

Социологические тесты: сущность и валидизация

Аннотация

В статье сделана попытка концептуализации такого направления эмпирической социологии, как социологическое тестирование. Для этого решается ряд задач. Во-первых, с точки зрения социологического тестирования оценивается релевантность четырех ключевых теорий, касающихся разработки измерительных инструментов в социальных науках, — классической теории тестов, математической теории измерения (Item Response Theory), теории надежности и теории валидности. Во-вторых, на основании опыта отечественной социологии очерчиваются сущностные особенности социологического тестирования (применимость в массовых опросах, нацеленность на изучение социальных феноменов, диагностический потенциал при анализе социальной системы). В-третьих, рассматривается специфика валидизации социологических тестов, включающей три этапа: теоретическая валидизация, выбор разновидности измерительного инструмента, эмпирическая валидизация. В завершении статьи рассмотрены примеры социологических тестов: Интегральный индекс социального самочувствия (Е.Головаха, Н.Панина), Типы политической культуры (Е.Головаха), SCL-9-NR (С.Дембицкий).

Ключевые слова: социологический тест, измерение, валидизация

В известной “Энциклопедии измерения и статистики” под редакцией Н.Салкаинда [Encyclopedia, 2007] термин “тест” встречается в названиях статей, преимущественно принадлежащих к трем тематическим разделам — понятия и проблемы в измерении, статистические техники, статистические тесты. В данной статье социологические тесты рассматриваются в первую очередь в контексте понятий и проблем измерения психологических, социально-психологических и социальных феноменов.

Интересно то, что несмотря на частое использование термина “тест” в названиях различных статей упомянутой энциклопедии, собственно его

определение не приводится. При этом дается определение шкалирования, под которым понимается деятельность с целью измерения или квантификации психологических установок (аттитюдов) или свойств (атрибутов) [Parker, 2007: р. 862]. Само измерение рассматривается как с классической (С.Стивенс), так и с современной точки зрения. В первом случае под ним понимается “присвоение числовых значений объектам или событиям согласно правилам”; при этом измерение рассматривается как процедура подготовки данных к статистическому анализу. Во втором случае измерение и статистика считаются фундаментально разными понятиями с собственными допущениями и результатами использования [Stemler, 2007: р. 582–584]. В рамках данной статьи социологические тесты рассматриваются в качестве измерительных инструментов, удовлетворяющих требованиям применения в социологических опросах.

Очерчивая такую перспективу, следует кратко рассмотреть четыре ключевые теории, касающиеся разработки измерительных инструментов в социальных науках, — классическую теорию тестов, математическую теорию измерения¹ (Item Response Theory), теорию надежности и теорию валидности, с тем чтобы оценить их релевантность в плане социологического тестирования.

Классическая теория тестов. С точки зрения этой теории, значение, полученное в результате использования теста, включает два компонента — истинное значение и ошибку измерения. Истинное значение рассматривается как среднее идентичных значений, полученных в результате повторяемых без ограничения измерений. Таким образом, ключевое допущение этой теории заключается в том, что полученное в рамках тестирования значение является суммой истинного значения и ошибки. При этом истинное значение всегда неизвестно. Измерительные ошибки — это любые аспекты измерения, кроме истинного значения. Их источниками может быть отбор пунктов методики, руководство теста, подсчет результатов и любые систематические ошибки измерения. Первые три источника приводят к неожиданным и несогласующимся эффектам, то есть оказывают несистематическое воздействие на результаты измерения и, соответственно, являются случайными ошибками. Такие ошибки не связаны как с истинным значением, так и с ошибками других измерений этого же свойства. В свою очередь, систематические ошибки возникают в тех случаях, когда тест измеряет что-то отличное от того, что он должен измерять (проблема валидности). Второе допущение касается параллельных тестов. Оно заключается в том, что каждый индивидуальный пункт методики сам по себе может рассматриваться как тест. Это связано с тем, что каждый из них является производным от латентной переменной. При этом влияние латентной переменной и величина случайных ошибок являются одинаковыми во всех случаях [Alexopoulos, 2007: р. 140–142].

Не отрицая общей фокусировки теории (истинное значение + ошибка измерения), сложно согласиться с теми акцентами, которые расставлены в вопросах о случайных ошибках. Систематическое смещение может вноситься любым элементом теста, а не только его невалидностью. Сомнитель-

¹ Данный вариант перевода позаимствован из статьи В.Аванесова [Аванесов, 2007].

ным выглядит и допущение об одинаковом влиянии латентной переменной на все индикаторы теста, особенно когда речь идет не об измерительной шкале, а об интегральном индексе (этот вопрос рассматривается далее).

Математическая теория измерения. В рамках этой теории утверждается вероятностный характер взаимосвязи между ответом на тот или иной пункт теста и латентной переменной. Математическая теория измерения включает статистические модели оценки качества пунктов теста. Наиболее распространенными и популярными являются модели для оценки тестов, предназначенных для изучения одномерных характеристик и использующих пункты с дихотомической шкалой. Более комплексные модели существуют и для случаев, когда пункты могут включать шкалы ответов с более чем двумя категориями и/или когда изучаемые характеристики являются многомерными [Keller, 2007: p. 493–494].

Данная теория выглядит более приемлемой аналитической схемой по сравнению с классической теорией тестов. Вместе с тем, учитывая акцент математической теории измерения на таких свойствах индикаторов теста, как уровень подготовленности испытуемого, показатель уровня трудности заданий, способность теста отличать подготовленных испытуемых от неподготовленных и т.д. [Аванесов, 2007], она подходит прежде всего для разработки педагогических, а не социологических тестов.

