (на диагностику знаний, умений и навыков).

2. Методики стандартизованного самоотчета, *которые в свою очередь включают:*

- а) тесты-опросники;
- б) открытые опросники, предполагающие последующий контент-анализ;
- в) шкальные техники, построенные по типу семантического дифференциала Ч. Осгуда, и методики классификации;
- г) индивидуально-ориентированные техники типа ролевых пертурбарных решеток.

3. *Проективные техники* (пятна Г.Роршаха, ТАТ и др.).

4. *Диалогические (интерактивные) техники* (беседа, интервью, диагностические игры).

Рассмотрим несколько более подробно каждую из выделенных групп [15].

Объективные тесты. Эту категорию составляет большинство тестов на интеллект и специальные способности. Результаты обрабатываются с помощью ключа, заданного в форме объективного социокультурного норматива, т.е. имеются объективно «правильные» и «неправильные» ответы.

Методики стандартизованного самоотчета. Общим для этой группы методик является использование вербальных способностей испытуемого (например, обращение к его мышлению, воображению, памяти), а также стандартизованная (математическая) процедура обработки результатов. Различаются следующие методики:

А. Тесты-опросники: содержат набор пунктов (вопросов, утверждений), относительно которых испытуемый выносит какие-либо суждения (как правило, используется двух или трех альтернативный выбор ответов). Это такие же, как и объективные, тесты с заданными вариантами ответа на вопрос. Но ключ к каждому пункту в этом случае определяется не социокультурным нормативом, а с помощью особых психометрических процедур (например, таких, как надежность, валидность, стандартизация). Тесты-опросники применяются, как правило, для диагностики личностных черт, установок, ценностных ориентаций, самооценки.

Б. Открытые опросники: не предусматривают стандартизованных ответов; стандартизация обработки достигается путем отнесения произвольных ответов испытуемого к стандартным категориям; как правило, при этом применяется *контент-анализ* (анализ содержания). Эта техника применяется при наличии однозначно зафиксированного каким-либо образом материала наблюдения. Подсчитывается частота появления определенных элементов (фактов, единиц анализа) в материале наблюдения, а затем по соотношению этих частот делаются психологические выводы [20].

Первоначально метод был разработан для социально-психологического анализа газетных текстов, но его принципы можно распространить на любые продукты деятельности (в том числе тексты свободных сочинений учащихся, продукты их творчества и т.п.).

При появлении совершенной техники аудио- и видеозаписи приемы подобного анализа стали применяться к аудио и видеолентам, в которых живое поведение однозначно зафиксировано и может быть многократно воспроизведено как для экспертной оценки независимых экспертов, так и для формально-статистического анализа частоты появления однозначно регистрируемых фактов.

Например, если на видеозаписи урока зафиксировано, что ученик Сидоров три раза выстрелил из рогатки, то частота этого элементарного и однозначно интерпретируемого события является практически таким же объективным «протокольным» фактом, как и ответы испытуемого на объективный тест.

Поэтому для так называемого «кодирования» единиц контент-анализа (подсчет частоты их появления в материале наблюдения) привлекаются исполнители сравнительно невысокой квалификации («кодировщики»). Результаты их работы затем интерпретируются подготовленным экспертами-психологами.

В. Шкальные техники: предполагают оценку тех или иных объектов по выраженности в них качества, заданного шкалой (например: «холодный-горячий», «сильный-слабый»).

В этом случае шкальные оценки выносит сам испытуемый, а не психолог-диагност. Испытуемый оценивает внешние объекты или понятия, а выводы делаются о нем самом. Сдвиг субъективной точки отсчета при таком шкалировании является информативным признаком для вывода о ценностных ориентациях испытуемого.

Методики, где для диагностики используются системы представлений субъекта, следует считать такими же объективными, как и тесты. Ведь в этом случае речь идет о субъективности испытуемого, а не экспериментатора-диагноста. Последний практически не влияет на результаты, и мы получаем картину субъективного мира личности как таковую.

Но если данные методики применяются для диагностики личности (установок поведения и мотивов), то они оказываются технологически близкими к проективным, и в этом случае требуют высокой психологической квалификации.

Примерами проективных методик с использованием шкалирования являются (упоминавшиеся выше) «Тест цветовых предпочтений» (Люшер), «Тест предпочтения фотопортретов» (Сонди), «Тест юмористических фраз» (Болдырева, Шмелев) и другие.

Г. *Индивидуально-ориентированные техники* типа ролевых репертуарных решеток: могут по форме совпадать со шкальными, опросными методами, приближаться к беседе или интервью. Эти техники в наибольшей мере приспособлены для изучения личности конкретного индивида.

Проективные техники основаны на том, что недостаточно структурированный материал, выступающий в качестве «стимула», при соответствующей организации всего эксперимента в целом, *порождает процессы фантазии, воображения*, в которых раскрываются *те или иные характеристики субъекта*. В клинике употребление проективных методик часто основывается на интуиции и теоретической подготовке психодиагноста, которые оказываются необходимыми на этапе интерпретации данных.

Проективные техники - это, как правило, тесты со свободным ответом, в которых инструкция и тестовые стимулы столь неопределенны, что допускают выбор ответа в очень широких пределах. Направление этого выбора определяется структурой субъективного опыта, установками и мо-

тивами испытуемого. Именно это и позволяет использовать проективные методики для диагностики указанных свойств.

Понятие «ответ» следует трактовать максимально широко. Ответами в проективных методиках могут быть целостные рисунки или тексты. Когда тестовый стимул фактически отсутствует, а спонтанная активность испытуемого задана только свободной инструкцией (типа «Напишите сочинение на свободную тему»), то проективный метод сближается с так называемым *методом анализа продуктов деятельности* (подробнее см. далее).

Важно подчеркнуть, что в структуру проективной методики всегда входит более или менее структурированная и формализованная экспертная оценка результатов. Более надежную и валидную информацию с помощью этих методик можно получить только тогда, когда их результаты оценивают несколько независимых экспертов (обладающих как общей психологической квалификацией, так и специальным опытом обучения по данной конкретной методике). Затем эти оценки проверяются на согласованность.

К сожалению [15], в силу низкого уровня методической культуры среди неподготовленных пользователей до сих пор популярны паранаучные проективные тесты, при проведении (и главное - при интерпретации результатов) которых не соблюдены необходимые атрибуты научной технологии. Еще раз подчеркнем, что с научно-технологической точки зрения проективные методики - разновидность метода экспертных оценок и должны удовлетворять всем требованиям этого метода.

При применении проективных методик особенно повышается риск таких артефактов, как *инструментальные ошибки*, вызванные неправильным выполнением заданий испытуемым, и *ошибки психодиагноста*.

Диалогические техники учитывают, что психолог вступает в контакт с обследуемым и достигает наилучших диагностических результатов

за счет такой организации диалога, которая в наибольшей степени соответствует поставленной диагностической задаче.

К диалогическим техникам относят различные виды бесед (интервью), анкетирование.

Психологическая беседа (интервью) - является существенным звеном психологического исследования, когда она спланирована исследователем в соответствии с задачами данного исследования. Этот метод, один из наиболее древних и популярных в паранаучной психологии, очень часто кажется новичкам и дилетантам самым простым, естественным и эффективным. На самом деле он является самым субъективным из всех перечисленных методов. Ведь интервьюер добывает сведения об испытуемом, наблюдая и анализируя его ответы в момент оказания на него воздействия (поскольку вопросы интервьюера следует считать именно воздействием).

В случае непосредственного контактного интервьюирования (не по переписке и не по телефону) на испытуемого воздействуют все коммуникативные особенности интервьюера - не только его слова, но и поза, жесты, мимика, интонации. Результаты раскрытия личностных проблем оказываются совершенно разными в зависимости от того, устанавливается или нет доверительный контакт между интервьюером и интервьюируемым.

Метод психологической беседы чаще используют психологи-консультанты, чем психологи-диагносты. Во всех случаях профессиональные психологи не один год овладевают только каким-то одним видом беседы. Подчеркнем еще раз, что беседа относится к классу интерактивных методов, т.е. методов прямого взаимодействия и в этом качестве всегда сочетает сбор информации с оказанием психолого-педагогического воздействия. Определенной модификацией и формализацией метода беседы можно считать письменное заочное *анкетирование*, хотя сама по себе раз-

работка анкеты предъявляет совершенно иные требования к профессиональным и личным качествам исполнителя, чем проведение живой беседы. В школе анкетные опросы могут успешно проводиться только в старших классах.

Беседа ни в коем случае не может сводиться к простой регистрации непосредственных данных самонаблюдения. Высказывания изучаемого лица должны соотноситься с объективными данными, со всей ситуацией, в которой проходила беседа, и подвергаться опосредованному толкованию.

Вопросы, задаваемые в беседе, могут представлять собой задания, направленные на выявление качественного своеобразия изучаемых процессов. Каждый последующий вопрос должен ставиться с учетом той измененной ситуации, которая создалась в результате ответа испытуемого на предшествующий вопрос.

Будучи плановой, беседа не должна носить шаблонно- стандартного характера; она всегда должна быть максимально индивидуализированной. Следует учитывать, что беседа получает в психологии различное методологическое оформление в зависимости от различия исходных установок исследователя.

Например, Фрейд ввел понятие «психоаналитическая беседа». Ее задача заключалась в том, чтобы привести собеседника к осознанию и преодолению вытесненных из сознания влечений. Другим вариантом беседы является «клиническая беседа» Пиаже. Эта беседа строилась так, чтобы выявить исключительно спонтанные представления ребенка.

Анкетирование (анкеты, опросники). Применяется для опроса по заданной схеме некоторого круга лиц с целью сбора материала для решения определенной психологической проблемы. Эта схема фиксируется в анкете или опросном листе. Сферой применения анкетного метода служат по преимуществу массовые явления.

Ведущим основанием для классификации методик может также яв-

ляться *мера «объективности-субъективности»*. В случае объективности методики, влияние психолога-диагноста на результаты минимально. В случае субъективности методики, результаты, напротив, зависят от его опыта и интуиции. Заметим, что при проведении объективных или субъективных методик психолог-диагност должен выполнять совершенно различные технологические операции.

Следует также отметить, что не существует жесткой границы между объективными и субъективными методиками. Между ними находится ряд промежуточных вариантов методик, обладающих как определенными признаками объективности, так и определенными признаками субъективности. К ним относятся следующие методики [15, 16].

А. Стандартизированное аналитическое наблюдение. В отличие от свободного исследовательского наблюдения, в стандартизированном наблюдении психолог следует строгой методике: знает, какие факты в поведении наблюдаемого объекта регистрировать и как на основании этих фактов производить оценку латентных диагностических переменных.

Б. Включенное наблюдение с последующим рейтинг-шкалированием. Обычно оно применяется в педагогической практике. Очень часто практически невозможно организовать сколько-нибудь продолжительное наблюдение с привлечением независимых наблюдателей, не включенных в педагогический процесс (их рабочее время тоже надо оплачивать и это слишком дорого). Поэтому к вынесению суждений об объектах наблюдения часто привлекаются участники педагогического процесса. В общем случае это могут быть ученики, но, конечно, более квалифицированную информацию может и должен дать сам учитель. Мера выраженности оцениваемого свойства здесь обеспечивается не частотой появления определенного факта (например, определенного поступка) в материале наблюдения, а градацией субъективной оценки по определенной рейтинг-шкале (семи-

балльной, пятибалльной и т.д.).

При интерпретации результатов данного метода, безусловно, следует учитывать, что оценки включенного наблюдателя могут быть гораздо более субъективными, чем оценки независимого наблюдателя, не включенного в процесс взаимодействия с объектом наблюдения. Здесь резко повышается риск ошибок психодиагноста - в силу определенной заинтересованности, которую проявляет включенный наблюдатель. Сократить этот риск возможно только путем выяснения согласованности оценок, полученных от разных независимых наблюдателей.

В. Ролевая игра. Особой разновидностью интерактивных методов, особенно эффективных для диагностики детей, следует считать ролевую игру. Сам психолог, если даже он и не участвует в игре в определенной роли, так или иначе включен в процесс в качестве ведущего (игротехника). Правила игры, распределение ролей и поведение конкретных участников очень часто сразу же оказывают столь серьезное воздействие на субъекта, что не только его психическое состояние, но и установки, тактика поведения и другие долговременные психические свойства трансформируются (корректируются). Таким образом ролевая игра, также как и беседа, является комплексным, коррекционно-диагностическим методом.

Г. Активный (обучающий) эксперимент. С психологической точки зрения данный метод диагностики весьма важно отличать от стандартного теста, так как в этом случае дополнительными тестовыми стимулами служат не вполне стандартизированные частные инструкции, а также обучающие воздействия (помощь), исходящие от экспериментатора.

Этот тип эксперимента широко применяется в области возрастной психологии для определения диапазона отклонения от нормы и перспективы его компенсации (в некоторых случаях это позволяет выявить «зону ближайшего развития»). Это типичная диалоговая техника, требующая от

экспериментатора высокопрофессионального уровня рефлексии (самоучета). Особое значение обучающий эксперимент имеет для диагностики детей с недостатками развития [16].

Кроме указанных выше, существуют и другие частные методы, используемые в психодиагностике, например, патологический метод и метод анализа продуктов творчества.

Патологический метод: изучение патологии, нарушений психической жизни для познания закономерностей нормальной психики, оказало большие услуги психологии. Каждая функция и процесс могут быть изучены и в их патологической форме: восприятия - на галлюцинациях, память - на амнезиях, речь - на афазиях и т.д.

Однако, как ни велико значение психопатологии для психологии, не следует его преувеличивать и механически переносить полученные на патологическом материале результаты на нормальную психику.

Изучение продуктов творчества: занимает в системе психологического исследования существенное место. Этот метод обращается к изучению продуктов деятельности, чтобы по ним косвенно судить о психологических особенностях действующего субъекта и о его деятельности (поэтому его еще иногда называют методом косвенного наблюдения - Рибо).

Этим методом широко пользуются в исторической психологии для изучения психологии людей, живших в другие исторические эпохи и уже недоступных для непосредственного наблюдения или экспериментирования. Очень успешно применяют его и в *детской психологии*, используя продукты детского творчества для психологического изучения личности ребенка.

Этот метод занимал практически доминирующее положение в отечественной психологии (особенно в столице) в 30-70-е гг. XX в. - в период действия запрета на тестирование [16].

Следует сказать еще о двух типах диагностических методик, которые характеризуются *наименьшей вовлеченностью* психолога в процедуру психодиагностики и минимальным влиянием его личности и опыта на результаты психодиагностики.

1. Приборные психофизиологические методики. Диагностически значимые показатели (дыхание, пульс, кожное сопротивление, мышечный тонус и т.п.) регистрируются с помощью приборной регистрации. Но это не сами поведенческие реакции, а их физиологические индикаторы. Это *косвенный вид* диагностики. Эти методики чаще применяются для диагностики функционального состояния.

2. Аппаратурные поведенческие методики. Классическим примером являются хронометрические методики, позволяющие регистрировать скоростные параметры тестируемых реакций и ответов. Диагностируемые параметры буквально считываются с приборной шкалы соответствующего прибора. В некоторых случаях компьютерные игровые тесты также можно отнести к данной категории методик.

С помощью аппаратурных методик диагностируют элементарные психические функции (например, чувство равновесия или психомоторную координацию) и свойства нервной системы. Но не только. Можно привести примеры аппаратной диагностики психологической совместимости и срабатываемости («гомеостат Горбова»). Разновидность аппаратурных методик представляют собой тесты-тренажеры для диагностики определенных профессиональных умений. В них создается имитация реальных условий профессиональной деятельности (пульт водителя или пилота) [15, 16].

Следует отметить, что описанная выше классификация методик в большей мере отражает так называемый *номотетический подход* к описанию личности человека, в противовес *идеографическому подходу*. Рассмотрим несколько подробнее суть этих подходов.

Разделение номотетического и идеографического подходов началось с публикации Виндельбандом (1904) монографии «История и наука». В ней разделяются два направления научного анализа: с естественнонаучной и с гуманитарной ориентацией [22].

Идеографический подход ориентирован на описание и объяснение сложного целого (например, личности испытуемого). Описание должно быть максимально полным и конкретным, единичный элемент представляется при этом как уникальный феномен.

Номотетический подход ориентирован на открытие общих закономерностей, справедливых для любого частного случая.

Виндельбанд считал, что любой объект можно изучать каждым из этих способов. Оба они имеют как свои достоинства, так и недостатки. Основные недостатки идеографического подхода: 1) отсутствие объективности (т.е. полученные с его помощью результаты в определенной степени зависят от теоретической ориентации психодиагноста и его опыта); 2) невозможность открыть общие законы. В рамках идеографического подхода данные о таких показателях групповой статистики, как среднее значение, коэффициенты корреляций и другие статистики оказываются, как правило, бесполезными.

Номотетический подход также подвергался критике: действительно, на его основе можно описывать общие закономерности, однако, зная их, нельзя составить достаточно полное представление о конкретной личности, поскольку каждая личность уникальна [22].

Рассмотрим еще одну проблему, которая часто ставит начинающего психолога в тупик - *проблему объективности психодиагностических методик*.

Каковы критерии объективности и требования, предъявляемые к психодиагностическим методикам? Как отмечает А.Г.Шмелев [16], суще-

ствуется принципиальная разница между научно-обоснованными и паранаучными приемами и методами. Как правило, создатели паранаучных теорий и методик не говорят, что они обращаются к «непознаваемым» связям. Напротив, они часто слишком поспешно объявляют некоторые гипотетические и сомнительные связи «познанными» и «доказанными», в то время как никаких серьезных научных доказательств их идеи не получили*.

*Однако следует отметить, что среди многих сомнительных выдумок в паранаучных теориях могут встречаться остроумные догадки, опережающие научные знания своего времени. Один из известных примеров дает хиромантия. Изученное нейрофизиологией только в XX столетии явление соматотопической проекции (соответствие участков коры головного мозга и определенных зон руки, насыщенных афферентно-сенсорными и эфферентно-моторными нервными окончаниями) дает основание полагать, что между рисунком ладони и мозговой организацией существуют в принципе познаваемые материальные связи.

Профессиональный разработчик научного психологического теста при его конструировании следует *особому, научно обоснованному, технологическому алгоритму*. Этот алгоритм известен и другим специалистам - его коллегам, поэтому они могут всегда проверить обоснованность авторского ключа к тесту. Принципиальная воспроизводимость алгоритма разработки теста и проверки его «ключей» (связывающих тестовые задания и латентные факторы) вводит научную психодиагностику в мир большой науки, подчиняющейся объективному критерию «независимой воспроизводимости».

В чем суть алгоритма создания научных тестов? При создании научных тестов по поводу *каждого ключа к каждому вопросу выдвигается статистическая гипотеза*, которая затем обязательно проверяется экспериментально на большой выборке испытуемых. Здесь применяется точно такая же логика, как и во всех других экспериментальных науках.

Важно подчеркнуть, что в научной психодиагностике этап разработки тестов - это трудоемкий и строгий научный статистический эксперимент. В этом психометрическом эксперименте психологи проверяют, ре-

ально ли данный тест измеряет то, что он предназначен измерять. В ходе подобного эксперимента нужно не только дать возможность выполнить тест большому числу людей (минимум 50-60 человек), но и собрать объективную информацию о них (о том психическом свойстве, которое нужно измерить). Почти всегда эта задача трудна не только методически, но и организационно.

Итак, *кардинальное различие* научного и паранаучного методов состоит в том, что паранаучный метод игнорирует принципы внутренней самокритики и ценность отрицательного результата [16].

Таким образом, критерии объективности научных психодиагностических методик заложены в принципах и методах их разработки.

Профессиональный психолог при создании тестов использует, как правило, *внешние критерии*, не зависящие от его воли и желаний. Результаты ответов испытуемых на каждое отдельное задание теста сравниваются с этими внешними критериями. Те задания, которые дают результаты, совпадающие (или статистически значимо коррелирующие) с критерием, считаются *валидными* (т.е. обоснованными) по данному критерию. Если после отсева невалидных заданий в тесте остаются только валидные задания, тест в целом считается валидным.

Все вышесказанное касалось в основном измерительных, стандартизованных методик психодиагностики - тестов. Несколько иначе обеспечивается объективность при применении экспертных методов.

Например, рассмотрим возможный эпизод со знаменитой методикой чернильных пятен Роршаха. Напомним, что испытуемый, разглядывая эти пятна, дает свободные словесные описания любых фантазий, которые приходят ему в голову. И вот ребенок усматривает в одном из чернильных пятен «страшное чудовище». Говорит ли это о повышенной тревожности как устойчивой личностной черте ребенка?

Дилетант в психодиагностике может посчитать, что так и есть. Но компетентный специалист-эксперт, проанализировавший сотни протоколов методики Роршаха, прочитавший несколько толстых руководств (а только по одной этой методике выпущена в разных странах в

течение прошлого столетия целая «роршахиада» - сотни и тысячи книг и статей), никогда не будет торопиться. Он сравнит этот ответ с ассоциациями, возникшими у ребенка по поводу других пятен Роршаха; с его результатами по другим методикам, включая стандартные тесты; с данными, известными ему от родителей; с материалами собственных наблюдений за ребенком и беседы с ним. И только проанализировав такой комплекс данных, эксперт приходит к осторожному предположению о возможном повышении уровня тревожности (между прочим, минимально необходимый уровень тревожности - нормальная адаптивная черта любого человека - и ребенка, и взрослого).

Как отмечалось ранее, методики, в которых используется определенная, не вполне формализованная процедура анализа свободных фантазий испытуемого, называются *«проективными тестами»* (хотя, в строгом смысле, это, конечно, не тесты, а экспертные методы).

Однако и для экспертных методов в последние годы разработаны статистические процедуры, повышающие их объективность. Например, к ним относится статистическая проверка согласованности независимых экспертных оценок, которая составляет суть так называемого *«метода независимых судей»*.

Несколько профессионально подготовленных людей (не менее трех) независимо друг от друга оценивают какой-то объект (например, видеопленку с записью определенных событий на уроке по определенному набору шкал). Если после обработки их оценок выясняется, что они оказались статистически значимо согласованными, то суммарный результат считается более свободным от субъективизма каждого отдельного эксперта.

Итак, в психодиагностике существуют объективные методики, но только в том случае, когда их разрабатывает и применяет профессионально подготовленный психолог.

Ситуацию с использованием психологических тестов дилетантами можно образно сравнить с использованием кардиографа дикарями, которые просто не знают, куда надо ставить электроды и как расшифровывать

кардиограмму. Всем понятно, что использование кардиографа для диагностики сердца требует предварительных научных знаний о работе последнего, о клетках крови и др. Но далеко не все понимают, что грамотная постановка и интерпретация (трактовка) результатов теста на интеллект требует предварительных научных знаний о природе человеческого интеллекта.

Психологические тесты опасны тем, что, выражая результат в простой количественной форме, они создают у дилетантов иллюзию легкости и простоты использования, подталкивают их к паранаучному стилю работы с тестами.

Итак, объективные тесты существуют. Зависимость результата от профессионализма исполнителя не означает, однако, зависимости от его личностных качеств. Другое дело, что прогноз на основании психологического диагноза по результатам теста возможен сегодня только с вероятностной точностью. Но подобное положение имеет место в любых других, в том числе точных технических науках. Различие заключается лишь в размере вероятностной ошибки прогноза.

Поэтому научное психометрическое обоснование теста обязательно включает расчет и опубликование «стандартной ошибки измерения» для данного теста. Невысокая точность психологических прогнозов не уменьшает их ценности. Ведь выбирать приходится не между точным и приближенным прогнозом, а между приближенным прогнозом и случайным гаданием, а это уже дает иной раз ощутимые выгоды [16].

Рассмотрим основные требования, которые предъявляются к психодиагностическим методикам (учитывая при этом, что к психодиагностическим методикам разного типа рекомендуется предъявлять разные требования [15]).

1. Требования к измерительным методам (тестам):

а) должны быть однозначно сформулированы цели, обозначены предмет и область применения методики, причем предмет (диагностический конструкт) - в теоретических понятиях, релевантных (соответствующих) понятиям определенной системы теоретического знания. Должна быть четко выделена область применения теста - особая социальная среда (или сфера социальной практики), контингент испытуемых; конкретизированы цели использования результатов: для прогноза, для психологического вмешательства, для принятия определенного типа решений (правовых, административных и пр.) и т.д.;

б) процедура проведения должна быть задана в виде однозначного алгоритма, с которым может быть ознакомлен лаборант, не имеющий специальных психологических знаний; он может быть также введен в ЭВМ, которая используется для предъявления заданий и анализа ответов;

в) процедура обработки должна включать статистически обоснованные методы подсчета и стандартизации тестового балла (по статистическим или критериальным тестовым нормам). Выводы (диагностические суждения) на основе тестового балла должны сопровождаться указанием на вероятностный уровень их статистической достоверности;

г) тестовые шкалы должны быть проверены на репрезентативность, надежность и валидность в заданной области применения. Другие разработчики и квалифицированные пользователи должны иметь возможность повторить стандартизационные исследования в своей области и разработать частные стандарты (нормы);

д) процедуры, основанные на самоотчете, должны снабжаться средствами контроля за достоверностью, позволяющими автоматически отсеивать недостоверные протоколы;

е) головная методическая организация (определенного ведомства области применения) должна вести банк данных, собранных по тесту, и

производить периодическую коррекцию всех стандартов методики.

2. Общие требования к экспертным методам:

а) данный пункт повторяет содержание пункта а) для тестов. Дополнение: инструкция по привлечению экспертов содержит указание на их квалификацию; также указывается, сколько требуется экспертов для получения надежных данных по методу независимых оценок;

б) инструкции к проведению должны пройти специальные испытания на однозначность их выполнения экспертами по отношению к некоторому эталонному набору данных (текстов, рисунков, звуковых или видеозаписей и т.п.);

в) процедура обработки результатов должна включать такое документирование промежуточных этапов обработки, которое позволило бы другому эксперту перепроверить конечный результат;

г) пользователи-разработчики должны иметь возможность воспроизвести (повторить) нормативное исследование по измерению экспертной согласованности на эталонном наборе данных;

д) головная организация должна вести банк данных, обеспечивая подготовку и переподготовку (в соответствии с пересмотренными стандартами методики) пользователей.

Любая методика, не удовлетворяющая перечисленным выше требованиям, не может считаться профессиональной психодиагностической методикой.

Методики должны проходить аттестацию в рамках головных методических организаций, в обязанности которых входит составление библиотек аттестованных психодиагностических методик. Инструктивная литература по методикам, не прошедшим аттестацию, не пригодна для применения в практической психологии. Это не исключает возможности применения не аттестованных методик в исследовательских целях

Контрольные вопросы для самопроверки

§3. Системы классификации диагностических процедур

1. Что называется "методом" в психодиагностике?
2. Для чего необходима классификация методов психодиагностики?
3. Какие основания можно использовать для построения классификации методов психодиагностики?
4. Что понимается под термином "объективный тест"?
5. Что понимается под "стандартизованными самоотчетами"?
6. Что понимается под "проективными техниками"?
7. Какие существуют основные виды проективных техник?
8. Что такое "инструментальная ошибка" и чем она отличается от "ошибки психодиагноста"?
9. В чем суть патологического метода психодиагностики?
10. Какие существуют другие методы психодиагностики?
11. В чем разница между "номотетическим" и "идеографическим" методами описания личности?
12. Каковы критерии объективности психодиагностических методик?
13. Каким требованиям должны удовлетворять тесты (измерительные методы)?
14. Каким требованиям должны удовлетворять экспертные методы?

Литература к теме

1. Анастаси А. Психологическое тестирование. В 2-х кн. М., 1982.
2. Бурлачук Л.Ф., Морозов С.М. Словарь-справочник по психодиагностике. - СПб.: Изд-во "Питер", 1999. - 528 с.
3. Общая психодиагностика / Под ред. А. А. Бодалева, В. В. Столина.- М., 1987.
4. Основы психодиагностики / Под ред. А. Г. Шмелева. Ростов-на-Дону., 1996.

§4. Условия тестирования и психодиагностические задачи.

Выше мы отмечали, что психолог, планирующий провести диагностическое обследование, должен хорошо представлять себе задачу, стоящую перед ним, поскольку задача по существу определяет выбор методов диагностики. Однако на этот выбор влияет не только психодиагностическая задача, но и *условия*, в которых будет проводиться обследование.

В связи с этим, принято различать не только *задачи*, но и *ситуации* (условия) психодиагностики. Прежде всего, это *различение "ситуации клиента" и "ситуации экспертизы"*.

В первой ситуации человек обращается за помощью к психологу: он охотно идет на сотрудничество, старается выполнить инструкции как можно более точно, не имеет сознательных намерений исказить результаты обследования, например, приукрасить себя.

Во второй ситуации человек знает, что подвергается экспертизе, старается выдержать "экзамен", а для этого вполне осознанно контролирует свое поведение и свои ответы так, чтобы выглядеть в максимально выгодном свете.

В ситуации клиента к диагностическому инструменту можно предъявлять гораздо менее жесткие требования относительно его защищенности от фальсификации вследствие сознательной стратегии, чем в ситуации экспертизы.

Психодиагностические задачи можно различать также с точки зрения того, *кто и как* будет использовать диагностические данные и *какова ответственность* психодиагноста в выборе способов вмешательства в ситуацию обследуемого. Опишем кратко эти наиболее типичные ситуации [15].

1. *Данные используются специалистом-смежником для постановки непсихологического диагноза или формулирования административного решения.* Эта ситуация типична для использования психодиагностических данных в медицине. Психолог выносит суждение о специфических особенностях мышления, памяти, личности больного, а врач ставит медицинский диагноз. Психолог не несет ответственности ни за диагноз, ни за то, какое именно лечение будет проведено больному врачом.

По той же схеме происходит использование психодиагностических

данных при психодиагностике по запросу суда, комплексной психолого-психиатрической экспертизе, психодиагностике профессиональной компетентности работника или профпригодности по запросу администрации.

2. *Данные используются самим психодиагностом для постановки психологического диагноза, хотя вмешательство в ситуацию обследуемого осуществляется специалистом другого профиля.* Например, такова ситуация психодиагностики применительно к поиску причин школьной неуспеваемости: диагноз имеет психологический (или психолого-педагогический) характер, а работу по его реализации в жизнь проводят учителя, родители, другие воспитатели.

3. *Данные используются самим психодиагностом для постановки психологического диагноза, а последний служит ему основанием (или основанием для действий его коллеги-психолога) для разработки путей психологического воздействия.* Такова ситуация психодиагностики в условиях психологической консультации.

4. *Диагностические данные используются самим обследуемым в целях саморазвития, коррекции поведения и т. п.* В этой ситуации психолог несет ответственность за корректность данных, за этические, деонтологические аспекты "диагноза" и лишь частичную ответственность - за то, как этот диагноз будет использован клиентом.

Хотя и не существует жесткого соответствия между характером задачи и психодиагностическим методом, все же отмечается некоторая предпочитаемость тех или иных методов в тех или иных случаях [15].

Например, в ситуациях 1 и 2 методы должны давать "стратегическую" информацию о клиенте, т. е. обеспечивать более или менее долгосрочный прогноз, они также должны позволять соотнесение обследуемого с другими людьми, т. е. предполагать стандартизацию. Поэтому в данных ситуациях наибольшее употребление получили объективные тесты и тес-

ты-опросники, причем последние иногда основаны не на психологических категориях, а на категориях (системе понятий) заказчика. Таковы, например, известный Миннесотский многофакторный личностный опросник (ММРІ) и его модификации.

В ситуации 3 информация зачастую рассчитана на регулирование тактики практической работы самого психолога, соотнесение с "нормой" имеет меньшее значение, поэтому чаще используются идеографические техники, проективные и диалогические методы.

В ситуации 4 главное требование к методам - легкость перевода получаемых с их помощью данных на язык самого обследуемого. Этому условию удовлетворяет, например, тест 16-ти факторов Р. Кеттелла, но плохо соответствует ММРІ, диагностические описания которого рассчитаны на психиатра [15].

§5. Профессиональные и этические принципы психологического тестирования.

Также как и медицинское обследование, любое психодиагностическое обследование может существенно повлиять на судьбу человека. Поэтому во всех странах, где психодиагностическая практика широко развивается, она регламентируется как общим профессионально-этическим кодексом, так и специально разработанными профессионально-этическими стандартами в области тестирования [16].

Любой профессиональный психолог должен строго соблюдать профессионально-этические принципы психодиагностики. Воссозданное в 1994 году Российское психологическое общество подготавливает совместно с отраслевыми министерствами и ведомствами (прежде всего с министерствами образования здравоохранения РФ) серьезные законопроекты, регламентирующие доступ к психодиагностической инфор-

мации и порядок использования психологических тестов в стране.

Ведомственные организационные нормативы и правила проведения психодиагностики отражены, в частности, в "Положении о службе практической психологии образования в Российской Федерации" (Вестник образования, 1995, N7).

Любой пользователь должен знать общие принципы, которым следует его коллега психолог в области соблюдения профессионально-этических норм в психодиагностике. Без соответствующих знаний он не сможет понять, почему в некоторых случаях психолог просто не имеет права доверять кому-либо не только формулирование заключения по тесту, но и сбор первичной информации и должен проводить все психодиагностическое обследование от начала до конца самостоятельно.

Именно психолог несет всю полноту ответственности за возможный моральный и косвенный материальный ущерб, в том числе ущерб для соматического и нервно-психического здоровья, который может быть причинен испытуемому при неправильном проведении обследования.

Общие требования к профессиональной этике можно разделить на [14]:

- требования к разработчикам тестов;
- требования к психологу-пользователю;
- требования к специалистам (не являющимся психологами).

Изложим кратко основные профессионально-этические принципы в том виде, в каком они распространяются на всех пользователей сертифицированных психодиагностических методик. Это изложение дополнено необходимыми комментариями, поясняющими, как должно строиться взаимодействие пользователей с психологами в связи с тем или иным изложенным здесь принципом [16].

1. *Принцип специальной подготовки и аттестации лиц, использующих психодиагностические методики.*

Требование, чтобы диагностические методики использовались только достаточно квалифицированными пользователями, является первым шагом по защите индивида от их неправильного использования. Конечно, необходимая квалификация меняется в зависимости от типа диагностической методики.

Например, для правильного применения индивидуальных тестов интеллекта и большинства личностных тестов и опросников и особенно проективных техник требуется относительно длинный период интенсивного обучения. В то время как для тестирования достижений в учебной и профессиональной деятельности нужна *минимальная* специальная подготовка.

Психолог-диагност должен выбирать методики, которые подходят как для *конкретной цели*, с которой он проводит обследование, так и для *конкретного человека*, которого он обследует. Психолог также должен знать соответствующую научную литературу, касающуюся выбранной методики, и быть способным оценить технические параметры таких ее характеристик, как репрезентативность норм, надежность, валидность, достоверность.

Известно, что результаты диагностики «чувствительны» к большинству условий ее проведения. Психодиагност в подписываемом им заключении должен зафиксировать, что все необходимые методические условия были соблюдены. Психодиагност делает выводы или дает рекомендации только после рассмотрения диагностической информации в свете другой касающейся индивида информации.

В тех случаях, когда результаты стандартизированных методик явно противоречат другим источникам информации о человеке, психолог должен взять на себя ответственность за то, чтобы признать результаты проведенного обследования недостоверными, но обосновать это свое заклю-

чение указанием на объективные факты, свидетельствующие о недостоверности этих результатов.

Психолог должен знать границы своей компетенции и ограниченность используемых методов и не предлагать свои услуги, а также не использовать методики, которые не удовлетворяют профессиональным стандартам, установленным в отдельных областях практики и для определенных категорий методик.

В большинстве развитых стран получения диплома о высшем психологическом образовании еще недостаточно для получения права на психологическую практику. Для получения соответствующей лицензии кандидат проходит процедуру *аттестации*, включающую специальные экзамены и анализ опыта работы по использованию психодиагностических методик.

Такого рода процедуры аттестации с использованием стандартной системы требований к любым исполнителям, предусматриваются в работе соответствующих комиссий Российского психологического общества. Профильное образование в данном случае служит инструментом более адекватной подготовки к подобным профессиональным испытаниям, служащим своего рода вступительным экзаменом для новых членов психологического общества.

Таким образом, для того, чтобы быть свободным от ограничений в доступе к психодиагностическим методикам и информации, пользователь должен:

- пройти соответствующую психологическую подготовку;
- получить диплом государственного образца;
- пройти процедуру аттестации и стать членом Российского психологического общества.

И уже в этом качестве получить право на приобретение и исполь-

зование любых методик, распространяемых специализированными предприятиями и изданиями общества среди членов общества, а также федеральными и региональными центрами службы практической психологии образования.

Пользователь, не прошедший процедуру аттестации, должен получать психодиагностическую информацию "из рук" профессионального психолога. В некоторых случаях он может сам провести обследование, но под руководством психолога. Для разъяснения того, как разделяется ответственность в этом случае, сформулируем следующий дополнительный принцип [16]:

1а. Принцип личной ответственности.

Если психолог привлекает к определенным работам по выполнению психодиагностического обследования каких-то исполнителей, не являющихся аттестованными специалистами в области психодиагностики, то вся полнота ответственности за правильность проведения методики и корректность интерпретации и использования ее результатов целиком лежит именно на психологе.

При проведении группового тестирования по разветвленной психодиагностической программе (независимыми исполнителями в разных помещениях в разное время) психолог должен лично завизировать разработанную им программу обследования и лично подписать отчет, содержащий проинтерпретированные им результаты обследования.

2. Принцип ограниченного распространения психодиагностических методик (принцип "профессиональной тайны").

