

ТЕХНОЛОГИЯ РАЗРАБОТКИ ТЕСТОВ: ЧАСТЬ I

Н.А. Батурин, Н.Н. Мельникова

Впервые в отечественной психодиагностике описан универсальный процесс создания психодиагностической методики любого типа. Процесс состоит из 14 шагов, сгруппированных в 8 этапов. Реализация задач каждого шага является необходимым условием разработки методики. Пропуск шагов или несвоевременное выполнение соответствующих им задач отразится на качестве методики и не сократит, а увеличит затраты времени и средств. Изложение материала разделено на части. В первой части статьи описаны первые два начальных этапа: организационный и содержательный. Эти этапы – фундамент создания любой методики. При их описании выделены наиболее существенные и мало освещенные проблемы разработки психодиагностических методик.

Ключевые слова: психодиагностические методики, процесс разработки методик, пошаговая технология, планирование теста, содержание теста.

Введение

По мнению большинства авторитетных специалистов, в настоящее время отечественная психодиагностика переживает системный кризис (см. Н.А. Батурин, 2008). Одна из причин кризиса – малое количество отечественных тестов, удовлетворяющих современным международным требованиям (стандарты тестов APA, BPS, EFPA). Главная причина этого, на наш взгляд, заключается в слабой подготовке отечественных разработчиков в недостаточной освоенности технологий создания тестов. Достаточно сказать, что за всю историю психодиагностики в России было издано только одно детальное «Руководство по конструированию тестов» (П. Клайн, 1994), которое в настоящее время несколько устарело.

Разработка тестов в России – это удел одиночек (реже небольших групп), которые, как правило, создают методику «для себя», чтобы провести задуманное исследование (чаще всего в рамках работы над диссертацией). На серьезную работу по созданию диагностического инструментария для использования его другими, не хватает ни времени, ни сил, ни средств.

За рубежом уже давно тесты разрабатываются в исследовательских центрах, корпорациях издателей, большой группой разработчиков с разделением функций, фактически «на потоке» или конвейере. Поэтому многие

из тестов не имеют авторов, а только указание на фирму-разработчика (например, у знаменитого OPQ 32 автор – SHL group). Естественно, что за рубежом уже разработана подробная технология создания тестов. В огромных томах каждой детали разработки посвящены главы и даже книги (см., например, Handbook, 2006). В отечественной психодиагностике принцип поэтапной разработки тестов предложен достаточно давно, и, в частности, был реализован в одной из работ автора 8 лет назад (Батурин, Курганский, 2001).

Создание в Лаборатории психодиагностики ЮУрГУ первого в России Компендиума психодиагностических методик показало, что за прошедшие 100 лет в СССР и России было разработано всего около 700 методик, как сугубо исследовательских, так и психодиагностических (Н.А. Батурин, А.В. Пичугова, 2008). Это ничтожно мало по сравнению с существующими потребностями и масштабными разработками за рубежом (где их насчитывается более 20 000). Но больше всего настораживает, что только у 25 % отечественных методик есть хотя бы упоминание о проверке валидности, надежности и стандартизации. И совсем редко методики проходят полный процесс психометрической проверки. Необходимым компонентом западной методики являются «Технические отчеты», которые не попадают в поле зрения пользователя

(только по требованию), но они позволяют провести рецензирование и сертификацию в соответствующих органах, поскольку в них содержится вся информация о процессе создания методики, о проведенных психометрических экспериментах, о психометрических проверках готового к изданию теста.

Одна из задач данной статьи – сформировать новое отношение к самой деятельности по разработке тестов, способствовать изменению самого подхода к этому виду профессиональной деятельности. На наш взгляд, в России назрела необходимость перехода от инициативных разработок тестов к широкомасштабным коллективным проектам, занимающим свое место в одном ряду с разработками социально значимых практических технологий в других отраслях современной науки и практики. А для этого необходимо освоение современных технологий создания тестов.

Наиболее полно процесс разработки тестов представлен в работе М. Даунинга (Handbook, 2006). Им предлагается пошаговая технология, предназначенная для создания тестов достижений. Однако ключевые моменты, обозначенные М. Даунингом, являются важными и для создания других видов тестов. Это сделало возможным переработку технологии, предложенной М. Даунингом, с тем чтобы создать универсальную схему последовательности действий по разработке психодиагностического инструментария. Эта усовершенствованная схема предлагается в данной работе.

Общая схема процесса разработки тестов

Предлагаемая схема процесса разработки тестов представлена в таблице. Весь процесс создания психодиагностической методики состоит из 14 шагов, сгруппированных в 8 этапов. Каждый шаг имеет отдельную цель, предполагающую получение вполне конкретного конечного продукта, на который опираются дальнейшие действия. Цель каждого шага отражена непосредственно в его названии. Подразделение на этапы выделяет качественно различные области деятельности в процессе разработки теста. Действия разработчиков на отдельных этапах требуют разных компетенций, разных акцентов в подготовке, и, в отдельных случаях, могут осуществляться даже разными командами специалистов. Каждый шаг предполагает специфические задачи, которые будут описаны ниже. Более подробно планируется остановиться на ключевых зада-

чах каждого шага, а также на вопросах, требующих особого внимания.