Теория надежности. Надежность измерения связана с его согласованностью в различных контекстах. В общем виде выделяют три аспекта надежности: 1) результаты надежного инструмента не зависят от того, кто проводит исследование; 2) использование одного и того же надежного инструмента в разное время дает одинаковые результаты; 3) все части надежного инструмента взаимосвязаны. Соответственно, можно выделить три вида надежности: надежность оценщика или кодировщика (результаты оценки разных экспертов одного и того же объекта с помощью одного и того же инструмента приводят к одинаковым результатам), тест-ретестовая надежность (изменения в результатах повторных измерений соответствуют теоретическим ожиданиям), внутренняя согласованность (взаимосвязанность пунктов теста является высокой) [Juni, 2007: p. 834].

Конечно же, проверка надежности является важным условием разработки социологического теста. При этом актуальность того или иного вида надежности может существенно варьировать в соответствии с теми задачами, для решения которых разработан конкретный тест.

Теория валидности. Рассматривая данную концептуальную область с точки зрения количественных исследований, можно (но лишь с определенной долей условности) говорить о единой теории. Если же принимать во внимание импликация качественного и смешанного подходов, то речь должна идти о целом наборе теорий валидности. И поскольку, на мой взгляд, теория валидности в рамках количественного подхода является существенно лимитированной, далее уделяется внимание собственно проблеме валидизации.

Понятие социологического теста

Понятие “социологический тест” было введено в научный обиход советским, а ныне российским, социологом и педагогом В.Аванесовым в 1982 году. Он дал социологическому тесту два определения. Согласно первому,

социологический тест — это “система отобранных социологическими методами высказываний, предъявляемых респондентам с целью получения надежной и валидной информации об интересующих признаках”, согласно второму — “система высказываний, позволяющих получить обоснованное отображение интересующей эмпирической системы с отношениями в числовую систему с отношениями” [Аванесов, 1982: с. 41]. В рамках этой статьи я буду придерживаться первого определения.

Двумя аргументами В.Аванесова в пользу выделения социологического теста в качестве самостоятельного метода исследования являются следующие [Аванесов, 1982: с. 38–39]:

1) необходимость изучения в социологических исследованиях не только психологических (для этого используются психологические тесты), но и социологических феноменов¹, таких как ценностные ориентации, мировоззренческие установки, социальные потребности и т.п.;

2) актуальность этого направления исследований на современном этапе развития методологии социальных наук.

На мой взгляд, спустя более чем 30 лет оба этих аргумента не утратили своего веса. Социологические или социальные феномены по-прежнему находятся в фокусе интереса исследователей, а методология разработки и использования социологических тестов за эти годы попала в поле интересов лишь двух украинских социологов — Наталии Паниной и Евгения Головахи, разработавших ряд социологических тестов², которые сегодня используются в массовых опросах.

Н.Панина и Е.Головаха распространили понимание социологического теста и на психологические методики, специально адаптированные для использования в массовых опросах. И это полностью оправдано, поскольку большинство психологических тестов являются весьма объемными, что накладывает существенные ограничения на их использование в социологических исследованиях. Кроме того, Н.Панина и Е.Головаха сформулировали ряд требований, которым должен удовлетворять социологический тест [Головаха, 1997: с. 5]³:

- универсальность — методика должна включать наиболее общие существенные индикаторы изучаемого явления, что открывает возможность ее использования практически в любом социологическом исследовании;
- интегральность — возможность сведения различных измерений социологического теста к одному интегральному показателю;
- качественность — социологический тест должен быть “паспортизирован” статистическими показателями надежности и валидности;

1 Далее я буду говорить о социальных, а не социологических феноменах.

2 Е.Головаха и Н.Панина разработали такие тесты, как Интегральный индекс социального самочувствия и Типы политической культуры, а также адаптировали ряд зарубежных методик с целью применения их в массовых опросах в Украине.

3 Представленные ниже требования сформулированы в рамках разработки Интегрального индекса социального самочувствия, но, по сути, их можно рассматривать в качестве универсальных принципов социологического тестирования.

- стандартизированность — установление “норм” изучаемого явления, что позволит интерпретировать полученные результаты с точки зрения выраженности изучаемого феномена;
- компактность — социологический тест должен быть достаточно компактным, чтобы его можно было использовать в массовых опросах;
- “чувствительность” — учитывая требование компактности, социологический тест должен обладать приемлемой точностью измерения;
- объективность — цель измерения должна быть скрытой от респондента, что позволяет избежать фальсификации результатов тестирования.

Учитывая изложенное выше касательно сущности социологического тестирования, можно выделить два его измерения. Первое и принципиальное касается необходимости обеспечить методические требования, которые позволяют полноценно использовать измерительную методику (не важно — психологическую или социологическую) в рамках массовых опросов. Уже в этом случае можно говорить о социологическом тестировании. Второе измерение связано с теми аспектами социальной действительности, на изучение которых направлено использование социологического теста. Методики, отвечающие этому измерению, направлены на изучение социальных феноменов. В этом случае подготовка концептуальной базы социологического теста осуществляется преимущественно социологическими средствами.