В кратком изложении этот принцип гласит: сертифицированные психологическим обществом профессиональные психодиагностические методики могут распространяться *только среди аттестованных специалистов.*

Этот принцип тесно связан с предыдущим и имеет двоякую цель: неразглашение содержания методик и предупреждение их неправильного применения. Доступ к таким методикам должен быть ограничен теми людьми, кто имеет профессиональную заинтересованность и гарантирует их правильное использование, то есть они должны быть доступны только тем специалистам, которые будут проводить психологическую диагностику и отвечать за достоверность ее результатов.

Специально подчеркнем, что сохранение профессиональной тайны имеет значение не столько для поддержания авторитета и социального статуса психологов, сколько для обеспечения корректной и точной психодиагностики. *Профессиональные психодиагностические методики не должны поступать в открытую продажу.*

Вскоре после создания новой методики ее распространяют только для исследовательских целей. Это условие должно быть ясно указано в руководстве и, соответственно, применение этой методики должно быть ограничено только кругом лиц, участвующих в накоплении банка данных по этой методике и не использующих ее для вынесения отдельных выводов без параллельного использования других надежных методик.

Обязанностью автора и издателя диагностических методов является их проверка и повторная стандартизация, проводимые так часто, как это требуется, чтобы предотвратить их старение.

Быстрота, с которой методика устаревает, очень различна и зависит от ее содержания. Сертифицированные психодиагностические методики и их основные части нельзя печатать в газете, журнале или популярной брошюре ни в целях описания, ни для использования их при самооценке. Исключение составляют устаревшие тесты достижений, задания этих тестов можно публиковать в специальных брошюрах для тренировки учащихся в выполнении подобных тестов.

Исключения, касающиеся тестов достижений и методов экспертной оценки, формулируются следующим образом:

2а. Сертифицированные тесты достижений, разработанные в центрах, аккредитованных органами народного образования, могут использоваться педагогами как для оперативного, так и для аттестационного контроля уровня знаний. Любой педагог может самостоятельно разработать тест достижений, но использовать его он может только для оперативного контроля за уровнем знаний своих учеников.

2б. Методы экспертной оценки личностных качеств учащихся могут использоваться любым педагогом, но их результаты могут иметь применение только в оперативной педагогической работе с данными учащимися и не могут иметь аттестационного значения, то есть передаваться другим лицам и организациям.

3. Принцип обеспечения суверенных прав личности.

Вопрос, возникающий в особенности в связи с использованием личностных тестов, касается посягательства на тайну личности. Это сложный вопрос, поскольку о сущности некоторых тестов, раскрывающих эмоциональные и мотивационные особенности, а также установки личности, человеку не сообщается, и он может проявить такие черты в ходе тестирования, которые для него самого нежелательны, не осознавая, что он это делает.

Основной этический принцип в данном случае гласит: *человек не должен подвергаться какому-либо обследованию обманным путем. Это значит, что человек перед обследованием должен быть предупрежден, кто будет иметь доступ к результатам обследования и какие решения могут быть и приняты.*

В том случае, если обследованию подвергаются несовершеннолетние дети, такое предупреждение должны совершенно официально полу-

чить родители. Поясним здесь, что проинформированный о целях и общем смысле обследования (но, конечно, не о содержании тестовых заданий) испытуемый получает следующие возможности:

а) психологически мобилизоваться для его наилучшего выполнения;

б) отказаться от обследования, осознавая при этом, от каких возможностей он отказывается в случае положительных результатов обследования;

в) не формулировать своего отказа в явной форме, но применить собственную тактику выполнения методики, которая, с его точки зрения, позволяет ему скрыть истинную информацию о себе.

Данный этический принцип, который охраняет право индивида отказываться от участия в обследовании и, следовательно, защищает тайну его личности, осложняет работу психолога и повышает требования к его квалификации. При правильных взаимоотношениях и взаимном уважении психолога и обследуемого число отказов от участия в диагностировании может быть сведено к незначительному числу.

За. Учащиеся проходят сертифицированные тесты достижений в обязательном порядке, если эти тесты утверждены органами народного образования. В качестве возможной формы приема выпускных экзаменов в средних учебных заведениях и вступительных экзаменов в средние специальные и высшие учебные заведения.

4. Принцип объективности.

Обследование должно быть абсолютно беспристрастным. На его исполнителя не должны оказывать влияние общие впечатления о личности обследуемого: симпатии или, наоборот, антипатии, равно как и собственное состояние или настроение.

К сожалению, этот принцип слишком часто нарушается лицами, не имеющими специальной подготовки для проведения психодиагностиче-

ских обследований и не обладающими необходимыми элементами самоуважения в своем профессиональном самосознании и практическими навыками корректного, доброжелательно-нейтрального общения с испытуемым.

5. Принцип конфиденциальности.

Вся информация, получаемая в процессе обследования, должна быть *строго конфиденциальной*: она должна быть доступна только для тех, для кого она предназначена.

В отличие от предыдущей проблемы в данном случае вопрос стоит о том, кому и в какой форме психолог может сообщать информацию по результатам обследования.

Один из ответов на этот вопрос вытекает из принципа 4: можно сообщать информацию тем лицам, о которых был предупрежден испытуемый. В этой связи, перед психодиагностом встают следующие вопросы. Имеет ли право сам испытуемый получать эти сведения или родители обследованного ребенка? Если сообщать кому-либо результаты тестирования, то как это делать?

Психолог должен обеспечить такое преподнесение информации, чтобы родители осознавали личную ответственность за возможные невысокие результаты их ребенка по тестам, а не видели в этом лишь вину ребенка, которого за это "можно и нужно наказывать". Но если ребенок не был предупрежден о том, что результаты по тесту узнают учителя и родители, то психолог вообще не имеет права сообщать их ни тем, ни другим.

Особые ситуации возникают, когда диагностические результаты запрашиваются новыми людьми, как, например, в тех случаях, когда будущий наниматель или колледж просят предоставить им данные тестового обследования индивида. В таких случаях требуется полу-

чить согласие индивида на передачу данных. Это же относится и к обследованию в клинике или консультации, а также к тестированию, осуществленному с исследовательскими целями.

Другая проблема относится к сохранению диагностических данных в учреждениях. В случаях, когда данные получены либо для длительного использования в интересах индивида, либо для научных целей, для предотвращения неправильного их применения от психолога категорически требуется обеспечить строгий контроль за доступом любых лиц к этим данным.

Требование конфиденциальности психологической информации может нарушаться только в тех случаях, когда нераскрытие диагностических данных представляет опасность для обследуемого индивида или для общества.

Применительно к обследованию детей на психолого-педагогических консультациях (комиссиях) в целях выявления недостатков развития следует отметить, что школа получает общее заключение об итогах исследования, но официальные данные о результатах выполнения примененных тестов могут быть переданы только с согласия самого обследованного, его родителей или представляющих его лиц (например, опекунов).

Положение о конфиденциальности диагностической информации введено, в частности, в проект Закона Российской Федерации "Об образовании лиц с ограниченными возможностями здоровья".

б. Принцип психопрофилактического изложения результатов.

В последнее время все шире во многих странах распространяется следующий этический норматив: испытуемый также должен иметь возможность комментировать содержание своих результатов и в случае необходимости разьяснять или исправлять фактическую информацию.

Очевидно, что в этом случае мы получаем, как минимум, *два серьезных производных требования:*

ба. при сообщении результатов самому испытуемому должны быть соблюдены соответствующие меры предосторожности, направленные против их неправильного использования, неверной интерпретации или возможного появления невротически депрессивных реакций или обострения депрессивного состояния (то есть результат должен быть преподнесен по возможности в ободряющем, не травмирующем психику и самооценку контексте);

бб. если испытуемый настаивает на том, чтобы "переделать" тестовое задание, психолог должен иметь наготове "запасной" вариант практически эквивалентной методики (ряд профессиональных методик специально имеют так называемые "параллельные формы"), чтобы убедить самого испытуемого, что ему предоставлен дополнительный шанс, с одной стороны, и что результаты первого тестирования были достаточно достоверны, с другой стороны.

Примечание: в тех случаях, когда школьному психологу заведомо известно, что данные тестирования могут быть использованы родителями или учителями школьника как дополнительный инструмент в конфронтации с ним, (причем, формация носит психотравмирующий характер), то психолог имеет право вообще не сообщать эту информацию.

Во всех своих профессиональных действиях психолог следует главному императиву своего этического кодекса "Не повреди!". И знание об этом должно сделать понятным для всех, почему психолог ограничивает распространение психодиагностической информации: он делает так не ради сохранения этой информации в тайне, а ради благополучия, психического и физического здоровья всех людей [16].

Контрольные вопросы для самопроверки

§4, 5. Условия тестирования и психодиагностические задачи. Профессиональные и этические принципы психологического тестирования

1. Чем принципиальное отличие "ситуации клиента" от "ситуации экспертизы"?

2. Какие типичные ситуации диагностики существуют в зависимости от того, кто будет использовать данные психодиагностики? Дайте их краткую характеристику.

3. Какие виды профессионально-этических принципов выделяются в психологической диагностике? Дайте краткую содержательную характеристику каждому из них.

Литература к теме

1. Анастаси А. Психологическое тестирование. В 2-х кн. М., 1982.

2. Бурлачук Л.Ф., Морозов С.М. Словарь-справочник по психодиагностике. - СПб.: Изд-во "Питер", 1999. - 528 с.

3. Общая психодиагностика / Под ред. А. А. Бодалева, В. В. Столина.- М., 1987.

4. Основы психодиагностики / Под ред. А. Г. Шмелева. Ростов-на-Дону., 1996.

17. Нормы профессиональной этики для разработчиков и пользователей психодиагностических методик. Стандартные требования к психологическим тестам. 2-е издание. Ярославль; НПЦ "Психодиагностика", 1998.

Тема 2: "История развития психодиагностики как науки".

§1. Предыстория психодиагностики.

Психодиагностика, как система способов раскрытия субъективных особенностей психики, с самого начала развивалась в направлении от "внешнего", видимого, наблюдаемого, к "внутреннему", скрытому, подлежащему выявлению. За тысячелетия существования современного человеческого сообщества психодиагностика проделала путь от наивных, ненаучных интерпретаций астрологов и хиромантов до научных методов пренатальной диагностики девиаций в психическом развитии индивида [12].

Возникновение на рубеже XX века психологической диагностики было подготовлено с одной стороны, новыми направлениями в развитии психологии, с другой - запросами практики.

Одним из таких новых направлений стала *экспериментальная психология*, это вполне объяснимо, так как экспериментальный подход лежит в основе большинства психодиагностических методик. Возникновение экспериментальной психологии в 50 - 70-е годы прошлого века связано с возросшим влиянием естествознания на область исследования психических явлений, состоявшем в переводе изучения психических фактов в русло эксперимента и точных методов естественных наук [16].

Начало современного этапа развития психодиагностики как науки, связано с экспериментами по измерению различных проявлений человеческой психики, которые были организованы в конце XIX - начале XX века немецким психологом В. Вундтом и англичанином Ф. Гальтоном.

Именно В. Вундт в 1878 году основал в Германии первую лабораторию экспериментальной психологии. В его лаборатории в основном изучались ощущения и вызываемые ими двигательные акты - реакции, а также периферическое и бинокулярное зрение, цветоощущение и пр. По

образцу лаборатории Вундта создаются подобные экспериментальные лаборатории и кабинеты не только в Германии, но и в других странах (Франции, Голландии, Англии, Швеции, Америке).

В Англии кузен Ч. Дарвина Фрэнсис Гальтон (1883) впервые включил в состав предложенной им новой комплексной науки "антропометрики" особые измерительные испытания не только физических характеристик человека, но также пробы на остроту зрения и слуха, время моторной и словесной ассоциативной реакции и т.п. Ф. Гальтон предложил термин "тест", и с его именем по праву связывается начало уже не предыстории, а собственно истории психодиагностики.

Чуть позже к плеяде первых психодиагностов присоединяются российские психологи Г.И. Россолимо (1910) и А. Рыбаков (предложивший в 1913 году в качестве тестового средства использовать отпечатки чернильных пятен); швейцарец Г. Роршах (который ввел в 1921 году в научный обиход сам термин "психодиагностика", сделав его названием своей книги о проективном тесте на основе "чернильных пятен") и французы А. Бине и С. Симон, создавшие свою версию шкалы измерения интеллекта. С конца двадцатых годов наиболее значительные психодиагностические исследования стали проводиться в США, в лабораториях Стэнфордского и Колумбийского университетов [12, 16].

Таким образом, *первоначально психодиагностика зародилась как наука о методах экспериментальной дифференциальной психологии, изучающая психологические различия между людьми экспериментальным путем.*

Но дифференциально-психологическое изучение человека не было простым логическим следствием развития экспериментального метода. Оно складывалось под воздействием *запросов практики*, сначала медицины и педагогики, а затем и промышленного производства

(сфера индустриальной психологии).

Основоположники и последователи разных психологических школ по-разному отвечали на запросы практики. Рассмотрим коротко, как формировались некоторые важные психодиагностические методы в рамках основных психологических школ.

Тестовые методики были тесно связаны с теоретическими принципами *бихевиоризма*. Методологическая концепция бихевиоризма основывалась на том, что между организмом и средой существуют однозначные причинные (детерминационные) отношения. Организм, реагируя на стимулы внешней среды, стремится изменить ситуацию в благоприятную для себя сторону и приспосабливается к ней. Бихевиоризм ввел в психологию в качестве ведущей категорию поведения, понимая его как *совокупность доступных объективному наблюдению реакций на стимулы*.

В соответствии с основной бихевиористской концепцией цель диагностики сводилась первоначально к фиксации поведения. Именно этим занимались первые психодиагносты, разработавшие метод тестов.

Тестовая диагностика, или тестология, основателем которой считается Дж. Кеттелл (1890), долгое время оставалась центральной областью психодиагностики. Тесты оказались незаменимым инструментом изучения скрытого от внешнего наблюдателя мира психической реальности. Тестовые измерения из области научных исследований практически сразу же перешли в распоряжение психологической практики, где использовались для дифференциации людей с точки зрения различных девиаций, оценки развития разнообразных психических функций, задач профессионального отбора и пр.

§2. Развитие тестов интеллекта.

Первым исследователем, использовавшим понятие "интеллектуальный тест", был американский психолог Джеймс Маккин Кеттелл. Этот термин после статьи Кеттелла "Интеллектуальные тесты и измерения", опубликованной в 1890 году в журнале "Mind", приобрел широкую известность [16, 17].

В своей статье Кеттелл писал о том, что применение серии тестов к большому числу индивидов позволит открыть закономерности психических процессов и тем самым приведет к преобразованию психологии в точную науку. Вместе с тем он высказал мысль о том, что научная и практическая ценность тестов возрастет, если условия их проведения будут однообразными.

Уже первые психологические эксперименты выявили необходимость строгого контроля условий проведения наблюдений. Например, формулировка инструкций, которые давались испытуемому в эксперименте на время реакции, могла существенно изменить ее скорость. Подобным образом, была доказана важность проведения наблюдений испытуемых в стандартных условиях, и стандартизация процедуры в конце концов стала одним из отличительных признаков психологических тестов.

Так была провозглашена необходимость *стандартизации* тестов для того, чтобы стало возможным сравнение их результатов, полученных разными исследователями на разных испытуемых [1, 16].

Кеттелл предложил в качестве образца 50 тестов, включавших различного рода измерения чувствительности, времени реакции, времени, затрачиваемого на называние цветов, количества звуков, воспроизводимых после однократного прослушивания, и др.

Вернувшись в Америку после работы в лаборатории Вундта и чтения лекций в Кембридже, он немедленно стал применять тесты в

устроенной им при Колумбийском университете лаборатории (1891). Вслед за Кеттеллом и другие американские лаборатории начали применять метод тестов. Возникла необходимость организовать специальные координационные центры по использованию этого метода. В 1895 - 1896 гг. в США были созданы два национальных комитета, призванных объединить усилия тестологов и придать общее направление тестологическим работам.

Первоначально в качестве тестов использовались приемы экспериментально-психологического исследования. По форме они походили на приемы лабораторного исследования, но смысл их применения был принципиально иным. Ведь задачей психологического эксперимента является выяснение зависимости психического акта от внешних и внутренних факторов, например, характера восприятия от внешних раздражителей, запоминания - от частоты и распределения во времени повторений и т. д. При тестировании же психолог регистрирует индивидуальные различия психических актов, ни в коем случае не изменяя внешних условий осуществления этих психических актов.

Метод тестов получает широкое распространение. Новый шаг в его развитии был сделан французским врачом и психологом Альфредом Бине (1857 - 1911), создателем самой популярной для своего времени серии тестов.

В 1904 г. министерство образования Франции ввело А. Бине в Комиссию по изучению методов обучения умственно отсталых детей. Это побудило его заняться разработкой методик, с помощью которых можно было бы отделить детей, способных к учению, но ленивых и не желающих учиться, от страдающих прирожденными дефектами и не способных учиться в нормальной школе. Нужда в этом возникла в связи с введением всеобщего образования. Одновременно потребовалось создание

специальных школ для умственно неполноценных детей [1, 16].

А. Бине в сотрудничестве с Т. Симоном провел серию экспериментов по изучению внимания, памяти, мышления у детей разного возраста (начиная с трех лет). Проведенные на многих испытуемых экспериментальные задания были проверены по статистическим критериями и стали рассматриваться как средство определения интеллектуального уровня.

Первая батарея тестов или, как она тогда называлась, «шкала Бине - Симона» появилась в 1905 г. Затем она несколько раз пересматривалась авторами, которые стремились изъять из нее все задания, требующие специального обучения. Бине исходил из представления о том, что развитие интеллекта происходит независимо от обучения, в результате биологического созревания.

Шкала А. Бине в последующих редакциях (1908 и 1911 гг.) была переведена на немецкий и английский языки. Задания в тесте Бине были сгруппированы по возрастам (от 3 до 13 лет). Для каждого возраста подбирались определенные тесты. Они считались соответствующими данной возрастной ступени, если их решало большинство детей данного возраста (80 - 90%). Детям до 6 лет предлагалось по 4 задания, а детям старше 6 лет - 6 заданий. Задания подбирались путем исследования большой группы детей (300 человек).

Таким образом, стандартизация тестов Бине опиралась на достаточно репрезентативную (представительную) выборку испытуемых.

Показателем интеллекта в шкалах Бине был *умственный возраст*, который мог расходиться с хронологическим. Умственный возраст определялся по успешности выполнения тестовых заданий. Испытание начиналось с определения тестовых заданий, соответствующих хронологическому возрасту ребенка. Если он справлялся со

всеми заданиями, и ему предлагались задания более старшей возрастной группы. Если он решал не все, а некоторые из них, то испытание прекращалось. Если же ребенок не справлялся со всеми заданиями своей возрастной группы, то ему давались задания, предназначенные для более младшего возраста.

Испытания проводились до тех пор, пока не выявлялся возраст, все задания которого решаются испытуемым. Максимальный возраст, все задания которого решаются испытуемым, назывался *базовым умственным возрастом*.

Если, кроме того, ребенок выполнял также некоторое количество заданий, предназначенных для более старших возрастных групп, то каждое задание оценивалось числом "умственных" месяцев. Тогда к числу лет, определяемых базовым умственным возрастом, прибавлялось и некоторое число месяцев.

Пример: ребенок решил все задания, предназначенные для семилетнего возраста, и 2 задания, рассчитанные на восьмилеток. Число месяцев рассчитывается так. Вначале определялась "цена" одного задания в месяцах: $12 \text{ месяцев} / 6 \text{ (число заданий для восьмилеток)} = 2 \text{ месяца}$. После этого эта "цена" умножалась на число решенных заданий: $2 * 2 = 4 \text{ месяца}$. Итак, умственный возраст ребенка определялся равным 7 годам и 4 месяцам.

Несовпадение умственного и хронологического возраста считалось показателем либо умственной отсталости (если умственный возраст меньше хронологического), либо одаренности (если умственный возраст больше хронологического).

Вторая редакция шкалы Бине послужила основой работы по проверке и стандартизации, проведенной в Стэнфордском университете (США) коллективом сотрудников под руководством Л. М. Термена. Этот вариант тестовой шкалы Бине был предложен в 1916 г. и имел так много серьезных изменений, по сравнению с основным, что был назван *шкалой Стэнфорд-Бине*.

Основных нововведений по сравнению с тестами Бине было два: *введение в качестве показателя по тесту коэффициента интеллектуальности (IQ) и введение статистической тестовой нормы.*

Коэффициент *IQ* был предложен В. Штерном. Он предложил определять частное, получаемое при делении умственного возраста на хронологический. Этот показатель, умноженный на 100, он и назвал коэффициентом интеллектуальности. Используя этот показатель, можно классифицировать нормальных детей по степени умственного развития.

Шкала Стэнфорд-Бине рассчитана на детей в возрасте от 2,5 до 18 лет. Она состоит из заданий разной трудности, сгруппированных по возрастным категориям. Для каждого возраста наиболее типичный, средний показатель выполнения (x) равен 100, а статистическая мера рассеяния (среднее квадратическое, или стандартное отклонение) индивидуальных значений от этого среднего (σ) равняется 16. Все индивидуальные показатели по тесту, попавшие в интервал $x \pm \sigma$, т. е. ограниченные числами 84 и 116 считаются нормальными, соответствующими возрастной норме выполнения теста. Если тестовый показатель выше тестовой нормы (более 116), ребенок считается одаренным, а если ниже 84 - то умственно отсталым.

Шкала Стэнфорд-Бине получила популярность во всем мире. Она имела несколько редакций (1937, 1960, 1972, 1986 гг.). Показатель *IQ*, получаемый по шкале Стэнфорд-Бине, на долгие годы стал синонимом интеллекта. Вновь создаваемыми интеллектуальные тесты стали проверяться на валидность путем сопоставления с результатами шкалы Стэнфорд-Бине. И многие из них также используют стандартную шкалу с параметрами: $x=100$, $\sigma=16$. (Иногда $\sigma=15$).

Групповое тестирование. Следующий этап развития психологиче-

ского тестирования характеризуется *изменением формы проведения* тестового испытания. Все тесты, созданные в первом десятилетии XX века, были *индивидуальными* и позволяли вести опыт только *с одним испытуемым*. Использовать их могли лишь специально подготовленные люди, имеющие достаточно высокую квалификацию.

Эти особенности первых тестов ограничивали их распространение. Практика же требовала диагностировать большие массы людей с целью отбора наиболее подготовленных к тому или иному виду деятельности, а также распределения по разным видам деятельности людей в соответствии с их индивидуальными особенностями. Поэтому в США в период первой мировой войны появилась новая форма тестовых испытаний - *групповое тестирование*.

Необходимость как можно быстрее отобрать и распределить полутора миллионную армию рекрутов по различного рода службам, школам и училищам заставила специально созданный комитет поручить Артуру С. Отису разработку новых тестов. Так появились две формы так называемых армейских тестов - "Альфа" и "Бета". Первая из них предназначалась для работы с людьми, знающими английский язык вторая - для неграмотных и иностранцев. После окончания войны эти тесты и их модификации продолжали широко применять [16].

Групповые тесты не только делали реальными испытания больших групп, но наряду с этим допускали упрощение инструктирования, процедуры проведения и оценки результатов тестирования. К тестированию начали привлекаться люди, не имеющие настоящей психологической квалификации, а всего лишь обученные проведению тестовых испытаний.

В то время как индивидуальные тесты, такие как шкалы Стэнфорд-Бине, в основном применялись в клинике и для консультирования,

групповые тесты использовались преимущественно в системе образования, в промышленности и в армии. 20-е годы нашего столетия характеризовались настоящим тестовым бумом. Быстрое и широкое распространение тестологии было обусловлено, прежде всего, ее направленностью на оперативное решение практических задач.

Измерение интеллекта с помощью тестов рассматривалось как средство, позволяющее научно подойти к вопросам обучения, профотбора, оценки достижений и т. д. На протяжении первой половины XX века специалистами в области психологической диагностики было создано множество разнообразных тестов. При этом, разрабатывая методическую сторону тестов, они доводили ее поистине до высокого совершенства. Все тесты тщательным образом стандартизировались на больших выборках; тестологи добивались того, что все они отличались высокой надежностью и определенной валидностью (соответствием из меряемому психическому свойству).

Тем не менее, им свойственны известные *недостатки*. Валидизация выявила ограниченные возможности тестов интеллекта: необходимая *точность прогнозирования* на их основе успешности выполнения конкретных, достаточно узких видов деятельности часто не достигалась. Требовалась, помимо знания общего уровня интеллекта, дополнительная информация об особенностях психики человека.

Возникло новое направление в тестологии - тестирование *специальных способностей*, которое вначале призвано было лишь дополнить оценки тестов интеллекта, а впоследствии выделилось в самостоятельную область.

Контрольные вопросы для самопроверки

§1, 2. Предыстория психодиагностики. Развитие тестов интеллекта

1. Что послужило источником развития психодиагностики?
2. Кто первый из психологов ввел понятие "психологический тест"?

3. Какой вклад в развитие психодиагностики внес А. Бине?
4. В чем достоинства и недостатки "тестов интеллекта"?

Литература к теме

1. Анастаси А. Психологическое тестирование. В 2-х кн. М., 1982.
2. Бурлачук Л.Ф., Морозов С.М. Словарь-справочник по психодиагностике. - СПб.: Изд-во "Питер", 1999. - 528 с.
3. Общая психодиагностика / Под ред. А. А. Бодалева, В. В. Столина.- М., 1987.
4. Основы психодиагностики / Под ред. А. Г. Шмелева. Ростов-на-Дону., 1996.
5. Психологическая диагностика: Проблемы и исследования / Под ред. К. М. Гурвича. М., 1981.

§3. Развитие тестов специальных способностей и достижений.

Применение тестов интеллекта в практике показало их ограниченность, особенно в сферах профессионального образования и консультирования. Постепенно психологи приходили к мнению о необходимости дополнить общие тесты интеллекта тестами специальных способностей.

Стали появляться тесты механических, канцелярских, музыкальных, артистических способностей. Создавались тестовые батареи (комплекты) для отбора поступающих в медицинские, юридические, инженерные и другие учебные заведения.

Критическая оценка тестов интеллекта обнаружила тот факт, что успешность выполнения отдельных частей теста весьма индивидуальна (даже, если субъекты набирают в итоге одинаковый балл IQ). Например, индивид может иметь достаточно высокие баллы по вербальному субтесту и довольно низкие по числовому (или наоборот). Однако использовать тесты интеллекта для профотбора или профконсультирования не представлялось возможным, так как они не были рассчитаны на дифференциальный анализ способностей. Часто сравниваемые субтесты содержали слишком мало

заданий, чтобы по ним можно было судить о развитии той или иной способности. Все эти недостатки тестов интеллекта поставили перед психологами задачу создания тестов специально предназначенных для выявления индивидуальных различий.

А. Анастази [1] особо подчеркивает, что следует четко различать терминологию: термины - «тест способностей» и «тест интеллекта». Первый обозначал тесты, измеряющие относительно простые способности; под вторым термином объединялись достаточно однородные серии тестов, дающие один суммарный показатель типа IQ.

Другое различие состоит в том, что тесты специальных способностей измеряют, как правило, одну способность. Комплексные батареи способностей измеряют ряд способностей и дают профиль (в отличие от суммарного бала IQ), образуемый показателями каждой способности.

Было разработано около довольно много комплексных батарей способностей для использования в образовании, консультировании и распределении персонала. Различаясь составом, методическими качествами они сходны в одном - *их характеризует низкая дифференциальная валидность* (учащиеся, выбирающие разные области образования или профессиональной деятельности, незначительно различаются своими тестовыми профилями).

Такие батареи способностей строились с помощью *факторного анализа*, который по своей сути являлся их экспериментально-теоретической основой. Факторный анализ позволял точнее определить и классифицировать специальные способности.

Английский психолог Чарльз Спирмен, наблюдая зависимость между разными проявлениями интеллектуальной активности людей, предпринял попытку ответить на вопрос о том, существует ли общий фактор, лежащий в основе разнообразных проявлений человеческого ума и харак-

теризующий некую общую (генеральную) способность. Для проверки своей гипотезы он предложил математический аппарат, позволяющий на основе анализа матрицы коэффициентов корреляции вычислять общие закономерности, или факторы.

В своей статье (1904 г.) Ч. Спирмен пришел к выводу, что положительная корреляция между тестами различных способностей (например, математических и литературных) выявляет некоторый *общий генеральный фактор*. Он обозначил его буквой G (от англ. General - общий).

Помимо фактора, общего для всех видов деятельности, в каждой конкретной деятельности при ее осуществлении обнаруживается *специфический фактор, свойственный только данному виду деятельности* ("S-фактор") [1].

Например: после Ч. Спирмена многие другие исследователи неоднократно подтверждали существование генерального фактора. В том числе вполне остроумно это делалось с помощью факторного анализа таблицы обыкновенных отметок учеников по разным школьным предметам (М. Рошлэн). Существование фактора "G" на этих данных пояснить очень просто: как правило, отличники (ученики с высоким G) и двоечники (ученики с низким G), отличающиеся соответственно более высокими или более низкими баллами по всем предметам, встречаются чаще, чем ученики с резко нестабильным профилем достижений - высокими баллами по одним предметам и низкими баллами по другим предметам.

В дальнейшем выяснилось, что отдельные тестовые показатели объединяются в родственные подгруппы и за каждой такой подгруппой скрывается определенный *"латентный фактор"*, определяющий общий уровень достижений испытуемого по всей подгруппе тестов. Причем в разные тесты один и тот же "латентный фактор" вносит разный вклад (*вес, факторную нагрузку*).

Например, "вербальный фактор" может иметь больший вес в тесте на словарный запас, меньший - в тесте словесных аналогий и совсем незначительный - в тесте на математическое мышление.

Взаимосвязи тестов между собой являются результатом «нагруженности» их, родственным им всем, латентным фактором. Важно

подчеркнуть, что факторы, которые выделялись в результате факторного анализа, не придумывались и не задавались учеными из теоретических соображений, а возникали как бы сами собой - в результате появления группировок скоррелированных тестов.

Дальнейшую разработку и применение многофакторного анализа тестовых данных осуществил американский психолог Л. Л. Терстоун (основные работы вышли в 20 - 30-е гг.). На основе факторных исследований создавались многофакторные батареи тестов способностей, позволяющие измерять индивидуальный уровень каждой из способностей. Наиболее известны среди них Батарея тестов общих способностей (GATB), включающая тесты способностей для конкретных профессий.

Факторный анализ не обозначает один особый метод, а является, скорее, совокупностью различных способов статистического анализа, которые отражают специфику исходной психологической теории и интерпретационной установки исследователя. Кроме этого, результаты факторизации зависят от природы и количества переменных, подвергающихся анализу. От этих параметров, как и от изучаемой теоретической реальности, зависит решение о выборе формы факторного анализа и тех переменных, которые должны быть включены в исходную матрицу [12].

Многолетний опыт использования факторного анализа для изучения психических явлений показал, что факторный анализ - это обобщение линейных корреляций. Однако *линейные корреляции не являются универсальной формой выражения функциональной связи между психическими процессами. Следовательно, отсутствие линейных корреляций не может толковаться как отсутствие связи вообще.* Поэтому факторный анализ и добываемые посредством этого анализа факторы не всегда верно отражают зависимости между психическими процессами.

В настоящее время разработаны более сложные алгоритмы многомерной статистической группировки тестовых данных, например дисперсионный анализ.

Примечание: к возможностям традиционного факторного анализа и к выделенным с его помощью факторам современные ученые относятся с определенной осторожностью и не считают этот анализ универсальным инструментом изучения психики, а следовательно, и безошибочным базисом для конструирования психологических тестов.

Наряду с тестами интеллекта, специальных и комплексных способностей возник и еще один тип тестов, широко применяемых в учебных заведениях, - *тесты достижений*. Этим термином психометристы называли педагогические тесты знаний по определенным учебным предметам, а также профессиональные тесты - на специальные профессиональные умения и навыки.

Эти тесты отличаются от тестов интеллекта тем, что они отражают не столько влияние многообразного накопленного опыта, сколько влияние специальных программ обучения на эффективность решения тестовых заданий.

История развития этих тестов может быть прослежена с момента смены в Бостонской школе устной формы экзаменов на письменную (1845 г.). В Америке тесты достижений используются при отборе сотрудников на государственную службу, начиная уже с 1872 г., а с 1883 г. их применение становится регулярным. Наиболее значительная разработка элементов техники конструирования тестов достижений выполнена в течение первой мировой войны и сразу после нее [16].

Тесты достижений относятся к наиболее многочисленной группе диагностических методик. Главную роль при конструировании таких тестов играет то содержание, которое должно быть усвоено (учащийся или претендент на какое-либо рабочее место должен продемонстрировать какими знаниями и умениями он владеет).

Как отмечает Гуревич К. М. [17, 18], история развития тестов достижений начинается с момента смены в Бостонской школе (1845) устной формы проведения экзаменов на письменную. Этим самым как бы стандартизировалась сама процедура проведения экзамена, что позволяло существенно снизить элемент случайности и субъективности в оценках.

Наиболее известным тестом достижений и широко применяемым в США до сих пор, является Стэнфордский тест достижений (*SAT*) впервые опубликованный в 1923 г. Он оценивает уровень обученности в разных классах в средних учебных заведениях. Значительное число тестов специальных способностей и достижений было создано в рамках техники (индустриальной психологии) под воздействием практических запросов со стороны промышленности и экономики [16].

Дальнейшее развитие тестов достижений привело к появлению в середине XX века *критериально-ориентированных тестов*. В целом развитие исследований и сбор статистических данных о применении тестов в психодиагностике значительно продвинуло тестологию, составляющую в психологии ядро психометрики.

С 1936 года в США выходит специализированный журнал "Психометрика". Другой ведущий журнал в этой области - "Образовательные и психологические измерения". В этих журналах еще до второй мировой войны и вскоре после нее выпущены ставшие классическими работы таких психометристов, как Рюлон, Гилфорд, Кронбах, обосновавших, в частности, широко используемые ныне методы измерения надежности, валидности и репрезентативности тестов.

Контрольные вопросы для самопроверки

§3. Развитие тестов специальных способностей и достижений

1. В чем отличие тестов интеллекта от тестов специальных способностей и достижений?
2. Что такое "генеральный" или "G" фактор?
3. В чем отличие тестов достижений от тестов специальных способ-

ностей?

4. Каковы достоинства и недостатки применения факторного анализа в психодиагностике?

Литература к теме

1. Анастаси А. Психологическое тестирование. В 2-х кн. М., 1982.
2. Бурлачук Л.Ф., Морозов С.М. Словарь-справочник по психодиагностике. - СПб.: Изд-во "Питер", 1999. - 528 с.
3. Общая психодиагностика / Под ред. А. А. Бодалева, В. В. Столина.- М., 1987.
4. Основы психодиагностики / Под ред. А. Г. Шмелева. Ростов-на-Дону., 1996.
5. Психологическая диагностика: Проблемы и исследования / Под ред. К. М. Гурвича. М., 1981.
6. Психологическая диагностика: Учебное пособие / Под ред. К.М. Гуревича и Е.М. Борисовой. - М.: Изд-во УРАО, 1997. - 304 с.

§4. Развитие методов психодиагностики личности.

Базу для разработки личностных тестов составляют тесты, позволяющие измерять различные аспекты поведения личности, которые напрямую не связаны с уровнем интеллектуального развития. Среди психологов до сих пор не выработано единое, общепринятое понятие «тест личности». Обычно в психологической диагностике этот термин включает сферу измерения характеристик, подобных эмоциональной устойчивости, особенностей межличностных отношений, мотивации, интересов, ценностей и пр.

Условно можно выделить три группы тестов, измеряющих личностные особенности: *опросники (основанные на самоотчете), ситуативными тестами и проективными методиками* [17].

Известно [1, 15, 16, 19], что *опросники* - это большая группа методик, задания которых представлены в виде вопросов или утверждений, а задачей испытуемого является самостоятельное сообщение некоторых

сведений о себе в виде ответов.

Теоретической основой этого метода можно считать интроспекционизм - психологию самонаблюдения. Однако при заданных вариантах ответа самонаблюдение, которому придается стандартизированный характер, по многим формальным признакам сближается с объективным тестированием. Прототипом личностных опросников можно считать разработанный американским психологом Робертом Вудвортом в 1919 г. "Бланк данных о личности". Этот опросник был предназначен для выявления и отсеивания с военной службы лиц с невротической симптоматикой.

В последующие десятилетия опросники получили широкое распространение в качестве психодиагностического метода исследования личности. Искусная косвенная постановка вопросов, маскирующая их оценочную направленность, стандартизированная процедура предъявления и подсчета баллов во многом сблизили современные опросники с объективными тестами, основанными на объективных заданиях. Здесь речь идет скорее о тест-опросниках, а не о методе стандартизированного самонаблюдения как таковом [16].

Ситуативными называются такие тесты, которые нацелены на выполнение испытуемым какой-то деятельности. Как правило, испытуемому предлагают выполнить деятельность, моделирующую элементы его повседневной жизни.

Ситуативные тесты особенно интенсивно развивались во второе и третье десятилетие нашего века, а стандартизация их проводилась, в основном, на школьниках. Как отмечает К.М. Гуревич [17], с помощью этих тестов диагностируются такие стороны личности, как: склонность к мошенничеству, лжи, воровству, упорству и пр. Ситуативные тесты широко использовались также в период второй мировой войны. Обычно подобные

тесты носят комплексный характер и включают в себя наряду с изучением социальных, и эмоциональные черты личности.