Особо акцентировать внимание следует на том, что последовательность шагов, представленных в таблице, является не просто рекомендуемым алгоритмом, удобным для пользования, а отражает сложный процесс. И как для любого процесса здесь существует закономерность в последовательности действий, вытекающих одно из другого. Все задачи, решаемые на разных этапах процесса разработки тестов, взаимосвязаны. Результат каждого этапа становится основой и условием успешности выполнения задач на последующих. Если какой-то этап пропущен, или соответствующие ему задачи решены неадекватно, то это снижает эффективность действий на последующих этапах и, в конечном итоге, влияет на качество конечного продукта. Хотя особенности работы, длительность и сложность каждого этапа могут быть различными и зависят от специфики конкретного теста. Важно, чтобы все этапы процесса были основательно проработаны, и соблюдалась пошаговая последовательность действий.

Всем хорошо известно, что современная практика разработки психодиагностических методик в России в этом плане далека от совершенства. Спонтанность, отрывочность и разрозненность действий, связанных с разработкой тестов – скорее норма, чем исключение; и это серьезное препятствие к повышению качества отечественного инструментария. Характерно, что обозначенные шаги неодинаково представлены в современной практике разработки тестов в России. Среди них есть такие, которые считаются традиционно необходимыми и такие, которые часто опускаются или прорабатываются очень поверхностно.

Так для большинства авторов наиболее привычными являются шаги, связанные с проверкой валидности, надёжности теста и его стандартизацией (шаги 8-й и 9-й в таблице). Иногда все психометрические процедуры сводят только к ним. Тест считается «готовым», если он стандартизирован и к нему прилагаются сведения о коэффициентах валидности и надёжности. Однако, как видно из дальнейшего изложения, эти шаги являются следствием серьезной предварительной работы, и усилия здесь могут быть потрачены впустую, если отсутствует тщательная проработка содержательной области или недостаточно основательно выполнены действия,

Общая схема процесса разработки тестов

Этап	Основные задачи
Этап I. Организационный	
Шаг 1. Планирование проекта	<ul style="list-style-type: none"> – Формулирование предварительных характеристик теста – Определение основных источников валидности – Выбор психометрических и исследовательских моделей – Составление подробного плана разработки теста – Организационные решения
Этап II. Содержательный	
Шаг 2. Определение содержания	<ul style="list-style-type: none"> – Определение исходного теоретического концепта – Формирование репрезентативной выборки содержания – Описание ключевой феноменологии
Шаг 3. Составление спецификации теста	<ul style="list-style-type: none"> – Операционализация содержания – Выбор и фиксация формальных характеристик теста – Утверждение спецификации
– Этап III. Подготовительный	
Шаг 4. Разработка пунктов	<ul style="list-style-type: none"> – Разработка пунктов – Профессиональная редакция пунктов – Создание банка пунктов
Шаг 5. Сборка теста	<ul style="list-style-type: none"> – Компоновка рабочих версий теста – Утверждение пробной версии теста
Этап IV. Исследовательский	
Шаг 6. Апробация, анализ и коррекция пунктов	<ul style="list-style-type: none"> – Эмпирическая апробация пунктов – Отбор и переформулирование пунктов – Проверка внутренней согласованности и дискриминативности шкал – Повторный пилотаж (при необходимости) – Утверждение состава и структуры теста
Шаг 7. Уточнение процедуры тестирования	<ul style="list-style-type: none"> – Уточнение процедуры и времени – Разработка инструкций – Утверждение рабочей версии теста
Шаг 8. Изучение и проверка валидности и надёжности	<ul style="list-style-type: none"> – Исследование ретестовой надёжности – Изучение конструктивной валидности – Проверка критериальной валидности – Утверждение окончательной версии теста – Корректорская проверка окончательной версии теста
Этап V. Стандартизационный	
Шаг 9. Массовые обследования для стандартизации	<ul style="list-style-type: none"> – Формирование выборки стандартизации – Массовые обследования по плану – Стандартизация теста для различных групп – Фиксация способов перехода к нормам
Этап VI. Интерпретационный	
Шаг 10. Разработка схем интерпретации и диагностических отчётов	<ul style="list-style-type: none"> – Описание алгоритмов обработки данных и схем анализа результатов – Описание принципов интерпретации результатов – Разработка критериев для диагностических выводов – Описание моделей составления отчётов – Разработка компьютерных программ обработки данных и составления отчётов
Этап VII. Технический	
Шаг 11. Написание технических отчётов	<ul style="list-style-type: none"> – Подготовка подробной документации о результатах, полученных на всех этапах разработки теста – Составление технических отчётов о разработке и психометрической проверке теста
Шаг 12. Юридическое оформление	<ul style="list-style-type: none"> – Юридические процедуры – Рецензирование методики – Сертификация методики
Шаг 13. Издание	<ul style="list-style-type: none"> – Подготовка руководства для пользователя – Издание методики
Этап VIII. Эксплуатационный	
Шаг 14. Сопровождение использования теста	<ul style="list-style-type: none"> – Проверка основных психометрических характеристик методики другими исследователями – Дальнейшее изучение конструктивной валидности – Создание новых форм и модификаций теста – Совершенствование пользовательских характеристик теста

связанные с формулированием пунктов и сборкой теста. Далеко не все разработчики понимают, что стандартизация, даже на очень представительных выборках, сама по себе ничего не даёт: стандартизировать можно любой случайный набор данных. Ключевым моментом разработки теста является валидность и надёжность – те характеристики теста, которые должны прорабатываться ещё задолго до их проверки. Результат проверки в виде некоторого статистического показателя – это лишь закономерный итог всех предыдущих действий, их успешности или неадекватности.