Кроме указанных, полагаю, следует выделить и третье измерение социологического тестирования, которое связано с потенциалом использования полученных результатов. В идеале социологический тест должен быть не просто разработан как валидная и надежная методика, направленная на изучение социального или психологического феномена, но и быть связанным с более широкими теоретическими положениями о строении и функционировании социальной системы. Благодаря этому открывается возможность диагностических выводов о состоянии ее различных частей, возможных перспективах ее функционирования и развития. Но даже если соответствующие теоретические положения недоступны, но при этом методика удовлетворяет требованиям первого измерения социологического тестирования, соответствующие положения могут быть сформулированы в процессе использования методики в эмпирических исследованиях.

Последнее, о чем хотелось бы сказать в этой части статьи, связано с зарубежной практикой использования социологических тестов. Прежде всего следует отметить, что такого термина в зарубежной социологии я не встретил. Вместе с тем, используя очерченные выше измерения социологического тестирования, можно легко обнаружить, что ряд используемых в западной социологии измерительных методик удовлетворяет требованиям социологического тестирования. Приведу два общеизвестных примера. Первым измерительным инструментом, который можно рассматривать в качестве социологического теста, является методика изучения ценностей Ш.Шварца¹. Она имеет приемлемый для массовых опросов размер, направлена на изучение социального феномена, а результаты ее применения повсеместно используются для оценки состояния различных обществ. В качестве второго примера

¹ Более подробно о методике Ш.Шварца см.: [Магун, 2007].

можно привести известный тест для оценки психологического дистресса, разработанный Р.Кесслером (К6)¹. Этот тест повсеместно используется в эпидемиологических опросах. Интересен тот факт, что несмотря на психологическую направленность, ввиду своей краткости эта методика пригодна именно в социологических, а не психологических исследованиях.

Здесь у читателя может возникнуть вопрос, надо ли в таком случае² вообще говорить о социологическом тестировании. Мой утвердительный ответ на него связан с тем, что специфика социологического измерения полностью не охватывается методологическими средствами современных измерительных подходов, которые опираются преимущественно на психологию и педагогику. Соответственно, специфичные для социологии проблемы решаются в большей степени интуитивным путем, а такое состояние для методологии эмпирических исследований нельзя назвать приемлемым.

Процесс валидизации социологического теста

Здесь я хочу коснуться именно специфики валидизации социологических тестов, а не многочисленных техник, которые могут применяться при их разработке. Это связано с тем, что к упомянутым техникам относятся многочисленные процедуры формулировки и отбора индикаторов, формирования выборки (начиная от пилотных опросов и заканчивая репрезентативными исследованиями с целью выделения норм), статистической обработки результатов использования теста на различных этапах его разработки. Эти техники слишком многочисленны и не могут быть охвачены в рамках одной статьи. Кроме того, многие из них являются специфичными и анализ ситуаций их применения не представляется возможным. Поэтому я исхожу из того, что читатель знаком со многими из них, а также имеет все необходимое для их дальнейшего освоения³.

Комплексный подход к процессу валидизации социологического теста предполагает выделение трех его этапов: теоретической валидизации, выбора разновидности измерительного инструмента, эмпирической валидизации.

Теоретическая валидизация. Чтобы избежать проблемы “глупой собаки, гонящейся за своим хвостом” (где “глупая собака” — это теория, которую намереваются верифицировать с помощью измерительных процедур, а “хвост” — измерительная методика, построенная на базе еще не проверенной теории), в основу социологического теста должны быть положены уже валидные теоретические положения. Такая постановка вопроса связана с тем, что “ломаные” теоретические конструкты вполне могут приводить к приемлемым результатам статистического анализа. С целью подтверждения этого положения я провел методическое исследование, в рамках которого сравнивались результаты использования методики, построенной на основании уже валидных теоретических положений, с результатами той же методики, кото-

1 Более подробно о методике Р.Кесслера см.: [K10 and K6 Scales, s.a.].

2 Имеется ввиду ситуация, когда социологические тесты разрабатываются несмотря на отсутствие соответствующего понятия и специальной концептуальной области.

3 Более подробно о соответствующих техниках см., напр.: [DeVellis, 2012; Анастаси, 2007; Клайн, 1994].

рая была намерено искажена таким образом, чтобы нарушить ее содержательную валидность [Дембицкий, 2014]. С помощью двух этих вариантов измерительной методики преподаватели и старосты одного из киевских вузов оценивали поведенческие стратегии своих студентов (или одногруппников). Далее полученные результаты были сопоставлены со средними баллами студентов, которые они получили на всех сессиях бакалаврата (см. табл.). Как видно, что касается статистики, оба варианта равнозначны, чего нельзя сказать о теоретических конструктах, которые за ними стоят.

Таблица

Сила и направление связи результатов экспертного оценивания со средними значениями оценок, полученных на всех сессиях бакалаврата (N = 23)

Эксперты	Сила связи	
	Использование “ломаной” шкалы	Использование валидной шкалы
Первый преподаватель	0,52*	0,63**
Второй преподаватель	0,80**	0,74**
Старосты	0,75**	0,74**
Среднее значение трех оценок	0,80**	0,81**

* $p < 0,05$;

** $p < 0,01$.

Конечно же, с точки зрения статистического анализа (например, нахождения взаимосвязей с другими конструктами или предсказания значений других переменных) оба варианта шкалы можно оценить как равноценные. Но с точки зрения социологической интерпретации, “ломаная” шкала теряет какую-либо ценность, так как отказ от содержательных элементов ведет к потере информации о двух типах студентов из трех¹.

