

ТАТЬЯНА НИКИТИНА,

социолог отдела теории, истории и методологии социологии Института социологии НАНУ, аспирантка Киевского национального университета им. Тараса Шевченко

Теоретическая и эмпирическая интерпретация и операционализация понятия “валидность”¹

Abstract

Validity is one of the basic properties of empirical measurement. A common definition of validity is the extent to which the survey measure accurately reflects the intended construct. But in different sociological source of information there are different understandings of the domain of validity. Some researchers understand validity as a property of measurement, while others — as a property of measurement instrument. Researchers determine different types of validity. There is no standardization of notions. Such situation complicates the process of validation of particular measurement. In fact, there is no concrete pattern of measurements validation. The purpose of this article is to make theoretical and empirical interpretation and operationalization of the validity notion and as a result develop the concrete pattern of measurements validation.

Вопрос качества социологического измерения и социологической информации как результата такого измерения — один из важнейших для эмпирической социологии. Без доказательства качества социологического измерения мы не можем быть уверены в содержательных выводах, к которым приходим на основе данных, полученных в ходе измерения. В.Паниотто в работе “Качество социологической информации” писал: “Претензии социологии на участие в решении реальных практических задач, стоящих перед нашим об-

¹ Эта статья была бы невозможной без стипендии университета Констанц и помощи Агнежки Войты, благодаря чему здесь цитируется большинство первоисточников. Признательна всем, кто согласился прочитать и прокомментировать предыдущие версии статьи. Особая благодарность моему научному руководителю Андрею Петровичу Горбачику, а также Ксении Урсуленко, Татьяне Любивой и Алексею Сергеевичу Серее.

ществом, могут быть обоснованными лишь в том случае, если рекомендации социологов будут базироваться на надежной (прежде всего — валидной) социологической информации. Повышение надежности социологической информации — важный фактор повышения эффективности социологической науки и увеличения ее практической отдачи” [Паниотто, 1986: с. 158]. Валидность является одной из характеристик качества социологического измерения. В обобщенном виде валидность — **это мера того, насколько мы измеряем то, что запланировано измерить** [Bohrsted, 2000: с. 3207].

В социологической литературе существуют разные взгляды на то, что собственно характеризует валидность — социологическое измерение в целом или измерительный инструмент. Исследователи выделяют несколько видов валидности. Стандартизация понятий отсутствует. Различаются и взгляды на соотношение таких ключевых показателей качества социологического измерения, как валидность и надежность. Разные ответы на эти вопросы диктуют различные схемы проверки валидности.

Такая ситуация терминологической неопределенности, “размытости границ” понятия валидности усложняет проверку валидности конкретного измерения. Итак, необходимо разработать четкую схему проверки, которая предполагает теоретическую и эмпирическую интерпретацию и операционализацию понятия “валидность”, то есть определение его границ и содержания и очерчивания методов ее проверки.

Теоретическая интерпретация понятия “валидность”

Сегодня в украинской социологии параллельно существуют две традиции оценки качества социологического измерения в целом и оценки валидности — советская и западная, которые разнятся терминологически, но при более глубоком анализе становится понятно, что по сути они очень похожи.

Кратко отмечу, что имею в виду, говоря о “советской” и о “западной” традициях.

Под западной традицией оценки качества социологического измерения я понимаю подходы американских и западноевропейских авторов начиная с 30-х годов XX века до настоящего времени. Объединить их работы в одну общую традицию позволяет их преемственность и общность взглядов в принципиальных вопросах.

Западная традиция берет начало из психологических тестов, из работ по оценке качества этих тестов. К первым исследованиям надежности и валидности относятся монографии Л.Терстоуна “Надежность и валидность теста” [Thurstone, 1931] и П.Гилфорда “Психометрические методы” [Guilford, 1936]. Позже появились статьи в новом журнале “Образовательные и психологические измерения” (Educational and Psychological Measurement). В этих двух книгах и журнальных статьях формально было определено понятие “валидность”. Здесь его связывали с корреляцией данных с определенным критерием. В современной терминологии такой вид валидности называют валидностью по критерию.