В этой статье предлагается сконцентрировать внимание на целостном системном видении всего процесса по разработке тестов. Его преимущество в том, что становятся различимыми и даже очевидными линии развития закладываемых психометрических характеристик, их взаимосвязи друг с другом и с конкретными действиями, производимыми на различных этапах создания теста.

Процессуальное видение даёт возможность некоторого прогнозирования результата, понимания того, как конкретное действие отразится на том или ином показателе через два-три шага. Ориентация на процесс заставляет также разработчика постоянно оглядываться назад, на каждом шагу сверяя свои действия с предыдущими промежуточными результатами, их успешностью.

В последовательном продвижении по этапам разработки теста в статье будет показано значение каждого шага для итогового результата, для обеспечения базовых психометрических характеристик теста. Поскольку формат и объём статьи не позволяет детально описывать каждую из конкретных задач, решаемых на последовательных шагах разработки, более подробно будут обсуждаться те из них, которые представляют наибольшую сложность и требуют особого внимания разработчика.

Этап I. Организационный

Шаг 1. Планирование проекта

Первый этап – организационный является основой любого серьёзного проекта. На этом этапе принимаются ключевые решения, влияющие на выбор методов и моделей, используемых на разных этапах разработки. Шаг 1 не случайно назван «Планирование проекта». Проект – это несколько иной подход к самой деятельности по разработке тестов. Традиционно в нашей стране создание

методик воспринимается больше как индивидуальное творчество (если не сказать, «сочинительство»), а не как сложный практический проект, требующий стратегического планирования, организационных решений и даже подготовки бизнес-плана. В большинстве случаев этап планирования опускается или же оформляется постфактум, как формальный документ для отчёта. Это говорит о том, что культура разработки тестов в нашей стране пока находится на сравнительно низком уровне и требует изменения самого подхода к этому виду профессиональной деятельности.

На организационном этапе закладываются основы всей дальнейшей работы, определяются базовые позиции, влияющие на выбор психометрических и исследовательских моделей, которые будут использованы в проекте. Многие важные решения должны быть приняты ещё до начала непосредственных практических действий. На этом этапе формулируются предварительные характеристики будущей методики и составляется подробный план, включающий описание исследовательских программ и методов, а также принимаются организационные решения, призванные обеспечить практическое продвижение проекта.

По сути, работа на всех этапах проекта связана с последовательным «оттачиванием» основных психометрических характеристик теста, главной из которых, подчеркнем еще раз, является валидность. Чтобы внести некоторую ясность, необходимо ненадолго остановиться на современном понимании валидности. Мнение о том, что валидность достаточно «проверить» на конечном этапе разработки – лишь распространённая иллюзия. Серьёзная работа по достижению валидности начинается ещё задолго до её проверки. Здесь правильнее будет обозначить три сквозных задачи: обеспечение, изучение и, только потом, проверка валидности. И они проходят сквозь все этапы разработки методики, получая на каждом своё выражение.

Хотя мы привыкли к тому, что существуют множественные «виды» валидности (которых в разных источниках выделяется до десяти), следует особо отметить, что *валидность как способность инструмента измерять то, для чего он предназначен, по существу, одна*. Используемые наименования: содержательная, критериальная, конструктивная, прогностическая, конкурентная и др. – отражают лишь её разные грани, источники и способы выявления. В последнее время намечается тенден-

ция к сокращению количества видов валидности, которые, по сути, *можно свести к 3-м качественно различным формам: содержательной, критериальной и конструктивной* (см., например, стандарты АПА и форму рецензии на тест EFPA). Эти формы валидности предполагают разные методы осуществления и актуальны на разных этапах разработки.

Так, основы содержательной валидности, предполагающей соответствие закладываемого в тест содержания изучаемому концепту, формируются на этапе теоретической разработки и закладываются в тест при формулировании пунктов. Поэтому к содержательной валидности больше подходит термин «обеспечение», чем «проверка».

Критериальная валидность определяется на основе соотнесения результатов теста с внешним (по отношению к тесту) критерием. При этом такие критерии могут быть весьма различны: это данные наблюдаемого поведения, успешность последующей деятельности, результаты аналогичного теста. Разные критерии дают названия разным вариациям критериальной валидности: прогностическая, конкурентная и т. д. Именно к критериальной валидности больше всего подходит термин «проверка», которая проводится, как правило, на завершающих этапах разработки теста. Однако и эта форма валидности может временно «обеспечиваться», например, на этапе анализа пунктов, если выбрана модель отбора пунктов на основе соотнесения с критерием.

Конструктивная валидность, возможно, наиболее важная, базируется на теоретически обоснованной модели взаимосвязей измеряемых тестом параметров с другими явлениями. Эта форма валидности требует постепенного накопления информации из разных источников и «изучается» как в процессе разработки теста, для чего планируются специальные исследования, так и после его издания.