Следующий вопрос здесь касается того, каковы источники тех теоретических положений, которые могут рассматриваться в качестве валидных. Поскольку феномены, попадающие в фокус социологического измерения, могут быть социальными (например, отношение к властным институтам), социально-философскими (например, справедливость), социально-психологическими (например, психологический климат в коллективе) и психологическими (например, тревожность), постольку к таким источникам могут быть отнесены отдельные концепции в рамках социальных наук, предварительные специальные исследования, профессиональный практический опыт, а в определенных случаях и универсальный опыт, то есть знания, присущие всем представителям определенной социальной общности. Рассмотрим их более подробно.

Если говорить о научных концепциях, то речь следует вести о тех из них, построение которых опиралось на исследовательскую стратегию кейс-ста-

¹ С помощью оставшейся части шкалы все студенты могут быть отнесены только к “безразличному” типу, в то время как “декларативно заинтересованный” и “ответственный” типы выпадают из фокуса исследования.

ди¹. При этом я отношу к таковым и некоторые концепции исторической макросоциологии. К последним, например, принадлежат результаты, представленные в работах Ч.Тилли [Тилли, 2009] и Б.Андерсона [Андерсон, 2001]. Эти теории хоть и выглядят весьма далекими от уровня социологического измерения, но, тем не менее, задают правильную фокусировку относительно наиболее релевантных процессов в сегодняшнем мире национальных государств, увеличивая тем самым теоретическую чувствительность исследователя.

Отличительной чертой проведения кейс-стади с целью получения валидных теоретических положений является акцент на изучении социальных феноменов, проявляющихся в социальной действительности², с дальнейшим переходом к теоретическим положениям, требующим не верификации, а инспектирования и обогащения в рамках тех социальных контекстов, которые не были охвачены инициальным исследованием.

Что касается предварительных комплексных исследований, то речь опять-таки идет об использовании кейс-стади. Отличие лишь в том, что здесь исследование “заточено” под построение требуемой для дальнейшего социологического измерения теоретической базы. Соответствующим примером является мое исследование поведенческих практик студентов, совмещающих работу с учебой [Дембицкий, 2016: с. 99–124].

В случае профессионального опыта все значительно проще. Выполнение ежедневных функциональных обязанностей позволяет специалистам получить понимание различных аспектов социальных феноменов естественным путем. Поэтому такой опыт является потенциальным источником валидных теоретических положений. Примером использования профессионального опыта при построении измерительной методики является шкала психологического дистресса SCL-90-R Л.Дерогатиса [Derogatis, 2004], который отталкивался от тех симптомов различных психологических расстройств, которые были зафиксированы соответствующим профессиональным сообществом.

Наконец, в качестве отправной точки могут использоваться и аксиоматические научные положения, валидность которых не вызывает сомнений. Так, это могут быть утверждения, рассматривающиеся в качестве базисных в ключевых социологических теориях (функционализм, символический интеракционизм, теория конфликта, феноменология, теория обмена) или же универсальные социологические аксиомы. Примером последних является положение, использованное Е.Головахой и Н.Паниной для конструирования Интегрального индекса социального самочувствия: “социальное самочувствие человека определяется степенью удовлетворения его социальных потребностей, которые, в свою очередь, являются производными от существующей в обществе системы социальных благ, их производства и распределения” [Головаха, 1997: с. 23–24].

1 Более подробно см., напр.: [Yin, 2009; Дембицкий, 2016: с. 125–149].

2 Под “социальной действительностью” я понимаю человеческие поступки на различных уровнях социального взаимодействия, направленные на создание и поддержание социальных институтов.

Рассмотренные источники могут вносить различный вклад в достижение теоретической валидности. Поэтому естественно использовать несколько из них в рамках одного исследовательского проекта.

Говоря непосредственно о теоретической валидности, следует понимать, что теоретическая база исследования должна обеспечивать три ее компонента: очевидный, содержательный и композиционный. Очевидная валидность связана с аргументацией, в рамках которой изучаемый феномен вписывается в социальную действительность и, соответственно, не является “плодом” чьих-то домыслов; содержательная валидность обеспечивает значимость и представительность элементов теоретического конструкта; композиционная – правильность его внутренней структуры [Дембицкий, 2016: с. 22–31]. В таком виде теоретическая база предоставляет все необходимое для дальнейшей операционализации теоретических положений, а также оценки полученных в рамках эмпирической валидации результатов. Но прежде чем перейти к последней, необходимо определиться со спецификой измерительного инструмента.

Выбор разновидности измерительного инструмента. После утверждения валидной теоретической базы специфика разработки измерительного инструмента обуславливается выбором его разновидности. На основании работ С.Садмена и Н.Брэдберна [Садмен, 2002] и Р.ДеВеллиса [DeVellis, 2012], к разновидностям измерительных инструментов в социальных науках я отношу следующие:

- а) анкетные вопросы различной степени сложности (предмет изучения — мнения, поведение, знания);
- б) измерительные шкалы (предмет изучения — латентные переменные);
- в) интегральные индексы (предмет изучения — латентные переменные, социальные феномены).

Останавливаясь на том, что такое анкетные вопросы, считаю излишним. Если же говорить о различии между измерительными шкалами и интегральными индексами, то в первом случае речь идет об измерении некоторого непосредственно не наблюдаемого свойства, обуславливающего те аспекты его проявления, которые заложены в основу индикаторов методики; в случае же интегральных индексов латентная переменная или социальный феномен обуславливаются теми факторами, на основании которых сформулированы индикаторы. Принципиальным методическим результатом такого разделения является нерелевантность внутренней надежности в рамках обоснования качества интегральных индексов. В этом случае между различными причинными факторами может быть связь, а может и не быть.