В конце 1940-х — в начале 1950-х годов в научный оборот был введен еще один вид валидности — валидность по содержанию, или содержательная валидность (content validity). Определение содержательной валидности было приведено Э.Куретоном в 1951 году в монографии “Измерения в сфере образования” (Educational Measurement) [Cureton, 1951].

В 1955 году Л.Кронбах и П.Мел опубликовали статью под названием “Конструктная валидность в психологических тестах” [Cronbach, Meehl, 1955]. Из этой статьи в научный оборот попал еще один вид валидности — конструктная валидность. В 1959 году Д.Кемпбелл и Д.Фиске разработали стратегию проверки конструктной валидности, выделив два подвида конструктной валидности — конвергентную и дискриминантную. При процедуре проверки конструктной валидности они использовали мультипризнаковую—мультиметодную матрицу (МТММ)¹.

В 1954 году Американская психологическая ассоциация издала первые стандарты психологического тестирования, в которых давалось определение валидности и надежности и видов валидности и надежности [Technical recommendations, 1954]. В дальнейшем эти стандарты переиздавались в 1966, 1974, 1985 и 1999 годах [Standards, 1966; 1974; 1985; 1999]. На эти стандарты до сих пор ссылаются как психологи, так и социологи. На основании Стандартов выделяют три основных вида валидности: критериальную (имеющую два подвида: прогностическую и конкурентную), содержательную и конструктную.

В 1979 году в популярной серии “Университетские издания “Сейдж” вышла книга “Оценка надежности и валидности” Дж.Карминеса и А.Зеллера. Ее авторы также выделяют три вида валидности, как и в Стандартах. Большинство современных западных авторов работают в русле указанных выше классических исследований, расширяя и углубляя основные разработки классиков.

Советская традиция оценки качества социологического измерения основывалась на базовых принципах и понятиях метрологии — науки об измерении, методах и средствах обеспечения необходимого уровня его качества. Метрология опирается на методологические основы измерения и решает следующие задачи: создание эталонов, мер и инструментов измерения; установление средств оценки точности и правильности измерений [Докторов, 1979, с. 7]. Большинство советских авторов ссылались на метрологическую литературу. К “советской традиции оценки качества социологического измерения” относятся работы А.Здравомыслова (1969), В.Ядова (1972), В.Воловича (1974), Б.Докторова (1979), Г.Саганенко (1983), В.Паниотто (1986) и др. В большинстве случаев советские авторы использовали термин “обоснованность” вместо термина “валидность” (за исключением Паниотто, который применяет термин “валидность”). Но по смыслу валидность и обоснованность в этих исследованиях означают примерно одно и то же. В советской традиции, в противовес западной, не уделяли существенного внимания рассмотрению отдельных видов валидности (обоснованности) (опять-таки кроме работы Паниотто, в которой выделены теоретическая и эмпирическая валидность).

Нельзя говорить о прямой преемственности западной и советской традиций, поскольку они имели разные источники (психологическое тестирование и метрология) и развивались более или менее автономно, однако и об абсолютной изолированности советских авторов от западных нельзя говорить.

¹ С этим методом можно ознакомиться в статье: [Campbell, Fiske, 1959].

Что касается современной украинской литературы, то здесь встречаются ссылки как на советских, так и на западных авторов. В связи с этим наблюдается определенная терминологическая путаница.