Конечно, эти три формы валидности взаимосвязаны как три грани одного целого: если «смещена» или «заужена» содержательная область, то это выявится при изучении конструктивной валидности и отразится на показателях критериальной, которые будут неизбежно снижены. На самом деле, работа над валидностью, как над единым явлением, проходит сквозной линией через весь процесс разработки теста.

В начале организационного этапа задаются предварительные характеристики будущего

теста (такие как цель тестирования, область применения, контингент испытуемых и т. д.), которые в общих чертах обрисовывают желаемый конечный результат. Только ясное представление о конечном результате позволяет спланировать конкретные действия, выбрать методы и исследовательские модели, которые будут привлечены в процессе разработки.

На организационном этапе важно определить наиболее адекватные и информативные источники валидности для конкретного теста. А это напрямую зависит от цели тестирования (А. Анастаси, 2001). Несколько примеров позволяют проиллюстрировать это положение.

Так, например, если цель тестирования звучит как прогноз успешности деятельности, то основным источником валидности будет соотнесение результатов теста с показателями будущей деятельности. В этом случае акцент будет сделан на критериальной валидности. Соответственно, при составлении плана действий на IV-м этапе может быть избрана модель отбора заданий по критериальному принципу, а далее должны быть запланированы исследования, позволяющие получить конкретные показатели прогностической валидности.

Если цель тестирования – оценка степени усвоения знаний после прохождения некоторого обучающего курса (как в традиционных тестах достижений), то основным источником валидности – это соответствие содержания теста содержанию курса. Следовательно, акцент переносится на содержательную валидность, работа по обеспечению которой происходит на II-м этапе, где тщательно выверяются пропорции содержания и широта его охвата. Для отбора пунктов в этом случае может быть выбрана классическая модель анализа заданий или же IRT-модель.

Иногда целью разработки бывает диагностика целого комплекса сложных личностных переменных. В этом случае источники валидности – это тщательная проработка содержания и расширенные эмпирические исследования, способные выявить место изучаемых переменных среди других психологических явлений. Здесь акцент смещается на конструктивную валидность, изучение которой реализуется на IV-м этапе разработки через серию тщательно спланированных программ эмпирических исследований и психометрических экспериментов. Однако, вклад в изучение конструктивной валидности может быть внесён и на

начальных этапах разработки. Например, на содержательном этапе теоретический анализ может быть дополнен специально организованными исследованиями, уточняющими ключевую феноменологию. А для отбора заданий может быть рекомендована факторная модель с применением эксплораторного факторного анализа.

Таким образом, основные характеристики теста во многом задают ключевые источники валидности, которые, в свою очередь, определяют психометрические и исследовательские модели, привлекаемые к разработке. Приведённые примеры не являются универсальными и не претендуют на полноту охвата всех возможных вариантов. Конкретные решения зависят от типа теста, характера содержательной области тестирования и множества частных факторов. Однако именно чёткое представление о конечном результате позволяет грамотно выстроить стратегию разработки теста.

Итогом организационного этапа является поэтапный план проекта. Этот план содержит в краткой форме все существенные задачи, которые должны быть выполнены на каждом из этапов. Здесь же конкретизируются привлекаемые к разработке методы и модели.

Составление плана требует организационных решений, касающихся того, кто будет выполнять те или иные задачи, определения временных рамок, вопросов организации массовых исследований, финансирования и т. д. Здесь же должны быть приняты решения, связанные с привлечением или подготовкой специалистов различного профиля (например, разработчиков пунктов, экспертов по содержанию, программистов и др.). Успех всего предприятия в значительной степени зависит от проработанности такого плана, способности организаторов правильно выстроить стратегию действий, определить оптимальную линию исследований и обеспечить выполнение проекта грамотной финансовой и юридической базой.

Этап II. Содержательный

Второй этап – начало непосредственной работы над тестом. На данном этапе проводится серьёзная теоретическая работа по определению измеряемого концепта, и закладываются основы валидности будущего теста. Если содержание плохо определено или недостаточно тщательно описано, то никакие последующие действия не компенсируют этого

упущения. Можно потратить огромное количество сил и средств на массовые исследования и статистические процедуры, проводимые на последующих этапах, и потерпеть неудачу из-за недочётов в проработке содержания.

К сожалению, важность содержательного этапа разработки тестов традиционно недооценивается, и он неоправданно опускается, сокращается или же превращается в абстрактный теоретический обзор. Довольно часто отечественный разработчик полагается на свою интуицию или некоторую имплицитную теорию. Однако именно содержательный этап служит мостиком между теорией и практикой, где простраивается соответствие содержательных особенностей изучаемого концепта с формальными характеристиками будущего теста (такими, как тип теста, формат вопросов и др.).

На наш взгляд, недооценка содержательного аспекта разработки во многом вызвана непониманием того, что должно быть сделано на данном этапе. Поэтому рассмотрим подробнее специфику задач содержательного этапа разработки тестов, останавливаясь на тех моментах, где возникают наиболее серьёзные проблемы.