Эмпирическая валидизация. В идеале на основе результатов эмпирического исследования или их серии проверяются конструктивная, предсказательная, критериальная и конкурентная валидность.

Конструктивную валидность я рассматриваю в терминах, предложенных в работе Д.Кэмпбелла и Д.Фиске [Campbell, 1959]. На уровне простых анкетных вопросов, а также отдельных индикаторов измерительных шкал и интегральных индексов ее проверка осуществляется посредством сравнительного анализа распределения ответов и базовых описательных статистик, полученных при использовании различных формулировок и шкал ответов. На уровне итогового инструмента используется корреляционный анализ (в

широком смысле), если речь идет об измерительной шкале (с вниманием к внутренней согласованности) или интегральном индексе (без внимания к внутренней согласованности). Если же апробируется социологическая анкета, акцент делается на пилотных опросах с целью проверки контекстуальной обусловленности ответов при использовании различных последовательностей одних и тех же вопросов.

Критериальная валидность¹ основывается на получении ожидаемых различий при опросе “контрастных” групп и, таким образом, показывает общую “чувствительность” измерительного инструментария.

Предсказательная валидность связана с возможностью методики предоставлять информацию, необходимую для прогнозирования будущих состояний объекта исследования. С точки зрения социологического тестирования, это, возможно, наиболее труднодостижимый вид валидности — как в силу требуемой краткости инструментария, так и в силу комплексности изучаемых в социологии процессов и явлений. В связи с этим результаты социологического тестирования — лишь один из возможных источников эмпирической информации, на основании которой строится социологический прогноз.

Конкурентная валидность понимается как сравнительное преимущество измерительного инструмента в ряду схожих методик. Таким образом, ее обоснование связано с необходимостью сравнения показателей конструктивной, критериальной и предсказательной валидности различных социологических тестов, направленных на изучения одного и того же явления.

Примеры социологических тестов: социальное самочувствие, политическая культура, психологический дистресс

Далее анализируются три социологических теста, конструирование которых так или иначе связано с именами известных киевских социологов Е. Головахи и Н. Паниной. Для каждого теста кратко рассматриваются его предназначение, особенности конструирования, статистической проверки и валидизации, а также социологическая специфика в разрезе трех описанных ранее измерений социологического тестирования (применимость в массовых опросах, направленность на изучение социальных феноменов, диагностический потенциал, с точки зрения анализа социальной системы).

Социологический тест “Интегральный индекс социального самочувствия” (далее — ИИСС). ИИСС (авторы — Е. Головаха, Н. Панина²) представляет собой стандартизированную тестовую процедуру, которая была разработана в связи с необходимостью оценки массовых эмоционально-оценочных состояний в постоянно меняющихся условиях нестабильного переходного общества [Головаха, 1997: с. 11]. Социальное самочувствие понимается авторами методики как “эмоционально-оценочное отношение индивида к системе социальных отношений и к своему месту в этой системе” [Головаха, 1997: с. 20].

¹ Критериальная, предсказательная и конкурентная валидности интерпретируются мною достаточно свободно, что связано с авторским пониманием специфики социологических исследований.

² Статистический анализ был осуществлен А. Горбачиком.

При конструировании теста было выделено 11 сфер социальной жизнедеятельности индивида: социальных отношений, социальной безопасности, национальных отношений, социально-политическая, профессионально-трудовая, информационно-культурная, рекреационно-культурная, материально-бытовая (первого и второго уровня), межличностных отношений, личностная. Внутри каждой сферы было выделено по 4 индикатора, что дало в итоге 44 пункта методики, описывающие социальные блага. Такая детализация обеспечила содержательную валидность, обоснование которой было осуществлено на логическом уровне.

В качестве альтернатив ответа была сформулирована так называемая «шкала достаточности»: *не хватает; трудно сказать, хватает или нет; хватает; не интересует*. В целях статистической обработки было принято решение приравнять четвертый пункт шкалы (не интересует) ко второму (трудно сказать, хватает или нет) [Головаха, 1997: с. 21–24].

На основании результатов опроса ($N = 1810$), репрезентативного для взрослого населения Украины (от 18 лет и старше), проведена оценка валидности и надежности методики в целом, проанализировано использование каждого из предложенных в первоначальном варианте индикаторов социального самочувствия, построен краткий вариант теста, осуществлены стандартизация и нормализация шкалы [Головаха, 1997: с. 27–28].

На основании результатов корреляционного анализа каждого пункта теста с суммарным индексом, оценки влияния исключения отдельных пунктов на внутреннюю согласованность и проверки нормальности распределения суммарного индекса все 44 пункта методики были включены в финальный вариант. Успешную проверку прошли такие виды надежности, как расщепление теста по полам, ретестовая и внутренняя согласованность.

Также была проверена конструктивная валидность на основании сопоставления суммарного индекса для различных категорий респондентов, выделенных в соответствии с ответами на вопросы о различных аспектах жизненной удовлетворенности [Головаха, 1997: с. 40–43].

Сокращенный вариант (20 из 44 пунктов) был построен на базе формальных (сохранение внутренней согласованности) и содержательных оснований (сохранение в основных чертах первоначальной многоаспектности измерений). В результате был получен измерительный инструмент, отвечающий стандартам надежности и валидности [Головаха, 1997: с. 46–47].