Попытаюсь соотнести термины “советской традиции” с терминами “западной традиции”. Поиск терминологического соответствия этих двух традиций — не самоцель, это лишь один шаг на пути к пониманию сути понятия валидности и очерчиванию конкретных методов ее проверки. Следует отметить, что и советская, и западная традиции разработали весьма эффективные способы оценки качества измерения, и нельзя утверждать, что одна из этих традиций лучше или эффективнее. Но сегодня, когда Украина интегрируется в мировое социологическое пространство, участвует в международных сравнительных исследованиях (например, Европейское социальное исследование (European Social Survey — ESS) или Международной программе социального исследования (International Social Survey Programm — ISSP), более адекватно использование подходов именно западной традиции, поскольку они уже стали стандартными для многих стран и их использование облегчает международную исследовательскую коммуникацию, ведь “без унификации терминологии затруднено взаимопонимание между исследователями” [Паниотто, 1986: с. 7]. Тем более, что данная тенденция уже наметилась. Примером использования терминологии западной традиции оценки качества социологического измерения являются работы Н.Паниной [напр.: Панина, 1996], а также последние статьи в журнале “Социология: теория, методы, маркетинг”: “Субъективная надежность: теория и методы измерения (ИСН)” [Головаха, Горбачик, Любивая и др., 2008] и “Теоретическая валидность измерительной процедуры и смещение данных в социологическом исследовании” [Дембицкий, 2008].

В советской традиции интегральной характеристикой качества социологического измерения и информации¹ как результата этого измерения, считали надежность. Такую точку зрения разделяли В.Ядов, Г.Саганенко, В.Волович, Б.Докторов, В.Паниотто. Б.Докторов в 1979 году и В.Паниотто в 1986-м проанализировали разработки своих коллег по вопросам качества социологического измерения. На их основе можно определить основные показатели надежности социологической информации в советской традиции: обоснованность, правильность, устойчивость.

Обоснованность (валидность) — отсутствие теоретических, методологических, логических ошибок, правильность референций при разработке методики исследования, соответствие числовой модели исследуемой социальной системе [Паниотто, 1986: с. 27]. Б.Докторов называл это также отсутствием квазиошибки согласованности, то есть соответствием модели объекта самому объекту [Докторов, 1979: с. 28].

Правильность — мера отсутствия систематических ошибок измерения [Паниотто, 1986: с. 31].

Устойчивость — мера отсутствия случайных ошибок измерения [Паниотто, 1986: с. 35].

¹ Информацию, полученную в результате качественного измерения, можно считать качественной. Далее вопрос различения понятий качества социологического измерения и качества социологической информации не будет рассматриваться, поскольку социологическая информация является результатом социологического измерения.

В западной традиции нет интегральной характеристики качества социологического измерения; здесь чаще всего выделяют два показателя качества социологического измерения — валидность (validity) и надежность (reliability). Данная позиция отражена во всех Стандартах психологического и образовательного тестирования [Standards, 1966; 1974; 1985; 1999], в классической работе “Оценка надежности и валидности” Дж.Карминеса и А.Зеллера [Carmines, 1979], в таком современном издании, как “Справочник “Сейдж” по методам в социальных исследованиях” (Sage Handbook of social research methods) [Sage Handbook, 2008]. В западной традиции валидность и надежность понимаются так:

Валидность (validity) — это соответствие измерения его цели [Carmines, 1979: р. 12]. Валидность является мерой того, насколько мы действительно измеряем то, что запланировано [Bohrsted, 2000: р. 3207]. Валидность связывают с отсутствием систематических ошибок; измерение, относительно свободное от систематических ошибок, считают валидным [Carmines, 1980: р. 14].

Надежность (reliability) — это воспроизводимость измерения. Надежность касается того, в какой мере эксперимент, тест или любая другая процедура измерения дает те же результаты при повторном измерении. Надежность связана с тенденцией повторяемости в случае неоднократного измерения одних и тех же социальных феноменов одним и тем же измерительным инструментом. Чем устойчивее результаты измерения при повторных попытках, тем надежнее измерительная процедура, и наоборот, чем менее схожими оказываются результаты измерения при повторном измерении, тем менее надежной является процедура измерения [Carmines, 1979: р. 11–12]. Надежность связывают с отсутствием случайных ошибок — измерение, относительно свободное от случайных ошибок, считается надежным [Carmines, 1980: р. 14].