Интерпретация ситуативных тестов достаточно субъективна, в связи с чем возможность их применения вне чисто исследовательских целях при массовых обследованиях, считается весьма проблематичной.

Другим известным методом диагностики личности являются *проективные методики*, которые представляют собой специфическую, достаточно неоднородную группу психодиагностических методик, в основном, клинической ориентации.

Их родоначальником традиционно считается метод словесных ассоциаций К. Юнга, созданный им в 1904-1905 гг. и возникший на базе ассоцианистского течения в психологии.

Несмотря на то, что этот метод известен со времен Ф. Гальтона (1822 - 1911), который еще 1879 г. опубликовал результаты своих ассоциативных экспериментов, предлагая испытуемому отвечать на слово раздражитель первой пришедшей в голову словесной ассоциацией, именно К. Юнгу принадлежит открытие и доказательство феномена, лежащего в основе всех проективных методик - возможность посредством косвенного воздействия на значимые области переживания и поведения человека вызвать эксплицитные изменения в экспериментальной деятельности [15, 16]. Позднее эта методика получила дальнейшее развитие в исследованиях Э. Крепелина (1899), Г. Кента и А. Розанова (1910) и других.

Однако следует отметить, что интерпретация получаемых результатов определяется теоретическими взглядами психодиагноста. *Поэтому вопрос о валидности методики не может быть решен однозначно вне соотнесения с теоретическими позициями ее разработчиков и пользователей.* Ассоциативный эксперимент стимулировал появление таких

проективных методик, как "неоконченные предложения" или "незаконченные рассказы").

Считается, что для того, чтобы психологу прорваться в область бессознательного, понять скрытые в нем тенденции, необходимо в эксперименте направить сознание на решение особых заданий, которые позволили бы непроизвольно проявиться бессознательному в особой проективной продукции - бессюжетных словесных ассоциациях, сюжетных фантазиях, образах, выраженных в рисунках (как это делается в рисуночных проективных методиках) и т. п. Такого типа задания и включались в проективные методики [16].

Наибольший успех проективной диагностики связан с одним из самых популярных проективных тестов, разработанным в 1921 г. швейцарским психиатром Германом Роршахом, (который, одним из первых ввел в обиход термин "психодиагностика").

Создавая эту методику, Роршах экспериментировал с большим количеством чернильных пятен, которые он предъявлял различным группам психически больных людей. В результате своих наблюдений Роршах постепенно объединил те характеристики ответов, которые можно было соотнести с различными психическими заболеваниями, в относительно стройную систему показателей. В дальнейшем эта методика использовалась и анализировалась многими исследователями как за рубежом, так и у нас в стране.

Другим старейшим и наиболее распространенным в мире проективных тестов является - Тест тематической апперцепции (ТАТ) - была создана в США в 1935 г. Х. Морган и Г. Мюррем. Завершая краткий обзор истории развития и становления психологической диагностики на Западе, отметим, что она отличается широким разнообразием используемых методик, как в отношении формы, так и их содержания [15,

16].

Одновременно с обилием новых разработок для практической западной психодиагностики после второй мировой войны стал характерным определенный консерватизм, выражающийся в приверженности к работе с немногими тестами, которые прошли серьезную многолетнюю психометрическую адаптацию, по которым собраны десятки тысяч протоколов, выполнены сотни и тысячи диссертационных работ (по проверке этих тестов на особых контингентах испытуемых и в особых условиях, созданию частных тестовых норм, дополнительных шкал и т. п.).

Любому новому тесту, каким бы он ни был передовым в научном отношении, в этих условиях очень трудно конкурировать с "классическими" методиками, по которым накоплена огромная методическая литература. Даже новые компьютерные тесты, обладающие массой объективных достоинств (например, гибкими возможностями настройки на конкретного испытуемого - свойствами, так называемого, адаптивного тестирования), с трудом пробивают себе дорогу и до сих пор не могут сравниться в популярности с "классическими" методиками.

Не случайно многие образцы современных компьютерных тестов являются не более, чем компьютерными версиями существовавших до них буклетных, или "карандашно-бумажных", методик.

Медленно завоевывают себе приверженцев на Западе и такие новые научные направления в области личностной психодиагностики, порожденные опять же компьютерной революцией второй половины XX века, которые основываются на комбинации методов стандартизированного опроса и проективных техник. Это, прежде всего, *техники семантического шкалирования* (Ч. Осгуд, 1952; Дж. Келли, 1955), позволяющие в результате компьютерной многомерной обработки данных реконст-

руировать так называемые "субъективные семантические пространства" [16, 21].

Контрольные вопросы для самопроверки

§4. Развитие методов психодиагностики личности

1. Какие группы тестов относятся к методам психодиагностики личности? Дайте их краткую содержательную характеристику.
2. Кто был основателем проективной личностной психодиагностики?
3. Каковы основные особенности применения проективных тестов для диагностики личности?

Литература к теме

1. Анастаси А. Психологическое тестирование. В 2-х кн. М., 1982.
2. Бурлачук Л.Ф., Морозов С.М. Словарь-справочник по психодиагностике. - СПб.: Изд-во "Питер", 1999. - 528 с.
3. Общая психодиагностика / Под ред. А. А. Бодалева, В. В. Столина.- М., 1987.
4. Основы психодиагностики / Под ред. А. Г. Шмелева. Ростов-на-Дону., 1996.
5. Психологическая диагностика: Проблемы и исследования / Под ред. К. М. Гурвича. М., 1981.
6. Психологическая диагностика: Учебное пособие / Под ред. К.М. Гуревича и Е.М. Борисовой. - М.: Изд-во УРАО, 1997. - 304 с.
7. Шмелев А. Г. Психодиагностика личностных черт. – СПб.: Речь, 2002. – 480 с.

§5. Психологическая диагностика в России *

Особенностью развития отечественной психологии в последней четверти прошлого века было внесение в нее экспериментальных методов исследования, идущих от физиологии высшей нервной деятельности, с одной стороны, а с другой - из лаборатории В. Вундта, в которой проходили стажировку первые русские психологи-экспериментаторы (Н. Н. Ланге, например).

* **Примечание:** данный параграф написан по материалам книги: А.Г. Шмелев и коллектив. Основы психодиагностики. Учебное пособие для студентов педвузов. - Москва, Ростов-на-Дону: «Феникс», 1996. - 544с.

У истоков первого направления находились два величайших корифея отечественной науки - И. М. Сеченов (1829 - 1905) и И. П. Павлов (1849 - 1936).

Важной вехой стало открытие в Петербурге психоневрологического института В. М. Бехтеревым. Первая в России экспериментальная психологическая лаборатория была открыта в 1885 г. при клинике нервных и душевных болезней Казанского университета.

В 1895 г. по инициативе крупнейшего русского психиатра С. С. Корсакова была создана психологическая лаборатория при психиатрической клинике Московского университета. Заведовать ею стал ближайший помощник Корсакова А. А. Токарский.

Во всех этих лабораториях работали врачи невропатологи и психиатры, совмещавшие свои психологические исследования с врачебной практикой в клинике, а также студенты-медики. Исключение составляла психологическая лаборатория в Новороссийском университете (в Одессе). В отличие от других она была создана на историко-филологическом факультете профессором философии Н. Н. Ланге.

В этих лабораторных исследованиях изучались объективные признаки тех или иных психических явлений (например, изменение пульса и дыхания как отражение эмоций), доказывалась предметность, объективность наших восприятий, выяснялась зависимость памяти и внимания от условий опыта - и т. д. Кроме того, во всех экспериментальных лабораториях проводились исследования скорости протекания психических процессов.

Итак, во второй половине XIX века в отечественную психологию введен эксперимент. Но для возникновения психологической диагностики необходимо было, кроме того, чтобы практике потребовалось знание об индивидуально-психологических особенностях человека. Первые отечест-

венные работы по психологической диагностике были выполнены в первые десятилетия XX столетия.

Вероятно, одна из первых значительных дореволюционных отечественных работ по психологическому тестированию, представляющая законченное самостоятельное исследование, была выполнена Г. И. Россолимо в 1909 г. в Московском университете. Г. И. Россолимо - крупнейший невропатолог и психиатр, поставил своей целью найти метод количественного исследования психических процессов в нормальном и патологическом состояниях.

По существу этот метод, получивший широкую известность как в России, так и за рубежом, был одним из ранних оригинальных вариантов системы тестов для измерения умственной одаренности.

Эта система обследования, названная *методикой индивидуального психологического профиля*, сводилась к определению 11 психических процессов, которые оценивались по десятибалльной шкале на основании ответов на 10 специально подобранных вопросов.

Психические процессы, измеряемые методикой Россолимо, в целом составляли три группы: внимание и воля, точность и прочность восприятия, ассоциативная деятельность. Он предложил *графическую форму представления измерений психических процессов - вычерчивание "психологического профиля"*, который наглядно демонстрировал соотношение указанных процессов.

Труды Россолимо были с интересом встречены как психологами, так и психиатрами, специализирующимися по проблемам умственной отсталости. Подобные "профили" с того времени прочно вошли в психологическую диагностику.

Как считал П. П. Блонский, положительным в методике Россолимо было то, что в отличие от западного тестирования он стремился

к целостной оценке личности, к синтетическому способу изображения ее сильных и слабых сторон. Лишь впоследствии структурный способ исследования личности, к которому стремился Россолимо, начал укрепляться в психологической диагностике в Западной Европе и в США.

Еще один русский психолог, придерживавшийся сходных взглядов на изучение личности, А. Ф. Лазурский примерно в то же время создал новое направление в дифференциальной психологии - *научную характерологию*. Строго придерживаясь опыта и эксперимента как основных методов исследования, он в то же время стоял за создание научной теории индивидуальных различий.

Основной целью дифференциальной психологии он считал "построение человека из его наклонностей", а также разработку возможно полной естественной классификации характеров. Но работы в этом направлении были не закончены, им помешала преждевременная кончина исследователя (в 1917 г.).

С развитием дифференциально-психологических исследований психология в целом обогатилась рядом новых методов и подходов. Стали вполне осуществимы ее связи с практикой. Все это послужило основой для возникновения психологической диагностики. Собственно психодиагностические работы в России за малым исключением начали развиваться в послереволюционный период. Особенно много таких работ появилось в 20 - 30-е годы в области педологии и психотехники в связи с ростом популярности метода тестов в Советской России, как и на Западе в это же время.

Подавляющее большинство методик было копированием западных психологических тестов. Незначительные отличия проявлялись в форме проведения тестовых испытаний, в обработке и интерпретации экспериментального материала.

Определенный интерес с точки зрения развития новых форм тестирования представляет "Измерительная шкала ума" А. П. Болтунова (1928), положившая в основу своей работы шкалу Бине - Симона, переведенную и адаптированную П. П. Соколовым для испытания умственной одаренности русских школьников.

По сути дела шкала Болтунова представляет собой самостоятельную разработку нового набора тестов. Несмотря на известную аналогию со шкалой Бине - Симона, шкала Болтунова имеет специфические особенности: в ней модифицировано большинство заданий, введены совершенно новые задания, предложены новая инструкция и форма ее использования, определено время решения тестовых заданий, разработаны показатели возрастных степеней.

Принципиальное отличие шкалы Л. П. Болтунова от шкал Бине - Симона состоит в возможности проводить групповые испытания. И тем не менее данная работа типична для традиционного психологического тестирования того времени. Само понимание инструкций к тестовым заданиям требует от детей достаточно высокого уровня развития речевого мышления (вербального интеллекта).

Особое место в отечественных тестологических исследованиях занимают работы М. Ю. Сыркина, 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 между слоями и классами общества для тестологии становился все более острым и значимым.

В 20-е годы нынешнего века в нашей стране значительное развитие получила психология труда и психотехника (труды И.Н. Шпильрейна, С.Г. Геллерштейна, Н.Д. Левитова, А.А. Толчинского и др.). В рамках этих отраслей психологии развивалась психодиагностика, результаты которой нашли применение в ряде направлений народного хозяйства, прежде всего в промышленности, на транспорте, в системе профессионального обучения и профотбора. Если в советской педологии относительно больше внимания уделялось тестам интеллекта, то в психотехнике - тестам специальных способностей.

Психодиагностика в советский период. Педология и психотехника подвергались критике, в особенности за формальное использование теоретически не обоснованных тестов. Следует признать, что частично эта критика была обоснованной. Вызывал протест массовый перевод вполне умственно-сохранных, но педагогически запущенных детей (с низким уровнем развития речевого мышления или вербального интеллекта) из нормальных школ в школы для умственно отсталых детей.

Это было особенно несправедливо в стране, унаследовавшей от времен гражданской войны массовую разруху, неполные семьи и просто беспризорность, лишавшую детей своевременной полноценной практики речевого общения со взрослыми.

Эти объективные ошибки послужили поводом для большевистского руководства страны наложить фактический запрет на использование тестов в школе и на производстве. В 1936 г. вышло постановление ЦК ВКПб "О педологических извращениях в системе Наркомпроса", которое остановило более чем на 40 лет не только практику применения тестов, но фактически и научные разработки в области психодиагностики.

Резкая критика педологии сопровождалась отрицанием всего положительного, что было сделано учеными, так или иначе связан-

ными с педологией, в области психологии и психологической диагностики.

Новым идеологическим наступлением на психодиагностику и на психологию в целом явилась в 1950 г. печально-известная "павловская" сессия двух Академий (педагогических и медицинских наук).

Согласно агрессивно-материалистической идеологии (в духе сталинской интерпретации марксизма-ленинизма) легальное право на существование и развитие сохранялось не за "дифференциальной психологией", а за дифференциальной психофизиологией, связывающей многообразие психических явлений с механизмом условного рефлекса, а происхождение индивидуальных психологических различий исключительно с типами нервной системы в духе учения И. П. Павлова.

Разработанная в трудах Б. М. Теплова, Б. Г. Ананьева, В. С. Мерлина, их учеников и последователей, отечественная дифференциальная психофизиология на долгие годы оказалась единственной легальной формой существования психодиагностики в России.

Принятые на вооружение советскими педагогическими и детскими психологами так называемый *каузально-генетический метод* (Д. Б. Эльконин, В. В. Давыдов и другие), а также *метод формирования умственных действий* (П. Я. Гальперин, Н. Ф. Талызина) обозначали фактический отказ от применения психодиагностики как таковой в практике образования, ее замену *лонгитюдным исследованием* или исследованием в процессе формирующего воздействия, которые могут быть качественно выполнены только квалифицированными специалистами в рамках специальной научной программы.

В период "оттепели" 50 - 60-х годов относительно более свободные (от идеологического сверхконтроля) условия для развития психодиагностики появились скорее в Ленинграде, Киеве и региональных цен-

трах России (Перми, Казани, Ярославле, Курске и др.) и в прикладных отраслях, чем в московской академической психологии, где "табу", наложенное на сам термин "тест", ревностно поддерживалось идеологами от психологии вплоть до 70-х годов.

Значительная часть публикаций советских психологов вплоть до периода горбачевской либерализации и перестройки (вторая половина 80-х годов) была так или иначе отмечена признаками вынужденного самобичевания (что выражалось в признании справедливости постановления 1936 г.).

Они изобиловали огульной критикой западного, прежде всего психометрического подхода к психодиагностике, в сочетании с крайне осторожными попытками использования "качественных тестов" в патопсихологии (см. С.Я. Рубинштейн, 1970) и нестандартизированных "проб" в нейропсихологии (Л.Р. Лурия, 1973), а также в тех специальных областях, в которых на то было дано "высочайшее изволение" (авиационная и космическая медицина, спортивная психология и ряд других областей, где СССР вынужден был - для обеспечения реальной конкурентоспособности в мире применять научные методы отбора и оценки кандидатов - прежде всего летчиков и операторов на сложных и ответственных производствах).

Первые специализированные научные конференции по психодиагностике были организованы только в 70-е годы. Показательно, что они состоялись в Таллинне, а не в Москве. Только в 1982 г. под редакцией К. М. Гуревича и В. И. Лубовского впервые вышел в русском переводе учебник Л. Анастаси "Психологическое тестирование".

В это же время начали появляться и вполне современные работы по клинической психодиагностике (Е.Т. Соколова, Б.Ф. Бурлачук), по психометрике (В.С. Аванесов, В.М. Блейхер, В.К. Гайда, Ю.З. Гильбух, А.

Г. Шмелев, Л.Т. Ямпольский), первые частично адаптированные версии западных тестов (Ф.Б. Березин, Л.Н. Собчик, И.Н. Гильяшева), оригинальные методики диагностики одаренности и интеллектуального развития (Д.Б. Богоявленская, Л.А. Венгер, Я.А. Пономарев), психологической совместимости (Ф.Д. Горбов, Н.Н. Обозов), мотивации (Ю.М. Орлов), характера (Л.Е. Личко) и самосознания личности (В.В. Столин), социально-психологических отношений в коллективе (А.В. Петровский и его последователи), других важных психических свойств и отношений.

В науке наметился позитивный сдвиг. Но разрыв сферы науки и сферы практики, характерный для всей советской науки, пагубно отразился на психодиагностической практике. В СССР не производилось практически никакой систематической работы по селекции и сертификации психологических тестов. Тесты публиковались и распространялись без всякой системы, зачастую по принципам "самиздата" и являлись объектом пиратского копирования.

Наряду с относительно качественными и психометрически корректными версиями западных тестов (например, 16PF Кеттелла в адаптации А. Г. Шмелева) среди практических психологов, имеющих крайне низкий уровень психометрической подготовки, широко распространились многочисленные "параллельные" переводы западных тестов, не опирающиеся не только на проверенные "ключи" к этим тестам, но не имеющие даже отечественных норм.

Только после распада СССР в широкой публицистической литературе нашла отражение та истинная оценка, которую давали отечественные психологи политике КПСС в области психодиагностики. Принятая в однопартийном идеологизированном государстве модель кадрового менеджмента была, безусловно, по определению несовместима с использованием научно-разработанных

психодиагностических методик.

Но дело даже не в искусственной полной занятости, в отсутствии рынка избыточной рабочей силы (резервной "армии труда"). Дело в целенаправленном и сознательном насаждении "номенклатурных" методов выдвижения кадров "сверху", при котором принцип личной преданности выдвигаемых по отношению к выдвигающим ставился безусловно выше объективных деловых и личностных качеств.

Искусственное элиминирование легальных форм; конкуренции позволило десятилетиями скрывать фактический процесс перерождения элиты, чрезвычайного падения интеллектуального и исполнительского уровня руководителей прежде всего высшего управленческого звена. В этих условиях всякий намек на появление каких-то методов объективной оценки деловых и личных качеств вызывал неизбежное солидарное сопротивление правящего клана номенклатурных начальников.

Последствия такого беспрецедентно низкого, почти нелегального статуса в обществе, какого психодиагностика не имела, по-видимому, ни в одной из ныне развитых стран, безусловно, сказываются на уровне ее развития в нашей стране до сих пор.

Контрольные вопросы для самопроверки

§5. Психологическая диагностика в России

1. Кто является родоначальником отечественной психодиагностики?
2. Какие основные причины вызвали развитие психодиагностики в России?
3. Какие отечественные психологи внесли наибольший вклад в становление психодиагностики в России? Дайте каждому из них краткую содержательную характеристику?
4. В чем особенности развития психодиагностики в Советский период?

Литература к теме

1. Бурлачук Л.Ф., Морозов С.М. Словарь-справочник по психодиагностике. - СПб.: Изд-во "Питер", 1999. - 528 с.
2. Общая психодиагностика / Под ред. А. А. Бодалева, В. В. Столина.- М., 1987.
3. Основы психодиагностики / Под ред. А. Г. Шмелева. Ростов-на-Дону., 1996.
4. Психологическая диагностика: Учебное пособие / Под ред. К.М. Гуревича и Е.М. Борисовой. - М.: Изд-во УРАО, 1997. - 304 с.

Тема 3: "Основы конструирования психодиагностических методик".

§ 1. Статистическая природа тестовых шкал.

Любые первичные результаты, полученные с помощью какого-либо психологического теста, остаются бессмысленными, если отсутствует дополнительная информация, позволяющая их интерпретировать. Чаще всего результаты психологического теста можно проинтерпретировать, сопоставляя их с нормами выполнения этого теста группой лиц, которую принято называть *выборкой стандартизации*. Следовательно, в психодиагностике нормы устанавливаются опытным путем, исходя из того, как выполняет задания теста некоторая (репрезентативная) группа испытуемых. Зная распределение показателей в такой репрезентативной группе, можно выяснить место каждого индивидуального показателя (т.е. определить, отвечает ли он среднему значению, выше или ниже его).

Для того чтобы более точно определить место испытуемого относительно выборки стандартизации, полученный результат необходимо перевести в некую относительную меру – найти *производный показатель*. Производные показатели выполняют *две основные функции* [1]:

- позволяют *сравнивать между собой результаты, полученные по разным тестам* (или по разным шкалам теста).
- позволяют *определить положение каждого обследуемого относительно выборки стандартизации* и таким образом оценить его выполнение теста на фоне других испытуемых;

Примером простейшего производного показателя является обычный процентный показатель, однако в психодиагностике он не нашел применения, так как сам по себе процент правильно выполненных испытуемым заданий теста нельзя интерпретировать без другой дополнительной информации, например, без учета трудности заданий теста.

Производные показатели могут выражаться в одних и тех же единицах и относиться к одним и тем же или похожим нормативным выборкам для различных тестов. Поэтому, оказывается возможным сравнение индивидуальных относительных данных по многим различным психическим функциям.

Производные показатели можно получить разными путями (в соответствии с их функциями):

- определение достигнутого уровня развития;
- установление относительного положения индивида в некоторой группе.

Существующие типы производных показателей, будут рассмотрены ниже. Прежде необходимо дать дефиниции некоторым понятиями (в том числе и статистическим), лежащими в основе разработки и использования тестовых норм.

По определению, данному в словаре [3], *психологический тест* –это фиксированное во времени испытание, предназначенное для установления количественных (и качественных) индивидуально-психологических различий.

Наиболее типичное понятие теста в психодиагностике можно представить так [15]: *тест* - это последовательность кратких заданий (или пунктов), которая после выполнения их испытуемым, подвергается однозначной количественной интерпретации.

Исходы (ответы) испытуемых зависят от вида теста.

Например, в интеллектуальных тестах, состоящих из отдельных задач, ответы могут интерпретироваться так: "правильное решение", "ошибочное решение", "отсутствие ответа" (пропуск задачи из-за недостатка времени). Примеры интерпретации в личностных опросников, состоящих из высказываний, предлагаемых для подтверждения испытуемым: "подтверждение" (ответ "верно"), "отвержение" (ответы "не согласен", "неверно").

Суммарный балл по тесту подсчитывается с помощью *ключа*: ключ

устанавливает числовое значение исхода по каждому пункту. Например, за правильное решение задания - "+1", за неправильное или пропуск - "0". Тогда балл буквально выражает количество правильных ответов.

Известно [1, 15], что исход по отдельному заданию подвержен воздействию не только со стороны измеряемого фактора - способности или черты испытуемого, но и побочных шумовых факторов, которые являются иррелевантными по отношению к задаче измерения.

Примеры случайных факторов: колебания, вызванные неожиданными отвлекающими событиями (шум на улице, стук в дверь и т. п.), трудности в понимании смысла задания (вопроса), вызванные особенностями опыта данного конкретного испытуемого, и т. п.

Последовательность ответов оказывается последовательностью событий, содержащей постоянный и случайный компоненты, причем случайные компоненты могут оказывать как положительное, так и отрицательное влияние на результаты выполнения теста. Как известно, основным приемом, позволяющим устранить искажающее влияние случайных факторов на результат (суммарный балл), является балансирование этого влияния с помощью повторения. При этом фактически предполагается, что повторение обеспечивает *рандомизацию* (случайное варьирование) неконтролируемого фактора, в результате чего при суммировании исходов положительные и негативные эффекты случайных факторов взаимопоглощаются [15].

В оптимальном тесте набор заданий организуется таким образом, чтобы повысить долю постоянного компонента и сократить долю случайного в величине суммарного балла. Необходимо признать, однако, что несмотря на различные статистические ухищрения, суммарный балл в психологических измерениях содержит несравненно большую долю случайного компонента, чем в обычных физических измерениях. Поэтому, суммарный балл, набранный испытуемым по тесту, можно определить лишь в известных пределах, заданных ошибкой измерения.

Прежде, чем перейти к анализу типов кривых распределений суммарных баллов вспомним, что понимается в психологии под измерением и какие типы измерительных шкал существуют.

В психологии различают три основные процедуры измерения. Основанием для различения является *объект измерения*. Различают [7, 9]: "*измерение испытуемых*" (например, особенности их поведения); "*измерение, как задача для испытуемого*" (например, задача испытуемого состоит в ранжировании или классификации внешних объектов); *процедуру совместного измерения "стимулов и людей"* (например, поведение субъекта можно рассмотреть как проявление взаимодействия личности и ситуации). В узком смысле слова *психологическим измерением является только "измерение испытуемых"*.

С точки зрения математики, измерением называется операция установления взаимно однозначного соответствия множества объектов и символов (чисел); причем символы (числа) приписываются объектам по определенным правилам.

Правила, на основании которых происходит приписывание чисел объектам, определяют *шкалу измерения*. Различают множество различных видов шкал, однако наибольшее распространение в психологии получили следующие [1, 5, 15]: "*шкала наименований*", "*шкала порядка*", "*шкала интервалов*" и "*шкала отношений*". Основные их характеристики можно представить в виде таблицы 1.

Основная цель применения статистики - представить количественные данные в систематизированной и сжатой форме для того, чтобы облегчить их понимание.

Чтобы навести порядок в хаосе цифр, полученных в результате тестирования какой-то группы (например, в 1000 человек) нужно, прежде всего, составить таблицу *частотного распределения* (этих тестовых бал-

лов). Для этого показатели распределяются по заранее выбранным интервалам (классам) значений. После того, как тестовые показатели распределены по интервалам, подсчитывается их количество в каждом из них. Полученное таким способом число и есть *частота* (количество случаев) для соответствующего интервала.

Таблица 1. Измерительные шкалы.

Тип шкалы	Измерительная операция	Описательные статистические величины	Примеры
Наименования	Приписывание одинаковых чисел - наименований объектам, имеющим общий признак.	Число объектов в классе (категории).	Приписывание числового кода лицам с определенными социально-демографическими характеристиками.
Порядка	Ранжирование объектов по выраженности определенного признака.	Медиана, процентиля.	Ранжирование специалистов по степени профессиональной пригодности.
Интервалов	Определение величины различий между объектами.	Среднее арифметическое, стандартное квадратичное отклонение, коэффициент корреляции.	Шкала температуры Цельсия.
Отношений	Определение равенства отношений величин.	Коэффициент вариации.	Измерение длины, массы и т.д.

Как известно [1, 7, 9, 15], любое частотное распределение суммарных баллов можно представить графически в виде кривых распределений: *гистограммы, полигона частот и кумуляты*.

На *гистограмме* по оси абсцисс откладываются "границы классов", на которые разбиваются первичные показатели суммарных баллов, а высота столбцов (ось ординат), вычерченных над каждым интервалом, соответствует числу испытуемых, чьи результаты попали в этот интервал. (Их количество определяет высоту столбца в относительной частоте встречаемости показателя или в процентах). В *полигоне частот* число испытуемых

указывается точкой, расположенной над *серединой* интервала на высоте, соответствующей его частоте, а сами точки последовательно соединяются прямолинейными отрезками.

Если на оси абсцисс нанести срединные значения классов, а по оси ординат - *накопленные частоты* классов, то получится график называемый *кумулятой* [11].

Идеальная нормальная кривая это такая кривая, которая обладает важными математическими свойствами, на ней основаны многие виды статистического анализа. По существу эта кривая означает, что число случаев максимально в середине распределения и постепенно спадает к ее краям. Кривая симметрична и имеет единственный пик в центре. Причем, чем больше группа, тем ближе распределение к теоретической нормальной кривой.

Как отмечает А. Анастаси [1], группа тестовых показателей может быть описана в терминах той или иной *меры центральной тенденции*. Такая мера показывает наиболее типичный (или репрезентативный) результат, характеризующий выполнение теста всей группой. Самой известной из таких мер является *среднее (среднеарифметическое)* значение (X_{cp}). Другой мерой центральной тенденции является *мода* (M_o) или наиболее часто встречающийся результат. В частотном распределении мода определяется как середина интервала, для которого частота максимальна.

Третья мера центральной тенденции - это *медиана* (M_e), т.е. результат, находящийся в середине последовательности показателей, если их расположить в порядке возрастания или убывания. Медиана есть точка, делящая распределение ровно пополам, причем одна половина результатов лежит справа от нее, а другая слева.

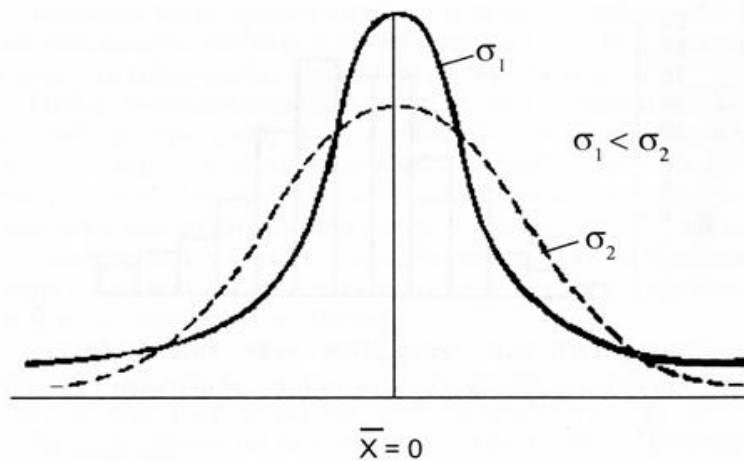


Рис. 1. Частотные распределения с одним и тем же средним значением, но с разным разбросом σ [по 1].

Для наиболее полного описания результатов теста используются *меры разброса* данных, показывающие степень отклонений индивидуальных показателей от центральной тенденции. Наиболее наглядным и известным способом представления разброса является *размах распределения*, т. е. разность между самым высоким и самым низким результатом. Но эта мера крайне точна и неустойчива, поскольку она определяется только двумя показателями. Единственный необычно высокий или низкий результат может заметно повлиять на величину размаха. Более точный метод измерения разброса основан на учете разности между каждым индивидуальным результатом и средним значением по группе.

Такой мерой разброса является *стандартное отклонение*, обозначаемое буквой " σ " (сигма - при ее вычислении отрицательные знаки устраняются благодаря возведению каждого отклонения в квадрат), которая вычисляется исходя из *дисперсии* (σ^2) или среднего квадрата отклонения.

Если *дисперсия* очень удобна при выяснении влияния различных факторов на индивидуальное выполнение тестовых заданий, то *стандартное отклонение* широко применяется как мера сравнения разбросов результатов измерений в различных группах. На рисунке 1, в качестве

примера, представлены два распределения, имеющие одно и то же среднее значение, но отличающиеся разбросом. Распределение, характеризующееся большими индивидуальными различиями, имеет большее σ .

Контрольные вопросы для самопроверки:

§1. Статистическая природа тестовых шкал.

1. Что понимается под термином "тест" в психодиагностике?
2. Для чего применяется повторное тестирование?
3. Что понимается под термином "измерение" в психодиагностике?
4. Какие типы шкал используются в психометрике?
5. Дайте определение, что в психометрике понимается под "гистограммой", "полигоном частот" и "кумулятой"?

Литература к теме.

1. Анастаси А. Психологическое тестирование. В 2-х кн. М., 1982.
2. Бурлачук Л.Ф., Морозов С.М. Словарь-справочник по психодиагностике. - СПб.: Изд-во "Питер", 1999. - 528 с.
3. Общая психодиагностика / Под ред. А. А. Бодалева, В. В. Столина.- М., 1987.
4. Основы психодиагностики / Под ред. А. Г. Шмелева. Ростов-на-Дону., 1996.

§2. Взаимосвязь свойств пунктов теста и видов распределений тестовых баллов.

Знание основных понятий и методов анализа заданий, а также этапов построения теста является необходимым и важным не только при конструировании новых тестов, но и при оценке опубликованных тестов. Кроме того, анализ заданий теста имеет прямое отношение к составлению неформальных, локальных тестов подобных тем, которые, например, готовят учителя для работы в классе в виде опросов или контрольных работ. К анализу заданий, включенных в тест, можно подходить как с точки зрения качественной оценки, так и с количественной точки зрения (т.е. изучать их статистические свойства).

Качественный анализ включает рассмотрение валидности по содержанию и оценку эффективности заданий (на них мы остановимся позже). *Количественный анализ* означает, в основном, измерение трудности (и некоторых видов валидности) заданий.

Валидность и надежность теста как целого зависят от свойств заданий, и их предварительный анализ позволяет повысить эти свойства теста. Устраняя, добавляя, заменяя или пересматривая отдельные задания можно усовершенствовать тест в целом [1].

Трудность заданий. Во всех физических измерениях калибровка шкалы измерительного инструмента производится на основе какого-либо эталона, роль которого выполняет физический объект, стабильно сохраняющий заданную величину измеряемого свойства (*например, метр, килограмм и пр.*).

В психометрике такие *физические* эталоны отсутствуют, так как мы не располагаем индивидами, которые были бы постоянными носителями заданной величины измеряемого свойства. Поэтому *роль косвенных эталонов в психометрике* отводится самим тестам: в том смысле, в каком трудность заданий теста можно рассматривать как величину, прямо пропорционально сопряженную со способностью (которую измеряют) - *чем труднее задание, тем выше должен быть уровень способности, требуемый для ее решения* [15].

По А. Анастази [1], чаще всего трудность задания определяется процентом испытуемых, давших правильный ответ. Чем легче задание, тем выше этот процент. *Например, слово, значение которого правильно указали 70% выборки стандартизации ($p=0,70$), считается более легким, чем слово, которое знают только 15% ($p=0,15$).* Обычно задания располагаются в порядке нарастания трудности, так, чтобы обследуемый начинал с относительно легких заданий и затем переходил к другим более слож-

ным.

Такое расположение дает индивиду больше уверенности снижает вероятность того, что он, затратив слишком много времени на задания, которые для него слишком трудны, упустит из вида те, которые ему по силам.

При составлении теста задания подбираются так, чтобы они соответствовали определенному уровню трудности. Большинство стандартизованных тестов способностей предназначены для как можно более точной оценки индивидуального уровня развития способности, и если в таком тесте никто не может справиться с заданием, (оно является слишком сложным), то это задание оказывается просто лишним в данном тесте. То же можно сказать и о заданиях, с которыми могут справиться все (очевидно, это будут слишком легкие задания).

Ни те, ни другие задания не несут никакой информации об индивидуальных различиях. А поскольку такие задания не влияют на вариативность тестовых результатов, они ничего не прибавляют к показателям надежности или валидности теста. Следовательно, чем ближе трудность задания к 1,0 (очень трудное задание) или к 0 (очень легкое задание), тем меньшей способностью к дифференцировке испытуемых обладает тест. И наоборот, чем ближе уровень трудности заданий к 0,5, тем выше его разрешающая способность.

Таким образом, выходит, что для максимальной дифференциации испытуемых все задания должны быть на уровне трудности 0,5. Решение, однако, осложняется тем фактом, что в пределах одного теста задания могут коррелировать друг с другом. Чем однороднее тест, тем выше эти корреляции.

В экстремальной ситуации, если все задания скоррелированы и имеют уровень трудности 0,50, то одни и те же 50% испытуемых спра-

вятся со всеми заданиями теста. В итоге одна половина обследованных покажет 100% результат, а результатом другой половины будет ноль. Ввиду этого, задания лучше всего выбирать так, чтобы уровень трудности отдельных заданий имел некоторый умеренный разброс, но в среднем составлял 0,50.

Трудность пунктов теста можно выявить только с помощью процедуры проведения теста. При этом, показателем трудности оказывается "процентильная мера": процент испытуемых, справившихся с заданием теста. Чем меньше процент, тем выше трудность.

Распределение результатов теста. Известно [1], что трудность теста в целом, зависит от трудности отдельных заданий, из которых он составлен. Проверку трудности всего теста применительно к популяции, для которой он предназначен, можно произвести с помощью построения распределения частот суммарных баллов (см. параграф выше).

Поскольку материал, описывающий различного рода «кривые распределения» относится к области математической статистики (и должен быть известен читателю), рассмотрим влияние трудности заданий на кривые распределения тестовых баллов в общих чертах.

Кривая распределения тестовых баллов отражает свойства пунктов, из которых составлен тест. Известно [11], что у нормальной кривой мода, медиана и средняя арифметическая совпадают. Наряду с практически симметричными (нормальными) распределениями встречаются и скошенные, асимметричные. Математически они характеризуются нарушением равенства между модой, медианой и средней арифметической выборки.

Принято различать [11], *правостороннюю*, или *отрицательную*, асимметрию и *левостороннюю*, или *положительную* асимметрию.

Если кривая имеет *левостороннюю положительную асимметрию*, это значит, что в тесте преобладают трудные задания. При положи-

тельной асимметрии частоты распределения суммарных баллов накапливаются больше в левой части распределения; вершина такого распределения сдвинута влево, мода оказывается больше средней арифметической.

Если кривая имеет *правостороннюю отрицательную асимметрию*, то значит большинство пунктов в тесте - легкие (слабые). При правосторонней отрицательной асимметрии варианты (набранные испытуемыми баллы) накапливаются больше в правой части распределения; вершина такого распределения сдвинута вправо, мода оказывается меньше средней арифметической.

Тесты первого типа плохо дифференцируют испытуемых с низким уровнем способностей: все эти испытуемые получают примерно одинаковый низкий балл. Тесты второго типа, наоборот, хуже дифференцируют испытуемых с высоким уровнем способностей.