Содержательный этап состоит из двух шагов (шаги 2-й и 3-й – в общем процессе). Сначала определяется и детально прописывается содержание изучаемого явления. Затем на основе такой проработки проводится операционализация содержания и составляется спецификация теста, которая будет инструкцией ко всем дальнейшим действиям, вплоть до разработки схем интерпретации.

Шаг 2. Определение содержания

Определение содержания предполагает решение трёх весьма серьёзных задач. Это: (1) определение исходного теоретического концепта, (2) формирование репрезентативной выборки содержания и (3) описание ключевой феноменологии. Их последовательное оформление осуществляет корректный переход от теоретических построений к операциональным единицам содержания, которые затем найдут своё выражение в конкретных пунктах теста.

Определение концепта – исходная точка конструирования теста и краеугольный камень в фундаменте будущего здания валидности. Явлению, для измерения которого планируется разработать тест, должно быть дано чёткое определение, в котором бы отражалась концептуальная позиция автора, его понимание и трактовка явления.

Следует отличать термин «*концепт*» (от лат. *conceptus* – содержание понятия) от термина «*конструкт*» (от лат. *construction* – составление, построение) (Философский энциклопедический словарь, 2000). Термин «конструкт», который является более привычным для разработчиков тестов, был введён в обиход в 1955 году в связи с появлением такого понятия, как конструктивная валидность (Cronbach & Meehl). Под конструктами понимаются «...категории, выводимые логическим путём из общих признаков, свойств или черт, обнаруживающих себя в непосредственно наблюдаемых эмпирических переменных» (по А. Анастаси, 2001, с. 134). При этом отмечается, что сами конструкты часто недоступны прямому наблюдению, а могут быть идентифицированы лишь по косвенным эмпирическим признакам. Понятие «конструкт» показало свою продуктивность, прежде всего, при работе со сложными, недоступными прямому наблюдению реальностями, например, такими, которые получены посредством факторного анализа. Однако позже термин конструкт получил более широкое применение, и в настоящее время используется в качестве обозначения практически любой переменной, которую измеряет некоторый тест (П. Клайн, 1994, К. Купер, 2000). При этом основная задача изучения конструктивной валидности теста обычно формулируется как *идентификация конструкта* (А. Анастаси, 2001; К. Купер, 2000).

Более подробно эти вопросы будут рассматриваться во второй части статьи при обсуждении методов изучения конструктивной валидности. Здесь же важно подчеркнуть, что на начальном этапе разработки теста при определении содержания имеется в виду всё же не конструкт как идентифицированная посредством эмпирических исследований категория, а именно *концепт* как определение содержания понятия и обозначение исходной теоретической позиции автора теста. От того, как мы определим концепт, зависит – что именно мы будем измерять. Изначальное грамотное и ясное определение концепта позволит позже сформулировать продуктивные гипотезы, которые лягут в основу исследований конструктивной валидности теста. И, при удачной разработке, изучение конструктивной валидности на IV-м этапе процесса должно показать соответствие полученного конструкта исходному теоретическому концепту.

Однако задача определения концепта не

так проста как кажется на первый взгляд. Чётко определённых и детально проработанных концептов в современной психологии не так уж много. Здесь разработчик может столкнуться с серьёзными проблемами, связанными с *неясностью и размытостью концептуальной области, выбранной для разработки*. Например, относительно некоторого явления могут сосуществовать несколько взглядов, подходов, позиций, выражающихся в разных дефинициях. Такая ситуация достаточно распространена, и наиболее ярко иллюстрируется примером недавно развернувшейся в научном сообществе дискуссии вокруг понятия «компетенции».

На самом деле, различие мнений – это естественный и конструктивный процесс, однако, разработчик теста в подобном случае ставится перед необходимостью выбора. Он обязан чётко и однозначно определить концепт, который закладывает в основу теста, поскольку выбранное определение диктует предмет диагностики, методические подходы, источники валидности и многие другие характеристики теста. Например, если мы рассматриваем компетенции как комплекс некоторых знаний, умений, навыков, это предполагает одну линию разработки, если как область полномочий должностного лица – совсем другую, если же ядром определения компетенций становятся сложные качества, ответственные за способность к мобилизации имеющихся знаний и готовность к эффективной работе – третью, в корне отличающуюся от двух предыдущих. Меняется предмет диагностики и, соответственно, результат применения методики на практике. Поэтому разработчику *важно обозначить свою исходную позицию, хотя бы для того, чтобы пользователи понимали, с чем они имеют дело*.

Таких «спорных» и многозначных концептов в современной психологической науке немало. Ещё один актуальный пример – «эмоциональный интеллект», с диагностикой которого связано немало трудностей. В последнее время одну из возможных причин этих затруднений стали видеть в самой формулировке концепта. Если используется родовое понятие «интеллект», то это диктует определённое содержательное поле и определённые методы и модели диагностики. Неточно выбранный ракурс при теоретическом осмыслении концепта смещает его содержание, что, при разработке методик, неизбежно проявится в трудностях их валидации. Неверная ис-

ходная теоретическая позиция заводит в тупик множество практических проектов по разработке тестов. Поэтому следует ещё раз подчеркнуть важность и необходимость серьёзной теоретической проработки исходного концепта на самых начальных стадиях разработки теста.