Из приведенного анализа советской и западной традиций видно, что по содержанию показатели качества измерения в них не противоречат друг другу. Основные отличия заключаются в терминологии. Отличие первое — в советской традиции есть интегральная характеристика качества социологической информации — надежность, в западной традиции такой интегральной характеристики нет. Но это не меняет сути понимания качества измерения. В советской традиции показателями надежности как интегральной характеристики качества измерения являются обоснованность, правильность и точность, в западной показателями качества измерения служат валидность и надежность. Исходя из содержания приведенных выше понятий, попытаемся их сопоставить (табл.).

Таблица

Терминологическое сопоставление советской и западной традиций

СОВЕТСКАЯ ТРАДИЦИЯ	ЗАПАДНАЯ ТРАДИЦИЯ
Надежность — интегральная характеристика качества измерения (информации)	Качество информации (quality of measurement)
Обоснованность	Валидность (validity)
Правильность	—
Устойчивость	Надежность (Reliability)

В западной традиции не разграничивают показатели отсутствия теоретических и систематических ошибок, объединяя их в одну категорию — валидность. В советской традиции эти показатели разграничивают как обоснованность и правильность. На мой взгляд, более адекватно разделять эти категории. Но, с другой стороны, как Б.Докторов, так и В.Паниотто отмечают, что четко разграничить обоснованность и правильность информации не всегда удается, поскольку в одних случаях теоретические ошибки совпадают с систематическими, а в других случаях это расхождение выражено достаточно ярко [Паниотто, 1986: с. 31; Докторов, 1979: с. 28]. Как замечает В.Паниотто “отделить обоснованность от правильности — это, в конечном счете, значит определить, где кончается методология и начинается методика... Образно говоря, обоснованность характеризует направление, в котором мы идем, а правильность — траекторию движения в данном направлении” [Паниотто, 1986: с. 32]. Устойчивость в советской традиции по содержанию соответствует надежности в западной традиции.

Как уже отмечалось, ныне, учитывая интегрированность украинской социологии в международный контекст, удобнее пользоваться международной терминологией, но при этом вводить в нее наработки советской традиции. Еще в 1986 году В.Паниотто писал: “Из двух вариантов — использовать как “имя” понятие “надежность” или вообще не присваивать имени качеству информации — более удобным я считаю последний вариант (использовать надежность и обоснованность как равноправные составляющие качества). Но для того, чтобы не возникло расхождений с уже устоявшимся в отечественной социологической литературе словоупотреблением ... в интересах ускорения процесса стандартизации социологической терминологии ... я буду понимать надежность как интегральную, обобщающую характеристику качества социологической информации, как ее “имя”” [Паниотто, 1986: с. 25–26]. Я, из тех же соображений, что и В.Паниотто, предлагаю для согласованности отечественной терминологии с международной понимать валидность и надежность как равнозначимые аспекты качества измерения.

В процессе развития понятия “валидность” в научный оборот были введены разные виды валидности. Однако валидность как показатель меры соответствия измерения тому, что запланировано измерить, является унитарным концептом. Если валидность — это унитарный концепт, корректнее говорить не о разных видах валидности, а о разных подходах к проверке валидности. Анализируя разные виды валидности, описанные в литературе, мы видим, что виды валидности предполагают разные подходы, разные методы проверки одного свойства — валидности.

Итак, следует рассматривать валидность как унитарный концепт, а разные виды валидности, описанные в литературе, — как разные методы проверки валидности.

Определив содержание понятия валидность путем соотнесения терминов советской и западной традиций, показав соотношение валидности с надежностью и выяснив, что валидность является свойством, наличие которого можно проверять разными методами, я завершаю теоретическую интерпретацию понятия валидность и перехожу к эмпирической интерпретации данного понятия.

Эмпирическая интерпретация понятия “валидность”

В рамках эмпирической интерпретации следует показать, к каким этапам измерения нужно применять понятие валидность, конкретнее — какие этапы измерения следует проверять на валидность, а какие этапы выходят за пределы валидности.