Если пункты обладают оптимальным уровнем трудности, то форма кривой распределения будет зависеть от того, насколько пункты однородны. При этом если выборка репрезентативна срезу популяции, то можно ожидать, что эти результаты распределятся приблизительно по нормальной кривой.

Если тест содержит разнородные задания примерно равного уровня трудности (а именно такие задания и подбираются для измерения интегральных свойств личности - с широкой областью применения), то нормальность распределения суммарных баллов возникает автоматически - как артефакт самой процедуры подсчета суммарных баллов.

При этом, конечно, форма кривой распределения баллов не позволяет говорить о реальной форме распределения измеряемого свойства, каким оно является само по себе - в широкой популяции испытуемых. Нормальность распределения есть артефакт, прямое следствие направленного отбора пунктов с заданными свойствами [15].

Это понимается так: когда распределение результатов теста, полученное на выборке стандартизации, заметно отличается от нормального, обычно производится корректировка трудности заданий, пока не достигается приблизительно нормальная кривая. В процессе создания теста нормальность распределения рассматривается скорее как методологическая проблема, чем как эмпирически наблюдаемый факт [2]. Всякий раз, когда группа, на которой проводится стандартизация теста, показывает распределение, которое нельзя признать нормальным, обычной реакцией психолога являются действия по корректировке теста. Таким образом, большинство тестов, видоизменяется до тех пор, пока они не дадут в популяции, для которой они предназначены, распределение, приближенное к нормальному.

Некоторые пункты шкалы удаляются или добавляются, другие перемещались по шкале вверх или вниз; при этом исследователь каждый раз оценивает, к чему приводят подобные изменения, и, в конце концов, добивается желаемой приближенности распределения к нормальному виду. Поэтому, говоря, что данное распределение нормальное, чаще всего подразумевается, что был проведен процесс тщательной стандартизации данного теста. И наоборот, говоря, что данное распределение не соответствует норме, подразумевается только то, что тест получился неудачным или что тест применялся к группе, для которой он был не пригоден.

Если подбираются пункты, *высоко положительно коррелирующие между собой* (ответы на пункты не являются статистически независимыми), то в распределении баллов возникает отрицательный эксцесс. Максимальных величин отрицательный эксцесс достигает по мере возрастания вогнутости вершины распределения - до образования двух вершин - двух мод (с "провалом" между ними).

Бимодальная конфигурация распределения баллов указывает на

то, что выборка испытуемых разделилась на две категории (с плавными переходами между ними): одни справились с большинством заданий, другие - не справились.

Такая конфигурация распределения свидетельствует о том, что в основе пунктов лежит какой-то один общий им всем признак; соответствующий определенному свойству испытуемых: если у испытуемых есть это свойство (способность, умение, знание), то они справляются с большинством пунктов, если нет этого свойства - то не справляются.

В некоторых редких ситуациях пункты могут отрицательно коррелировать друг с другом. В этом случае на кривой возникает *положительный эксцесс*: вся масса эмпирических точек скучивается вблизи среднего значения.

Такое возможно в двух случаях:

во-первых, когда ключ составлен неверно - объединены при подсчете отрицательно связанные признаки, которые обуславливают взаимоуничтожение баллов;

во-вторых, когда испытуемые применяют, разгадав направленность опросника, специальную тактику "медианного балла" - искусственно балансируют ответы "за" и "против" одного из полюсов измеряемого качества.

Есть несколько причин, по которым разработчики тестов обычно стремятся получить нормальное распределение. Если делать допущения относительно распределения какого-либо человеческого качества, нормальный график в большинстве ситуаций является наиболее вероятным. Известная сложность и множественность факторов, определяющих то, как данное качество проявляется у индивида, заставляет ожидать, что оно будет распределяться в соответствии с теорией вероятности.

Другой причиной стремления к поиску соответствия с нормальным

распределением является та, что нормально распределенные данные позволяют подвергать их различным типам статистического анализа, который в противном случае оказывается неприменимым. Однако следует иметь в виду, что другие типы распределения могут быть предпочтительны для некоторых специальных целей и соответствующим образом использоваться [2].

Итак, когда в качестве единственного эталона измерения психодиагностами рассматривается сам тест, то в качестве меры измеряемого свойства выступает местоположение балла на кривой распределения.

Контрольные вопросы для самопроверки

§2. Взаимосвязь пунктов теста и видов распределений тестовых баллов

1. Что понимается под трудностью заданий в психодиагностике?
2. Какая должна быть оптимальная трудность заданий? Дайте обоснование Вашему ответу.
3. Как связана трудность заданий с формой распределения суммарных тестовых баллов?
4. Для чего и всегда ли необходимо преобразовывать "сырые" баллы в производные показатели?
5. Почему большинство тестов разрабатывается так, чтобы распределение тестовых баллов в выборке стандартизации подчинялось нормальному закону?

Литература к теме

1. Анастаси А. Психологическое тестирование. В 2-х кн. М., 1982.
2. Анастаси А. Дифференциальная психология. Индивидуальные и групповые различия в поведении / Пер. с англ. – М.: Апрель Пресс, Изд-во ЭКСМО-Пресс, 2001. – 752 с.
3. Бурлачук Л.Ф., Морозов С.М. Словарь-справочник по психодиагностике. - СПб.: Изд-во "Питер", 1999. - 528 с.
4. Общая психодиагностика / Под ред. А. А. Бодалева, В. В. Столина.- М., 1987.

§3. Понятие нормы. Проблемы стандартизации показателей.

Одним из основных понятий в психометрике является понятие *нормативного оценивания*, под которым понимается подход к оценке и интерпретации измеряемых тестом показателей, отражающих какие-либо особенности личности, путем сравнения индивидуальных результатов со статистическими значениями нормативной выборки. По сравнению с альтернативным принципом *ипсативного оценивания*, *нормативное оценивание* более распространено. Здесь наиболее полно проявляется измерительный, характер психометрических техник. Заметим, что под ипсативным оцениванием понимается измерение «относительно себя». Например, ипсативный тест личности, мог бы показать, что у индивида выше потребность достижения, чем потребность в аффилиации.

Основным *достоинством* нормативного оценивания является возможность строгого ранжирования испытуемых по результатам относительно выборочных данных, использования метрических шкал интервалов, относительная доступность интерпретации оценок пользователям психодиагностической информации.

Недостаток нормативного оценивания - эмпиричность рассчитываемых показателей, определенная условность перенесения выборочных данных на индивидуальное обследование [3].

Понятие нормы. Из основных понятий нормативного оценивания вытекает другое понятие – понятие *нормы*. Никакие результаты, полученные с помощью психологических тестов не имеют смысла до тех пор, пока они не будут сравнены с тестовыми нормами, которые возникают в процессе стандартизации нового теста, при прохождении через него большого количества испытуемых, представляющих ту популяцию, для которой данный тест был разработан. Затем полученные данные используются *как стандарт* для оценки показателей индивидов. Нормы могут быть выраже-

ны по-разному, например: как интеллектуальный возраст, как проценты или как стандартные значения, - но все они позволяют психологу определить «положение» испытуемого в выборке стандартизации.

Тестовые нормы - количественные и качественные критерии оценки результатов теста, позволяющие определить уровень достижений или степень выраженности психологических свойств, которые являются объектами измерения. В качестве таких критериев могут выступать как статистические показатели выборки стандартизации, так и различные признаки-симптомы, свидетельствующие о том или ином уровне выраженности диагностируемых качеств.

В психодиагностике наибольшее распространение получили *количественные тестовые нормы*, рассчитанные на основе определения средних величин (X_{cp}) и дисперсии (σ^2) в выборке стандартизации. Количественные тестовые нормы, упорядоченные в шкалы на основе процедур Z-преобразования (будут рассмотрены далее), содержатся в специальных таблицах, прилагаемых к руководствам по проведению тестирования. В этом виде тестовые нормы позволяют установить относительное место каждого индивидуального результата по сравнению с выборочными данными, выраженными в долях дисперсии. Такие количественные тестовые нормы наиболее типичны для тестов интеллекта, опросников личностных и др.

В проективных техниках, в связи с затруднениями, возникающими при формализации первичных оценок и вследствие сложности учитываемых диагностических показателей, количественные тестовые нормы мало распространены.

В виде качественных тестовые нормы могут выступить, например, стандартизированные наборы квалификационных требований к испытуемому, аналогичные шкалам умственного развития, либо специально разра-

ботанные для конкретного теста комплексы диагностических признаков. Качественные критерии выступают как нормативы, позволяющие отнести индивида к той или иной диагностической группе. Комплексы критериев тестовые нормы могут быть упорядочены в нормативные или порядковые шкалы. Одна и та же психодиагностическая методика может иметь количественные и качественные тестовые нормы, позволяющие взаимодополнить и обогатить интерпретацию результатов.

Для каждой возрастной группы тестовые нормы обычно рассчитываются отдельно (обязательное условие для тестов общих способностей, при этом закономерное изменение нормативных показателей здесь служит своеобразным критерием *валидности* теста). В клинической психодиагностике, иногда разрабатываются разные тестовые нормы для отдельных контингентов больных (например, для больных разных нозологий и нормальных). Реже встречается дифференциация норм в зависимости от пола, профессиональных особенностей, уровня образования и т. д.

Рестандартизация и определение норм для новой социоэтнической популяции является одним из основных и обязательных элементов *адаптации теста*.

Наиболее сложной проблемой определения тестовых норм является отбор и комплектация выборки нормирования. Выборка должна соответствовать по своему объему назначению методики. В психодиагностике действует следующая закономерность: *чем более генерализованной (широкой) является область применения теста, тем большим должно быть число испытуемых в нормативной выборке*. При этом центральное внимание должно уделяться обеспечению *репрезентативности выборки* по отношению к обследуемой тестом популяции. Поясним это положение.

Если ставится задача простого уменьшения погрешности измерения, то для этой цели вполне достаточна выборка стандартизации из 500 испы-

туемых. Однако, нельзя вынести никакого суждения относительно объема выборки не зная, для какой категории лиц предназначен тест. Если тест, например, рассчитан на общую популяцию старших школьников, то объем выборки стандартизации должен составлять как минимум около 10000 человек. Если тест рассчитан на какую-то специфическую выборку, например, спасателей или водолазов, то объем выборки стандартизации может быть значительно меньше. Можно утверждать, что *репрезентативность выборки является более важной характеристикой, чем ее размер*.

При одинаковых объемах стратифицированная выборка (выборка, разделенная на страты, классы) более эффективна, чем случайная (рандомизированная). Наиболее частыми критериями стратификации являются: социальное положение, регион проживания, возраст, пол, образовательный уровень и т.д. [3].

Итак, при формировании выборки стандартизации следует учитывать два ее важных фактора: *репрезентативность и объем*.

Рассмотрим несколько более подробно некоторые виды норм [1, 2].

1. Нормы развития.

Возрастные нормы. Независимо от способа выражения, показатели, основанные на возрастных нормах, довольно грубы и плохо поддаются точной статистической обработке. Тем не менее, они достаточно наглядны, особенно при клиническом обследовании, а также при решении ряда научных проблем.

Умственный возраст. Возрастные нормы используются и в тестах, не подразделенных на возрастные уровни (например, шкалы Бине-Симона). В этом случае нормы устанавливаются для значений первичного результата теста, такого, как общее число правильных ответов, время выполнения заданий, число ошибок или же некоторая комбинация таких показателей. Значения первичных результатов, полученных на выборке стан-

дартизации для каждого возрастного уровня, и составляют возрастные нормы такого теста. Например, средний показатель 8-летних детей является нормой для возраста 8 лет. Если показатель обследуемого равен среднему значению первичного результата для этого возраста, то его умственный возраст составляет 8 лет. Подобным образом могут быть выражены любые первичные показатели такого теста.

Эквивалентный класс. Показатели тестов достижений в обучении часто интерпретируются с помощью понятия эквивалентный класс, введение которого объясняется тем, что все тесты этого типа применяются для обследования учащихся. Например, говорят, что ученик достиг уровня VII класса по орфографии, уровня VIII класса по технике чтения и уровня V класса по арифметике. Подобная характеристика оказывается такой же наглядной, как и умственный возраст в тестах интеллекта. Нормы классов определяются подсчетом среднего первичного результата, полученного детьми соответствующего класса. Так, если среднее количество правильно решенных задач арифметического теста в выборке стандартизации четвероклассников равно 23, то число 23 соответствует эквивалентному IV классу. Промежуточные эквивалентные классы, представляющие как бы доли класса, обычно определяются путем интерполяции, хотя их можно получить и непосредственно, тестируя детей несколько раз в учебном году.

Несмотря на популярность, классные нормы имеют ряд *недостатков*. *Во-первых*, содержание обучения меняется от класса к классу. Поэтому такие нормы применимы только к общеобразовательным предметам, обучение которым ведется на всех уровнях, охватываемых данным тестом. *Во-вторых*, они неприменимы в старших классах, где многие предметы изучаются только один или два года. Даже если предмет преподается на протяжении всего обучения в школе, его значение может ме-

няться от класса к классу и, следовательно, скорость его изучения может быть различной. Иными словами, единицы шкал эквивалентных классов явно не равны друг другу, причем отсутствует определенная закономерность в их изменении для разных предметов.

Шкалы порядка. Другой подход к нормам развития связан с исследованиями по детской психологии. Наблюдения за развитием младенцев и дошкольников дали обширный материал, позволяющий описать типы поведения для различных актов по таким функциям, как моторика, сенсорика, речевое общение, формирование понятий. Например, можно отразить по месячный уровень развития ребенка (в каждой из четырех основных сфер его поведения: моторики, языка, адаптивного и личностно-социального поведения) в специально составленных таблицах развития.

Шкалы, разработанные по этим данным, являются *шкалами порядка* по тому, что этапы развития наступают в определенной последовательности, причем каждый новый этап предусматривает предварительное усвоение типов поведения, характерных для предыдущих этапов.

В целом, шкалы порядка предназначены для выяснения того, на какой стадии развития находится та или иная конкретная психическая функция. Слово «порядок», входящее в название данного типа шкал, указывает на существование единообразия в развитии, проходящем через последовательные стадии.

2. Внутригрупповые нормы.

Описания большинства современных стандартизованных тестов в той или иной форме содержат данные о внутригрупповых нормах. С их помощью индивидуальный результат оценивается в соответствии с выполнением наиболее сопоставимой нормативной группы. Например, первичный показатель ребенка сравнивается с показателями детей того же возраста или того же года обучения. В качестве примера таких показате-

лей можно упомянуть Z-показатели, станайны, процентиля и пр. (суть некоторых из них будет рассмотрена ниже). Внутригрупповые показатели имеют единый и четко определенный количественный смысл и могут быть использованы в рамках большинства методов статистического анализа.

Выбор конкретного вида показателя определяется удобством, привычностью и легкостью разработки норм. Большинство типов внутригрупповых производных показателей в принципе подобны друг другу, если они корректно введены и интерпретируются надлежащим образом. При соблюдении определенных статистических условий каждый из этих показателей может быть легко переведен в любой другой.

1. Специфические нормы.

Один из подходов к решению проблемы сопоставимости тестов (согласованию норм) заключается в составлении таблиц эквивалентности показателей разных тестов. Такие таблицы могут быть составлены эквипроцентильным методом, и тогда показатели считаются эквивалентными, если они имеют равные процентиля в данной группе. Другим возможным решением этой проблемы является стандартизация тестов для более узкой популяции, релевантной специфическим целям каждого теста. В таких случаях границы нормативной популяции должны быть четко определены и приведены вместе с нормами. Например, можно указать, что данные нормы относятся к управленческому персоналу крупных производственных фирм или к первокурсникам технических институтов.

Достаточно специализированные нормы желательно иметь для многих целей тестирования. Даже когда имеются репрезентативные нормы для широкой популяции, часто оказываются полезными так называемые *подгрупповые нормы*. Потребность в них возникает, когда показатели теста заметно меняются от одной подгруппы к другой. Сами подгруппы могут формироваться по признаку возраста, года обучения, типа школьной

программы, пола, географического региона, проживания в городе или в сельской местности, социоэкономического уровня и т.д. Именно характер использования теста определяет и наиболее существенный признак формирования подгрупп, и предпочтительность общих или специфических норм.

Следует также упомянуть о *локальных нормах*, которые нередко разрабатываются пользователями тестов для конкретных социальных единиц. Группы, к которым относятся такие нормы, еще более специфичны, чем даже обсуждавшиеся выше подгруппы. Например, предприниматель может выработать нормы, лучше отвечающие специфике предлагаемой работы; администрация университета - нормы, рассчитанные на обучение в данном университете; школа может оценивать выполнение тестов своими учениками в соответствии с собственным распределением показателей. Локальные нормы такого типа в большей степени, чем какие-либо другие, отвечают таким задачам тестирования, как предсказание достижений в труде или учебе, сравнение относительного уровня знаний детей по различным предметам, исследование динамики их развития [1].

Относительность норм. Относительность норм становится очевидна при попытках осуществить сравнения между тестами. Хорошо известно, что независимо от вида показателей тестов, их всегда следует приводить вместе с названием теста, в котором они получены, поскольку тестовые показатели нельзя интерпретировать в отрыве от конкретного теста. Например, если в результате тестирования Иванов получил $IQ = 90$, а Сидоров - $IQ = 110$, то без последующих разъяснений эти данные ни о чем не говорят. Взаимное положение результатов этих испытуемых может оказаться обратным, если им придется "поменяться" тестами, которые они проходили.

Точно так же позиция индивида по отношению к диагностируемым

психическим функциям может быть неверно истолкована из-за несопоставимости тестовых норм. Допустим, учащемуся были даны тесты на понимание слов и пространственное восприятие для определения уровня развития некоторых его навыков. Если первый из этих двух тестов стандартизован на случайной выборке учеников старших классов, а второй на группе мальчиков из ремесленного училища, то можно ошибочно заключить, что индивид гораздо более развит в вербальном, чем пространственном отношении, тогда как на самом деле может иметь место обратное соотношение.

Существуют три основные причины систематических изменений результатов, полученных одним и тем же индивидом в различных тестах.

Во-первых, тесты, даже если они одинаково называются, могут различаться по содержанию. Примеры тому - тесты интеллекта, обычно фигурирующих под одним и тем же именем. И это, несмотря на то, что одни из них включают в себя только вербальные задания, другие связаны с пространственными навыками, а третьи могут содержать вербальные, пространственные и числовые задания.

Во-вторых, могут оказаться несравнимыми единицы измерения. Например, если IQ одного теста построен при $\sigma = 12$, а другого, при $\sigma = 18$, то испытуемый, который в первом тесте получит IQ= 112, во втором, скорее всего, будет иметь IQ=118.

В-третьих, характер выборок стандартизации, использованных при определении норм для разных тестов, может оказаться различным. Очевидно, один и тот же индивид будет выглядеть лучше на фоне более слабой, чем более сильной группы.

Несопоставимость содержания тестов или единиц измерения обычно выявляется при рассмотрении самого теста или руководства по его использованию. Однако несоответствие нормативных выборок заметить гораздо труднее, по всей вероятности, оно и является причиной многих, не

поддающихся иному объяснению, расхождений в результатах теста.

Обычно показатели конкретного испытуемого сравниваются с нормативными показателями посредством какого-либо преобразования, позволяющего определить место положения этого испытуемого в группе.

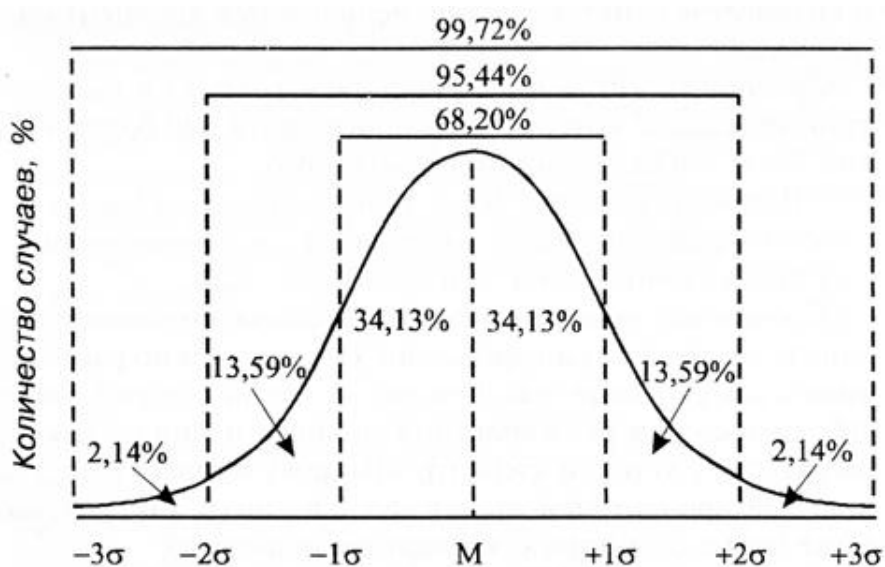


Рис. 2. Процентное распределение случаев по нормальной кривой

Прежде, чем рассматривать такие преобразования, напомним кратко некоторые свойства нормального распределения (см. рисунок 2).

Известно [1], что, используя показатель σ (стандартное отклонение) применительно к нормальной кривой распределения, можно представить прямое соответствие между σ и относительным количеством случаев (допустим правильных ответов на тест):

Например, для $X = 40$ и $\sigma = 4,9$ имеем интервал равный $+1\sigma = 44,9$; $+2\sigma = 49,8$. Процент случаев, приходящихся на интервал между X и $+1\sigma$, для нормального распределения равен 34,13. Так как кривая симметрична, 34,13% случаев приходится также и на интервал от X до -1σ , таким образом, в диапазоне от -1σ до $+1\sigma$ приходится 68,26% случаев. Почти все (99,72%) случаи лежат в пределах $+3\sigma$ относительно среднего значения.

Различают *первичные* или "*сырые*" показатели, которые получаются непосредственно после того, как проведен тест и подсчитаны суммарные баллы, и "*производные*" показатели – полученные из первичных, "сырых" показателей путем применения к ним каких-либо математических проце-

дур.

Процентиль. В качестве одного из универсальных производных показателей, пригодных для разных (по своей качественной направленности и количеству пунктов) тестов, используется "процентильная мера".

Процентиль - процент испытуемых из выборки стандартизации, которые получили равный или более низкий балл, чем балл данного испытуемого.

Например, если 30% людей правильно решают 5 задач в тесте на пространственное воображение, то первичному показателю 5 соответствует 30-й процентиль (P30).

Процентили указывают на относительное положение индивида в выборке стандартизации. (Их еще можно рассматривать как ранговые градации, общее число которых равно 100, однако, при ранжировании принято начинать отсчет сверху, т.е. с лучшего члена группы, получающего ранг 1; в то время как, в случае процентилей отсчет ведется снизу, так что чем ниже процентиль, тем хуже позиция индивида).

50-й процентиль (*P50*) соответствует медиане. Процентили свыше 50 представляют показатели выше среднего, а те, которые лежат ниже 50, - сравнительно низкие показатели.

25-й и 75-й процентили известны также под названием *1-го и 3-го квартилей (Q1 и Q3)*, поскольку они выделяют нижнюю и верхнюю четверти распределения. Как и медиана, они удобны для описания распределения показателей и сравнения с другими распределениями.

Процентили не следует смешивать с обычными процентными показателями, которые являются *первичными* показателями и представляют собой процент правильно выполненных заданий, тогда как *процентиль* - это *производный показатель*, указывающий на долю от общего числа членов группы.

Процентильные показатели обладают рядом достоинств. Их легко

рассчитать и понять даже сравнительно неподготовленному человеку. Их применение достаточно универсально, они одинаково применимы как к детям, так и взрослым и подходят к любому типу теста, измеряет ли он способности или свойства личности.

Процентильные показатели имеют существенные недостатки. Первый - связан с неравенством их как единиц измерения, особенно на краях распределения. Если распределение первичных показателей приближается к нормальной кривой, (что справедливо для большинства тестовых показателей) то различия между первичными показателями вблизи медианы (или центра) распределения в процентильном выражении преувеличены, тогда как аналогичные различия вблизи краев распределения сильно занижены.

Напомним, что в нормальной кривой случаи тесно сгруппированы в центре и по мере приближения к краям рассеиваются. Следовательно, каждый данный процент случаев вблизи центра соответствует более короткому расстоянию по оси абсцисс, чем тот же процент ближе к краям распределения.

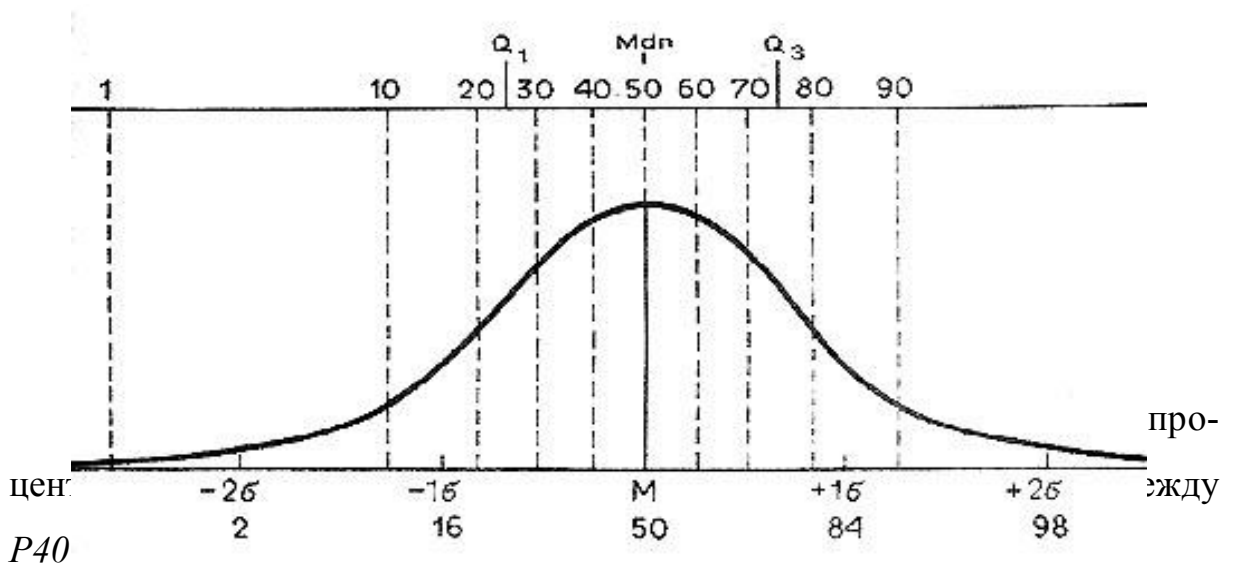


Рис. 3. Ранги процентилей при нормальном распределении

Еще более явно расхождение между этими расстояниями при *P10*

и $P1$. (В теоретической нормальной кривой нулевой процентиль достигается лишь в бесконечности и поэтому не может быть показан на графике).

То же соотношение получится, если процентилю отмечать интервалами одинаковой длины σ , откладывая их влево и вправо от пика нормальной кривой. Такие процентилю обозначены в нижней части рисунка 3. Видно, что разность процентилю между пиком и $+1\sigma$ равна 34 (84-50), а между $+1\sigma$ и $+2\sigma$ - всего 14 (98 - 84).

Второй недостаток: как показатели, процентилю нельзя использовать для последующего статистического анализа, поскольку они являются значениями порядковой шкалы.

Итак, процентилю показывают относительное положение каждого индивида в нормативной выборке, а не величину различия между результатами.

Стандартные показатели. Преимущественное использование в тестах стандартных показателей объясняется их пригодностью во многих отношениях. Такие показатели выражают отклонение индивидуального результата от средней нормы в единицах, пропорциональных стандартному отклонению распределения [1].

Стандартные показатели могут быть получены как линейным, так и нелинейным преобразованием первичных показателей. Если используется *линейное преобразование*, то при этом сохраняются соотношения между первичными показателями, поскольку они вычисляются вычитанием из каждого первичного показателя одной и той же величины с последующим делением результата на другую постоянную величину.

Относительная величина разницы между стандартными показателями, полученными при таком линейном преобразовании, в точности соответствует относительной величине различия первичных показате-

лей.

Все свойства первоначального распределения показателей полностью воспроизводятся в распределении стандартных линейных показателей. По этой причине любые вычисления, которые можно производить с исходными данными, могут также выполняться и с линейными стандартными показателями без какого-либо искажения конечных результатов.

Линейно преобразованные стандартные показатели часто именуется просто как стандартный показатель или Z - показатель. Чтобы вычислить Z, находят разность между индивидуальным первичным результатом и средним значением для нормативной группы и затем делят эту разность на σ нормативной группы (см. таблицу 2).

Таблица 2. Пример вычисления значений стандартного показателя Z

$Z_i = \frac{x_i - \bar{X}}{\sigma}$	$\bar{X} = 60$	$\sigma = 5$
Испытуемый 1 $x_1 = 65$		Испытуемый 2 $x_2 = 58$
$Z_i = \frac{65 - 60}{5} = +1,0$		$Z_i = \frac{58 - 60}{5} = -0,4$

Очевидно, что получающиеся при таком вычислении отрицательные показатели означают, что выполнение тестов индивидом было ниже среднего. Более того, поскольку для большинства групп область значений уместается в пределах от 3σ ниже и выше среднего значения, удовлетворительное различие индивидуальных показателей возможно, только если Z вычисляется с точностью хотя бы до одной десятой. Все это делает показатель Z неудобным для вычислений и сообщений результатов.

Поэтому обычно применяется еще одно линейное преобразова-

ние, единственная цель которого придать показателям более удобную форму. Стандартное преобразование Z-показателей можно представить в виде формулы:

$$Z_t = \bar{X} + \sigma Z \quad (**)$$

где Z_t - преобразованный Z-показатель, \bar{X} – среднее значение преобразованного распределения, σ – стандартное отклонение преобразованного распределения.

Таким образом, чтобы перевести Z-показатель в новую шкалу, необходимо просто умножить его на выбранную величину σ , и полученное произведение прибавить (с учетом знака при Z) к выбранному среднему значению \bar{X} .

При желании в качестве \bar{X} и σ можно выбрать любые удобные значения; например, показатели отдельных субтестов в шкалах интеллекта Векслера преобразуются так, что $\bar{X}=10$, а $\sigma=3$. Считается (16), что для тестов с распределением баллов, если не полностью нормальным, то более или менее симметричным, преобразованные Z_t -показатели со средним значением $\bar{X} = 50$ и с $\sigma = 10$ являются точной значимой нормой.

Все эти меры служат образцами линейного преобразования стандартных показателей.

Напомним, что *одной из причин введения производной шкалы вместо первичных показателей является стремление к сопоставимости показателей различных тестов*. Значения только что рассмотренных линейно преобразованных стандартных показателей сопоставимы только, *если их исходные распределения имеют приблизительно одну и ту же форму*.

В этих условиях результат, соответствующий, скажем, $+1\sigma$ над средним в каких-либо двух тестах, означает, что индивид занимает по отношению к обеим нормативным группам одно и то же положение.

Чтобы добиться сопоставимости результатов, принадлежащих к распределениям различной формы, можно применить *нелинейное преобразование*, позволяющее придать распределению форму заданной кривой. В качестве эталона обычно используется нормальное распределение, хотя при определенных обстоятельствах другой тип распределения может оказаться более пригодным.

Одним из главных доводов в пользу такого выбора является то, что большинство распределений первичных показателей ближе к нормальному, чем к какому-либо иному. Более того, физические характеристики организма, такие, как рост и вес, измеряющиеся в шкалах с равными единицами, определенными на основе некоторых физических операций, обычно имеют нормальное распределение.

Нормализованные стандартные показатели - это стандартные показатели, соответствующие распределению, преобразованному так, что оно принимает вид нормальной кривой.

Их значения могут быть найдены с помощью таблиц, в которых приводится процент случаев различных отклонений в единицах σ от среднего значения для нормальной кривой. При этом сначала определяется процент лиц в нормативной выборке с тем же или более высоким первичным результатом. Затем этот процент отыскивают в таблице нормального распределения частот и по нему находят соответствующее значение нормализованного стандартного показателя.

Нормализованные стандартные показатели имеют ту же форму, что и линейно преобразованные стандартные показатели, т.е. при среднем значении они равны 0, а при стандартном отклонении равны 1.

Таким образом, значение 0 нормализованный показатель принимает в случае, если индивидуальный результат приходится на самую середину нормальной кривой, т.е. превосходит 50% результатов группы. Ре-

результат «-1» означает, что он превосходит приблизительно 16% результатов группы, а «+1» - 84%. Эти проценты соответствуют точкам, лежащим на 1σ ниже и выше среднего значения нормальной кривой.

Как и при линейном преобразовании, нормализованным стандартным показателям можно придать любую удобную форму (используя приведенную выше формулу **).

Например, в руководстве по конструированию психологических тестов (США) считается [8], что типичным преобразованием ненормализованных стандартных показателей должно быть приведение их к распределению со средним значением $\bar{X} = 50$ и стандартным отклонением $\sigma = 10$. В результате получается Т-показатель, предложенный впервые Мак-Коллом: на шкале Т число 50 соответствует среднему значению; 60 показатель в 1σ над средним и т.д.

Таким образом, Т-показатели это нормально распределенные стандартные показатели со стандартным отклонением 10.

Еще одним достаточно известным преобразованием является *шкала станайн*, разработанная во время второй мировой войны для использования военно-воздушными силами США.

В этой шкале используются только однозначные числа. Среднее значение показателя равно 5, а σ - примерно 2. Название станайн (сокращение от *standart nine*, т. е. стандартная девятка) связано с тем, что этот показатель принимает значения от 1 до 9. Использование однозначных чисел удобно для машинной обработки, поскольку каждый показатель занимает на перфокарте всего один столбец.

Первичные показатели легко преобразуются в станайны упорядочиванием их числовых значений и приписыванием им новых значений в соответствии с нормальной кривой процентов, приведенной в таблице 3.

Например, если в группе ровно 100 человек, то 4 из них имеющие низшие показатели, получают станайн 1, следующие 7 - станайн 2, следующие 12 - станайн 3 и т. д. Если группа состоит из большего или меньшего числа случаев, то предварительно выясняется, скольким из них соответствует каждый из выписанных в табл. 4 процентов. Так, при 200

случаях станаин 1 будет приписан 8 случаям (4%, 200), а при 150 случаях -6 (4% от 150).

Таблица 3. Значения процентов нормального распределения для перевода первичных результатов теста в станаины.

Процент	4	7	12	17	20	17	12	7	4
Станайн	1	2	3	4	5	6	7	8	9

Эта таблица была составлена для перевода рядов случаев непосредственно в станаины для любой группы от 10 до 100 случаев. Станаины, ввиду их практических и теоретических достоинств, находят достаточно широкое применение, особенно в тестах способностей и достижений.

Хотя нормализованные стандартные показатели отвечают основным целям тестирования, тем не менее, имеются определенные технические возражения против нормализации всех распределений подряд. Такое преобразование следует проводить при наличии большой и репрезентативной выборки, когда есть основания считать, что отклонение распределения от нормального произошло в силу определенных дефектов теста, а не особенностей выборки или действия других факторов, влияющих на исследуемую функцию. Следует также отметить, что, когда исходное распределение первичных показателей приближается к нормальному, линейные и нормализованные стандартные показатели мало будут отличаться друг от друга. Хотя методы получения этих двух типов показателей совершенно различны, сами показатели в таких условиях будут почти тождественны [1, 15].

В целом, если это возможно, следует предпочесть такую нормализацию распределения, которая достигается надлежащей коррекцией уровней трудности тестовых заданий, а не путем последующего преобразования явно ненормального распределения. При наличии приблизительно нормального распределения первичных показателей линейные стандартные

показатели будут служить тем же целям, что и нормализованные стандартные показатели.

Итак, в качестве источника конкретной меры выступает *нормативная выборка* (выборка стандартизации), на которой построено нормативное распределение тестовых баллов. Как *базовые*, процентильные шкалы лежат в основе всех традиционных шкал, применяемых в тестологии (Т - очки ММРІ; баллы IQ; стенов 16 PF и др.).

Приведем примеры параметров наиболее популярных стандартных шкал:

1) Т-шкала Мак-Колла (тест-опросник ММРІ и др. тесты, где среднее равно $X = 50$, а $\sigma = 10$)

2) Шкала IQ: $X=100$ и $\sigma=15$,

3) Шкала "стеннайнов" (целочисленные значения от 1 до 9 - стандартная девятка): $X=5,0$ и $\sigma=2$,

4) Шкала "стенов" (стандартная десятка, 16PF Кеттелла) предложена Р.Б. Кеттеллом. Этот способ представляет собой перевод исходных тестовых оценок в 10-балльную шкалу, путем разбиения оси значений тестовых оценок на 10 интервалов, соответствующих долям стандартного отклонения.

Для этого среднее арифметическое по группе принимается за среднюю точку и ей приписывается значение, равное 5,5 балла по стандартной десятибалльной системе. Всякая оценка в интервале $(\bar{X} + 0,25\sigma)$ переводится в 6 стенов, а оценка $(\bar{X} - 0,25\sigma)$ дает стенов равный 5. Любое дальнейшее увеличение или уменьшение тестовой оценки на $0,5\sigma$ увеличивает или уменьшает стандартную оценку на 1 стенов.

При такой системе стандартизации к среднему диапазону (или к норме) принято относить стандартные оценки от 4 до 7 стенов. Только при получении стандартных оценок до 3 стенов и выше 8 стенов следует гово-

ритель о значимых отклонениях, выходящих за границы средней нормы [13].