Проанализируем социологическую специфику теста. ИИСС разрабатывался и оценивался прежде всего как инструмент социологического измерения для массовых опросов. Поэтому он удовлетворяет принципиальному требованию социологического тестирования — применимости в массовых опросах. Сам феномен социального самочувствия включает как психологические, так и социальные компоненты, со значительным преобладанием последних. Следовательно, показатели ИИСС зависят прежде всего от специфики общественных процессов, связанных с удовлетворением потребностей индивидов. С одной стороны, это является аргументом в пользу диагностического потенциала теста при анализе социальной системы нашего общества. Вместе с тем здесь остается ряд нерешенных вопросов, среди которых, например, учет изменения актуальности различных потребностей в контексте более масштабных социальных процессов (так, война на Востоке Украины должна была сказаться на актуальности сфер социальной безопасности и национальных отношений) и разработка критериев оценки

успешности различных социальных институтов, отвечающих за предоставление основных социальных благ в нашем обществе (хотя потенциал теста вполне позволяет рассчитывать на решение этих задач).

Обобщая представленную информацию, можно сделать вывод, что ИИСС является полноценным социологическим тестом, который удовлетворяет высоким стандартам, выдвигаемым к разработке измерительных инструментов в социальных науках.

Социологический тест “Типы политической культуры” (далее – ТПК). ТПК (автор – Е.Головаха¹) является интегральным тестом, предназначенным для определения типа политической культуры респондента. При этом под политической культурой подразумевается “совокупность устойчивых форм политического сознания и поведения людей, которые определяют особенности функционирования разных субъектов политического процесса в рамках определенной политической системы” [Щербак, 2006: с. 155].

При детализации теоретического конструкта было выделено два измерения политической культуры: демократия / тоталитаризм, активность / пассивность (в смысле гражданского участия). Для каждого из измерений было сформулировано по 50 утверждений, при этом для сбалансированности было разработано по 25 утверждений для демократии и тоталитаризма, а также для активности и пассивности.

Для регистрации ответов была использована шкала Лайкерта: “полностью согласен”, “скорее согласен”, “трудно сказать, согласен или нет”, “скорее не согласен”, “абсолютно не согласен”.

С целью сокращения методики до приемлемого для социологических опросов размера, проверки конструктивной валидности и ретестовой надежности был проведен пилотный опрос среди респондентов старше 18 лет ($N = 130$). Для оценки ретестовой надежности каждого пункта использовался как корреляционный анализ (надежным считался пункт, который показывал силу связи не менее 0,3), так и расчет шкальных оценок по формуле средней арифметической погрешности (не более 1).

Далее на основании факторного анализа были определены суждения, которые дают однофакторное решение на шкалах демократия / тоталитаризм и активность / пассивность. Также была оценена внутренняя надежность (α Кронбаха) и сила связи между отдельными индикаторами и аддитивным индексом. Учитывая все полученные результаты из исходного набора суждений было отобрано по восемь для каждой из шкал.

Этот промежуточный вариант был включен в опрос в рамках “Европейского социального исследования”. Дальнейшая апробация методики также была проведена с использованием проверки внутренней согласованности и факторной структуры. Для обеих шкал были получены двухфакторные решения, с разделением на шесть и два индикатора. После удаления четырех утверждений, не входивших в число основных факторов, был получен финальный вариант теста, включающий по шесть суждений для шкал демократии / авторитаризма и активности / пассивности.

Финальный вариант ТПК был использован для выделения типов политической культуры населения Украины с использованием кластерного анализа методом k -средних. Исходя из теоретических ожиданий наличия четы-

¹ Статистический анализ был осуществлен А.Горбачиком.

рех типов политической культуры (демократическая активная, демократическая пассивная, авторитарная активная, авторитарная пассивная), было построено четыре кластера, которые подтвердили выдвинутые ранее теоретические допущения [Стегний, 2015: с. 29–32].

Рассмотрим тест с точки зрения социологии. Как и в предыдущем случае, ТПК был спроектирован и реализован как социологический тест. Касательно краткости он даже превосходит ИИСС, что обеспечивает его соответствие первому измерению социологического тестирования. Две концептуальные дихотомии, лежащие в основании этого инструмента, являются всецело социальными, что подтверждает его направленность на изучение именно социальных феноменов. Также нет никаких сомнений, что ТПК измеряет результирующие эффекты институциональной динамики и демонстрирует потенциал демократических процессов в обществе. Соответственно, его результаты в высшей степени применимы для оценки состояния социальной системы.

При этом я хочу отметить один аспект, вызывающий вопросы. Так, все шесть суждений для шкалы демократия / авторитаризм сформулированы таким образом, что касаются именно демократических установок. Следовательно, вопрос относительного того, насколько по ним можно судить об авторитарных установках, остается открытым. Особенно актуален этот вопрос с точки зрения кризисных процессов в современной Украине, увеличивающих вероятность развития авторитарных взглядов среди населения нашей страны.

Тем не менее, тест ТПК является высококачественным социологическим инструментом, базирующимся на не подлежащих сомнению теоретических положениях и прошедшим многоуровневую процедуру отбора эмпирических индикаторов, что делает его постоянное использование в социологических опросах по-настоящему актуальным.

Социологический тест “SCL-9-NR”. SCL-9-NR (автор — С.Дембицкий) является модификацией известного психологического теста SCL-90-R¹, разработанного Л.Дерогатисом [Derogatis, 2004]. Отличительными чертами SCL-9-NR по сравнению с SCL-90-R являются его краткость (для каждого из девяти симптоматических измерений оставлен только один индикатор), а также принципиальные изменения в задании методики и шкале ответов (но не в формулировках самих индикаторов). Главное предназначение теста — оценка психологического дистресса респондентов в рамках массовых опросов.