Пределы любого понятия можно очертить в том или ином более общем контексте. Таким контекстом для валидности является измерение. За основу возьмем определение измерения Х.Блейлока, на сегодняшний день чаще всего цитируемое в западной литературе. Блейлок понимал измерение как теоретически нагруженный процесс и усматривал принципиальную методологическую проблему в анализе связи между содержательной социологической теорией и открытой применительно к этой теории концепцией измерения. По его определению, “измерение — это процесс соотнесения абстрактных понятий, которые нельзя прямо наблюдать, с эмпирическими индикаторами, которые подлежат прямому наблюдению” [Blalock, 1969(a): p. 6]. Блейлок указывает на невозможность проведения измерения без наличия теоретических рамок и выступает против противопоставления теории и эмпирических исследований [Blalock, 1969(b), p. vi].

Опираясь на это определение, выделю этапы измерения (на примере количественного опроса).

1. Создание теоретической модели¹ изучаемого явления, которая адекватно отражает само изучаемое явление.
2. Отбор эмпирических индикаторов в соответствии с разработанной теоретической моделью.
3. “Перевод” индикаторов в конкретные вопросы и выбор наиболее адекватных шкал.
4. Разработка репрезентативной выборки.
5. Подготовка необходимой для проведения полевого этапа документации (инструкции интервьюерам, маршрутные листы), инструктаж интервьюеров.
6. Проведение полевого этапа (сбор информации), кодировка полученной информации.

Если рассматривать валидность как характеристику социологического измерения или как характеристику социологической информации (результата социологического измерения), то в процесс проверки валидности следует включать проверку правильности проведения всех этапов социологического измерения, обозначенного выше. В таком случае валидность становится интегральной характеристикой качества социологического измерения и социологической информации. Но общепризнано, что качество измерения предполагает и надежность, и репрезентативность выборки, и качественный отбор, и обеспечение работы интервьюеров. А это, в свою очередь,

¹ Поскольку большинство объектов социологического исследования нельзя непосредственно измерять, то исследователи, прежде чем проводить исследования, должны четко описать изучаемый объект — выделить его основные свойства, дифференцировать составные части, если это возможно. Именно такое теоретическое описание, с выделением характерных свойств и составляющих, я имею в виду, говоря о создании теоретической модели изучаемого объекта. Синонимом теоретической модели изучаемого явления можно назвать созданный исследователем теоретический концепт этого явления.

слишком расширяет и “размывает” рамки процедуры проверки валидности конкретного измерения.

Таким образом, исходя из стремления четко очертить “границы” понятия валидности и разработать схему проверки валидности, чтобы не вносить терминологическую путаницу в уже и без того запутанный дискурс социологического измерения, а также учитывая предшествующий опыт, считаю уместным рассматривать валидность как характеристику адекватности измерительного инструментария и корректности применения определенного инструмента в конкретной ситуации исследования.

Согласно этому выводу “границы” проверки измерения на валидность включают проверку следующих аспектов:

- насколько адекватно разработанная теоретическая модель изучаемого явления отражает именно изучаемое явление;
- насколько адекватно подобраны эмпирические индикаторы;
- насколько адекватно эмпирические индикаторы “переведены” в конкретные вопросы и насколько адекватно подобраны шкалы.

То есть валидность — это показатель качества измерения, которое необходимо обеспечить *до проведения самого процесса измерения*, до сбора данных. Валидность должна достигаться на этапах разработки теоретической модели исследуемого явления и инструментария измерения, *а проверяться после завершения процесса измерения* на основе собранных данных.

Если принять такой взгляд на валидность, то валидность, наряду с надежностью, репрезентативностью и качеством проведения полевого этапа является одним из показателей качества социологического измерения, а значит, и качества социологической информации (рис. 1).