Следующая задача в рамках определения содержания – **формирование репрезентативной выборки содержания** (А. Анастаси, 2001; К. Купер, 2000). Она является первым шагом на пути перехода от теоретического концепта к конкретным тестовым процедурам.

В некоторых случаях решение этой задачи не представляет особых трудностей, если мы, например, имеем дело с тестами учебных достижений. В этом случае в содержательное поле, которое будет представлено в тесте, входят все основные разделы подлежащего проверке материала. *Важно, чтобы был обеспечен полный охват содержания* и все разделы пройденного курса были представлены в будущем тесте. Иногда при планировании теста предпочитается равномерное распределение пунктов по всем содержательным областям. Однако во многих случаях важно учитывать и относительную значимость тех или иных разделов для общей оценки знаний. Эта значимость может определяться экспертами – специалистами в соответствующей области. (В тесте более значимые разделы могут быть представлены большим количеством пунктов, или же оцениваться большим количеством баллов).

Принцип формирования выборки содержания, проиллюстрированный на примере тестов достижений и заключающийся в полном охвате релевантных содержательных областей, универсален для всех видов тестов. Однако в случае разработки методик, направленных на диагностику личности, эта задача сопряжена с гораздо большими сложностями (А.Г. Шмелёв, 2002). В качестве областей содержания здесь должны быть выделены типы ситуаций, соответствующие искомому личностным проявлениям. Например, если говорить о чертах личности, то известно, что каждой черте соответствует специфическое *ситуативное поле*, в котором эта черта актуализируется и может проявиться с большей долей вероятности. Так, ситуативное поле для межличностной толерантности будет существенно отличаться от круга ситуаций, в которых мы можем ожидать проявлений такой характеристики, как ответственность.

При этом важно определить круг релевантных черт ситуаций во всей полноте. От этого во многом зависит как валидность будущего теста, так и его надёжность. Разнообразие содержательных областей вносит вариативность в содержание пунктов теста, что обеспечивает необходимый разброс в источниках ошибки измерения, способствуя усреднению и нивелированию ненужных отклонений от истинного показателя (К. Купер, 2000). Если же при составлении теста используется только небольшая часть содержательного поля, то в итоговом результате накапливается *систематическая* ошибка измерения, связанная с однотипностью посторонних влияний. Это не только сужает содержание, но и искажает его: мы получаем так называемое *«сопутствующее смешение»*, когда к искомому содержанию как бы примешивается некий «инородный» компонент, проистекающий из специфики ситуативной области содержания. Например, если мы используем как базу для измерения межличностной толерантности только ситуации межличностного конфликта, то итоговый результат будет во многом отражать также и стиль поведения в конфликте; и выделить здесь компонент собственно толерантности будет практически невозможно. Или же, если для измерения ответственности как личностной черты привлекаются лишь ситуации, связанные с учебной деятельностью, то неизбежно смешение с мотивацией учебных достижений. Таким образом, «сужение содержания» и ограничение ситуативного поля приводит к «смещению содержания» и, соответственно, к серьёзным нарушениям валидности теста.

Адекватность и полнота охвата содержания – ключевое условие валидности будущего теста. И оно обеспечивается на начальных этапах разработки – при формировании выборки содержания. Однако следует помнить, что содержательная валидность теста подвергается серьёзному риску на последующих этапах разработки, особенно в ходе проверки и отбора пунктов, когда изначально выверенное содержание может быть значительно сужено в погоне за показателем внутренней согласованности. Поэтому рекомендуемый состав и пропорция содержательных элементов закрепляются в *спецификации теста* (см. 3-й шаг разработки), и к этому документу следует постоянно обращаться в процессе дальнейшей работы.

Описание феноменологии – третья зада-

ча, связанная с определением содержания. Она напрямую вытекает из предыдущей и базируется на её основе. Когда выделены и описаны содержательные области, разработчик ищет ответы на вопросы: в чём конкретно проявляется интересующее его явление, каковы его характерные признаки, как его можно обнаружить? Например, в чём конкретно будут выражаться толерантность, ответственность, доброжелательность. Будут ли это какие-либо действия, мысли, чувства, отношения, и какие именно.

Описание феноменологии требует чёткого обозначения эмпирических феноменов, посредством которых выражаются искомые характеристики. Это должны быть конкретные показатели, которые можно наблюдать на практике и каким-либо образом зафиксировать. Например, наличие знания может выражаться в том, что человек может воспроизвести требуемую информацию, найти ответ на вопрос, увидеть изученную закономерность в эмпирическом материале, использовать знания для решения практических задач и т. д. Если таких индикаторов много, то важно выделить из них ключевые, наиболее показательные, – те, которые лучше отражают суть явления, определяют его уникальную специфику и позволяют дифференцировать (отличить) его от родственных, смежных или имеющих сходное ситуативное поле. Удачный выбор ключевой феноменологии – половина успеха будущего теста.

Феноменология служит своего рода мостиком, через который осуществляется непосредственный переход к диагностическому инструментарию. Именно феноменология будет закладываться в пункты теста. Поэтому описание феноменологии должно быть по возможности предельно конкретным. Это облегчит выбор формата пунктов и определение типа будущего теста. Иногда для прояснения феноменологии требуются дополнительные пилотажные исследования или же привлечение экспертов, имеющих опыт практической работы в соответствующей области.