Для наглядности приведем образец такой таблицы для фактора "А" - опросника 16PF:

Сырые баллы: 0-4 5-6 7 8-9 10-12 13 14-15 16 17-18 19-20

Стены: 1 2 3 4 5 6 7 8 9 10

Применение стандартных шкал позволяет прибегать на практике к более грубым, приближенным способам проверки типа распределения тестовых баллов. Если, например, процентильная нормализация с переводом в стены и линейная нормализация с переводом в стены по формуле («Z») дают совпадающие целые значения стен для каждого « Y_i », то это означает, что распределение обладает нормальностью с точностью до «стандартной десятки».

Применение стандартных шкал необходимо для соотнесения результатов по разным тестам, для построения "диагностических профилей" по батарее тестов и тому подобных целей.

Подчеркнем еще раз, что с точки зрения теории измерений процентильные шкалы относятся к *порядковым шкалам*: они дают информацию, у кого из испытуемых сильнее выражено измеряемое свойство, но ничего не позволяют судить о том, насколько или во сколько раз сильнее.

Для того, чтобы строить на базе таких шкал количественный прогноз, нужно повысить уровень измерения. Переход к *шкалам интервалов* производится или на базе эмпирического распределения, или на базе произвольной модели теоретического распределения. В большинстве случаев в роли такой теоретической модели оказывается модель нормального распределения, хотя в общем случае может быть использована любая модель [1, 15].

В дифференциальной психометрике нередко используются еще два

вида шкал (и соответственно два вида тестовых норм):

во-первых, это то, что можно условно назвать "абсолютными тестовыми нормами" - при этом, в роли шкалы выступает сама шкала "сырых" очков;

во-вторых, "критериальные" тестовые нормы, причем, применение таких норм можно считать оправданным в двух случаях:

а) когда сама тестовая "сырая" шкала имеет практический смысл (например, студент, изучающий иностранный язык, должен знать как можно больше слов этого языка, и сырой показатель лексического теста имеет практический смысл);

б) когда применяются "критериальные" тестовые нормы: сырой балл по тесту в результате эмпирических исследований связывается с заданной вероятностью успешности какой-то практической деятельности (например, вероятность успеха "критериальной" деятельности, таковой для упомянутого выше примера может быть синхронный перевод монолога и течение 30 минут).

Контрольные вопросы для самопроверки

§3. Понятие нормы. Проблемы стандартизации показателей.

1. Для чего необходимо рассчитывать тестовые нормы, и какие виды норм существуют?

2. Какие стандартные показатели вы знаете? Перечислите их и дайте краткую характеристику.

3. Чем отличаются линейные стандартные показатели от нелинейных?

4. Что понимается под "нормализованными" стандартными показателями?

5. Какие наиболее распространенные стандартные шкалы Вы знаете? Дайте им краткую содержательную характеристику.

Литература к теме

1. Анастаси А. Психологическое тестирование. В 2-х кн. М., 1982.

2. Бурлачук Л.Ф., Морозов С.М. Словарь-справочник по психодиаг-

ностике. - СПб.: Изд-во "Питер", 1999. - 528 с.

3. Общая психодиагностика / Под ред. А. А. Бодалева, В. В. Столина.- М., 1987.

4. Основы психодиагностики / Под ред. А. Г. Шмелева. Ростов-на-Дону., 1996.

§4. Устойчивость распределения.

Проверка устойчивости распределения. Общая логика проверки устойчивости распределения основывается на следующем индуктивном рассуждении: *если полученное по половине выборки распределение хорошо моделирует конфигурацию целого распределения, то можно предположить, что это целое распределение будет также хорошо моделировать распределение генеральной совокупности* [15].

Таким образом, доказательство устойчивости распределения означает *доказательство репрезентативности тестовых норм*. Традиционный способ доказательства устойчивости сводится к выяснению хорошего приближения эмпирического распределения к какому-либо теоретическому. Но если эмпирическое распределение не приближается к теоретическому, несмотря на значительное увеличение численности выборки, то приходится прибегать к более общему методу доказательства.

Его простейший вариант может быть сведен к следующим шагам:

- построить таблицу перевода "сырых" очков в нормализованную шкалу по данным всей выборки;

- затем, применить эту таблицу для каждого испытуемого из половины выборки (при этом под "половиной" выборки подразумевается случайная половина, в которую испытуемые зачисляются случайным образом);

- построить распределение нормализованных баллов для половины выборки;

- проверить приближается ли это распределение к нормальному (близость к нормальному распределению проверяется с помощью критерия Колмогорова, при $n < 200$ целесообразно использовать более мощные критерии "хи-квадрат" или другие);

- сделать вывод: если распределение нормализованных баллов из половины выборки хорошо приближается к нормальному, то это значит, что заданные таблицами нормализации тестовые нормы определены устойчиво.

В общем случае такой простейший метод установлении однородности двух эмпирических распределений может быть применен и при разбиении выборки по какому-либо систематическому признаку. Если, в частности, по какому-либо из популяционно-значимых признаков (пол, возраст, образование, профессия) психолог получает значимую неоднородность эмпирических распределений, то это значит, что относительно данных популяционных категорий тестовые нормы должны быть специализированы (одна таблица норм - для мужчин, другая - для женщин и т. д.).

Более корректный метод статистической проверки однородности двух распределений, полученных при расщеплении выборки на равные части, опять же связан с применением критерия Колмогорова. Для этого с табличным значением сравнивается величина:

$$K_e = \max |F_{i1} - F_{i2}| \sqrt{\frac{n}{4}},$$

где F_{i1} - кумулятивная относительная частота для i -того интервала шкалы по первой половине выборки;

F_{i2} - та же частота для второй половины;

n - численность полной выборки;

K_e - эмпирическое значение статистики Колмогорова.

Точные значения квантилей распределения Колмогорова для определения размеров выборки можно найти в справочниках по статистике. Применение критерия Колмогорова не зависит от нормальности целого распределения.

Итак, априорная предпосылка нормальности распределения тестовых баллов основывается скорее на принципах операционального удобства, чем на теоретической необходимости. Психометрически корректные процедуры получения устойчивых тестовых норм возможны также с помощью специальных методов непараметрической статистики (критерий "хи-квадрат и т. п.) для распределений произвольной формы.

Выбор статистической модели распределения - полностью зависит от психометриста до тех пор, пока сам тест выступает в качестве единственного эталона измеряемого свойства. В этом случае остается лишь тщательно следить за соответствием сферы применения диагностических норм той выборке испытуемых, на которых они были получены. Произвольность в выборе статистической модели шкалы исчезает, когда речь заходит о внешних по отношению к тесту критериях [1, 15]. Рассмотрим, в связи с этим, репрезентативность критериальных тестов.

§5. Репрезентативность критериальных тестов.

В тестах по критерию в качестве реального эталона применяется критерий (как правило, это какой-то социальный норматив), ради которого создается тест - целевой критерий.

Такая ситуация имеет место в обучении: тестирование, направленное на получение информации об уровне усвоения определенных знаний, умений и навыков (при профессиональном обучении), должно точно отражать уровень освоения этих навыков и тем самым давать надежный прогноз эффективности конкретной профессиональной деятель-

ности.

Рассмотрим операциональную схему шкалирования, применяемую при создании критериального теста [15]. Пусть имеется некоторый критерий C , ради прогнозирования которого психодиагност создает какой-то тест X .

Для простоты представим C как дихотомическую переменную с двумя значениями - «1» и «0» (например, поступит выпускник школы в вуз или нет). $C=1$ означает, что i -тый субъект достиг критерия (попал в "высокую" группу по критерию, т.е. поступил в вуз), $C=0$ означает, что i -тый субъект не достиг критерия (попал в "низкую" группу, т.е. не поступил).

Затем психолог проводит тест X на интересующей его выборке, и в результате каждый индивид (школьник) получает определенный тестовый балл X_i . После того как для каждого индивида из выборки становится известным значение C (например, поступил или нет в вуз), психодиагност располагает всех индивидов по порядку возрастания балла X_i и для каждого деления исходной шкалы сырых тестовых баллов подсчитывает с какой вероятностью (P) каждый индивид (школьник) может попасть в "высокую" группу по критерию.

На рисунке 4 показаны распределения вероятности $P(C=1)$ в зависимости от X_i . Очевидно, что кривая, представленная на рисунке 1, является эмпирической линией регрессии C по X_i .

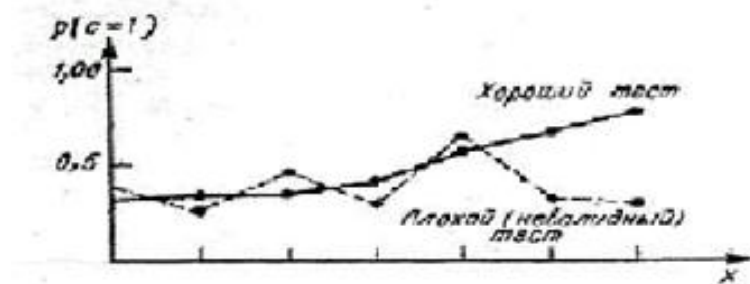


Рис. 4 Иллюстрация эмпирической зависимости между вероятностью критериального события $P(C=1)$ и величиной X тестового балла

Теперь можно сформулировать основное требование к критериальному тесту: линия регрессии должна быть монотонной функцией C от X . Иными словами, ни для одного более высокого значения X_i вероятность P не должна быть меньшей, чем для какого-либо менее высокого значения X_i .

Если это условие выполняется, то открывается возможность для критериального шкалирования сырых баллов X . При критериальном шкалировании к делениям сырой шкалы X применяется поточечный перевод прямо в шкалу P на основании эмпирической линии регрессии.

Например, если испытуемый А получил по тесту X 18 "сырых" очков и этому результату соответствует $P=0,6$, то испытуемому А ставится в соответствие показатель 60%.

Конечно, любая эмпирическая кривая является лишь приближенной моделью той зависимости, которая могла бы быть воспроизведена на генеральной совокупности.

Поэтому обычно предпринимаются попытки аппроксимировать эмпирическую линию регрессии какой-либо функциональной зависимостью, что позволяет затем производить прогноз с применением формулы (а не таблицы или графика).

Например, если линия регрессии имеет вид приблизительно такой,

какой изображен на рис.6, то применение процентильной нормализации позволяет получить простую линейную регрессию S по нормализованной шкале Z .

Это как раз тот случай, когда имеет место эквивалентность стратегии, использующей выборочно-статистические тестовые нормы, и стратегии, использующей критериальные нормы.

В завершении данного материала, рассмотрим операции по анализу распределения тестовых баллов, построению тестовых норм и проверке их репрезентативности.

Действия, которые последовательно должен произвести психолог при построении тестовых норм сводятся к такому алгоритму [15]:

1. Сформировать выборку стандартизации (случайную или стратифицированную по какому-либо параметру) из популяции, на которой предполагается применять тест. Провести на каждом испытуемом из выборки тест в сжатые сроки (чтобы устранить иррелевантный разброс, вызванный внешними событиями, происшедшими за время обследования).

2. Произвести группировку "сырых" баллов с учетом выбранного интервала квантования (интервала равнозначности). Интервал определяется величиной W/m , где $W = X_{\max} - X_{\min}$ - размах; m - количество интервалов шкалы. Необходимо иметь в виду, что группировка индивидуальных баллов, набранных испытуемыми по тесту, сказывается не только на распределении баллов по классам (интервалам квантования), но и на средних показателях, характеризующих то или иное измеряемое свойство. Экспериментальные данные показывают [11], что совершенно не безразлично, не только то, на какое число классов разбивается вся совокупность полученных тестовых баллов, но и как определяются границы первого класса.

Чтобы получились более адекватные результаты обработки статистических данных, рекомендуется:

- *во-первых*, не брать слишком большой интервал шкалы - он не должен превышать $1/3-1/4$ величины среднего квадратичного отклонения;

- *во-вторых*, определять границы первого класса следует так, чтобы минимальный показатель набранных по тесту баллов, попадал примерно в середину этого класса.

3. Построить распределение частот тестовых баллов в виде таблицы и в виде соответствующих графиков гистограммы и кумуляты.

4. Произвести расчет среднего и стандартного отклонения, а также асимметрии и эксцесса с помощью компьютера. Проверить гипотезы о значимости асимметрии и эксцесса. Сравнить результаты проверки с визуальным анализом кривых распределения.

5. Произвести проверку нормальности одного из распределений (можно с помощью критерия Колмогорова или с одним из указанных способов) или произвести процентильную нормализацию с переводом в стандартную шкалу, а также линейную стандартизацию и сравнить их результаты (с точностью до целых значений стандартных очков).

6. Если совпадения не будет - нормальность отвергается, тогда произвести проверку устойчивости распределения расщеплением выборки на две случайные половины. При совпадении нормализованных баллов для половины и для целой выборки считать нормализованную шкалу устойчивой.

7. Проверить однородность распределения по отношению к варьированию заданного популяционного признака (пол, профессия и т.п.) с помощью критерия Колмогорова. Построить графики гистограммы и кумуляты для полной и частной выборок. При значимых различиях разбить выборку на разнородные подвыборки.

8. Построить таблицы процентильных и нормализованных тестовых норм (для каждого интервала шкалы "сырого" балла). При наличии разно-

родных подвыборок для каждой подвыборки должна быть своя таблица.

9. Обсудить конфигурацию полученных распределений.

10. В случае негативных результатов - отсутствия устойчивых норм для шкалы с заданным числом интервалов - осуществить обследование более широкой выборки или отказаться от плана использования данного теста.

Контрольные вопросы для самопроверки

§4, 5. Подсчет параметров и оценка типа распределения

1. Что понимается под понятием "репрезентативности"?
2. Для чего необходима проверка устойчивости распределения?
3. На чем основана общая логика проверки устойчивости распределения тестовых баллов?
4. Как проверяется репрезентативность критериальных тестов?
5. Какие действия последовательно должен произвести психолог при построении тестовых норм?

Литература к теме

1. Анастаси А. Психологическое тестирование. В 2-х кн. М., 1982.
2. Бурлачук Л.Ф., Морозов С.М. Словарь-справочник по психодиагностике. - СПб.: Изд-во "Питер", 1999. - 528 с.
3. Гусев А. Н., Измайлов Ч. А., Михалевская М. Б. Измерение в психологии: общий психологический практикум. М.: Смысл, 1997. -287 с.
4. Дружинин В.Н. Экспериментальная психология: Учебное пособие - М.: ИНФРА-М, 1997. - 256 с.
5. Общая психодиагностика / Под ред. А. А. Бодалева, В. В. Столина.- М., 1987.
6. Основы психодиагностики / Под ред. А. Г. Шмелева. Ростов-на-Дону., 1996.

§6. Надежность теста. Теория надежности.

Надежность - одно из основных понятий психодиагностики и, вместе с тем, одно из самых сложных. К.М. Гуревич отмечает «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 изме-

рения - она должна показывать, какая часть изменчивости показателей ошибочна» [17, стр. 27].

На практике понятие "надежности" имеет два значения: одно связано с воспроизводимостью результатов тестирования, другое – с внутренней согласованностью теста [8].

На понятии "надежность" основывается вычисление "ошибки измерения", с помощью которой определяются вероятные пределы колебания измеряемой величины, возникающей под воздействием случайных факторов.

Можно сказать, что в широком смысле слова надежность теста показывает, в какой мере индивидуальные различия в тестовых результатах являются «истинными», а в какой мере они могут быть отнесены к случайным ошибкам. *Надежность это помехоустойчивость теста, независимость его результатов от действия всевозможных случайных факторов.* К числу таких факторов следует отнести:

- разнообразие внешних материальных условий тестирования (время суток, освещенность, температура, наличие посторонних звуков и т.п.);
- динамичные внутренние факторы, по-разному действующие на разных испытуемых в ходе тестирования (настроение, утомление, другие особенности состояния испытуемого);
- информационно-социальные факторы (контакт с психологом, наличие других людей, сама ситуация тестирования и пр.).

Разнообразие и изменчивость этих факторов так велики, что они обуславливают появление у каждого испытуемого непрогнозируемого по размеру и направлению отклонения измеренного тестового балла (т.е. такого, который можно было бы получить в идеальных условиях).

Поэтому, общий разброс (дисперсию) результатов произведенных измерений можно представить как результат суммы двух источников

разнообразия: самого измеряемого свойства и нестабильности измерительной процедуры, обуславливающей наличие ошибки измерения. Это нашло свое выражение в классической формуле, описывающей надежность теста в виде отношения истинной дисперсии к дисперсии эмпирически зарегистрированных баллов [1, 13, 15, 16]:

$$\alpha = 1 - \frac{\sigma_e^2}{\sigma_x^2}$$

где: α - надежность теста; σ_e^2 - дисперсия ошибки;

σ_x^2 - дисперсия теста.

Итак, эта формула читается так: надежность теста равняется единице минус отношение дисперсии ошибки к дисперсии эмпирически зарегистрированных баллов.

Из этой формулы получаем, что *стандартная ошибка измерения* равна: $\sigma_e = \sigma_x \times \sqrt{1 - \alpha}$. Можно сказать, что величина ошибки измерения - обратный индикатор точности измерения: чем выше ошибка, тем шире диапазон неопределенности на шкале (доверительный интервал), внутри которого оказывается статистически возможное расположение истинного балла данного испытуемого [15].

В общем случае можно сказать, что ошибку измерения в психометрике определяют с помощью *корреляционных методов*, которые дают возможность оценить *надежность* или *точность* через устойчивость и согласованность результатов, полученных как на уровне целого теста, так и на уровне отдельных его пунктов.

Рассмотрим вначале надежность теста, связанную с воспроизводимостью результатов тестирования – так называемую, "ретестовую" или "диахронную" надежность. Затем, проанализируем понятие "надежности",

связанное с внутренней согласованностью теста. В завершении этого параграфа проанализируем критический взгляд на такие понятия, как "ретестовая надежность", "синхронная надежность" и достаточно подробно остановимся на "теории надежности", позволяющей психологу, разрабатывающему или использующему тест, понять всю важность данного параметра теста.

1. Ретестовая надежность (или диахронная надежность; или надежность теста, как целого). Самый естественный способ определить надежность результатов теста - использовать тот же тест второй раз.

В этом случае коэффициент надежности просто равен корреляции между двумя результатами, полученными на одних и тех же испытуемых в каждом из двух случаев проведения теста. Приводя в руководстве к тесту его ретестовую надежность, всегда следует указывать, в каком интервале времени она измерена. Как правило, выбирают этот интервал исходя из следующих соображений.

На дисперсию ошибки тестовых результатов обычно влияют случайные колебания с периодом от нескольких часов до нескольких месяцев. Поэтому, определяя тип тестовой надежности, стараются придерживаться небольших временных интервалов.

(Например, при тестировании маленьких детей этот период должен быть еще короче, чем у взрослых испытуемых, поскольку в раннем детстве возрастные изменения происходят в течение месяца и даже быстрее.).

В целом для любого типа испытуемых интервал между двумя последовательными применениями теста обычно *не должен быть меньше двух недель и превышать 6 месяцев* [1].

Следует учесть, что данное понятие надежности в основном характеризует выполнение теста самого по себе, а не тестируемую область поведения (т.е. характеризует стабильность теста, как измерительного инструмента).

В качестве коэффициента корреляции для этого типа надежности обычно подсчитывают известный коэффициент корреляции произведения моментов Пирсона [1, 15]:

$$r = \frac{\sum XY}{N\sigma_x\sigma_y}, \quad \text{где}$$

r - коэффициент корреляции между результатами двух тестовых испытаний;

σ_x - эмпирическая дисперсия первого испытания;

σ_y - эмпирическая дисперсия второго испытания;

\sum - сумма произведения отклонений от средних значений каждого испытуемого в первом и втором испытаниях;

N - число испытуемых.

Оценка значимости этого коэффициента производится следующим образом:

- вычисляется квадратическая ошибка коэффициента корреляции по формуле

$$m_r = \pm \frac{1-r^2}{\sqrt{n}} ;$$

при малом числе наблюдений n берется "числом степеней свободы", обычно как $n-2$, и ошибка коэффициента корреляции вычисляется так

$$m_r = \pm \frac{1-r^2}{\sqrt{n-1}} ;$$

Ошибка коэффициента корреляции приближается к нулю, когда коэффициент корреляции приближается к единицы. Таким образом, при $r = 1$ независимо от знака, $m_r = 0$.

Значение коэффициента корреляции оценивается с помощью критерия достоверности, который представляет отношение этого коэффициента к своей средней квадратической ошибке, т.е.

$$t = \frac{r}{m_r} ;$$

Далее полученный критерий достоверности (t) сравнивается с таб-

личным (см. приложение 2).

Напомним из статистики, что если необходимо оценить достоверность различий, наблюдаемых между двумя коэффициентами корреляции, формула для расчета критерия достоверности принимает вид

$$t = \frac{r_1 - r_2}{3\sqrt{(m_1^2 + m_2^2)}} .$$

Знание надежности теста позволяет уточнить "истинное" значение тестового балла индивида, применяя формулу: $X_\infty = rX_i + (1 - r)\bar{X}$, где

X_∞ - истинный балл;

X_i - эмпирический балл i -го испытуемого;

\bar{X} - среднее для теста;

r - эмпирическая надежность теста;

Например, что испытуемый получил балл IQ по шкале Стенфорд-Бине равный 120 нормализованным очкам, $\bar{X} = 100$, $r = 0,9$. Тогда истинный балл будет равен: $X_\infty = 0,9 \times 120 + 0,1 \times 100 = 118$

Все выше сказанное касалось тестов с *интервальными шкалами*, для *шкал порядка* в качестве меры ретестовой надежности используется ранговый коэффициент корреляции Спирмена:

$$\rho = 1 - \frac{6\sum d^2}{n(n^2 - 1)}, \text{ где } d^2 - \text{разность рангов } i\text{-го испытуемого в первом и}$$

втором ранговом ряду (1, 4).

Оценка значимости коэффициента корреляции Спирмена производится аналогичным образом через расчет квадратической ошибки по формуле

$$m_r = \pm \frac{1 - \rho^2}{\sqrt{n - 1}};$$

Теперь проанализируем понятие "надежности", связанное с внут-

ренной *согласованностью* теста, и которое находит свое выражение в таких понятиях, как "одномоментная" или "синхронная" надежность.

2. Одномоментная надежность (или синхронная надежность, или согласованность пунктов теста). Этот тип надежности независим от устойчивости (поскольку не имеет временного интервала) и имеет особую содержательную и операциональную природу. Ее надо понимать именно как согласованность частей теста. В психотехнике этот вид надежности часто называют *коэффициентом внутренней согласованности теста*.

Для того, чтобы ее измерить следует скоррелировать между собой параллельные формы теста. Особо отметим, что проводится *только однократное применение теста*, а уже потом его искусственно разбивают на две параллельные формы. Чаще всего параллельные формы теста получают расщеплением составного теста на "четную" и "нечетную" половины: к первой относятся все четные пункты заданий теста, ко второй - соответственно, все нечетные. По каждой половине рассчитываются суммарные баллы и между двумя рядами баллов подсчитываются допустимые коэффициенты корреляции [1, 15]. Если параллельные формы тесты не нормализованы, то предпочтение отдается ранговой корреляции. При подобном расщеплении получается коэффициент корреляции, относящийся к половинам теста. Для того, чтобы найти надежность (согласованность) целого теста, пользуются формулой Спирмена-Брауна:

$$r_{xx} = \frac{2r_x}{1 + r_x}, \text{ где}$$

r_x - эмпирически рассчитанная корреляция для половин;

r_{xx} - надежность целого теста.

Следует отметить, что делить тест на две части можно разными способами, и каждый раз получаются несколько разные коэффициенты; по-

этому в психометрике предложен способ оценки синхронной надежности, который соответствует разбиению теста на такое количество частей, сколько в нем отдельных пунктов. В этом случае, синхронную надежность теста можно оценить с помощью формулы Кронбаха:

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum_{i=1}^k \sigma_i^2}{\sigma_x^2} \right), \quad \text{где}$$

α - коэффициент Кронбаха;

k - количество пунктов (заданий) теста;

σ_i^2 - дисперсия по i -му пункту теста;

σ_x^2 - дисперсия суммарных баллов по всему тесту.

В 1957 году Дж. Китс предложил следующий критерий для оценки статистической значимости коэффициента α (1):

$$\chi_{n-1}^2 = \frac{k(n-1)}{k(1-\alpha) + \alpha}, \quad \text{где}$$

χ_{n-1}^2 - эмпирическое значение статистики χ^2 с $n-1$ степенью свободы;

k - количество пунктов;

n - количество испытуемых;

α - надежность.

Вычисленная статистика χ_{n-1}^2 далее сравнивается с табличной (см. приложение 2).

Как видно, формула Кронбаха позволяет оценить взаимную согласованность пунктов теста, используя только подсчет дисперсий (вся важность этой формулы для психометрики станет понятна после анализа теории надежности).

Коэффициент α позволяет также оценить и среднюю корреляцию между i -тым и j -тым произвольными пунктами теста, так как он связан с

этой средней корреляцией следующей формулой:

$$\alpha = \frac{k\bar{r}_{ij}}{1 + (k-1)\bar{r}_{ij}}, \quad (***)$$

где \bar{r}_{ij} - средняя корреляция между пунктами теста.

Из всего вышесказанного вытекает возможность повысить синхронную надежность теста, увеличивая численность пунктов теста (в k раз). Из формулы видно, что при больших k малое значение r - может сочетаться с высокой надежностью (*например: пусть $r = 0,1$, а $k = 100$, тогда согласно формуле (***) имеем: $\alpha = 0,91$*).

При прочих равных условиях, *чем больше заданий содержит тест, тем выше его надежность.*

Все приведенные формулы относятся к оценке надежности *одномерного теста*, направленного на измерение *одной характеристики*.

Перейдем теперь к обсуждению "теории надежности" тестов [10], опираясь на понятия изложенные выше.

Теория надежности тестов. Известно [1, 2, 8, 10, 15, 16], что оценки полученные индивидуумами по психологическим тестам используются как численные выражения абстрактных психологических понятий. По этой причине важно иметь уверенность в том, что эти оценки по любому психологическому тесту, во-первых, имеют небольшую случайную ошибку измерения (что говорит о высокой статистической надежности) и, во-вторых, действительно измеряют то, что они ставят целью измерить (т.е. имеют высокую валидность, которая подробнее будет рассмотрена позже).

По определению К. Купера [10], анализ понятий «систематическая» и «случайная ошибка» измерения приводит к важному аспекту психометрики, известному как теория надежности.

Следует напомнить, что главнейшей характеристикой психологиче-

ских тестов является то, что каждая шкала теста должна оценивать одну (и только одну!) психологическую характеристику.

Большинство тестов (составленных по типу опросников) обрабатываются суммированием по ключу ответов на каждое задание и интерпретировать значение набранного балла можно только в том случае, если все задания в шкале измеряют одну и ту же базовую психологическую характеристику. Если все утверждения измеряют одну психологическую характеристику, то чем выше оценки испытуемых, полученные по данному тесту, тем более развита у них эта характеристика.

Однако если задания теста измеряют две и более совершенно разные характеристики, такая интерпретация невозможна. Поэтому очень важно убедиться в том, что все утверждения в определенной шкале оценивают одну (и только одну!) психологическую характеристику. Существуют два основных способа убедиться в этом.

Теория надежности - эта теория, в которой изначально предполагается, что все задания теста предназначены для того, чтобы измерять одну и ту же характеристику; кроме того, теория надежности позволяет проверить, так ли это в действительности.

Для того чтобы лучше понять суть систематической и случайной ошибок, рассмотрим аналогию психологических и физических измерений.

Физические измерения. Измерение предметов в повседневной жизни может проводиться с различной точностью, при этом всегда существует некоторая "ошибка измерения", связанная с определением физических параметров предмета, которая составляет небольшой процент от измеряемого параметра. Как правило, величина этой ошибки зависит от точности используемого измерительного инструмента. В частности, в случае измерения длины (например, стола) могут использоваться - рулетка, сантиметр или линейка, которые имеют свою, заданную ошибку измерения.

Ошибка, связанная с проведением каждого измерения, может рассматриваться как случайная в том смысле, что она будет изменяться случайным образом от одного измерения к другому.

К примеру, если бы измеряли некоторый стол одним и тем же инструментом 100 раз, его длина иногда могла оказаться равной, 155 см, в другой раз несколько меньше – 154. Однако если бы мы усреднили эти 100 измерений, они должны были бы дать более точную оценку его подлинной длины, чем та, которая получилась бы в результате одного измерения, поскольку случайные ошибки измерения имеют тенденцию нивелировать влияния друг друга при усреднении.

Таким образом, ошибки измерения могут быть сведены к минимуму; при этом измерительные инструменты должны быть сконструированы так, чтобы получаемые с их помощью оценки гарантированно находились под влиянием только одной физической переменной - именно той, которую хотят измерить. Например, показатели длины, получаемые по шкале измерительной рулетки, не должны зависеть от времени дня, когда осуществляется измерение, от температуры помещения, где происходит измерение, света или фактуры измеряемого объекта или еще чего-либо, за исключением его длины. Это прямой эквивалент принципа, который заложен при использовании любого теста: измерительные инструменты должны измерять только *одну (психологическую) характеристику* объекта.

На практике обеспечить это далеко не так просто. Допустим, что указанные выше измерительные инструменты, полностью свободны от ошибок измерения - «случайных ошибок», упоминавшихся выше. Дает ли это основание полагать, что длина стола может быть измерена с полной точностью? К сожалению, нет, поскольку ни один из этих инструментов не измеряет только длину. Сантиметр (особенно если он на тканевой основе), будет слегка вытягиваться или сжиматься в зависимости от изменения

температуры и влажности, и поэтому в холодный влажный день или в сухой и жаркий он будет давать несколько различающиеся показатели. Точность других измерителей может (хотя и в очень малой степени) также зависеть от иных параметров окружающей среды.

Таким образом, даже если мы примем, что все эти инструменты измеряют длину (и только ее одну), размеры, определяемые каждым из них, будут в действительности подвержены влиянию нескольких различных переменных. Назовем их *источниками "систематической ошибки"*. В отличие от обсуждавшихся выше *случайных ошибок*, источники систематических ошибок не обнаруживают тенденцию к устранению, когда проводятся повторные измерения при одних и тех же физических условиях. Если мы 100 раз измеряем длину стола с помощью сантиметра в жаркий влажный день, то показатели будут всегда слегка преувеличены, поскольку измерительная лента будет коробиться.

Если перечисленные три способа измерения длины стола (т.е. рулетка, сантиметр или линейка) находятся под влиянием различных физических переменных окружающей среды, то каким образом следует определять "подлинную" длину, руководствуясь этими тремя, слегка различающимися показателями? Решение, которое напрашивается само собой, состоит в том, чтобы усреднить эти три показателя, надеясь на интуитивную очевидность того, что среднее трех измерений окажется ближе к "подлинному" значению, нежели каждое из измерений, взятое поодиночке.

Теперь можно подвести итог сказанному, сформулировав несколько основополагающих принципов [10]:

- "Хорошие" измерительные инструменты - это такие, на которые мало влияет случайная ошибка.
- "Хорошие" измерительные инструменты не подвержены влияниям источников систематической ошибки.

- Проведение многократных измерений при разных физических условиях и усреднение результатов уменьшают вклад случайных ошибок.

- Усреднение измерений, полученных с помощью разных инструментов, будет вести к уменьшению вклада систематической ошибки.

При психологических измерениях ответ, который испытуемый дает на задание теста, представляет собой аналог измерения длины одним из методов, описанных выше, с одной лишь существенной разницей, имеющей практическое значение, особенно в случае личностных измерений.

Допустим, что в личностном опроснике задан вопрос: "Получаете ли вы удовольствие от шумных вечеринок?", на который можно ответить, отмечая по пятибалльной шкале ранги от "совершенно согласен" до "совершенно не согласен". Попытаемся назвать ряд очевидных факторов, которые могут повлиять на то, какие ответы будут даваться. Среди таких факторов можно отметить следующие:

- уровень экстраверсии (личностная черта);
- число вечеринок, на которых недавно побывал испытуемый (фактор усталости);
- возраст испытуемого;
- его религиозные убеждения;
- контекст, в котором задавался вопрос (диагностическая ситуация);
- способ, который испытуемый использует при работе с пятибалльной шкалой: некоторые индивидуумы используют оценки 1 и 5 довольно свободно, в то время как другие никогда не обращаются к полюсам шкалы;
- склонность соглашаться: установлено, что люди склонны соглашаться с утверждениями;
- настроение испытуемого;
- случайная ошибка: если задать ему тот же самый вопрос двумя ми-

нутами позже, можете получить несколько отличающийся ответ.

Этот список, вероятно, может содержать и другие важные переменные, поскольку множество посторонних факторов определяет, каким образом индивидуум будет отвечать на вопрос в личностном тесте.

Такое же заключение относится и к оценкам поведения испытуемого или к оценкам его способностей. Следовательно, любой фрагмент собранных данных при оценке индивидуальных различий, подвержен влиянию большого числа разнообразных факторов.

В экспериментах, можно определить меру влияния каждого из этих факторов на индивидуальный ответ, полученный на каждый вопрос теста. Например, если вопрос предназначен для измерения такой черты, как экстраверсия, "хорошим" будет вопрос, при котором эффекты всех других переменных окажутся малы, аналогично тому, как на "хороший" показатель длины влияет расстояние, а не температура, давление воздуха или что-либо еще. В предыдущем примере, касавшемся измерения длины стола, реальная длина стола оказывала решающее влияние на показатели, получаемые с помощью сантиметра.

К сожалению, в психологии это не так. Практически невозможно найти вопрос личностного теста, для которого диагностируемая черта объясняла бы более чем 20—30% вариативности индивидуальных ответов на вопросы. Большая часть вариативности обязана своим происхождением другим факторам.

Проблема действительно сложная. Кажется, что невозможно придумать вопросы, которые измеряли бы черту в чистом виде, поскольку ответы индивидуумов на каждый вопрос теста подвержены влияниям множества черт, состояний, аттитюдов, настроений и везения. Можно ли надеяться, что личность или способности могут быть оценены с какой-либо степенью точности?

Тем не менее, существует подход к решению этой проблемы. Например, можно привести некоторые другие вопросы, измеряющие экстраверсию, каждый из которых зависит от действия различного набора посторонних факторов. Известно [10], что Айзенк считает экстравертов социальными, оптимистичными, разговорчивыми, импульсивными и т.д., - значит, можно сформулировать вопросы, которые измеряли бы и эти переменные тоже. Вопрос типа "Ведете ли вы себя тихо во время общественных мероприятий?" был бы подвержен влиянию определенного числа посторонних факторов, но лишь некоторые из них оказались бы теми же, что и для первого вопроса.

Таким образом, если опросники конструировались из некоторого количества вопросов, на каждый из которых действует различный набор посторонних факторов, влияние последних будет иметь тенденцию к снижению, в то время как влияние черты будет накапливаться. Следовательно, чтобы разработать более точное измерение личностной черты, необходимо:

- сформулировать несколько вопросов, каждый из которых отражает *разные* аспекты черты и, следовательно, оказывается под воздействием *различных* наборов посторонних факторов;
- оценить ответы на эти вопросы;
- сложить эти оценки вместе.

Средний балл, полученный по опроснику, неизбежно будет лучшей оценкой черты индивидуума, чем ответ на один-единственный вопрос, поскольку посторонние факторы устраняют действие друг друга. Это тот же принцип, о котором говорилось в примере с измерением длины стола. Когда мы используем этот принцип, 80%, 90% (и более) вариативности в общей оценке теста будет обуславливаться личностной чертой, что намного лучше, чем 20 или 30%, которые можно было бы получить с помощью од-

ного отдельно взятого, даже самого хорошего, вопроса. Этот простой принцип составляет основу *теории надежности*.

Продолжим анализ теории надежности, необходимо рассмотреть еще один термин – "*специфическая вариативность*".

Вернемся к примеру с вопросом об экстраверсии. Вполне возможно, что некто, не являющийся экстравертом и не получающий удовольствия шумных компаний, и чей ответ не подвержен сильному влиянию любого другого постороннего фактора, может тем не менее просто извлекать удовольствие из "шумных" вечеринок.

Другими словами, может получиться так, что некоторые индивидуумы ответят на этот вопрос полным согласием, даже, несмотря на то, что такой вариант ответа невозможно предугадать исходя из знания их установок, личностных черт и прочих обстоятельств из числа "побочных факторов". Этот факт и учитывается с помощью понятия, называемого "специфической вариативностью".

Выше было показано, что отдельно взятый вопрос теста - плохое средство измерения черты и что значительно лучшую оценку ее выраженности можно получить, если мы сложим оценки, полученные по некоторому количеству вопросов, измеряющих различные аспекты черты.

Представим себе, что для измерения определенной черты разработано около k вопросов, и они предъявляются приблизительно n испытуемым. Пока мы только *допускаем*, что все вопросы измеряют одну и ту же черту (проблемы о том, как проверить это допущение и устранить вопросы, которые плохо ее измеряют, будут рассмотрены позже). Специализированные компьютерные программы (такие, как операция оценки "надежности" в SPSS, "STADIA") могут быть использованы, чтобы вычислить по этим данным статистическую характеристику, которую различные авторы называют как: "надежность" теста, "альфа", "коэффициент альфа", "KR-

20", "альфа Кронбаха" или "внутренняя согласованность". Деталей того, как вычисляется эта статистика, мы касаться не будем, но их можно найти в большинстве учебников по статистике.