Оригинальная методика, предложенная Л.Дерогатисом, базируется на таких симптоматических измерениях, как враждебность, обсессивно-компульсивное расстройство, депрессия, межличностная сензитивность, параноидальные идеи, тревожность, соматизация, фобическая тревожность, психотизм, каждое из которых включает набор индикаторов (от шести до тринадцати пунктов). При отборе индикатора для каждого симптоматического измерения учитывалась сила его связи с индексом своей подшкалы, а также с интегральным индексом SCL-90-N [Дембицкий, 2016а: с. 54–55].

¹ С целью адаптации методики в Украине SCL-90-R был включен Н.Паниной в мониторинговые опросы Института социологии НАН Украины в 1997 и 1999 годах.

Изменение формулировки тестового задания и, соответственно, самой шкалы ответов¹ было осуществлено на основании рекомендаций, приведенных в известной работе С. Садмена и Н. Брэдберна по разработке социологического инструментария [Садмен, 2002: с. 65–96]. Главной причиной такой коррекции послужило предполагаемое смещение в ответах респондентов, которое возникает при использовании оригинального варианта формулировки: респонденты были склонны выбирать меньшие значения на шкале ответов, что уменьшало и финальный уровень индекса психологического дистресса. Как показала дальнейшая проверка, такое предположение оказалось верным, и использование новых формулировок значительно улучшило финальное распределение индекса [Дембицкий, 2016а: с. 59–61].

На основании онлайн-опросов в Киеве ($N = 200$) и Львове ($N = 200$) была проверена внутренняя надежность и факторная валидность методики. В дальнейшем, поскольку тест направлен на изучение психологического дистресса, был проведен опрос в Черниговской областной психоневрологической больнице ($N = 181$) с целью проверки критериальной валидности и влияния эффекта социальной желательности. Для оценки последнего на обычной выборке совместно с кафедрой методологии и методов социологических исследований КНУ им. Тараса Шевченко был проведен онлайн-опрос среди пользователей социальных сетей ($N = 367$). Полученные результаты подтвердили как критериальную валидность теста, так и приемлемый уровень влияния эффекта социальной желательности на его результаты. Кроме этого, на основании опроса в психоневрологической больнице были выделены значения теста, характеризующие повышенный и высокий дистресс [Дембицкий, 2016b].

С целью проверки конструктивной и конкурентной валидности, а также выделения норм теста на основе результатов репрезентативного для взрослого населения Украины опроса SCL-9-NR был включен в мониторинговое исследование “Украинское общество” в 2016 году². На момент написания статьи соответствующие результаты еще не были доступны для анализа.

Оценивая методику в качестве социологического теста, следует признать, что таковой она может рассматриваться только в контексте первого измерения социологического тестирования — его применимости в социологических опросах. Собственно, на данный момент основная работа по разработке методики была проделана именно в этом направлении. При этом, несмотря на то, что это наиболее краткий из рассматриваемых в статье социологических тестов, он несколько уступает упомянутому ранее тесту психологического дистресса Р. Кесслера (К6). Однако SCL-9-NR базируется на девяти симптоматических измерениях, в то время как К6 только на двух — депрессии и тревожности.

Сам психологический дистресс нельзя назвать социальным феноменом, однако можно предположить его зависимость от ряда социальных характеристик, что потенциально делает SCL-9-NR инструментом анализа соци-

1 Была использована шкала: “никогда”, “изредка”, “периодически”, “почти постоянно”.

2 Организатором исследования выступает Институт социологии НАН Украины. Результаты опроса репрезентативны для населения Украины старше 18 лет (выборка — не менее 1800 человек). Анкета опроса содержит более 300 вопросов и охватывает различные аспекты жизни украинского общества.

альной системы. Для проверки этого допущения в дальнейшем будет проведена дополнительная исследовательская работа.

Краткие выводы

Учитывая опыт отечественной и зарубежной эмпирической социологии, можно сказать, что каждое из трех измерений социологического тестирования является областью научной практики, включающей множество научных проблем, требующих решения.

Первое измерение (применимость методики в массовых опросах) связано преимущественно с методическими аспектами разработки методики, что предполагает техническую осведомленность ученого и организационные навыки, необходимые для проведения методических исследований. Вместе с тем, если речь идет о втором измерении — изучении социальных феноменов, то техническим и организационным навыкам должны сопутствовать надежные теоретические знания и социологическое воображение. Наконец, если задействованы все три измерения, то следует также добавить достаточные ресурсы — как временные, так и финансовые, ведь результаты использования социологического тестирования должны быть тщательно изучены в привязке к соответствующему социальному контексту. Все эти аспекты делают социологическое тестирование более комплексной научной практикой по сравнению с психологическим.

Соответственно, методология социологического измерения не должна находиться в зависимости от психологического, пускай последнее и обладает непоколебимым авторитетом и является ядром измерительного подхода в социальных науках. Исходным пунктом социологического тестирования должен быть изучаемый феномен, а также тот социальный контекст, в котором он изучается. Что касается всей совокупности подходов и техник, используемых психологами, то ее следует рассматривать исключительно в качестве доступного, но не определяющего инструментария.