В целом, все задачи 2-го шага: определение исходного теоретического концепта, формирование выборки содержания и описание ключевой феноменологии направлены на общую цель – заложить основы валидности теста, обеспечив её в содержательной области. По сути, проработка этих задач даёт осмысленные ответы на вопросы: (1) «что» именно предполагается измерять тестом; (2)

«где» следует искать эмпирические проявления изучаемого явления; (3) «как», посредством каких эмпирических феноменов, это явление даёт о себе знать. Оформленные ответы на эти вопросы служат прочной основой всех последующих действий по разработке теста.

Шаг 3. Составление спецификации теста

Последовательный переход от одной задачи к другой, осуществляемый на 2-м шаге разработки подготавливает условия для **операционализации содержания**. Составление спецификации – шаг разработки, где продолжается работа по обеспечению содержательной валидности теста. И основная задача спецификации – это перевод содержания в операциональные единицы теста.

Спецификация – это схема теста, в которой определены все его **формальные характеристики**. В спецификации в обязательном порядке фиксируются:

- тип теста,
- формат пунктов,
- описание невербального стимульного материала (при необходимости),
- состав и количество шкал,
- количество пунктов (для теста в целом и для каждой шкалы в отдельности),
- соотношение содержательных элементов,
- процедура тестирования,
- способы обработки результатов,
- тип стандартизованных показателей,
- формы отчётов о результатах.

Спецификация выступает в роли своеобразного «технического задания» для разработчиков, к которому они будут обращаться на всех последующих этапах.

Многие элементы спецификации прямо вытекают из результатов предыдущих шагов. Так, например, соотношение содержательных элементов в тесте отражает особенности содержательного поля, а формат пунктов во многом определяется ключевой феноменологией. Поэтому, чем тщательнее проведена работа по определению содержания, тем легче составить хорошую спецификацию, и тем качественнее она получится.

Однако составление спецификации не следует рассматривать как формальную механическую работу. Она требует тонкого понимания специфики закладываемого в тест содержания, широты знаний в области конструирования тестов и опыта такой работы. Почти каждый пункт спецификации требует

принятия важных решений, взвешивания альтернатив и умения оценить влияние отдельных действий на основные психометрические характеристики теста.

В этом смысле, самое трудное и самое ответственное решение – это **выбор формата пунктов** для будущего теста. Обсуждение различных форматов тестовых пунктов – задача не одной статьи, поэтому здесь остановимся только на самых общих принципах. Формат пунктов тесно связан с ключевой феноменологией, описание которой было одной из задач предыдущего, 2-го шага разработки. По сути, должен быть найден формат, который был бы способен фиксировать ключевую феноменологию с минимальными искажениями.

В действительности, у разработчика всегда существует выбор. Одна и та же феноменология может быть зафиксирована различными способами. Эти способы базируются на регистрации разных типов данных (L, Q, T, S), и для каждого типа данных, в случае конкретного теста, существует несколько теоретически возможных форматов. Например, если в качестве ключевой феноменологии выделена некоторая модель поведения, то информацию о ней можно получить посредством: наблюдения за реальным поведением человека, через самооценку частоты поведенческих проявлений, выбор модели поведения из вариантов в заданной ситуации, а также через проекцию действий на персонаж, оценку различных форм поведения с помощью специально подобранных шкал... И это ещё далеко не полный список вариантов.

Понятно, что выбор не должен быть случайным. Каждый из возможных способов имеет «плюсы» и «минусы». Каждый формат привносит с собой свои, специфичные для него, источники ошибки, которые могут быть связаны с процедурной сложностью, социальной желательностью, ошибками саморефлексии, сопутствующим смещением и т. д. Здесь важно взвесить «опасность» искажающих влияний. Некоторые из перечисленных недостатков могут корректироваться особенностями инструкции, специальным стимульным материалом, нюансами шкалы; некоторые же форматы «отпадают» сразу из-за большого балласта искажающих факторов. Аргументами в пользу выбора того или иного формата также могут быть приближенность к эмпирической феноменологии, простота получения данных, возможность формализации. Перед

осуществлением выбора разработчику полезно иметь перед собой список если не всех, то большинства возможных вариантов. Это расширяет поле видения и создаёт необходимые условия для правильного выбора.

Надо отметить, что сама задача выбора формата пунктов содержит немалый элемент творчества. Конечно, существуют эталонные формы, которые давно описаны, однако, вариативность способов фиксации одной и той же феноменологии требует творческого с ними обращения. Задача разработчика не только выбрать подходящий образец, но и наилучшим способом «приспособить» его к специфике конкретного теста. В некоторых случаях на этапе составления спецификации может быть проведён пилотаж разных форматов. Утверждённый формат пунктов становится эталоном для разработчиков заданий.