Как можно ожидать исходя из выше изложенного, на коэффициент альфа влияют два фактора:

- *средняя величина корреляции между вопросами теста*. Поскольку ранее мы допустили, что различные задания теста подвержены действию разных посторонних факторов, единственная причина, по которой ответы индивидуумов на любую пару заданий должны коррелировать между собой, состоит в том, что оба вопроса измеряют одну и ту же скрытую черту. Поэтому, если все вопросы теста измеряют одну и ту же черту, корреляции между ними будут высокими и положительными;

- *количество вопросов в шкале*. Еще раз подчеркнем, что общая цель построения шкалы из нескольких вопросов состоит в том, чтобы попытаться устранить действие посторонних факторов.

Легко понять: чем больше вопросов в шкале, тем более вероятно, что все эти посторонние факторы будут устранены. В этом случае может оказаться полезной формула Спирмена - Брауна (рассмотренная выше). Она позволяет предсказать, как будет увеличиваться или уменьшаться надежность шкалы, если число вопросов в шкале меняется.

Итак, *надежность теста* - это статистическая характеристика, которая может быть вычислена на основе любого набора данных (при условии, что выборка составляет не менее 200 испытуемых). Напомним также, что максимально возможное значение надежности составляет 1,0 (минимальное значение, при определенных обстоятельствах, может быть меньше 0).

Для больших тестов квадратный корень из коэффициента альфа представляет очень близкую аппроксимацию к корреляции между оценка-

ми индивидуумов по определенному тесту и *подлинной оценкой* их черты. Например, коэффициент альфа равный 0,7, предполагает корреляцию равную 0,84 ($\sqrt{0.7}=0,84$), между оценками, полученными по тесту, и подлинными оценками испытуемых, в то время как величина коэффициента альфа, равная 0,9, подразумевает, что корреляция достигает такого высокого значения, как 0,95.

Основная цель использования психологических тестов - *попытаться достичь максимально возможного приближения к подлинной оценке черты личности*, из этого следует, что тесты должны иметь высокое значение коэффициента альфа.

Широко распространенное эмпирическое правило указывает на то, что тест не должен использоваться, если он имеет коэффициент альфа ниже 0,7, а применять его при принятии важных решений по поводу *конкретного индивидуума* (например, для оценки необходимости коррекционного обучения) можно только в том случае, если величина коэффициента альфа больше 0,9.

Теперь рассмотрим вопрос о *содержании заданий* теста. Проблема заключается в том, что довольно легко повысить среднюю корреляцию между заданиями теста, задавая несколько раз, по существу, один и тот же вопрос, слегка перефразируя его в каждом случае. Благодаря этому все посторонние факторы, которые влияют на первый вопрос, будут влиять и на второй. Поскольку оба вопроса имеют отношение к одному и тому же поведению, они будут также разделять большую часть своей специфической вариативности. Поэтому можно ожидать, что корреляция между двумя такими утверждениями будет близка к 1,0. Примерами двух таких утверждений могут быть:

"Мне нравятся шумные компании" и "Мне нравятся шумные вечеринки". Поскольку эти два задания, по сути, задают один и тот же вопрос,

трудно представить себе, что многие люди могли бы полностью согласиться с одним и столь же решительно не согласиться с другим. Ответы на эти два вопроса *обязаны* иметь высокую положительную корреляцию. При условии, что корреляции между заданиями теста обычно невелики (в лучшем случае порядка 0,2-0,4), корреляция 0,9, полученная в результате сопоставления двух фактически идентичных утверждений, существенно увеличит среднюю корреляцию по тесту. В результате этого произойдет значительное увеличение коэффициента альфа.

Однако должно быть ясно, что в этом случае нарушаются два главных условия составления теста: каждый вопрос должен быть подвержен влиянию различного набора посторонних факторов, и каждый должен иметь свою собственную "уникальную" вариативность, которая не разделяется другими вопросами. Поэтому, крайне важно убедиться, что задания в каждой шкале хорошо подобраны. В некоторых случаях сделать это не сложно.

Например, в случае разработки словарного теста просто необходимо подбирать задания из словаря (исключая те слова, которые встречаются ниже определенного порога частотности, или слова архаичные, специальные, т.е. профессиональные, термины). Когда это сделано, единственное, что будет влиять на корреляцию между ответами на пару заданий, - степень, с которой каждое из них измеряет скрытую черту (грамотность).

Не существует магической формулы для автоматической реализации этого принципа при конструировании тестов. Только от разработчика теста, зависит гарантия того, что единственной причиной, объясняющей корреляцию ответов на любую пару заданий, является та скрытая черта личности или способностей, которую они оба должны измерять. К сожалению, некоторые подходы к конструированию тестов, часто ведут к появлению большого количества искусственно завышенных корреляций, что да-

ет, в свою очередь, завышенную оценку коэффициента альфа.

Важно также обеспечить, чтобы выборка испытуемых, чьи тестовые оценки используются для вычисления коэффициента альфа, была репрезентативна группе, в которой будет применяться данный тест. Бессмысленно, например, установив величину коэффициента альфа в размере 0,9 на выборке студентов университета, затем считать, что этот тест будет годиться для использования на общей популяции, поскольку студенты университета - это не случайная выборка. Еще раз подчеркнем, что не существует количественного способа определить, будет ли тест, имеющий высокое значение коэффициента альфа на одной выборке, так же работать на другой, - это дело здравого смысла. Самое безопасное - вычислять коэффициент альфа во всех случаях использования теста, хотя в качестве предварительного условия обязательным будет тестирование большей выборки испытуемых (рекомендуется выборка - минимум 200 человек). При использовании должным образом коэффициент альфа может быть очень полезен.

Выше мы использовали понятие "*подлинная оценка*", но не определили его значение. Рассмотрим его несколько подробнее.

Любой тест можно рассматривать как комплекс заданий, выбранных из большого *набора* вопросов, которые потенциально могли быть заданы. Например, тест на правописание - это выборка большого количества слов из словаря. Тест, измеряющий тревогу, - это набор всех (многих!) вопросов, которые можно было бы перечислить с целью измерения множества аспектов тревоги. Тест математических способностей - это выборка из почти бесконечного числа математических заданий, которые только можно было бы написать.

Итак, *подлинное значение черты индивидуума* - это оценка, которую он получил бы, если бы ему предъявили каждое возможное задание из

полного набора.

Поясним это на примере. Если было бы необходимо оценить чью-либо способность правильно писать каждое слово из словаря, то можно было бы узнать точно, каковы способности этого человека к правописанию, поскольку отсутствовала бы ошибка измерения, обусловленная случайным выбором заданий. Однако в реальном тесте используется лишь небольшая выборка заданий из всего возможного набора.

Если задания теста формируют репрезентативную выборку по отношению к полному набору заданий, то квадратный корень из коэффициента альфа довольно точно оценивает корреляцию между оценкой, полученной испытуемыми при выполнении теста, и их подлинной оценкой (т.е. оценкой, которую они могли бы получить, если бы им были предъявлены все задания из полного набора).

Чем выше величина коэффициента альфа, тем меньше будет ошибка при измерении черты; зная надежность теста и стандартные отклонения тестовых оценок, можно найти статистическую характеристику, которая называется "стандартная ошибка измерения", которая показывает насколько ошибка измерения может быть связана с каждым измерением (формула для вычисления стандартной ошибки приведена ниже):

$$m_i = \sigma\sqrt{1-\alpha};$$

где σ – стандартное отклонение тестовых оценок; α – коэффициент надежности теста.

Таким образом, знание надежности теста позволяет делать некоторые заключения, касающиеся величины ошибки, которая, вероятно, имеется в любом измерении, при условии, конечно, что задания данного теста можно считать репрезентативными по отношению к полному набору заданий.

Ретестовая надежность - временная стабильность, имеет совер-

шенно другой смысл. Как следует из названия, она проверяет, в какой степени оценки черты остаются более или менее постоянными в течение времени. Большинство тестов разработано для измерения именно стабильных черт таких, как экстраверсия, способности к вычислениям или нейротизм, а в определении черты подчеркивается, что это относительно устойчивая диспозиция. Последнее предполагает, что индивидуумы должны иметь похожие оценки, когда они тестируются в двух ситуациях (например, с интервалом в несколько недель) при условии, что:

- с ними не случилось ничего значительного в интервале между двумя тестированиями (например, не было эмоционального кризиса, изменений, обусловленных развитием, или значительного образовательного опыта, который мог бы повлиять на черту);

- тест хорошо измеряет черту.

Если при первом тестировании тест показывает, что ребенок гениален, а через месяц его интеллект оказывается на среднем уровне, то либо концепция интеллекта характеризует в большей степени состояние, а не черту, либо тест несостоятелен.

Проблема, однако, заключается в том, что ретестовая надежность основывается на общей оценке, она ничего не говорит о том, как люди выполняют отдельные задания. В то время как коэффициент альфа показывает, измеряет ли набор заданий некоторую конкретную скрытую черту, *прекрасную ретестовую надежность может иметь набор заданий, которые не имеют ничего общего (!) между собой.*

Например, если вы попросили кого-то суммировать номер своего дома, размер обуви и год рождения в двух разных ситуациях, полученная статистическая характеристика будет обнаруживать впечатляющую ретестовую надежность, хотя эти три пункта не имеют между собой ничего общего.

Индекс «μ».

В связи с обсуждением понятия "*ретестовая надежность*" целесообразно остановиться на *возможности дифференцирования* таких психологических понятий как "*состояние*" и "*черта личности*" на основе статистической оценки.

Имея показатели ретестовой надежности и внутренней согласованности теста, представляется возможным математически проверить, является ли тестируемое психическое образование устойчивой личностной характеристикой или текущим состоянием. Говард Е. в своей работе [23] приводит формулу для вычисления индекса μ, по величине которого можно судить, что диагностируется данным тестом (черта личности или состояние субъекта в момент тестирования; например, тревожность как стабильная черта личности и как ситуативная характеристика).

Остер Т.Р. [24], приводит количественные данные для индекса μ, по которым можно отличить устойчивые характеристики личности от ситуативных. Если μ находится в пределах от 0,4 до 0,7, то данный тест определяет *ситуативную характеристику индивида*; при μ, равном 0,1-0,2, тест определяет *стабильную личностную характеристику*. Индекс μ вычисляется по формуле:

$$\mu = 1 - \frac{\kappa}{\alpha};$$

где: κ - показатель ретестовой надежности теста; α - коэффициент внутренней однородности теста.

Надежность параллельных форм. Ради полноты картины следует упомянуть надежность параллельных форм. Конструкторы тестов иногда создают из набора заданий несколько тестов. Чтобы создать две параллельные формы теста, задания предъявляются большой выборке испытуемых и подбираются пары заданий со сходным содержанием и уровнем

трудности.

Например, обе формы теста могут включать задание решить анаграмму из семи букв, и в обоих случаях ответом будут слова, одинаково часто встречающиеся в языке, и только около 25% выборки будут способны решить каждую из них. Одно задание будет, затем включено в форму "А" теста, а другое - в форму "Б". Эти два теста продаются отдельно, и (теоретически) несущественно, какой из них будет использоваться в определенных целях, поскольку обычно прилагаются специальные усилия, гарантирующие, что эти две версии дают одинаковое распределение оценок (и благодаря этому позволяют использовать одни и те же таблицы норм для обеих форм теста). Если оба теста измеряют одну и ту же черту, следует ожидать высокой положительной корреляции между оценками индивидумов по двум формам теста. Эта корреляция известна как *надежность параллельных форм*. Однако, поскольку из-за трудностей создания, параллельные формы имеют относительно немногие тесты, она используется редко [10].

Надежность отдельных пунктов. Очевидно, что надежность теста обеспечивается надежностью пунктов, из которых он состоит. Чтобы повысить ретестовую (диахронную) надежность всего теста, надо отобрать из исходного набора пунктов (проверенных в пилотажных психометрических экспериментах) такие пункты, на которые испытуемые дают устойчивые ответы.

		Второе тестирование	
		<i>Да</i>	<i>Нет</i>
Первое тестирование	<i>Да</i>	А	В
	<i>Нет</i>	С	Д

Для дихотомических пунктов с ответами типа "верно" или "невер-

но", устойчивость удобно измерять с использованием четырехклеточной матрицы сопряженности (см. выше).

Здесь в клетке А суммируются частота ответов "Да", данных испытуемым при первом и втором тестировании; в клетке В суммируются частота ответов "Да", данных испытуемым при первом и "Нет" при втором тестировании; клетке С суммируются частота ответов "Нет", данных испытуемым при первом и "Да" при втором тестировании; клетке Д суммируются частота ответов "Нет", данных испытуемым при первом и втором тестировании [1, 11, 15].

В качестве меры корреляции вычисляется известный [11] фи-коэффициент:

$$\phi = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}},$$

При этом, значимость фи-коэффициента определяется с помощью критерия хи-квадрат по формуле: $\chi_1^2 = \phi^2 n$. (а)

Если вычисленное значение хи-квадрат выше табличного с одной степенью свободы, то нулевая гипотеза (о нулевой устойчивости) отвергается (см. приложение 2).

Удобство в использовании фи-коэффициента состоит в том, что он одновременно оценивает степень оптимальности данного пункта по силе (трудности): фи-коэффициент оказывается тем меньше, чем сильнее частота ответов "Да" отличается от частоты ответов "Нет".

В некоторых случаях подобный анализ позволяет уточнить ключ для пункта: если пункт получает значимый положительный фи-коэффициент, то ключ определяется величиной "+1", если пункт получает значимый отрицательный фи-коэффициент, то величиной "-1". Если пункт получает незначимый фи-коэффициент, то его целесообразно исключить из батареи.

При «ручных» вычислениях фи-коэффициента удобно вначале с

помощью формулы (а) определить граничное значение значимого (по модулю) фи-коэффициента.

При постоянном (многократном) подсчете суммарных баллов - ключ для каждого пункта теста можно определять в виде более чувствительного, тетраэрического коэффициента корреляции (который также применяется для дихотомических пунктов) [4]:

$$R = \text{Cos}\left(\frac{180}{1 + \sqrt{\frac{BC}{AD}}}\right), \text{ где } A, B, C, D - \text{клетки четырехпольной таблицы.}$$

В этом случае в первом столбце таблички суммируются ответы испытуемых, из "высокой" группы (т.е. из группы испытуемых, набравших максимальное количество баллов), во втором столбце - из "низкой" (из группы испытуемых, набравших минимальное количество баллов).

При нормальном распределении частот суммарных баллов "высокая" и "низкая" группы отсекаются справа и слева 27% -ними средними квантилями.

	Группы испытуемых	
	<i>Высокая</i>	<i>Низкая</i>
Да	A	B
Нет	C	D

Для того, чтобы повысить *одномоментную* (синхронную) *надежность* теста, следует из исходной пилотажной батареи пунктов отбросить те, которые плохо согласованы с остальными.

Как отмечалось выше, согласованность всех пунктов можно вычислить, используя формулу Кронбаха (для расчета α), или же, в случае наличия дихотомических групп, определить с помощью точечного бисериального коэффициента корреляции, который учитывает амплитуду от-

клонения индивидуальных суммарных баллов от среднего балла [4]:

$$R = \frac{X_e - X_n}{\sigma} \sqrt{\frac{n_e - n_n}{N(N-1)}}, \text{ где}$$

X_e - средний балл по «высокой» группе;

X_n - средний балл по «низкой» группе;

σ - стандартное отклонение для суммарных баллов всех индивидов из выборки;

n_e - количество испытуемых в «высокой» группе;

n_n - количество испытуемых в «низкой» группе;

N - общее количество испытуемых в выборке.

А. Анастаси относит критерий внутренней согласованности теста к валидности, однако, по мнению В.В. Столина, если и можно в данном случае говорить о валидности, то только в смысле особой внутренней валидности теста. Однако следует учесть, что слишком высокая согласованность снижает внешнюю валидность теста по критерию, (мы рассмотрим эту зависимость более подробно при анализе различных видов "валидности" теста).

Последовательность действий при проверке надежности. Можно предложить следующий алгоритм действий для проверки надежности теста[15].

1. Узнать, существуют ли данные о надежности теста, предполагаемого к использованию, на какой популяции и в какой диагностической ситуации проводилась проверка.

Если проверки не было или признаки новой популяции и ситуации явно специфичны, провести заново проверку надежности с учетом указанных ниже возможностей.

2. Если возможности обследования испытуемых, смысл теста и возможности обработки позволяют, то произвести второе тестирование на

всей выборке стандартизации и подсчитать все коэффициенты, приведенные в настоящем параграфе, как для целого теста, так и для отдельных пунктов. Анализ полученных коэффициентов позволит понять:

- насколько пренебрежима ошибка измерения;
- дает ли данный тест интервальную шкалу или только диагностичен для крайних групп, насколько устойчиво измеряемое свойство во времени (возможен ли статистический прогноз);
- в каких своих частях (пунктах) тест менее надежен (анализ этих пунктов позволяет психологически осмыслить содержательный механизм взаимодействия пунктов с испытуемыми).

3. Если возможности ограничены, произвести повторное тестирование только на части выборки (не менее 30 испытуемых), подсчитать (вручную) ранговую или четырехклеточную корреляцию для оценки внутренней согласованности (методом расщепления) и стабильности целого теста.

Контрольные вопросы для самопроверки

§6. Надежность теста. Теория надежности.

1. Что понимается под надежностью теста?
2. Какие виды надежности существуют в психометрике? Дайте им краткую содержательную характеристику.
3. Как проверяется надежность отдельных пунктов теста?
4. На чем основана теория надежности?
5. Что понимается под согласованностью пунктов теста? Относится ли это понятие в равной степени и к надежности, и к валидности?
6. Какие действия последовательно должен произвести психолог при проверке надежности?

Литература к теме

1. Анастаси А. Психологическое тестирование. В 2-х кн. М., 1982.
2. Бурлачук Л.Ф., Морозов С.М. Словарь-справочник по психодиагностике. - СПб.: Изд-во "Питер", 1999. - 528 с.
3. Корниенко А.Ф. Теория и практика психологического исследования. Учебное пособие. Казанский пед. университет, Казань, 2000. -258 с.
4. Общая психодиагностика / Под ред. А. А. Бодалева, В. В. Столи-

на.- М., 1987.

5. Основы психодиагностики / Под ред. А. Г. Шмелева. Ростов-на-Дону., 1996.

§7. Дискриминативность теста.

В § 2 темы 1, посвященной содержанию и предмету психодиагностики, было отмечено, что к психологическим измерениям предъявляются определенные научные требования. С точки зрения психометрики имеются пять характеристик, которым должны удовлетворять все методики и тесты для психологических измерений, к ним относятся: надежность, дискриминативность, валидность, достоверность и репрезентативность.

Прежде чем переходить к обсуждению методов достижения валидности, рассмотрим способы установления дискриминативности. Под *дискриминативностью теста* будем понимать способность теста дифференцировать испытуемых в диапазоне от "максимального" до "минимального" результата, набранного по тесту.

Как отмечает П. Клайн [8], потенциально тест может быть гораздо более дискриминативным, чем другие средства измерения, например, интервью или рейтинги.

Показатели дискриминативности связаны по существу с ранжированием испытуемых. Основной показатель коэффициента дискриминативности, "δ" Фергюсона, который рекомендуется для оценивания дискриминативности тестов, достаточно прост в вычислении. Если не касаться процедуры вывода формулы, то ее можно представить в следующем виде:

$$\delta = \frac{(n+1)(N^2 - \sum f_i^2)}{nN^2} ;$$

где N - количество испытуемых, n - количество заданий,

f_i - частота встречаемости каждого показателя.

Коэффициент Фергюсона $\tilde{\delta}=0$, когда все испытуемые получили одинаковые показатели, (то есть, когда нет дискриминативности), и равно 1 при равномерном (прямоугольном) распределении.

Вычислить $\tilde{\delta}$ Фергюсона можно следующим образом.

1. Подсчитать, как часто встречаются значения показателей для данного теста.

2. Возвести эти числа в квадрат и просуммировать: $\sum f_i^2$;

3. Прибавьте 1 к количеству заданий: $n + 1$.

4. Возвести в квадрат количество испытуемых: N^2 .

5. Перемножить количество заданий на результат шага (4): $n N^2$.

6. Подставить все найденные элементы в формулу.

Разработчик тестов должен учитывать некоторые характеристики $\tilde{\delta}$ коэффициента. Поскольку для равномерного (прямоугольного) распределения (наиболее дискриминативного) необходимы задания, в которых бы наиболее полно были реализованы все возможные проявления измеряемого свойства, это означает, что дискриминативность до некоторой степени противостоит надежности, так как использование заданий с широким перечнем возможных проявлений измеряемого свойства уменьшает взаимную корреляцию между заданиями (см. § 6).

Конечно, распределение показателей, которое дает тест – это, прежде всего, функция трудности заданий, а это влияет не только на надежность, но также и на дискриминативность. Это означает, что при конструировании теста следует исходить из предназначения теста, поскольку именно этот факт определяет то, на что ориентироваться разработчику - на достижение максимальной надежности или максимальной дискриминативности.

Поскольку дискриминативность целого теста зависит от дискриминативности входящих в него заданий, рассмотрим проблему определения

дискриминативности отдельных заданий теста.

Аналогично с определением дискриминативности целого теста под дискриминативностью *отдельных пунктов* будем понимать способность отдельных пунктов (заданий) теста дифференцировать обследуемых относительно "максимального" или "минимального" результата теста [3].

Любой ответ испытуемого на конкретное задание можно оценить по двухбалльной шкале – "верно" (1 балл), "неверно" (0 баллов). Сумма баллов по всем пунктам представляет собой первичную ("сырую") оценку. Мера соответствия успешности выполнения одной задачи (одного пункта) всему тесту является показателем дискриминативности задания теста для данной выборки испытуемых и называется коэффициентом дискриминации (индексом дискриминации).

$$r_{pb} = \frac{\bar{x}_n - \bar{x}}{\sigma_x} \sqrt{\frac{N_n}{N - N_n}}$$

где \bar{x} - среднее арифметическое всех индивидуальных оценок по тесту;

\bar{x}_n - среднее арифметическое оценок по тесту у испытуемых, правильно выполнивших задание (в случае *опросника личности* - соответствие с "ключом");

σ - среднеквадратическое отклонение индивидуальных оценок по тесту для выборки;

N_n - число испытуемых, правильно решивших задачу (или тех, чей ответ на данный пункт опросника соответствует "ключу");

N - общее число испытуемых.

Коэффициент дискриминации может принимать значения от -1 до +1. Высокий положительный r свидетельствует об эффективности деления испытуемых. Высокое отрицательное значение r свидетельствует о непригодности данного пункта для теста, о его несоответствии суммарному результату.

Индекс дискриминативности задания теста может быть вычислен с помощью метода контрастных групп. Необходимым условием применения метода в этом случае является наличие близкого к нормальному распределению оценок по критерию валидизации. При этом, доля членов контрастных групп может изменяться в широких пределах в зависимости от величины выборки. Чем больше выборка, тем меньшей долей испытуемых можно ограничиться при выделении групп с высоким и низким результатами. Чаще из выборки "извлекают" по 27% или 33% испытуемых.

Индекс дискриминации вычисляется с использованием формулы четырехпольного коэффициента корреляции:

$$r_{phi} = \frac{f_g - f_d}{\sqrt{pq}} \quad \text{где:}$$

f_g - число лиц, правильно решивших задачу, по отношению к общему числу обследованных в группе с максимальным результатом;

f_d - число лиц, правильно решивших задание в группе с минимальным результатом;

p - общая пропорция правильно выполненных заданий ($f_g + f_d$);

q — число лиц, давших неверное решение ($1 - p$).

Критические значения этого коэффициента, свидетельствующие о диагностической ценности (на уровне $p < 0,05$), в зависимости от числа обследованных (n) приведены ниже:

n	25	50	100	200
r	0,39	0,28	0,20	0,14

Максимальная точность определения r достигается тогда, когда максимальная и минимальная группы составляют по 27% выборки.

При анализе дискриминативности задания теста особое внимание следует уделить определению статистической значимости коэффициентов

корреляции. В тех случаях, когда значение коэффициента дискриминации приближается к нулю и уровень значимости невысок, проверяемый пункт теста должен быть пересмотрен в связи с некорректностью формулировки задания или вариантов ответа на него [3].

§8. Валидность тестов.

Как отмечает А. Анастази [1], *валидность теста* это понятие, которое указывает на то, что измеряет тест и *насколько хорошо* он это делает. Однако проблемы валидизации психологических тестов хотя и являются центральными для психометрики, но, до сих пор решенными не до конца. Решение этой проблемы зависит не от статистического аппарата, а от уровня развития теоретического аппарата дифференциальной психологии. Валидность является комплексной характеристикой, которая включает в себя не только сведения о том, для чего она создана, но и какова ее действительность, эффективность [15, 18].

Соотношение валидности и надежности. Валидность (или обоснованность) всякой процедуры измерения состоит в однозначности (устойчивости) получаемых результатов относительно измеряемых свойств объектов, т. е. относительно предмета измерения. Отличие понятия валидности от надежности измерения удобно раскрывать с помощью различения "объекта" и "предмета" измерения.

Надежность - это *устойчивость процедуры относительно объектов (испытуемых)*. Надежность не обязательно предполагает валидность. В психологии довольно часто возникает такая ситуация, когда исследователь вначале предлагает определенную процедуру измерения, показывает ее надежность - способность устойчиво различать объекты, но вопрос о валидности остается открытым.

Устойчивость теста относительно объектов (испытуемых) является необходимым, но не достаточным условием его устойчивости относительно измеряемых атрибутов (свойств) объектов. *Надежность является необходимым, но не достаточным условием валидности.* Отсюда понятно основное соотношение психометрики:

$$\text{валидность} < \text{надежность}$$

Это означает, что валидность теста не может превышать надежность теста. Данное соотношение, однако, неверно трактовать как "чисто" математическое, как указание на прямо пропорциональную связь валидности и надежности. Повышение надежности отнюдь не обязательно приводит к повышению валидности. По А. Анастази [1], валидность определяется репрезентативностью теста относительно измеряемой области поведения. Если эта область поведения складывается из разнообразных феноменов, то содержательная валидность теста автоматически требует представленности в нем моделей всех этих разнообразных феноменов.

Например, возьмем глобальное понятие "речевая способность" (этому психолингвистическому термину в традиционной тестологии соответствует "вербальный интеллект"). Сюда относятся такие относительно независимые друг от друга навыки, как навыки письма и чтения. Если заботиться о содержательной валидности соответствующего теста, то нужно ввести в него группы заданий на проверку этих довольно разных по своему операциональному составу компонентов вербального интеллекта.

В результате введения разнородных пунктов и субшкал (субтестов) мы обязательно сокращаем внутреннюю согласованность, одномоментную надежность теста, но зато добиваемся повышения содержательной валидности.

Следовательно, для расширения области применения теста психодиагност должен избегать излишнего повышения внутренней согласованности. Одновременно с этим снижением внутренних корреляций

между различными пунктами теста (как это уже упоминалось ранее) обязательно исчезает отрицательный эксцесс на кривой распределения тестовых баллов, и она все более приближается по форме к нормальной кривой.

Рассматривая понятие валидности, следует иметь в виду, что про валидность нельзя говорить обобщенно (к примеру, нельзя сказать, что тест имеет высокую или низкую валидность), давая определение валидности всегда следует указать конкретное предназначение теста.

Существует *две стратегии*, при помощи которых можно доказать наличие связи между тестом и тем, что он должен измерять: *теоретическая и эмпирическая валидизация*. Процедура проверки валидности называется валидизацией.

Теоретическая валидизация имеет отношение к самой методике – это валидность измерительного инструмента (теста).

Эмпирическая валидизация относится не столько к самой методике, сколько к цели ее использования [6, 18].

Рассмотрим основные типы валидности и соответствующие им процедуры валидизации.

1. Эмпирическая валидность. Эмпирическая валидность измеряется всегда с помощью статистического коррелирования: подсчитывается корреляция двух рядов значений - баллов по тесту и показателей по внешнему параметру, избранному в качестве критерия валидности. Эмпирическая валидность теста соотносится, как правило, с какими-либо внешними для психологии социально-прагматическими критериями. Эти критерии являются показателями, обладающими непосредственной ценностью для определенных областей практики [6, 15].

Например, в области педагогической психологии это "успеваемость" (которую надо повысить), в психологии труда это "производительность труда" и "текучесть кадров", в медицине - "состояние здоровья пациента", в психологии управления - "совместимость", в юриди-

ческой психологии - "преступность" (которую надо понизить) и т. п.

Ориентируясь непосредственно на эти категории, психолог, пытающийся скоррелировать результаты теста с этими показателями, фактически решает сразу две задачи: задачу измерения валидности и задачу измерения практической эффективности своей психодиагностической программы. Если получен значимый коэффициент корреляции, то можно считать, что решены с позитивным результатом сразу две эти задачи. Но если корреляции не обнаружено, то остается неопределенность: либо невалидна сама процедура (тестовый балл не отражает например, стрессоустойчивости оператора), либо неверна гипотеза о наличии причинно-следственной связи между психическим свойством и социально значимым показателем.

Таким образом, социально-прагматические критерии являются комплексными: они позволяют измерить *валидность-эффективность*, но не каждое из этих двух свойств теста отдельно [15].

2. Процедура эмпирической валидации. Организация выборки при эмпирической валидации зависит от временного статуса критерия.

Если этот критерий - событие в прошлом (*ретроспективная валидация*), то к участию в психодиагностическом обследовании достаточно привлечь только тех испытуемых, которые оказались на экстремальных полюсах по этому критерию. В результате применяется метод экстремальных (контрастных) групп.

Коррелирование с суммарным баллом по тесту оценивается с помощью бисериального коэффициента корреляции. При этом в статусе дихотомической переменной (на месте отдельного пункта) оказывается сам критерий валидности.

Если критерий - будущее событие (*проспективная валидация*), то выборка должна быть составлена с запасом - с учетом вероятного

объема экстремальных групп в будущем.

Например, нужно выяснить, позволяет ли диагностика темперамента прогнозировать повышенный риск соматических заболеваний (гипертония, язва, астма и т. п.).

Пусть на основании эпидемиологических исследований известно, что в течении трех лет из 1000 здоровых людей этими болезнями заболевают, например, 57 человек. Это означает, что превентивной (предупреждающей) диагностикой должно быть охвачено около 2000 человек, чтобы получить численность "высокой" группы (заболевших) порядка 100.

Перспективная валидизация выявляет прогностическую эффективность диагностической процедуры. Высокая прогностическая валидность доказывает как валидность самого измерения, так и наличие предполагаемой причинной связи.

Ретроспективная валидизация позволяет в лучшем случае решить только первую из двух задач [15].

Например, если для исследования личностной предрасположенности к совершению краж проведено обследование лиц, находящихся под следствием (т. е. уже совершивших преступление), то выявление акцентированных черт "тревожности", "агрессивности" и т.п. еще не может интерпретироваться как свидетельство причинных факторов преступности - эти черты могут быть лишь следствием сложившихся обстоятельств: лишение свободы, угрызения совести и т.п.

3. Понятия "Содержательная" и "Конструктивная" валидность.

Их соотношение.

Содержательная валидность (внутренняя, логическая) - это комплекс сведений о том, насколько задания теста репрезентируют измеряемые свойства и особенности. Одним из основных требований при установлении содержательной валидности методики является *отражение в содержании теста* ключевых сторон изучаемого психологического феномена.

Если область поведения или особенность очень сложна, то *содержательная валидность* требует представления в заданиях теста всех важнейших составных элементов исследуемого явления. Так, при разработке теста «вербального интеллекта» необходимо ввести группы заданий (суб-

тестов) для проверки довольно разнородных по своему операциональному составу навыков письма и чтения.

Конструктивная валидность. Это один из основных типов валидности, отражающий степень представленности (репрезентации) исследуемого психологического конструкта *в результатах теста*. В качестве конструкта могут выступать практический или вербальный интеллект, эмоциональная неустойчивость, интроверсия, понимание речи, переключаемость и т. д. Иначе говоря, конструктивная валидность определяет область теоретической структуры психологических явлений, измеряемых тестом.

Поскольку проявления таких конструктов, как, например, интеллект в деятельности человека многообразны и неоднозначны с точки зрения их выделения, процедура установления конструктивной валидности по сравнению с валидностью критериальной или содержательной более сложна.

При объяснении связи полученных тестовых результатов с теоретическим конструктом необходимо постепенное накопление разнообразной информации о динамике развития измеряемого свойства, а также о его взаимодействии с другими психическими явлениями.

Среди конкретных методов характеристики конструктивной валидности, в первую очередь, необходимо назвать сопоставление исследуемого на предмет конструктивной валидности теста с другими методиками, конструктивное содержание которых известно. Наличие корреляции между новым и аналогичным по конструкту тестом указывает на то, что разрабатываемый тест «измеряет» примерно ту же сферу поведения, способность, личностное качество, что и эталонная методика. Такая процедура валидизации напоминает определение критериальной валидности в том смысле, что эталонный тест, валидность которого определена, выступает в качестве независимого критерия [3].

Следует, однако, заметить, что, в отличие от критериальной валидации, при анализе конструктивной валидности не требуется высокой степени связи результатов двух тестов. Если окажется, что новый и эталонный тесты практически идентичны по содержанию и результатам и разрабатываемая методика не обладает преимуществами краткости или легкости применения, это означает лишь дублирование теста, оправданное только с точки зрения создания параллельной формы теста. Смысл процедуры конструктивной валидности состоит в установлении одновременно как сходства, так и различия психологических феноменов, измеряемых новым тестом по сравнению с известным.

При анализе конструктивной валидности методики обычно формулируют ряд гипотез о том, как будет коррелировать разрабатываемый тест с широким кругом других тестов, направленных на конструкты, находящиеся в теоретически известной или предполагаемой связи с исследуемыми. При этом конструктивная валидность характеризуется не только связями проверяемого теста с близкородственными показателями, но и с теми, где, исходя из гипотезы, значимых связей наблюдаться не должно.

Подтверждение совокупности теоретически ожидаемых связей составляет важный круг сведений конструктивной валидности. В англоязычной психодиагностике такое операциональное определение *конструктивной валидности* обозначается как "*предполагаемая валидность*" (assumed validity).

Прямое отношение к характеристике конструктивной валидности имеет *факторный анализ*, позволяющий строго статистически проанализировать структуру связей показателей исследуемого теста с другими известными и латентными факторами, выявить общие и специфические для группы сопоставляемых тестов факторы, степень их представленности в результатах, т. е. определить факторный состав и факторные нагрузки ре-

зультата теста. Исключительная важность такой процедуры является основанием для выделения ее в *особый вид конструктивной валидности - факторную валидность*.

Важным аспектом конструктивной валидности *является внутренняя согласованность*, отражающая то, насколько определенные пункты (задания, вопросы), составляющие материал теста, подчинены основному направлению теста как целого, ориентированы на изучение одних и тех же конструктов. Как уже было отмечено, анализ внутренней согласованности осуществляется путем коррелирования ответов на каждое задание с общим результатом теста. Следует отметить, что критерий внутренней согласованности указывает лишь на меру связи всего содержания теста с измеряемым конструктом, давая лишь косвенную информацию о природе измеряемого свойства.

При определении конструктивной валидности важное место принадлежит изучению динамики измеряемого конструкта. При этом мы можем опираться на гипотезы о его возрастном развитии, влиянии тренировок, обучения, освоения профессии и т. д. Одним из таких подходов является применение критерия возрастной дифференциации, что предусматривает согласование результатов теста с ожидаемыми изменениями, которые претерпевает изучаемое психическое свойство в возрастной динамике при переходе на новый этап развития. Этот метод конструктивной валидности особенно важен для валидации тестов интеллекта, достижений в обучении.

В комплекс сведений о конструктивной валидности методики входят также данные, относящиеся к сфере критериальной и содержательной валидности. Так, критерии, используемые при валидации, несут информацию, позволяющую раскрыть область поведения, качества, представленные в тесте в виде конструкта. Для характеристики конструктивной валидности необходимы связь с практическими формами деятельности, достовер-

ность прогноза реального поведения.

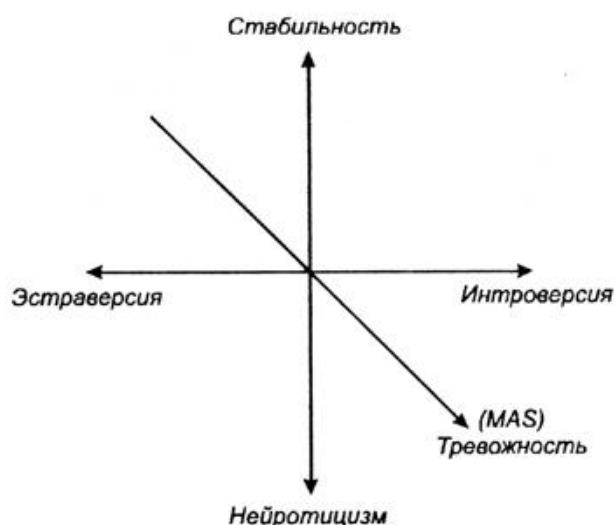
Однако конструктивная валидность является качественно более высоким и комплексным уровнем описания теста, характеризуя область измеряемого поведения в широких психологических понятиях. Благодаря данным конструктивной валидности мы можем с психологических позиций закономерно объяснить результаты теста и их дисперсию, обосновать диагноз, введя измеряемое свойство в систему психологических категорий, прогнозировать поведение в более широких пределах, чем это задается областью деятельности, для которой определялась содержательная валидность.