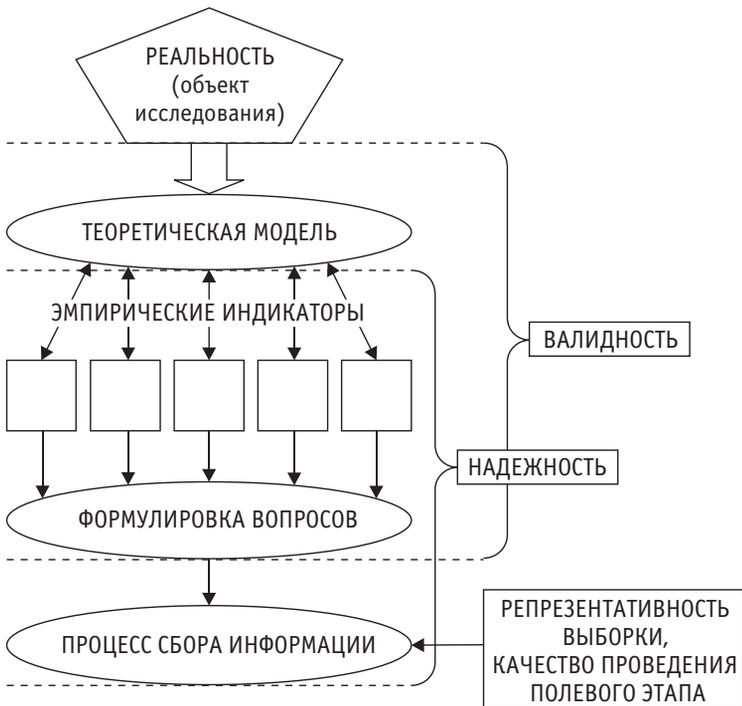


Рис. 1. Этапы процесса измерения

Определив “границы” понятия валидность путем выделения этапов измерения, на которых она должна достигаться, и очертив, что именно должно подлежать валидации, мы завершаем эмпирическую интерпретацию и переходим к операционализации.

Операционализация понятия “валидность”

Как отмечалось, валидность — это показатель свойства измерения — меры того, насколько измерительный инструмент измеряет то, что запланировано измерить. Операционализация понятия валидности — это поиск последовательных операций, которые позволяют установить наличие и интенсивность проявления свойства у носителя (носителем валидности является измерительный инструмент). Фактически речь идет о поиске конкретных методов проверки валидности.

Замечу, что словосочетание “методы проверки валидности” не слишком часто встречается в литературе. Чаще пишут о “видах валидности”. На мой взгляд, корректнее использовать именно выражение “методы проверки валидности”. Несмотря на это, для описания накопленного опыта проверки валидности здесь будет использован термин “виды валидности”, поскольку я буду ссылаться на источники, в которых он использовался.

В западной традиции (см. рис. 2) чаще всего выделяют *три вида* валидности: содержательную, критериальную (конкурентную и прогностическую) и конструктивную (конвергентную и дискриминантную). Некоторые исследователи выделяют также очевидную, или внешнюю, валидность (face validity).