Особое место в спецификации занимает раздел, где фиксируется соотношение содержательных элементов в тесте. Например, если некоторая черта личности проявляется в нескольких типах ситуаций, то здесь указывается, сколько конкретно пунктов теста должно соответствовать тому или иному типу ситуаций. Данный раздел спецификации может быть оформлен в виде таблицы, где фиксируется пропорция содержательных элементов, выраженная в процентном соотношении, и, затем, в пересчёте на абсолютное количество пунктов для каждой содержательной области. Такая таблица становится важным подспорьем на этапе разработки и отбора пунктов.

Выше мы уже говорили о том, что сужение выборки содержания неизбежно приводит к снижению валидности теста за счёт накопления систематической ошибки измерения и смещения содержания. Чтобы избежать такого эффекта, следует постоянно сверяться с таблицей спецификации, особенно при разработке заданий, сборке теста и на этапе отбора пунктов. Отсев отдельных пунктов по статистическим показателям часто существенно изменяет выверенную пропорцию содержательных элементов. Бывает так, что от изначально широкого и тщательно прописанного содержания остаются только две–три специфические области, пункты соответствующие которым удачно преодолели статистический барьер. Обращение к спецификации на этом этапе служит напоминанием о том, какие пункты нельзя просто отбросить, а надо обязательно переформулировать или создать новые. В окончательной версии теста должна быть

сохранена заданная в спецификации пропорция различных содержательных элементов.

Однако следует особо отметить, что бывают случаи, когда далеко не все разделы спецификации могут быть окончательно утверждены на II-м этапе разработки. Например, структура теста (в частности, состав и количество шкал) иногда проясняется только после специальных исследований, проводимых на IV-м этапе, например, после экспериментального факторного анализа. История создания общеизвестных многофакторных личностных опросников (начиная с Р. Кэттелла) – тому прекрасная иллюстрация. В подобных случаях более верным решением будет оставить этот пункт спецификации под вопросом, с пометкой о необходимости уточнения.

В связи с этим, первоначально составленная спецификация может незначительно корректироваться и уточняться в процессе разработки теста (например, как уже говорилось выше, может уточняться количество и наименование шкал, может несколько измениться общее количество пунктов теста, а иногда – даже формат пунктов, если он не оправдал себя на этапе эмпирической проверки).

Составление спецификации теста является закономерным завершением содержательного этапа: базируясь на тщательной проработке содержания, спецификация становится планом-программой всех дальнейших действий по разработке теста. Готовая спецификация оформляется как отдельный документ, который находится всегда «под рукой» у разработчиков.

В целом, II-й содержательный этап разработки закладывает базовые основы валидности теста. Без тщательной проработки содержания все последующие действия не имеют практического смысла. Каждый шаг этого этапа, каждая последовательная задача осуществляет постепенный переход от абстрактного теоретического концепта к конкретным

тестовым процедурам, разработке и проверке которых посвящены следующие два этапа.

Литература

1. Анастаси, А. *Психологическое тестирование* / А. Анастаси, С. Урбина. – СПб.: Питер, 2001. – 668 с.
2. Батулин, Н.А. *Пособие по разработке тестов достижений. Учебное пособие* / Н.А. Батулин, Н.А. Курганский. – Челябинск, 2001. – 75 с.
3. Батулин, Н.А. *Современная психодиагностика России* / Н.А. Батулин // *Вестник ЮУрГУ. Серия «Психология»*. – 2008. – Вып. 2. – С. 4–9.
4. Батулин, Н.А. *Компендиум психодиагностических методик России: описание и первичный анализ* / Н.А. Батулин, А.В. Пичугова // *Вестник ЮУрГУ. – Серия «Психология»*. – 2008. – Вып. 1. – С. 63–68.
5. Клайн, П. *Справочное руководство по конструированию тестов: Введение в психометрическое проектирование* / П. Клайн; под ред. Л.Ф. Бурлачука. – Киев: Изд-во ПАН Лтд, 1994. – 688 с.
6. Купер, К. *Индивидуальные различия* / К. Купер; под ред. И.В. Равич-Щербо. – М.: Аспект Пресс, 2000. – 527 с.
7. *Философский энциклопедический словарь* / под ред. Л.Ф. Ильичёва, П.Н. Федосеева и др. – М.: Энциклопедия, 2000. – 840 с.
8. Шмельёв, А.Г. *Психодиагностика личностных черт* / А.Г. Шмельёв. – СПб.: Речь, 2002. – 480 с.
9. *American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). Standards for educational and psychological testing. Washington, DC: American Educational Research Association. 101 p.*
10. *Handbook of test development / edited by Steven M. Downing, Thomas M. Haladyna. – 2006 by Lawrence Associates, Inc. 778 p.*

Поступила в редакцию 1 мая 2009 г.

Батулин Николай Алексеевич. Доктор психологических наук, профессор, декан факультета психологии, заведующий кафедрой психодиагностики и консультирования Южно-Уральского государственного университета: nikbat@list.ru.

Nikolay A. Baturin. PsyD, professor, the dean of the Faculty of psychology, head of chair «Psychological diagnostics and Counselling», South Ural State University: nikbat@list.ru.

Мельникова Наталья Николаевна. Кандидат психологических наук, доцент кафедры социальной психологии ЮУрГУ: MNN17@yandex.ru.

Natalia N. Melnikova. Candidate of psychological sciences, docent of department of social psychology of South Ural State University: MNN17@yandex.ru.