Следует учитывать, что понятие конструктивной валидности указывает на высокую зависимость эмпирических связей теста от теоретических представлений автора теста об измеряемом свойстве. Для иллюстрации приведем пример взаимоотношений между двумя популярными тест-опросниками: MAS Ж. Тейлор и EPI Г. Айзенка [15]. Многочисленные корреляционные исследования, проведенные на репрезентативных выборках, показали, что шкала (MAS) "тревожность" Тейлор положительно коррелирует со шкалой "нейротизм" Айзенка и отрицательно со шкалой "экстраверсия" Айзенка.

С точки зрения концепции Г. Айзенка, эти данные можно рассматривать как свидетельства низкой валидности шкалы Тейлор: MAS коррелирует не только с релевантным фактором "нейротизм", но и с иррелевантным фактором "интроверсия". С этой точки зрения опросник Тейлор оказывается просто нечувствительным к особой разновидности "нейротизма" - к нейротизму (тревожности) экстравертов: в перечне пунктов MAS отсутствуют высказывания, в которых могла бы проявиться тревожность экстраверта. Однако с точки зрения того теоретического смысла, который приписывают показателям MAS Ж. Тейлор, эта ситуация вполне законо-

мерна, желательна и никак не является артефактом - следствием дефекта их диагностического средства, поскольку MAS измеряет общий уровень драйва - неспециализированного побуждения, которое как раз достигает максимума при сочетании нейротизма (специфическая активация по Г. Айзенку) и интроверсии (неспецифическая активация).

Рис. 5. Векторная модель соотношения поазателя «тревожности» по тесту MAS с факторами (по осям X и Y) теста Айзенка (по В.В. Столину).



Таким образом, не всегда краткие названия тестов однозначно выражают теоретический статус диагностического конструкта - понятия об измеряемом свойстве.

4. Конвергентная и дискриминантная валидность. Известно [1, 6, 15, 16], что от того, как психолог определяет диагностический конструкт, зависит стратегия включения в тест определенных пунктов. Если Айзенк определяет свойство "нейротизм" как независимое от экстраверсии-интроверсии, то это означает, что в его опроснике примерно поровну должны быть представлены пункты, с которыми будут соглашаться невротичные интроверты и невротичные экстраверты.

Если же на практике окажется, что в тесте будут преобладать пункты из квадранта "нейротизм-интроверсия", то, с точки зрения теории Айзенка, это означает, что шкала "нейротизм" оказывается нагру-

женной иррелевантным фактором - "интроверсией". (Точно такой же эффект возникает, если появится перекося в выборке - если в ней будет больше невротичных интровертов, чем невротичных экстравертов).

Чтобы избежать таких сложностей, желательно иметь дело с такими эмпирическими показателями (пунктами), которые однозначно информируют только об одном факторе. Но это требование реально никогда не выполняется: всякий эмпирический показатель оказывается детерминированным не только тем фактором, который нам нужен, но и другими - иррелевантными факторами задачи измерения.

Поэтому основная задача состоит в специальном подборе пунктов так, чтобы все потенциальные иррелевантные факторы были уравновешены: ни один из них не встречался бы чаще других на множестве пунктов, включенных в тест. В связи с этим, можно дать следующее определение конвергентной и дискриминантной валидности.

Конвергентная валидность теста – это соответствие пунктов измеряемому фактору.

Дискриминантная валидность теста – это сбалансированность пунктов относительно иррелевантных факторов.

Эмпирически она выражается в отсутствии значимой корреляции с тестом, измеряющим концептуально независимое свойство. С точки зрения теории Айзенка, тест Тейлора не обладает дискриминантной валидностью по отношению к фактору "экстраверсия-интроверсия", хотя и обладает определенной конвергентной валидностью по отношению к релевантному фактору - "нейротизм".

5. Экспертная эмпирическая валидизация. В отсутствие какого-либо уже валидизированного теста, параллельно измеряющего изучаемое свойство, а также в отсутствие разработанного теоретического контекста, позволяющего проверять конструктивную валидность, психодиагност ока-

зывается перед необходимостью привлечения к валидации теста экспертов [1, 15].

В отличие от экспертного анализа содержания теста эмпирическая экспертная валидизация предполагает работу экспертов не с тестом (лучше, чтобы о нем эксперты вообще ничего не знали), а с испытуемыми из выборки стандартизации.

Экспертам следует обеспечить стандартные условия для наблюдения за испытуемыми. Но не всегда такое стандартизованное наблюдение удастся организовать. Даже если мы предприняли серьезные усилия по организации наблюдения за поведением испытуемых в какой-либо искусственной лабораторной ситуации, такое наблюдение все равно будет значительно уступать по информативности "полевому" наблюдению - в естественных условиях. Если измеряемое свойство теоретически определено как устойчивая универсальная черта личности - как диспозиция к инвариантному поведению в широком спектре ситуаций, то и отдельного полевого наблюдения окажется недостаточно для получения полноценного экспертного критерия валидности.

Поэтому на практике часто прибегают к оценкам особого типа, к субъективным оценкам, которые выносят испытуемому люди из его круга, имеющие опыт реального общения с ним [15]. С учетом этого процедура оценивания приспособляется к обычным людям, не являющимися психологами. На психолога падает большая нагрузка по составлению детальной инструкции оценщикам, однозначно задающей смысл оцениваемой характеристики. Лучшие условия для такой процедуры возникают при наличии группы испытуемых, тесно общающихся между собой, которые могут одновременно побывать и испытуемыми по отношению к тесту, и оценщиками по отношению друг к другу.

В отечественной литературе эта процедура получила сокращенное

обозначение ГОЛ - "групповая оценка личности" [1, 15]. Для того, чтобы групповая оценка была источником действительно валидной информации, оценщики должны согласованно оценивать испытуемых. Если в оценках разных оценщиков нет согласованности, то это означает, что либо оцениваемое свойство не проявилось у объекта оценивания, либо оценщики по-разному проинтерпретировали инструкцию.

Для измерения согласованности должна быть построена таблица с оценками (табл. 3). Методы анализа данных, содержащихся в такой таблице, формально совершенно эквивалентны тем методам, которые применяются для обработки таблиц "испытуемые - пункты". В частности, суммы по строкам дают суммарные баллы, полученные каждым испытуемым у всех K оценщиков. Таким образом, оценщики в данном случае оказываются формально в роли теста. Рассчитывая попарные корреляции между различными столбцами таблицы 4 можно получить коэффициенты согласованности для отдельных пар оценщиков. Глобальной мерой согласованности оценщиков может служить тот же коэффициент надежности α Кронбаха.

Если сама групповая оценка не обнаруживает надежности, то она не может использоваться в качестве критерия валидации при проверке валидности теста.

Таблица 4. Групповая оценка личности независимыми экспертами

Оценщики Испытуемые	O₁	O₂	...	O_K	Σ
I₁	X ₁₁	X ₂₁	...	X _{K1}	C ₁
I₂	X ₁₂	X ₂₂	...	X _{K2}	C ₂
...
I_n	X _{1n}	X _{2n}	...	X _{Kn}	C _n

Эмпирическое значение коэффициента валидности рассчитывается как линейная или ранговая корреляция между двумя рядами значений - тестовыми баллами и суммарными баллами экспертной оценки [1, 15].

6. Анализ пунктов по критерию валидности. Известно [1, 6, 15, 16], что валидность целого теста зависит от валидности входящих в него пунктов, причем максимальная валидность по критерию достигается за счет отбора таких пунктов из теста, которые, обладая значимой корреляцией с критериальным показателем, минимально коррелируют между собой.

Следует напомнить, что отбор пунктов именно по критерию валидности обеспечивает *максимальную прагматическую эффективность теста*.

Реально такой отбор можно произвести, рассчитывая бисериальную корреляцию (или фи-корреляцию) критерия с каждым пунктом из пилотажной батареи.

Таким образом, еще раз подчеркнем, что анализ валидности отдельных пунктов служит не только прагматическим целям, но может и должен служить целям углубления представлений о содержательно-теоретическом смысле измеряемого свойства.

На основании содержательного анализа пунктов, отобранных по критерию, психолог уточняет и корректирует свою первоначальную теоретическую схему, свое понимание измеряемого свойства [1, 15].

7. Порядок работы психолога при проверке валидности. Очень трудно выделить универсальный алгоритм работы психолога по проверке валидности, ибо существуют различные подходы к обеспечению валидности, обусловленные теоретико-методологическими различиями определенных психологических школ.

Для прагматически ориентированных тестологов (какими традици-

онно являлись до недавнего времени почти все англо-американские специалисты) главный момент - поиск операционально строго заданного социально-прагматического критерия валидности по отношению к которому диагностические тесты и их составные части (пункты) подбираются как бы "автоматически" - в ходе эмпирико-статистических процедур сбора и корреляционного анализа результатов.

Но, конечно, неверно было бы приписывать этому подходу "бездумность в опоре на статистику": ведь статистика только тогда позволяет выявить валидное подмножество тестов (пунктов), когда исходное множество подобрано не случайно, а с использованием априорных корректных содержательно-психологических представлений.

Современные методологи психологического тестирования фактически единодушно приходят к признанию (как наиболее оптимальной) рационально-эмпирической стратегии конструирования теста и проверки валидности [15].

1. *Прогнозирование результатов корреляционных экспериментов по проверке конструктивной валидности.* Провести теоретический анализ диагностического конструкта, разработать теоретическую концепцию тестируемого психического свойства. Выявить (с использованием литературы) системы взаимосвязанных диагностических конструктов, внутри которых новый диагностический конструкт характеризуется определенными структурно-функциональными связями и отношениями.

2. *Конструирование пунктов теста.* Выявить составные части теоретического конструкта, сформулировать системы "эмпирических индикаторов" - операционально однозначных показателей, фиксирующих проявление конструкта в различных поведенческих ситуациях.

3. *Формулирование релевантного социально-прагматического критерия* для проверки валидности - эффективности методики.

4. *Оценка валидности эмпирических индикаторов.* Спланировать и провести корреляционное исследование (или квазиэксперимент) на специально подобранной выборке испытуемых, для которых известно значение (или будет известно) критериального показателя, а также результаты по родственным психологическим тестам. При необходимости на этих испытуемых провести дополнительные тесты с целью - получить возможность корреляционной проверки конструктивной валидности теста (экспертные оценки в данном случае рассматриваются в статусе одной из возможных параллельных процедур получения критериальной или психологической информации).

5. *Оценка достоверности* эмпирических индикаторов.* Провести исследование достоверности результатов, если используется самоотчет и диагностическая ситуация может быть воспринята испытуемыми с настроенностью.

6. *Отсев пунктов (индикаторов), не удовлетворяющих критериям валидности и достоверности.* Измерить надежность для сокращенной шкалы, состоящей только из валидных пунктов. Если надежность оказывается невысокой, то психолог снова возвращается к этапу 1 и уточняет теоретические представления.

(*О достоверности см. §9).

Контрольные вопросы для самопроверки

§7, 8. Дискриминативность теста. Валидность теста.

1. Что понимается под дискриминативностью теста?
2. Как рассчитать дискриминативность отдельных заданий теста?
3. Что понимается под валидностью теста?
4. С чем больше связано понятие валидности - с предметом или объектом измерения?
5. Что понимается под термином "эмпирическая валидность"?
6. Что включает в себя процедура экспертной валидации?
7. Что понимается под проспективной валидностью теста?
8. Что понимается под ретроспективной валидностью теста?
9. Какие задачи решаются с помощью доказательства проспективной

и ретроспективной валидности теста?

10. Что понимается под содержательной валидностью теста?
11. Что понимается под конструктивной валидностью теста?
12. Дайте определение конвергентной валидности теста.
13. Дайте определение дискриминативной валидности теста.
14. Какова суть процедуры экспертной эмпирической валидации?
15. Что из себя представляет процедура ГОЛ "групповая оценка личности"?
16. Какие действия последовательно должен произвести психолог при доказательстве валидности теста?

Литература к теме

1. Анастаси А. Психологическое тестирование. В 2-х кн. М., 1982.
2. Бурлачук Л.Ф., Морозов С.М. Словарь-справочник по психодиагностике. - СПб.: Изд-во "Питер", 1999. - 528 с.
3. Корниенко А.Ф. Теория и практика психологического исследования. Учебное пособие. Казанский пед. университет, Казань, 2000. -258 с.
4. Клайн П. Справочное руководство по конструированию тестов. - Киев: ПАН ЛТД, 1994.
5. Общая психодиагностика / Под ред. А. А. Бодалева, В. В. Столина.- М., 1987.
6. Основы психодиагностики / Под ред. А. Г. Шмелева. Ростов-на-Дону., 1996.

§9. Достоверность самоотчета.

Внутри проблемы валидности целесообразно выделить вопрос об обеспечении валидности процедур стандартизованного самоотчета, к которым относятся различные техники шкалирования, классифицирования, сравнения и тест-опросники. Специфические *проблемы валидности, связанные с активностью человека как объекта в психодиагностике*, целесообразно обозначить особым образом - проблемы обеспечения "*достоверности*".

Вербальная форма подачи тестового материала порождает у испытуемого определенные размышления о цели тестирования. Если ситуация (например, экспертиза) диктует испытуемому необходимость фальсифи-

кации ответов, то испытуемый редко отказывается от этой возможности.

Психологические факторы, от которых зависит достоверность данных самоотчета, условно можно сгруппировать в следующие классы [15]:

1. *Факторы знания.* У испытуемого может быть более или менее осознанное представление о следующем:

- свойственно или нет ему в действительности тестируемое поведение (с некоторыми ситуациями, имплицитно подразумеваемыми в вопросе тест-опросника, испытуемый мог на практике никогда не встречаться);

- какое личностное свойство скрывается за тем или иным конкретным поведением, описанным в суждении;

- как это свойство соотносится с общепринятыми моральными нормами и признаками социального успеха.

2. *Фактор "социальной желательности".* Обозначает тенденцию испытуемого давать о себе социально одобряемую информацию. Сила этой тенденции зависит как от общей внеситуативной установки испытуемого на морализацию "Я-образа" и социальную успешность, так и от того, насколько эту установку актуализирует сама ситуация тестирования (диагностическая ситуация). Однако эта тенденция не будет давать систематического искажения, если испытуемые не смогут разгадать направленность опросника и связать диагностируемое свойство с тем или иным полюсом "социальной желательности".

Таким образом, действие этого фактора до некоторой степени опосредовано действием фактора "знание". Однако при диагностике личностных свойств, тесно связанных с психической "нормальностью" или "социальной успешностью", фактор "социальной желательности ответа" обуславливает очень серьезные искажения.

3. *Факторы "индивидуальной тактики".* Здесь подразумевается

действие "Я-концепции" ("Я" для себя) и "Я-образа" ("Я" для других) на ситуативную тактику испытуемого в момент тестирования. Выполняя тест-опросник, испытуемый всегда находится в невольном диалоге с самим собой и в своих ответах на вопросы раскрывает себя не только для других, но и для себя самого. Испытуемый стремится подтвердить "Я-концепцию" или фальсифицировать определенный "Я-образ" с заданными свойствами.

Как правило, в ситуациях высокого социального риска "Я-образ" полностью доминирует.

Например, преступник при экспертизе стремится, прежде всего, предстать больными или неприспособленным к жизни, хотя в действительности ему было бы приятно думать о себе как о вполне адаптированном здоровом человеке. Точно так же склонные подчеркивать свои трудности и проблемы клиенты, обратившиеся за помощью к психологу-психотерапевту (чтобы вызвать к себе повышенное внимание). В менее регламентированных ситуациях, наоборот, может доминировать мотивация самопознания: в этом случае испытуемый невольно стремится подтвердить с помощью теста свои гипотезы о самом себе.

Заслуживают внимания также особые формы отказа испытуемого от тестирования: позиционный стиль ответа ("соглашательство" или, наоборот, "отрицание"), случайные ответы. Для выявления подобных отказов обычно достаточно довольно простых мер:

- для исключения влияния соглашательства ("отрицания") применяются перечни с "прямыми" (ответ "верно" в пользу измеряемого свойства) и "обратными" (ответ "неверно" в пользу измеряемого свойства) пунктами. Кроме того, производится подсчет баланса "подтверждающих" и "отвергающих" ответов: если баланс резко нарушается, то протокол признается бракованным;

- для выявления случайных ответов в большие перечни вводят вопросы-дубли (синонимические перифразы) или прямые повторы: если испытуемый слишком часто по-разному отвечает на одинаковые вопросы, значит он применяет случайную тактику. Вводят также и крайне ред-

кие утверждения, с которыми испытуемые, как правило, соглашаются только по ошибке.

Более сложные методы требуются для борьбы с социальной желательностью. Рассмотрим три наиболее часто используемых варианта.

1. *Введение особых шкал "лжи" в диагностический вариант методики.* Они состоят из "вопросов-ловушек": тот или иной ответ на эти вопросы явно предопределен социальной желательностью. Если испытуемый набирает слишком высокий балл по шкале лжи, его протокол бракуется. Более тонкий вариант - введение шкал "коррекции" (например, в ММРІ). Получение определенного балла по этим шкалам вызывает внесение поправок к баллу по другим шкалам, скоррелированным со шкалой коррекции. Величина поправок определяется коэффициентом линейной регрессии (измеренным в нормативном эксперименте) между баллами, полученными по шкале коррекции и основной диагностической шкале (шкале свойства).

2. *Использование инструкции на преднамеренную фальсификацию результатов для устранения или сбалансирования "социальной желательности".* Участникам предварительных замеров кроме обычной инструкции дается (во вторую очередь) дополнительная: "Заполните опросник от лица человека, желающего произвести самое благоприятное впечатление". Затем производится отбор пунктов на основании того, насколько процент ответов на них отличается от 50 процентов (где 50% - величина, ожидаемая для пунктов, являющихся нейтральными с точки зрения "социальной желательности").

В качестве меры желательности в данном случае можно воспользоваться следующим коэффициентом:

$$\hat{S} = \frac{N(+)-N(-)}{N(+)+N(-)} = \frac{2N(+)-n}{n} \quad \text{где:}$$

$N(+)$ - сумма ответов "верно" на данный пункт при инструкции на фальсификацию;

$N(-)$ - сумма ответов "неверно" в тех же условия;

n - численность испытуемых.

Значимость коэффициента приближенно оценивается по критерию "хи-квадрат" (см. стр. 197). В данном случае в правой части формулы вместо ϕ подставляется \hat{S} :

$$\chi^2 = \hat{S}^2 n.$$

Поскольку инструкция на преднамеренную фальсификацию создает условия, в которых социальная желательность суждений акцентируется, то значимыми следует считать такие высокие по модулю значения, при которых расчетное χ^2 превышает теоретическое (табличное) значение χ^2 . Такие пункты должны быть либо полностью исключены из опросника (что редко удается), либо число "положительных" и "отрицательных" должно быть уравновешено.

Подобным образом, могут быть отобраны и пункты для шкалы лжи. Суммарный балл по шкале лжи распределяется так, как показано на рисунке 6.



Рис. 6. смещение распределения тестовых баллов по шкале «лжи» к полюсу высоких значений X при инструкции на выполнение теста с позиции «идеального» человека (по В.В. Столину)

В качестве критерия разделения испытуемых на шкале лжи выбирается такая точка, которая позволяет минимизировать ошибки типа "пропуск" (зачисление "лживых" испытуемых в число "правдивых") и ошибки типа "ложная тревога" (зачисление "правдивых" в число "лживых"). Положение критической точки на шкале можно менять в зависимости от баланса цены двух типов ошибок: в некоторых случаях "пропуск" гораздо опаснее, чем "ложная тревога" [15].

Фальсифицирующая инструкция используется также и для исследования степени "скрытности-открытости" формулировок вопросов. Например, испытуемому дается инструкция на симуляцию высокой тревожности по опроснику Тейлор. В этом случае, ответы на многие пункты значительно изменяются.

Такие пункты являются слишком открытыми - информируют испытуемого об измеряемом свойстве и позволяют ему вносить тенденциозные искажения в результаты в своих интересах.

3. *Стратегия "сбалансирования социальной желательности"* стала применяться с распространением факторного анализа. Ставится цель обеспечить дискриминативную валидность теста относительно шкалы "социальной желательности". Это достигается с помощью факторного анализа корреляций между пунктами. Факторный анализ, в применении к одномерному тест-опроснику, как правило, выделяет два фактора: относящийся к самому измеряемому свойству и относящийся к социальной желательности.

Из перечисленных выше трех методов первый относится к отсеву испытуемых и требует введения в перечень вопросов для шкалы "лжи". Методы 2 и 3 позволяют отобрать только такие пункты, которые обеспе-

чивают устранение социальной желательности. Но они, как правило, трансформируют сам конструкт, который обязательно становится ортогональным (независимым) к социальной желательности.

При необходимости диагностики свойств, обязательно коррелированных с желательностью, единственный метод состоит в применении шкал коррекции и корректирующих поправок, но и этот метод нельзя считать вполне надежным. Так что диагностика свойств, сопряженных с социальной желательностью, в ситуациях экспертизы всегда рискованна.

Однако, в ситуациях, когда сам испытуемый заинтересован в точных результатах ("ситуация клиента"), психодиагност может пользоваться опросниками, не опасаясь артефакта социальной желательности.

При подготовке *особенно значимого* психодиагностического обследования, в котором надо принципиально исключить всякую возможность преднамеренной фальсификации результатов, следует дополнить критерий оценки достоверности с помощью особой шкалы "лжи" еще одним критерием, основанным на использовании "фальсифицирующей" инструкции" [15]. Для этого после обычной инструкции той же самой пилотажной выборке психолог дает инструкцию: "А теперь снова выполните задание, но старайтесь описать себя так, чтобы выглядеть максимально благоприятно в глазах большинства других людей". В результате применения такой инструкции психолог получает дополнительную таблицу, содержащую фальсифицированные данные.

В таком случае кроме подсчета очень грубого индекса "желательности" (\hat{S}) у психолога возникает возможность поэлементного сравнения ответов испытуемых на один и тот же вопрос в обычной и "фальсифицированной" диагностической ситуации. Очевидно, что недостоверным следует считать вопрос, ответы на который будут изменены в "фальсифицированной" ситуации в определенном систематическом направлении.

Здесь может пригодиться простейший критерий значимости изменений основанный на распределении "хи-квадрат". Для каждого пункта строится четырех клеточная матрица сопряженности:

		ПОСЛЕ инструкции	
		Ответы	Да
ДО инструкции	Да	A=40	B=36
	Нет	C=22	D=48

Здесь в клеточке А указана частота сохранения ответа "верно" на некоторый i-тый вопрос (например, из 76 ответивших "верно" таких оказалось только 40 испытуемых), в клетке В - частота изменения ответа "верно" на ответ "неверно" и т.п. Как видим, в этом примере показано, что смена инструкции привела к значительным изменениям.

Для оценки значимого направления этих изменений используется критерий Макнимара, который предложил сравнивать между собой по критерию хи-квадрат только элементы В и С этой таблички:

$$\chi_e^2 = \frac{(|B - C| - 1)^2}{B + C}$$

где χ_e^2 - вычисленное эмпирическое значение статистики хи-квадрат с одной степенью свободы. Для нашего примера $\chi_e^2 = 2,91$, что ниже, чем граничное значение $\chi^2 = 3,84$, и, следовательно, нулевая гипотеза об отсутствии значимых направленных изменений не может быть отвергнута - пункт не является значимо нагруженным артефактным фактором социальной желательности и может быть включен в диагностический вариант тест-опросника без изменений.

По результатам такого исследования удобно составить таблицу Кх2: в первом столбце - показатели корреляции пунктов со шкалой лжи,

во втором - показатели значимости изменений при переходе к фальсификации. *Безусловно достоверными* можно считать только те пункты, которые не получили значимых коэффициентов ни в первом ни во втором столбце.

Если таких пунктов оказывается слишком мало для составления надежной шкалы и если среди недостоверных пунктов достаточно много таких, которые обладают существенной внутренней (корреляция с суммарным баллом по основной шкале) или внешней (корреляция с критерием) валидностью, то следует прибегнуть к тактике балансирования: ввести в основную шкалу поровну "прямых" и "обратных" пунктов по шкале "социальной желательности" так, чтобы все четыре квадранта были заполнены пунктами равномерно (среди "прямых" по основной шкале было бы поровну "прямых" и "обратных" по "желательности", и среди "обратных" по основной - также поровну) [15].

Без указанных предосторожностей тест-опросник неизбежно будет давать систематические искажения результатов всякий раз, когда испытуемый будет квалифицировать диагностическую ситуацию как ситуацию экспертизы.

Другой путь - управление процессами категоризации в ходе самой диагностики. В так называемой "репертуарной модификации" теста-опросника испытуемому специально предлагают выполнять тест не только от своего имени, но и от имени определенного репертуара ролей - "большинство людей", "моральный человек", "преуспевающий человек" и т.п.

Однако применение таких методов требует сложной компьютерной обработки, а, следовательно, не всегда доступны.

Контрольные вопросы для самопроверки

§9. Достоверность теста

1. Что понимается под достоверностью методики?
2. Какие основные факторы влияют на достоверность данных само-

отчета?

3. Какие методы необходимо применить, для того чтобы существенно повысить достоверность методики?

4. Как определить ту критическую точку на "шкале лжи", при которой суммарный балл, набранный испытуемым по этой шкале, позволяет не браковать протокол по всему тесту?

5. В чем суть стратегии "сбалансирования социальной желательности"?

Литература к теме

1. Анастаси А. Психологическое тестирование. В 2-х кн. М., 1982.

2. Бурлачук Л.Ф., Морозов С.М. Словарь-справочник по психодиагностике. - СПб.: Изд-во "Питер", 1999. - 528 с.

3. Общая психодиагностика / Под ред. А. А. Бодалева, В. В. Столина.- М., 1987.

4. Основы психодиагностики / Под ред. А. Г. Шмелева. Ростов-на-Дону., 1996.

§10. Технология создания и адаптации методик.

Рассмотрим некоторые аспекты *целостной стратегии* конструирования, эмпирической апробации и внедрения методики в практику [15]. В реальном случае создание оригинальной методики или адаптация зарубежной методики не могут сводиться только к проверке (или перепроверке) отдельных психометрических свойств - репрезентативности, надежности, валидности, достоверности - в произвольной последовательности. В одних случаях целесообразно начинать с одного, в других - с другого этапа работы.

По существу любая реальная ситуация использования теста не является ситуацией только "конструирования" или "применения". Можно без преувеличения сказать, что существует континуум между крайними полюсами:

"создание"----- "применение"

и каждая ситуация до определенной степени удалена от обоих полюсов. Трудно назвать такой случай, когда бы конструирование совер-

шенно нового теста начиналось "с нуля", "на пустом месте". Также трудно найти и такие случаи, когда все аспекты тестирования были бы полностью неизменными и воспроизводили бы уже совершенно исследованную нормативную ситуацию применения готового теста.

Однако на практике, все многообразие ситуаций, всю комбинаторику независимых параметров психологи, как правило, пытаются свести к двум-трем типовым ситуациям [15].

1. *Ситуация применения.* Тест кем-то разработан (возможно, в других социокультурных условиях), известны тестовые нормы, полученные на представителях данной языковой культуры (несоответствие выборки стандартизации и выборки применения по половозрастной структуре и профессионально-культурным признакам признается несущественным).

2. *Ситуация адаптации.* Тест кем-то разработан - проверены надежность и валидность, но *отсутствуют* тестовые нормы (как правило, отсутствуют вообще для любых представителей данной языковой культуры). Задача адаптации сводится, таким образом, к построению тестовых норм.

3. *Ситуация создания.* Есть концепция психического свойства, но нет "подходящей" процедуры его измерения, удовлетворяющей требованиям места, времени, возможностям количественного анализа и ограничениям прочих ресурсов. Необходимо разработать измерительную процедуру, проверить ее надежность, валидность, построить тестовые нормы.

Но уже к середине 80-х годов стала совершенно очевидна несостоятельность такой упрощенной системы представлений.

Прежде всего, остановимся на вопросах адаптации "переводных" тестов. Данный путь быстрого пополнения репертуара методик кажется многим психологам наиболее экономичным, кратчайшим к надежной и валидной психодиагностике.

Однако, если при этом адаптация сводится только к построению нормативного распределения тестовых баллов, то это означает, что валидность и надежность адаптированной методики в новых условиях принимаются на веру, а теоретическая концепция автора теста и содержание использованных им критериев валидности просто *переносятся в наши условия без изменений* (ведь для любой, в том числе и для невалидной и ненадежной методики, можно получить распределение).

Подобный перенос дает пренебрежимые погрешности только для тестирования относительно элементарных психических свойств (таких, как свойства нервной системы, функциональные состояния, сенсомоторные параметры, элементарные когнитивные функции, т. е. структурно-динамические характеристики психической деятельности), причем с использованием объективных процедур (психофизиологическая регистрация, тесты с "физическими" критериями успеха и т. п.).

При тестировании интегральных психических свойств личности и индивидуального сознания (черты, мотивы, установки, самооценка, общие способности, стиль общения, ценностные ориентации, интересы и т. п.), а также употреблении любых лингвистических средств в самой процедуре тестирования, - во всех этих случаях *ограничиваться только сбором тестовых норм при адаптации недопустимо!*

Требуется серьезная эмпирическая работа по проверке надежности и валидности в новых социокультурных условиях, работа, фактически соответствующая по своему объему созданию оригинальной методики.

С этой точки зрения заимствование зарубежных диагностических тестов способностей, черт характера, интересов и т. п. вовсе не оказывается кратчайшим путем к психодиагностике. Рассмотрим в качестве примера необходимые этапы эмпирико-статистической работы при адап-

тации многомерного переводного тест-опросника [15].

1. *Анализ внутренней валидности, внутренней согласованности пунктов, из которых состоит тест-опросник.* Этот анализ совершенно необходим, если в зарубежной методике он применялся в качестве средства самого конструирования методики. Этот анализ призван показать, что некое (еще неясно какое именно) общее диагностическое свойство, "лежащее на пересечении всех эмпирических индикаторов", действительно существует.

Такой анализ обязателен по отношению ко всем тестовым шкалам, полученным с помощью факторного анализа, например к тест-опросникам Айзенка EPI и 16PF Кеттелла. В то время, как к опроснику "локус контроля" или ко многим основным клиническим шкалам ММРІ требование внутренней согласованности применять необязательно, поскольку пункты в эти шкалы подбирались по внешнему критерию и не связаны в один фактор. Анализ внутренней согласованности может быть применен и к одномерным, и к многомерным тестам.

2. *Проверка устойчивости к перетестированию.* Совершенно необходима при диагностике свойств, по отношению к которым теоретически ожидается инвариантность во времени.

Анализ ретестовой надежности может быть (так же как анализ надежности - согласованности) совмещен с исследованием информативности отдельных пунктов теста, а также, возможно, и устойчивости отдельных пунктов. Без сведений о ретестовой надежности *психолог не имеет права использовать тест* для построения любого элементарного статистического экстраполирующего прогноза.

3. *Анализ корреляций с релевантным внешним критерием.* Этот этап адаптации совершенно необходим, если тест разрабатывался изначально как критериально-ориентированный, т. е. отбор пунктов произво-

дился на основании их корреляций с каким-то критерием валидности. Например, подобная работа проделана для сокращенной модифицированной версии ММРІ коллективом Ф. Б. Березина.

4. *Рестандартизация тестовых норм.* До недавнего времени только этот этап работы по адаптации тестов признавался всеми психологами как необходимый. Но и в этом случае далеко не всегда воспроизводилась необходимая статистическая работа по проверке устойчивости полученного распределения тестовых баллов к расщеплению выборки.

5. *Проверка воспроизводимости структуры отношений между шкалами* является специфичным этапом для многомерных тестов. Например, для теста Айзенка принципиальна ортогональность, статистическая независимость факторов "экстраверсия - интроверсия" и "нейротизм - эмоциональная стабильность". На воспроизводимости структуры связей шкал между факторами 16PF основывается корректность подсчета вторичных факторов [13].

Пять указанных этапов адаптации позволяет убедиться в том, что адаптация зарубежных тестов мало чем уступает по объему эмпирико-статистической работы конструированию оригинальных методик.

Несколько другие задачи ставит перед психологом ситуация "внутрикультурного переноса" теста на новую популяцию, отличающуюся от выборки стандартизации половозрастными или профессионально-культурными особенностями. В этом случае необходимо [15]:

1. **Проверить валидность методики**, если методика чувствительна по своему содержанию к профессиональной или региональной специфике (могут ли отвечать пенсионеры, или школьники, или жители отдаленных регионов на вопросы, подразумевающие типичные ситуации из жизни студента, обучающегося в крупном городе европейской части?). Для того надо выбрать по возможности максимально экономичную процедуру про-

верки валидности. Размеры выборки в эксперименте по проверке валидности должны быть таковы, чтобы можно было бы надеяться на получение статистических значимых связей между тестовым показателем и критерием валидности (это, как правило, *не менее 30 испытуемых*).

Если результаты проверки валидности оказываются неудовлетворительными (коэффициент корреляции явно ниже 0,5 и дальнейшее наращивание выборки все равно не оправдывает применение теста со столь низким показателем валидности), то по собранным результатам следует произвести простейший "анализ пунктов" - посмотреть, не оказались вторые пункты явно неинформативными (все испытуемые отвечают одинаково), не оказались ли некоторые вопросы явно бессмысленными и слишком "прямыми", социально значимыми в данной ситуации. Не исключено, что при удалении неудачных пунктов из подсчета тестового балла необходимая валидность будет обнаружена.

В отсутствие доступного внешнего критерия необходимо прибегнуть к проверке внутренней валидности методики.

2. *Проверить соответствие тестовых норм.* Только после позитивного результата в проверке валидности целесообразно расширять тестовых баллов.

Квантили этого распределения необходимо сравнить с квантилями нормативного распределения: если расхождения оказываются пренебрежимыми (например, не превышают ошибки измерения), то можно принять вывод о приложимости к данной популяции универсальных тестовых норм. Однако на практике чаще возникают значимые отличия.

В этом случае психолог должен построить так называемые "частные" тестовые нормы.

Подчеркнем, что абсолютное большинство ситуаций, которые обычно расцениваются как ситуации "применения", на самом деле явля-

ются ситуациями более или менее серьезного "внутрикультурного переноса". Именно эти обстоятельства определяют высокий уровень требований к психометрической подготовке психологов-психометристов [15].

При создании новых диагностических методик решающее значение имеют методологическая ориентация и статус психолога.

Под статусом в данном случае понимаются существенные различия в нормативных (закрепленных в должностных инструкциях) требованиях к продукту деятельности психолога, работающего в исследовательском учреждении; и психолога, работающего в практическом учреждении.

Если в первом случае психолог имеет право считать своим "конечным продуктом" внутриспсихологическую валидизацию сконструированной методики, то во втором случае он, как правило, обязан обеспечить практическую эффективность своей психодиагностической программе - указать на статистически значимую связь результатов диагностики с величиной какого-то социально ценного показателя - критерия, а затем построить на основе этой связи психологическую концепцию "вмешательства" в ситуацию с целью ее изменения в желаемом направлении.

Эти изначальные различия в статусе психолога-исследователя и психолога-практика определяют оправданные различия в стратегии конструирования тестов и тестовых батарей.

Для психолога-исследователя ведущая ценность - это теоретическая обоснованность и эмпирическая однозначность диагностического конструкта, которую он достигает с помощью оправданной ориентации на конструктивную валидизацию.

У исследователя хватает ресурсов для того, чтобы обеспечить

множественность разнообразных по своему статусу критериев валидности - от субъективных оценок валидности включенных наблюдателей (L - данные по Кеттеллу) до экспериментального моделировании реальных ситуаций проявления измеряемого свойства (T-данные по Кеттеллу).

Для психолога-практика ведущая ценность - это эффективность, пусть даже она будет достигнута с помощью теоретически "эkleктичного" инструмента - эмпирического средства, не имеющего соответствующего научно-строгого конструкта [15].

Появление подобных прагматических тестов, очевидно, обусловлено действием объективного социально-исторического механизма, который можно было бы кратко назвать "*опережающим запросом практики*". Этого феномена не было бы, если бы все запросы практики можно было бы заранее предвидеть и рационально планировать, подготавливая релевантные научно обоснованные диагностические процедуры.

Но такие предвидение всегда удастся осуществить лишь в какой-то мере - реальная практика тем и отличается от дедуктивного движения в плоскости абстракций, что всегда приводит к столкновению с новыми и неожиданными явлениями, что обуславливает и появление новых методик.

Большинство ценных прагматических тестов становятся в дальнейшем предметом для изучения со стороны психологов-исследователей, апостериорно реконструирующих научную концепцию валидности этих тестов, что значительно улучшает их характеристики и активизирует сферу применения. Психодиагностика как специфический институт, реализованный в форме управляющего кольца "наука - практика - наука - практика...", достигнет оптимальных режимов в своем функционировании, если любые попытки конструирования и применения любых тестов, будут тщательно документироваться, будут доступны для повторения в

любом звене (исследовательском или практическом) психологической службы.

При обеспечении эффективной информационной базы, оперативных форм информационной связи между звеньями этой системы (как методическими, так и практическими) всякое практическое применение тестов станет одновременно и работой по конструированию - по подготовке новых, пересмотренных версий методик - как со стороны психометрики, так и со стороны теории.

Итак, при создании нового теста психолог должен придерживаться определенного алгоритма действий [13]. Рассмотрим коротко его суть. Алгоритм создания нового теста сводится к следующим основным этапам:

- формирование информационной базы исследования;
- конструирование интегральных показателей;
- классификация испытуемых на однородные группы.

Формирование информационной базы исследования. Сформировать информационную базу для создания теста - это значит выделить систему признаков (характеризующих измеряемый психический феномен), построить на их основе «пробный» вариант теста и с его помощью провести пилотажное исследование.