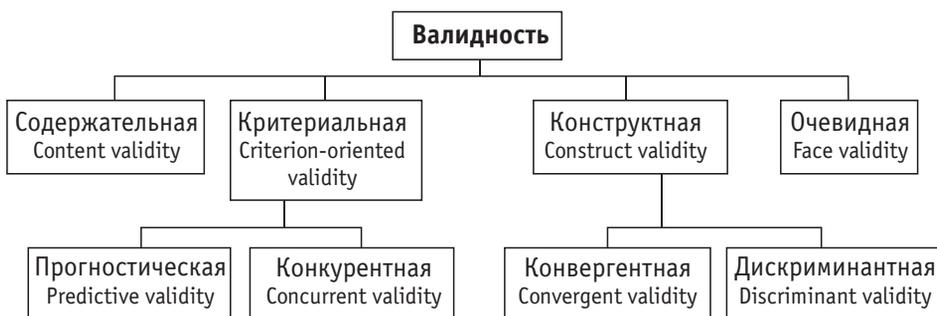


Рис. 2. Виды валидности

В советской традиции либо вообще не различают разные виды обоснованности (валидности), либо выделяют два вида обоснованности (валидности): теоретическую и эмпирическую [Паниотто, 1986: с. 109], понимая под теоретической валидностью конструктивную валидность, а под эмпирической — критериальную валидность. Содержательная валидность выпадает из данной классификации, поскольку ее автор рассматривает ее как касающуюся исключительно тестов, направленных на оценку некоторых знаний и умений [Паниотто, 1986: с. 106]. Поскольку я рассматриваю содержательную валидность шире (см. ниже), считаю этот вид валидности очень важным. Фактически, по содержанию теоретическая и эмпирическая валидности соответствуют конструктивной и критериальной.

Содержательная валидность. Существует много определений содержательной валидности, большинство из которых по содержанию соответствует следующей дефиниции: *содержательная валидность является мерой релевантности и репрезентативности элементов¹ измерительного инструмента измеряемому конструкту* [Haynes, Richard, Kubany, 1995: p. 238]. При этом *релевантность* означает соответствие элементов инструмента измеряемому явлению. Релевантным является инструмент, элементы которого соотношены со всеми аспектами исследуемого явления, следовательно, после анализа собранных данных мы должны получить информацию обо всех сторонах исследуемого явления. Репрезентативность же *измерительного* инструмента — это мера того, насколько пропорционально элементы измерительного инструмента соотносятся с измеряемым явлением, то есть элементы измерительного инструмента должны пропорционально представлять разные аспекты исследуемого явления.

Содержательная валидность может быть разной для разных групп, в частности национальных или культурных. Поэтому ее нужно оценивать конкретно для группы, в которой будет проводиться исследование. Например, измерительный инструмент может быть валидным для американцев и одновременно невалидным для азиатов, поскольку люди разных культур по-разному могут понимать содержание одного и того же вопроса [Marsella, Kameoka, s.a].

Содержательная валидность всегда должна определяться для конкретной сферы исследования, для которой был разработан измерительный инструмент [Haynes, Richard, Kubany, 1995: p. 241]. Разработка измерительных инструментов происходит в рамках имеющихся теорий. Поскольку теоретические аспекты схватывания явления развиваются, релевантность и репрезентативность элементов измерительного инструмента также подлежат пересмотру.

Таким образом, содержательная валидность всегда привязана к конкретному промежутку времени. Динамичная природа изучаемых в социологии явлений обуславливает ряд последствий для содержательной валидности [Haynes, Richard, Kubany, 1995: p. 242]:

1. Показатели содержательной валидности не могут быть стабильными в течение длительного промежутка времени.
2. Содержательную валидность измерительных инструментов нужно периодически пересматривать.
3. Измерительные инструменты должны адекватно отражать изменения в исследуемом явлении.
4. Без пересмотра (прежних) измерительных инструментов можно сделать неправильные выводы об исследуемом явлении.

Многие авторы рекомендовали методы содержательной валидации. Приведу основные из них.

¹ В стандартизированном интервью элементами измерительного инструмента (анкеты) являются вопросы, в контент-анализе — категории анализа, в глубинном интервью — обсуждаемые темы. Иногда для удобства объяснения вместо термина “элементы измерительного инструмента” здесь будет использоваться термин “вопросы измерительного инструмента”, реферирующий к стандартизированному интервью. Однако приведенные объяснения адекватны и для всех других методов сбора информации.

Следует четко определить сферу принадлежности явления и его аспекты, прежде чем переходить к формулировке вопросов измерительного инструмента [Northorp, 1947; Suen, 1990]. Этот шаг является основной и наиболее сложной фазой в разработке валидного по содержанию измерительного инструмента [Murphy, 1994].

Нужно изучить мнение экспертов и респондентов для начальной разработки вопросов (в случае интервью), для чего можно использовать глубинные интервью с экспертами и респондентами. Это способствует тому, что элементы (вопросы) измерительного инструмента будут релевантными и репрезентативными для измеряемого явления. В ходе этого процесса могут быть обнаружены новые аспекты явления, что потребует пересмотра его теоретической модели.