Данное требование подтверждается и тем, что все из пяти упомянутых в статье социологических тестов разработаны с использованием существенно отличающихся подходов и техник. Таким образом, при безусловном внимании к последним ключевым показателем качества социологического теста является его эффективность для изучения социальной действительности в фундаментальных, прикладных и оценочных исследованиях: право ключевого голоса принадлежит прагматическим соображениям.

Источники

Аванесов В. Истоки и основные понятия математической теории измерения (Item Response Theory) / В. Аванесов // Педагогические измерения. — 2007. — № 3. — С. 3–36.

Аванесов В. Тесты в социологическом исследовании / В. Аванесов. — М.: Наука, 1982. — 200 с.

Анастаси А. Психологическое тестирование / А. Анастаси, С. Урбина. — СПб.: Питер, 2007. — 688 с.

Андерсон Б. Воображаемые сообщества. Размышления об истоках и распространении национализма / Б. Андерсон. — М.: Канон-пресс-Ц: Кучково поле, 2001. — 288 с.

Головаха Е. Интегральный индекс социального самочувствия (ИИСС): конструирование и применение социологического теста в массовых опросах / Е. Головаха, Н. Панина. — К.: Ин-т социологии НАН Украины, 1997. — 64 с.

Дембицкий С. Критериальная валидизация экспресс-теста SCL-9-NR / С. Дембицкий // Социология: теория, методы, маркетинг. — 2016. — № 2. — С. 102–116.

Дембицкий С. Относительно проблематичности статистической проверки теоретических конструкторов / С. Дембицкий // Социология: теория, методы, маркетинг. — 2014. — № 2. — С. 190–194.

Дембицкий С. Теоретическая валидизация в социологическом исследовании / Дембицкий С. — М. : ЛЕНАНД, 2016. — 200 с.

Дембицкий С. Экспресс-тест SCL-9-NR: методика оценки выраженности психологического дистресса для массовых опросов / С. Дембицкий // Социология: теория, методы, маркетинг. — 2016. — № 1. — С. 52–64.

Клайн П. Справочное руководство по конструированию тестов: Введение в психометрическое проектирование / Клайн П. — К. : ПАН ЛТД, 1994. — 283 с.

Магун В. Жизненные ценности населения: сравнение Украины с другими европейскими странами / В. Магун, М. Руднев // Украинское общество в европейском пространстве / под ред. Е. Головахи, С. Макеева. — К. : Ин-т социологии НАН Украины : Харьков. нац. ун-т им. В.Н. Каразина, 2007. — С. 226–273.

Садмен С. Как правильно задавать вопросы / С. Садмен, Н. Брэдберн. — М. : Ин-т Фонда “Общественное мнение”, 2002. — 382 с.

Стегний О. Соціологічні підходи до вивчення типів політичної культури / О. Стегний // Український соціум. — 2015. — № 2. — С. 26–35.

Тилли Ч. Принуждение, капитал и европейские государства. 1990–1992 гг. / Тилли Ч. — М. : Территория будущего, 2009. — 328 с.

Щербак М. Социально-классовые характеристики политической культуры в Украине: опыт применения социологического теста “Типы политической культуры” / М. Щербак // Социология: теория, методы, маркетинг. — 2006. — № 3. — С. 153–160.

Alexopoulos D. Classical Test Theory / D. Alexopoulos // Encyclopedia of measurement and statistics / ed. by N.Salkind. — Thousand Oaks ; London ; New Delhi : SAGE, 2007. — P. 140–143.

Campbell D. Convergent and Diskriminant Validation by the Multitrait-multimethod Matrix / Donald Campbell, Donald Fiske // Psychological Bulletin. — 1959. — № 2. — P. 81–105.

Derogatis L. The SCL-90-R, the Brief Symptom Inventory (BSI), and the BSI-18 / L. Derogatis, M. Fitzpatrick // The use of Psychological Testing for Treatment Planning and Outcomes Assessment. Vol. 3: Instruments for Adults / ed. by Mark E. Maruish. — New Jersey ; London : Lawrence Erlbaum Associated Publishers, 2004. — P. 1–42.

DeVellis R. Scale Development: Theory and Applications / DeVellis R. — Los Angeles ; London ; New Delhi ; Singapore ; Washington DC : SAGE, 2012. — 216 p.

Encyclopedia of measurement and statistics / ed. by N. Salkind. — Thousand Oaks ; London ; New Delhi : SAGE, 2007.

Juni S. Reliability Theory / S. Juni // Encyclopedia of measurement and statistics / ed. by N.Salkind. — Thousand Oaks ; London ; New Delhi : SAGE, 2007. — P. 834–835.

K10 and K6 Scales [Electronic resource] — Access mode :
http://www.hcp.med.harvard.edu/ncs/k6_scales.php.

Keller L. Item Response Theory / L. Keller // Encyclopedia of measurement and statistics / ed. by N.Salkind. — Thousand Oaks ; London ; New Delhi : SAGE, 2007. — P. 493–496.

Parker S. Scaling / S. Parker // Encyclopedia of measurement and statistics / ed. by N.Salkind. — Thousand Oaks ; London ; New Delhi : SAGE, 2007. — P. 582–584.

Stemler S. Measurement / S. Stemler, D. Birney // Encyclopedia of measurement and statistics / ed. by N.Salkind. — Thousand Oaks ; London ; New Delhi : SAGE, 2007. — P. 862–865.

Yin R. Case Study Research: Design and Methods / Yin R. — London ; Thousand Oaks ; New Delhi : Sage, 2009. — 219 p.