При этом наиболее сложной и трудно формализуемой частью этого этапа является выявление системы исходных признаков, т. к. предмет измерения изначально определяется, как правило, в довольно общем виде. Как отмечает Л. Т. Ямпольский [13], наилучшим основанием для содержательной разработки предмета тестирования является наличие хорошо развитой психологической теории (что в психологии бывает довольно редко). В связи с этим, конечным результатом содержательного анализа должно быть четкое вербальное определение анализируемого конструкта и расчленение его на составные части.

После этого переходят к поиску системы признаков, релевантно отражающих основное содержание конструкта. Система признаков должна удовлетворять следующим общим требованиям:

- полнота описания - система должна охватывать все аспекты измеряемого понятия;

- экономичность описания - требование, которое диктуется временными ограничениями, трудностями сбора, обработки и анализа больших по объему массивов данных;

- четкая структурированность системы признаков - признаки должны группироваться достаточно равномерно относительно ранее выделенных аспектов измеряемого понятия;

- количественная определенность отбираемых признаков - признаки должны допускать последующую формализацию (математическую обработку).

На основе этих требований должен быть составлен список заданий (пунктов) и создан первоначальный вариант теста.

Поскольку качество отдельных заданий (вопросов, утверждений) является определяющим для всего опросника, рассмотрим трудности, возникающие при формулировании заданий, а также преимущества различных типов заданий.

При разработке заданий для личностных опросников необходимо учитывать следующие проблемы, которые, если их не обойти, неизбежно приведут к низкой валидности тестов [8].

1. *Установка на согласие.* Это тенденция испытуемого соглашаться с утверждениями или отвечать на вопросы "да" независимо от их содержания. Чаще всего проявляется, когда утверждения (вопросы) неоднозначны и неопределенны.

2. *Установка на социально одобряемые ответы.* Это тенденция ис-

пытуемых отвечать на вопросы теста так, чтобы выглядеть "социально положительным": если возможен "социально желательный" ответ, то весьма вероятно, что испытуемые будут его давать. Эта установка оказывает, в частности, влияние на результаты теста ММРІ.

3. *Установка на неопределенные или средние ответы.* Если в опроснике представлена средняя категория ответов, отражающая нерешительность или неуверенность в ответе (например, "не уверен", "не знаю", или "затрудняюсь ответить"), то многие испытуемые склонны к ней прибегать, как к безопасному компромиссу. Это приводит к снижению валидности заданий, поскольку большинство методов анализа вопросов основывается на крайних значениях показателей.

4. *Установка на "крайние" (расположенные по краям шкалы) ответы.* Эта установка может проявляться при использовании многоэлементной рейтинговой шкалы. Некоторые испытуемые, независимо от содержания вопросов, предпочитают выбирать крайние ответы.

5. *Выборка из генеральной совокупности вопросов.* В тестах интеллекта и специальных способностей относительно легко убедиться в том, что задания принадлежат или с высокой степенью вероятности выглядят принадлежащими к подразумеваемой для них генеральной совокупности. Лингвистические задания нелегко перепутать с математическими или какими-либо другими. Однако, в области особенностей диагностики личности и темперамента все гораздо сложнее.

В самом деле, иногда даже опытные разработчики вопросов обнаруживают, что вопросы нагружены не теми факторами, для выявления которых они были созданы или, что вопросы вообще не нагружены ни одним из факторов. Эта проблема с точки зрения классической теории тестов состоит в трудности определения конкретной выборочной совокупности вопросов. В результате возникает проблема подбора адекватной выборки из

выборочной совокупности вопросов или утверждений для теста, а без этого тест не может быть валидным.

6. *Выборка из генеральной совокупности испытуемых.* Известно [8], что в личностных тестах труднее обеспечить адекватность выборки из выборочной совокупности, чем в тестах способностей. При использовании тестов способностей обычно известна вполне определенная категория лиц (популяция), для которой данный тест предназначен, и таким образом, по крайней мере, в принципе, можно эффективно подбирать выборки.

В личностных тестах, в отличие от тестов, разрабатываемых для испытуемых с той или иной патологией, в идеале нужны выборки из общей популяции (т.е. всего населения страны), в которой встречаются все возможные показатели. Такие выборки, как показано, должны быть большими, и их обычно трудно получить.

7. *Проблемы в установлении адекватного критерия валидности.* Существует, как отмечалось выше в разделе о валидности, значительная трудность в нахождении адекватного критерия. Например, если мы пытаемся измерить авторитаризм, то будем вынуждены полагаться на рейтинги, поскольку нет никаких других внешних мерил (в отличие от, например, школьных экзаменов в случае диагностики способностей). Рейтинги - это неадекватный инструмент, а если бы они подходили, тест был бы не нужен. Аналогично, если существуют высокоэффективные тесты авторитаризма, которые могут быть использованы в качестве критерия валидности, вероятно, в новом тесте нет необходимости.

Таковы основные сложности, возникающие при конструировании личностных опросников, и следует отчетливо их себе представлять, прежде чем приступать к формулированию вопросов или утверждений. Две последние проблемы, хотя они и являются определяющими, выходят на первый план уже после того, как подобраны задания теста.

Рассмотрим далее основные принципы формулирования вопросов (утверждений) для личностных опросников. Как отмечается [8], для тестов личности имеет большое значение то, чтобы ответы на вопросы могли быть объективно и с высокой надежностью оценены. Прежде всего, этого необходимо добиваться путем разработки формы вопросов. Ниже приведены формы вопросов, используемых опросниках, которые оказались наиболее эффективными на практике.

1. *Вопросы с ответом типа "да-нет"*. Такие вопросы легко формулировать, они понятны испытуемым и ответы на них даются быстро. Такова, например, форма вопросов личностных тестов Айзенка. Типичный дихотомический вопрос: "Вам нравятся шумные компании?"

2. *Вопросы с ответами типа "да - не знаю - нет", "да-затрудняюсь ответить - нет"*. Это вариант вопросов типа "да-нет", описанных выше, с добавлением категории неопределенности, потому что некоторые испытуемые становятся раздражительными и несговорчивыми, если их заставляют отвечать либо "да", либо "нет" на вопросы, в ответах на которые они не очень уверены.

Трудность с такими вопросами состоит в том, что эта "средняя" категория очень привлекательна для испытуемых, однако редко бывает информативной. Практика показала, что дихотомические вопросы более предпочтительны, поскольку они заставляют преодолевать нежелание делать выбор.

3. *Альтернативные задания (с ответами типа "правда-ложь")*. Эти задания состоят из утверждений (часто от первого лица), которые испытуемые должны отметить как верные или неверные для них. Примером альтернативного задания является: "Я ненавижу втискиваться в переполненный автобус". Это форма утверждений, используемая в ММРІ. По существу, она ненамного отличается от дихотомических вопросов, хотя

формулирование их как альтернативных утверждений, а не как дихотомических вопросов, может до некоторой степени повлиять на языковую представленность задания.

4. *Задания с ответами типа "нравится - не нравится"* (одно слово или фраза). Испытуемые должны указать, нравится им предложенное или не нравится. Такой тип заданий чаще всего применяется в проективных методиках, однако, он может оказаться чрезвычайно полезным, хотя и необычным, и в обычных тестах-опросниках.

5. *Задания с рейтинговыми шкалами.* Эти задания состоят из предложений, к которым прилагаются рейтинговые шкалы. В них преодолеваются трудности, связанные с коррелирующими дихотомическими заданиями и испытуемым они кажутся более разумными, чем дихотомические задания, хотя и в этом случае, может проявиться установка на крайние ответы. Чаще всего используются две семибалльные шкалы, крайними значениями которых являются: "всегда-никогда" и "определенно да - определенно нет", в зависимости от формулировки задания. Типичным примером может быть: "Мне нравится смотреть мелодраматические фильмы: всегда, очень часто, часто, от случая к случаю, редко, очень редко, никогда". Очевидной проблемой для этой шкалы является различная интерпретация индивидуумами терминов частоты.

6. *Разнообразные трихотомические задания.* В действительности это варианты трихотомической формы вопроса типа "да-нет" (тип 1 в данном списке). Их использует тогда, когда они более подходят по смыслу в отношении некоторых утверждений, чем "да - затрудняюсь ответить - нет". Примерами трихотомий являются: "обычно - иногда - никогда", "верно - не знаю - неверно", "согласен - не уверен - не согласен". Эти варианты предоставляют большую гибкость при формулировании заданий, чем жесткие формы типа "да-нет" (дихотомическая) или альтернативные.

7. *Трихотомические задания с выбором.* Это варианты типа (б), позволяющие вложить в лаконичную форму задания почти любую мысль. В них используются три завершающих предложение фразы, одну из которых испытуемый должен выбрать. Вот типичный пример: "Когда мне нечего делать, я могу: (а) позвонить другу, чтобы поболтать; (б) заняться разгадыванием трудного кроссворда; (в) пойти на джазовый концерт".

8. *Другие формы.* Основные типы заданий, описанные в пунктах (1) - (7), - это те, которые в основном используются в наиболее известных личностных опросниках. Все они, с точки зрения формы (в отличие от содержания), довольно легко формулируются, лаконичны, просты в работе, а также для подсчетов.

Удовлетворение этим требованиям обеспечивает содержательную валидность теста. По другому говоря, содержательная валидность должна рассматриваться лишь как начальный этап в конструировании теста.

Конструирование интегральных показателей (шкал). На этом этапе решаются две основных подзадачи:

- оценка того, насколько удачно исходная система признаков описывает содержание измеряемого свойства;
- на базе отобранных признаков строится обобщающий показатель.

Обобщающий (интегральный) показатель представляет из себя, по сути, «шкалу», «ось» или «фактор», который содержит в себе весь континуум многообразных оценок по всем пунктам теста.

Эти две подзадачи решаются по-разному, в зависимости от того, имеется ли в наличии какой-либо внешний критерий или нет. При наличии такого внешнего критерия, интегральный показатель может быть построен с помощью регрессионного анализа, причем уравнение регрессии связывает внешний критерий с исходными признаками.

Если внешний критерий отсутствует (что бывает наиболее часто), то

используют метод факторного анализа. Именно с его помощью можно все множество первоначальных показателей свести к минимальному количеству основных факторов.

Классификация испытуемых на однородные группы. Эта задача возникает в связи с необходимостью сравнения испытуемых по уровню выраженности психических свойств. Прежде всего встает вопрос о том, какие различия можно считать существенным, а какими можно пренебречь. Здесь же решается и вопрос об уровнях развития исследуемого свойства у испытуемых.

Группировка испытуемых может быть осуществлена с помощью алгоритма автоматической классификации, который предназначен для выделения фиксированного числа классов, задаваемого психологом, исходя из содержательных соображений и вида распределения значений интегральных показателей.

После классификации каждому испытуемому может быть приписано число в соответствии с номером класса, к которому он был отнесен по индивидуально набранному баллу.

От одномерных классификаций довольно просто перейти к многомерной классификации испытуемых. Для этого все выделенные интегральные показатели объединяются на одном графике в «профиле». Подобный способ представления данных позволяет легче воспринять и интерпретировать полученные данные [13].

Контрольные вопросы для самопроверки

§10. Технология создания и адаптации методик

1. Какие основные ситуации встречаются при работе с тестами? Охарактеризуйте кратко суть каждой из них.
2. Назовите основные этапы эмпирико-статистической работы, которые необходимы при адаптации многомерного переводного тест-опросника.
3. Какие главные факторы необходимо учитывать при создании но-

вых психодиагностических методик?

4. Какие основные проблемы возникают при разработке личностных опросников?

5. Дайте краткую характеристику основным видам ответных шкал, которые используются при разработке личностных опросников?

Литература к теме

1. Анастаси А. Психологическое тестирование. В 2-х кн. М., 1982.
2. Бурлачук Л.Ф., Морозов С.М. Словарь-справочник по психодиагностике. - СПб.: Изд-во "Питер", 1999. - 528 с.
3. Гусев А. Н., Измайлов Ч. А., Михалевская М. Б. Измерение в психологии: общий психологический практикум. М.: Смысл, 1997. -287 с.
4. Дружинин В.Н. Экспериментальная психология: Учебное пособие - М.: ИНФРА-М, 1997. - 256 с.
5. Мельников В.М., Ямпольский Л.Т. Введение в экспериментальную психологию личности: Учеб. Пособие для слушателей ИПК преподавателей пед. дисциплин и пед. ин-тов. - М.: Просвещение, 1985. - 319 с.
6. Общая психодиагностика / Под ред. А. А. Бодалева, В. В. Столина.- М., 1987.
7. Практикум по экспериментальной и прикладной психологии / Под ред. А.А. Крылова. - Л.: Изд-во ЛГУ, 1990, 272 с.
8. Практическая диагностика. Методики и тесты. Учебное пособие. - Самара: Изд-во "Бахрах", 1998 - 672 с.
9. Психологический журнал. Т. 14, № 5, 1993. С. 90-98 ("Чужие рисунки").
10. Шевандрин Н.И. Психодиагностика, коррекция и развитие личности. - М.: "ВЛАДОС", 1998. - 512 с.

Список литературы, использованной при подготовке книги по психодиагностике

1. Анастаси А. Психологическое тестирование. В 2-х кн. М., 1982.
2. Анастаси А. Дифференциальная психология. Индивидуальные и групповые различия в поведении / Пер. с англ. – М.: Апрель Пресс, Изд-во ЭКСМО-Пресс, 2001. – 752 с.
3. Бурлачук Л.Ф., Морозов С.М. Словарь-справочник по психодиагностике. - СПб.: Изд-во "Питер", 1999. - 528 с.
4. Гласс Дж., Стенли Дж. Статистические методы в педагогике и психологии. - М.: Прогресс, 1979. - 500 с.
5. Гусев А. Н., Измайлов Ч. А., Михалевская М. Б. Измерение в психологии: общий психологический практикум. М.: Смысл, 1997. -287 с.
6. Диагностика психического развития. Под ред. Й. Шварцара. Прага, Авиценум, 1978. - 387 с.
7. Дружинин В.Н. Экспериментальная психология: Учебное пособие - М.: ИНФРА-М, 1997. - 256 с.
8. Клайн П. Справочное руководство по конструированию тестов. - Киев: ПАН Лтд, 1994.
9. Кулагин Б. В. Основы профессиональной психодиагностики.- Л., 1984.
10. Купер К. Индивидуальные различия / Пер. с англ. Т.М. Марютиной под ред. И.В. Равич-Щербо – М.: Апрель Пресс, 2000. – с.527
11. Лакин Г.Ф. Биометрия. М.: Высшая школа, 1968. - 284 с.
12. Либин А.В. Дифференциальная психология: На пересечении европейских, российских и американских традиций. 2 изд., переработанное. – М.: Смысл; Per Se, 2000. –549 с.
13. Мельников В.М., Ямпольский Л.Т. Введение в экспериментальную психологию личности: Учеб. Пособие для слушателей ИПК преподавателей пед. дисциплин и пед. ин-тов. - М.: Просвещение, 1985. - 319 с.
14. Нормы профессиональной этики для разработчиков и пользователей психодиагностических методик. Стандартные требования к психологическим тестам. 2-е издание. Ярославль; НПЦ "Психодиагностика", 1998.
15. Общая психодиагностика / Под ред. А. А. Бодалева, В. В. Столина.- М., 1987.
16. Основы психодиагностики / Под ред. А. Г. Шмелева. Ростов-на-Дону., 1996.
17. Психологическая диагностика: Проблемы и исследования / Под ред. К. М. Гуревича М., 1981.
18. Психологическая диагностика: Учебное пособие / Под ред. К.М. Гуревича и Е.М. Борисовой. - М.: Изд-во УРАО, 1997. - 304 с.

19. Психологический словарь / Под ред. В. П. Зинченко, Б. Г. Мещерякова. - М., 1996.
20. Хотинец В.Ю. Четыре метода анализа документов в социально-психологических исследованиях: Учеб. пособие / УдГУ. Ижевск, 2001. 62 с.
21. Шмелев А. Г. Психодиагностика личностных черт. – СПб.: Речь, 2002. – 480 с.
22. Ян тер Лаак. Психодиагностика: проблемы содержания и методов. - М.: Изд-во "Институт практической психологии", Воронеж, 1996. - 384с.
23. Howarth E. The index for differentiation of state and trait scales. – Psychol. Rep., 1978, Vol. 43, No. 2, p.474.
24. Oaster T.R.F. μ index differentiation of state and trait scales with new stressor scales. – Psychol. Rep., 1982, Vol. 51, No. 1, p.272.

Дополнительная литература по курсу психодиагностики.

1. Аванесов В.С. Тесты в социологическом исследовании. - М.: Наука, 1982.
2. Атлас для экспериментального исследования отклонений в психической деятельности человека/Под ред. И. А. Полищука, А. Е. Видренко. Изд. 2. Киев, 1979. 124 с.
3. Бажин Е.Ф. и др. Метод исследования субъективного контроля / Психологический журнал, - 1984, Т. 5, № 3. С. 152-162.
4. Березин Ф.Б. и др. Методики многостороннего исследования личности. - М.: Медицина, 1976.
5. Блейхер В.М., Бурлачук Л.Ф. Психологическая диагностика интеллекта личности. - Киев: Выща школа, 1978.
6. Бызова В.М. Основы психодиагностики, учебное пособие, Сыктывкар, Гос. университет, 1992, 59 с.
7. Гайда В.К., Захаров В.П. Психологическое тестирование: учебное пособие. - Л.: Изд-во ЛГУ, 1982. С. 13-18.
8. Ганзен В.А., Балин В.Д. Теория и методология психологического исследования. - СПб РИО, СПб ГУ, 1991, 75 с.
9. Гарбер Е.И., Казча В.В. Методика профессиографии - Саратов: Изд-во Саратовского ун-та, 1992, 196 с.
10. Гильбух Ю.З. Актуальные проблемы валидации психологических тестов. / Вопросы психологии, 1978, N 5.
11. Гильбух Ю. З. Психодиагностика в школе. М., 1989.
12. Гуревич К.М., Акимова А.К., Козлова В.Т. Статистическая норма или социально-психологический норматив? / Психологический журнал, 1986, N 3.
13. Забрамная С. Д. Наглядный материал для психолого-

педагогического обследования детей в медико-педагогических комиссиях. М., 1985.

14. Изучение школьника и составление его психологической характеристики: Методические указания к практике по психологии. Л.: ЛГПИ, 1986. 27 с.

15. Корнилова Т. В. Введение в психологический эксперимент: Учебник. М.: Изд-во Моск. ун-та, Че-Ро, 1997. 256 с.

16. Кроз М.В. Аннотированный указатель методов социально-психологической диагностики: учебное пособие. - М.: Изд-во МГУ, 1991, 55 с.

17. Кэмпбелл Д. Модели экспериментов в социальной психологии и прикладных исследованиях. М.: Прогресс, 1980. 392 с.

18. Левитан К.М. Знаете ли Вы себя? Сборник психологических тестов и рекомендаций. - Екатеринбург: Профтех, 1991.

19. Лучшие психологические тесты для профотбора и профориентации: описание и руководство к использованию / Отв. ред. А.Ф. Кудряшев. - Петрозаводск: Петроком, 1992, 318 с.

20. Методики изучения интеллекта. 4.1 / И.Н. Агафонова и др. - СПб: Гор. институт усовершенствования учителей, Кафедра педагогики и психологии, 1991, 221 с.

21. Методы социальной психологии / Под ред. Е. С. Кузьмина, В. Е. Семенова. Л.: ЛГУ, 1977. 175 с.

22. Платонов К. К. Психологический практикум. М.: Высшая школа, 1980. 165с.

23. Практикум по психодиагностике. Прикладная психодиагностика / И.Ю. Беляева и др. - М., Изд-во МГУ, 1992, 116 с.

24. Практикум по психодиагностике: дифференциальная психометрика / Под ред. В.В. Столина, А.Г. Шмелева. - М. : Изд-во МГУ, 1984. С. 16-17.

25. Практикум по общей, экспериментальной и прикладной психологии: Учебн. пособие / Ред. А. А. Крылов, С. А. Маничев. СПб.: Питер, 2000.

26. Психологические методы исследования в клинике. Л., 1967. 324 с.

27. Психодиагностические методы (в комплексном лонгитюдном исследовании студентов) / Под ред. А. А. Бодалева, М. Д. Дворяшиной, И. М. Палея. Л.: ЛГУ, 1976.248 с.

28. Психологический словарь / Под ред. В. П. Зинченко, Б. Г. Мещерякова. - М., 1996.

29. Проблема цвета в психологии / Отв. ред. А.А. Митькин, Н.Н. Корж. - М.: Наука, 1993, 204 с.

30. Психологический статус личности в различных социальных условиях: диагностика и коррекция / Межвузовский сборник / Под ред. В.С.

Мухиной. - М.: Прометей, 1992, 198 с.

31. Развитие и диагностика способностей / Л.Г. Алексеева и др. / Отв. ред. В.Н. Дружинин, В.Д. Шадриков. - М.: Наука, 1992, 179 с.

32. Реан А. А. Психология изучения личности: Учебн. пособие. СПб.: Изд-во Михайлова В. А., 1999. 288 с.

33. Рейнвальд Н.И. Личность и характер: учебное пособие. - М.: Изд-во РУДН, 1992, 88 с.

34. Романова Е. С. Графические методы в психодиагностике. СПб.: Речь, 2001.

35. Рыбалко Е. Ф. Возрастная и дифференциальная психология: Учебн. пособие. Л.: ЛГУ, 1990. 256 с.

36. Рубинштейн С.Я. Экспериментальные методики патопсихологии. - СПб.: Ленато, 1998. - 166 с.

37. Русалов В.М. Новый вариант адаптации личностного теста ЕРУ / Психологический журнал, 1987. № 1.

38. Русалов В.М., Гусева О.В. Сокращенный вариант личностного опросника Кеттелла (8PI) / Психологический журнал, 1990. Т. 11, №11. С. 34-48.

39. Семенов В. Е. Метод изучения документов в социально-психологических исследованиях: Учебн. пособие. Л.: ЛГУ, 1983. 104 с.

40. Слостенин В. А., Мажор Н. Е. Диагностика профпригодности. М., 1992.

41. Соколова Е.Т. Проективные методы исследования личности. - М.: Изд-во МГУ, 1987.

42. Способности и склонности: комплексные исследования / Под ред. Э.А. Голубевой. - М.: Педагогика, 1989, 199 с.

43. Суслов В.И., Чумакова Н.П. Психодиагностика. Учебное пособие СПб РИО, СПб ГУ, 1992, 59 с.

44. Черепанова Н.С., Дорожкин Ю.П. Графологическая диагностика. Методическое пособие. - М.: НМПИО "МАГ", 1992, 32 с.

45. Экспериментальная психология / Под ред. П. Фреса, Ж. Пиаже. - М.: Прогресс, 1975. Гл. 17.

Приложение 1.

Подсчет параметров и оценка типа распределения.

Для описания выборочного распределения, как правило, используются следующие известные параметры [1, 11, 15]:

1. Среднее арифметическое:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i ; \quad \text{где:}$$

X_i ; - балл i -того испытуемого;

n - количество испытуемых в выборке (объем);

Выборочная, или средняя, квадратическая ошибка средней арифметической:

$$m_x = \frac{\sigma}{\sqrt{n}} .$$

2. Среднее квадратическое (стандартное) отклонение:

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}} ;$$

Σ - сумма квадратов тестовых баллов для n испытуемых.

Средняя **ошибка** среднего квадратического (стандартного) отклонения:

$$m_\sigma = \frac{\sigma}{\sqrt{2n}} .$$

Стандартная ошибка среднего арифметического (математического ожидания) оценивается по формуле: $m_x = \frac{\sigma}{\sqrt{n}}$ на основе ошибки математического ожидания строятся доверительные границы, в которых заключена величина генеральной средней: $tm_x - \bar{X} + tm_x$.

Если тестовый балл какого-либо испытуемого попадает в границы доверительного интервала, то нельзя считать, что испытуемый обладает

повышенным (или пониженным) значением измеряемого свойства с заданным уровнем статистической вероятности.

Оценка достоверности вычисленных статистик производится по следующим формулам:

Критерий достоверности различий, наблюдаемых между средними арифметическими двух независимых выборок:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{m_1^2 + m_2^2}} .$$

3. Асимметрия:

$$A_s = \frac{1}{\sigma^3} (\theta - 3C^2 \bar{X} + 2\bar{X}^3);$$

где: \bar{X} - среднее арифметическое;

σ - стандартное отклонение;

$$\theta - \text{среднее кубическое: } \theta = \sqrt[3]{\frac{\sum P \bar{X}^3}{n}} ;$$

$$C - \text{среднее квадратическое: } C = \sqrt{\frac{\sum P \bar{X}^2}{n}} ;$$

P - частота появления варианты.

Коэффициент асимметрии величина не именованная. Он колеблется в пределах от нуля до единицы. При совершенно симметричных распределениях коэффициент асимметрии равен нулю. Асимметрия считается незначительной, если $A_s \leq 0,25$. При $A_s \geq 0,25$ скошенность распределения оказывается уже значительной.

$$4. \text{Эксцесс: } E_x = \frac{1}{\sigma^4} (Q^4 - 4\theta^3 \bar{X} + 6C^2 \bar{X}^2 - 3\bar{X}^4) - 3,$$

$$Q - \text{среднее значение четвертой степени: } Q = \sqrt[4]{\frac{\sum \bar{X}^4}{n}}$$

При положительном эксцессе показатель E_x - число положительное,

а при отрицательном - E_x - число отрицательное. Если $E_x \leq 0,2$, эксцесс практически отсутствует; если же $E_x \geq 0,5$, но ≤ 1 , эксцесс считается заметным, но небольшим. Предельное значение отрицательного E_x равно - 2, что указывает на наличие распределения с двумя вершинами.

Асимметрия и эксцесс нормального распределения должны быть равны нулю. Если они существенно отклоняются от нуля (хотя бы один из двух параметров), то это означает, что полученное эмпирическое распределение аномально.

Проверку значимости асимметрии можно произвести на основе общего неравенства Чебышева:

$$|A| \leq \sqrt{\frac{\sigma_a}{1-P}}; \quad ("a")$$

где:

σ_a - дисперсия эмпирической оценки асимметрии:

$$\sigma_a = \frac{6(n-1)}{(n+1)(n+3)}, \text{ где}$$

p - уровень значимости или вероятность ошибки первого рода; ошибки в том, что будет принят вывод о незначимости асимметрии, при наличии значимой асимметрии (в формулу подставляют стандартные $p=0,05$ или $p=0,01$ и проверяют выполнение неравенства).

Сходным образом оценивается значимость эксцесса:

$$|E| \leq \sqrt{\frac{\sigma_e}{1-P}}, \quad ("б")$$

где:

σ_e - эмпирическая дисперсия оценки эксцесса, определяется по формуле:

$$\sigma_e = \frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)}$$

Гипотезы об отсутствии асимметрии и эксцесса принимаются с вероятностью ошибки p (пренебрежимо малой), если выполняются неравенства ("а") и ("б").

Если проверка согласованности эмпирического распределения с нормальным дает положительные результаты, то это означает, что полученное распределение можно рассматривать как устойчивое - репрезентативное по отношению к генеральной совокупности - а, значит, на его основе можно определить репрезентативные тестовые нормы. Если проверка не выявляет нормальности на заданном уровне, то это означает, что или выборка мала и нерепрезентативна к популяции, или измеряемое свойство и устройство теста (способ подсчета) вообще не дают нормального распределения [15].

В принципе, требование нормальности распределения не является обязательным. Можно с равным успехом пользоваться другими хорошо разработанными моделями гамма-распределения, пуассоновского распределения и т.п. Критерий Колмогорова позволяет оценить близость эмпирического распределения к любому теоретическому распределению. При этом устойчивым и репрезентативным может оказаться распределение любого типа. Если из нормальности, как правило, следует устойчивость, то обратное неверно - устойчивость вовсе не обязательно предполагает нормальность распределения.

Наличие значимой положительной асимметрии свидетельствует о том, что в системе факторов, детерминирующих значение измеряемого показателя, преобладают факторы, действующие в одном направлении - в сторону повышения показателя.

На практике распределения такого рода преобразуют в нормальное

(приближенно нормальное) с помощью логарифмической трансформации: $Z = \ln y$. При этом говорят, что распределение показателей подчиняется "логнормальному" закону.

Подобную алгебраическую нормализацию тестовой шкалы применяют к показателям с еще более резко выраженной положительной асимметрией.

Например, в процедурах контент-анализа сам тестовый показатель является частотным - он измеряет частоту появления определенных категорий событий в текстах.

Приложение 2.

Статистические таблицы значимости*

Таблица 1. Значимость Т-распределения Стьюдента.

Число степеней свободы	Уровень значимости					
	0,10	0,05	0,02	0,01	0,002	0,001
1	6,314	12,71	31,82	63,66	318,3	636,6
2	2,920	4,303	6,965	9,925	22,33	31,60
3	2,353	3,182	4,541	5,841	10,21	12,92
4	2,132	2,776	3,747	4,604	7,173	8,610
5	2,015	2,571	3,365	4,032	5,893	6,869
6	1,943	2,447	3,143	3,707	5,208	5,959
7	1,895	2,365	2,998	3,499	4,785	5,408
8	1,860	2,306	2,896	3,355	4,501	5,041
9	1,833	2,262	2,821	3,250	4,297	4,781
10	1,812	2,228	2,764	3,169	4,144	4,587
11	1,796	2,201	2,718	3,106	4,025	4,437
12	1,782	2,179	2,681	3,055	3,930	4,318
13	1,771	2,160	2,650	3,012	3,852	4,221
14	1,761	2,145	2,624	2,977	3,787	4,140
15	1,753	2,131	2,602	2,947	3,733	4,073
16	1,746	2,120	2,583	2,921	3,686	4,015
17	1,740	2,110	2,567	2,898	3,646	3,965
18	1,734	2,101	2,552	2,878	3,610	3,922
19	1,729	2,093	2,539	2,861	3,579	3,883
20	1,725	2,086	2,528	2,845	3,552	3,850
21	1,721	2,080	2,518	2,831	3,527	3,819
22	1,717	2,074	2,508	2,819	3,505	3,792
23	1,714	2,069	2,500	2,807	3,485	3,767
24	1,711	2,064	2,492	2,797	3,467	3,745
25	1,708	2,060	2,485	2,787	3,450	3,725
26	1,706	2,056	2,479	2,779	3,435	3,707
27	1,703	2,052	2,473	2,771	3,421	3,690
28	1,701	2,048	2,467	2,763	3,408	3,674
29	1,699	2,045	2,462	2,756	3,396	3,659
30	1,697	2,042	2,457	2,750	3,385	3,646
∞	1,645	1,960	2,326	2,576	3,090	3,291

Таблица дает значения введенной Стьюдентом величины t для уровней значимости, наиболее часто применяющихся при нахождении критериев значимости и границ доверительного интервала.

Приложение 2.

Таблица 2. Критические значения критерия χ^2 для трех степеней доверительной вероятности

n \ p	0,95	0,99	0,999	n \ P	0,95	0,99	0,999
1	3,8	6,6	10,8	26	38,9	45,6	54,1
2	6,0	9,2	13,8	27	40,1	47,0	55,5
3	7,8	11,3	16,3	28	41,3	48,3	56,9
4	9,5	13,3	18,5	29	42,6	49,6	58,3
5	11,1	15,1	20,5	30	43,8	50,9	59,7
6	12,6	16,8	22,5	32	46,2	53,5	62,4
7	14,1	18,5	24,3	34	48,6	56,0	65,2
8	15,5	20,1	26,1	36	51,0	58,6	67,9
9	16,9	21,7	27,9	38	53,4	61,1	70,7
10	18,3	23,2	29,6	40	55,8	63,7	73,4
11	19,7	24,7	31,3	42	58,1	66,2	76,1
12	21,0	26,2	32,9	44	60,5	68,7	78,7
13	22,4	27,7	34,5	46	62,8	71,2	81,4
14	23,7	29,1	36,1	48	65,2	73,7	84,0
15	25,0	30,6	37,7	50	67,5	76,2	86,7
16	26,3	32,0	39,3	55	73,3	82,3	93,2
17	27,6	33,4	40,8	60	79,1	88,4	99,6
18	28,9	34,8	42,3	65	89,8	94,4	106,0
19	30,1	36,2	43,8	70	90,5	100,4	112,3
20	31,4	37,6	45,3	75	96,2	106,4	118,5
21	32,7	38,9	46,8	80	101,9	112,3	124,8
22	33,9	40,3	48,3	85	107,5	118,2	131,0
23	35,2	41,6	49,7	90	113,1	124,1	137,1
24	36,4	43,0	51,2	95	118,7	130,0	143,3
25	37,7	44,3	52,6	100	124,3	135,8	149,4

Приложение 2.

Таблица 3. Значимость коэффициента корреляции (по Пирсону)

Число степеней свободы	Уровень значимости				
	0,10	0,05	0,02	0,01	0,001
1	0,9877	0,99692	0,99951	0,99988	0,9999988
2	0,9000	0,9500	0,9800	0,9900	0,9990
3	0,805	0,878	0,9343	0,9587	0,9911
4	0,729	0,811	0,882	0,9172	0,9741
5	0,669	0,754	0,833	0,875	0,9509
6	0,621	0,707	0,789	0,834	0,9249
7	0,582	0,666	0,750	0,798	0,898
8	0,549	0,632	0,715	0,765	0,872
9	0,521	0,602	0,685	0,735	0,847
10	0,497	0,576	0,658	0,708	0,823
11	0,476	0,553	0,634	0,684	0,801
12	0,457	0,532	0,612	0,661	0,780
13	0,441	0,514	0,592	0,641	0,760
14	0,426	0,497	0,574	0,623	0,742
15	0,412	0,482	0,558	0,606	0,725
16	0,400	0,468	0,543	0,590	0,708
17	0,389	0,456	0,529	0,575	0,693
18	0,378	0,444	0,516	0,561	0,679
19	0,369	0,433	0,503	0,549	0,665
20	0,360	0,423	0,492	0,537	0,652
25	0,323	0,381	0,445	0,487	0,597
30	0,296	0,349	0,409	0,449	0,554
35	0,275	0,325	0,381	0,418	0,519
40	0,257	0,304	0,358	0,393	0,490
45	0,243	0,288	0,338	0,372	0,465
50	0,231	0,273	0,322	0,354	0,443
60	0,211	0,250	0,295	0,325	0,408
70	0,195	0,232	0,274	0,302	0,380
80	0,183	0,217	0,257	0,283	0,357
90	0,173	0,205	0,242	0,267	0,338
100	0,164	0,195	0,230	0,254	0,321

Приложение 2.

Таблица 4. Значимость коэффициента корреляции рангов (по Спирмену).

n	0,05	0,01	n	0,05	0,01	n	0,05	0,01
5	0,94	-	17	0,48	0,62	29	0,37	0,48
6	0,85	-	18	0,47	0,60	30	0,36	0,47
7	0,78	0,94	19	0,46	0,58	31	0,36	0,46
8	0,72	0,88	20	0,45	0,57	32	0,36	0,45
9	0,68	0,83	21	0,44	0,56	33	0,34	0,45
10	0,64	0,79	22	0,42	0,54	34	0,34	0,44
11	0,61	0,76	23	0,42	0,53	35	0,33	0,43
12	0,58	0,73	24	0,41	0,52	36	0,33	0,43
13	0,56	0,70	25	0,40	0,51	37	0,33	0,42
14	0,54	0,68	26	0,39	0,50	38	0,32	0,41
15	0,52	0,66	27	0,38	0,49	39	0,32	0,41
16	0,50	0,64	28	0,38	0,48	40	0,31	0,40

Примечание*. Таблицы приложения 2 взяты из книг [11, 18]:

- Лакин Г.Ф. Биометрия. М.: Высшая школа, 1968. - 284 с.

- Психологическая диагностика: Учебное пособие / Под ред. К.М. Гуревича и Е.М. Борисовой. - М.: Изд-во УРАО, 1997. - 304 с.

УЧЕБНОЕ ПОСОБИЕ

Г.С. ПРЫГИН
Основы психодиагностики

РЕДАКТОР М.И. ЧЕРКАССКАЯ
КОМПЬЮТЕРНАЯ ВЕРСТКА И ДИЗАЙН ОБЛОЖКИ А.И. ЧЕКАЛИНОЙ

УМК «Психология»

ЛИЦЕНЗИЯ № 00451 от 15 ноября 1999 г. ГИГИЕНИЧЕСКИЙ СЕРТИФИКАТ № 77.99.1 1 .953.Д.004699. 07.03
от 03.07.2003 г.

АДРЕС: 121069, г. МОСКВА, ТРУБНИКОВСКИЙ ПЕР., д. 22, СТР. 2.
ТЕЛ. (095)746-02-39, ТЕЛ./ФАКС (095)952-45-90
E-MAIL: COLLECT@MAIL.RU

ПОДПИСАНО В ПЕЧАТЬ 01.12.2003. ФОРМАТ 84x108/32. ТИРАЖ 2000. БУМАГА ОФСЕТНАЯ. ГАРНИТУРА "TimesET.

Усл. печ. л. 6. ЗАКАЗ № 3733

ОТПЕЧАТАНО В ПОЛНОМ СООТВЕТСТВИИ С КАЧЕСТВОМ ПРЕДОСТАВЛЕННЫХ ДИАПОЗИТИВОВ
В ОАО «ИЗДАТЕЛЬСКО-ПОЛИГРАФИЧЕСКОЕ ПРЕДПРИЯТИЕ «ПРАВДА СЕВЕРА».
163002, АРХАНГЕЛЬСК, ПР. НОВГОРОДСКИЙ, 32