Нужно исследовать пропорциональную репрезентативность элементов измерительного инструмента, представляющих разные аспекты изучаемого явления.

Элементы (вопросы) в измерительном инструменте должны быть распределены, или “взвешены” так, чтобы отражать важность и интенсивность разных аспектов исследуемого явления. Если отдельные вопросы чрезмерно или недостаточно репрезентируют аспекты объекта исследования (то есть определенные аспекты исследуемого явления представлены в измерительном инструменте большим или меньшим, чем нужно, количеством вопросов), то в полученных данных мы будем иметь сдвиг, что приведет к смещениям в выводах.

Итак, подытожим основные методы проверки содержательной валидности [Haynes, Richard, Kubany, 1995; p. 245]:

1. Детальное определение исследуемого явления.
2. Изучение мнения респондентов и экспертов в процессе создания модели исследуемого объекта.
3. Экспертная оценка элементов измерительного инструмента на предмет релевантности и репрезентативности измеряемого явления.

Содержательная валидность обеспечивается на этапе построения теоретической модели объекта исследования и разработки измерительного инструмента. Проверка содержательной валидности означает проверку соответствия созданной теоретической модели исследуемого явления самому явлению и проверку соответствия измерительного инструмента разработанной модели. Достижение содержательной валидности подобно теоретической и эмпирической интерпретации и операционализации основных понятий исследования, а проверка содержательной валидности схожа с выяснением корректности теоретической и эмпирической интерпретации и операционализации. Проверка содержательной валидности уже апробированного ранее инструмента может обнаружить необходимость коррекции этого инструмента.

Критериальная валидность состоит в соответствии информации об изучаемом явлении, полученной с использованием измерительного инструмента, валидность которого проверяется, и информации, полученной другими средствами и из других источников, служащей критерием валидности [Carmines, Zeller, 1980: p. 17]. Этот критерий должен определяться другим валидным относительно того же явления измерительным инструментом, таким как статистика.

Различают две разновидности валидности по критерию — конкурентную, то есть параллельную, и прогностическую.

Конкурентная валидность (concurrent validity) определяется путем сопоставления данных, полученных валидируемым инструментом, с данными, полученными по критерию валидности в тот же период времени [Carmine, Zeller, 1980: p. 18].

Прогностическая валидность (predictive validity) определяется путем изучения корреляции между данными, полученными валидируемым инструментом, и неким критерием, позволяющим определить состояние изучаемого свойства в будущем [Клайн, 1994: с. 27]. Например, прогностическая валидность инструмента, прогнозирующего результаты выборов, определяется путем изучения корреляции результатов измерения с результатами голосования. Валидность теста отбора работников проверяется путем изучения корреляции результатов теста с результатами работы отобранных работников.

Главная сложность в осуществлении критериальной валидации заключается в выборе валидных внешних критериев. К примеру, для инструмента, измеряющего электоральные предпочтения, внешним критерием являются выборы. Для инструмента, измеряющего материальное положение граждан, — данные налоговой службы. Но такие критерии валидности не всегда доступны, а зачастую их просто не существует. Поэтому данный метод проверки валидности не всегда применим для социологических измерений.

В.Аванесов выделял еще один вид валидности — *валидность по известным группам* [Паниотто, 1986: с. 104]. Суть метода заключается в том, что некоторым способом (например, по экспертным оценкам) выбираются две группы респондентов, которые отличаются по измеренному свойству, и проводится опрос этих респондентов с использованием валидируемого инструмента. Если в этих двух группах результаты измерения действительно различаются, то это будет свидетельством в пользу валидности измерительного инструмента. Если нет — то либо измерительный инструмент не является валидным, либо гипотеза о различии групп ошибочна.

В популярной среди американских исследователей книге “Методология исследования” [Survey Methodology, 2004] последний метод проверки тоже упоминается и приводится следующий пример: “Ожидается, что уровень консерватизма американцев, голосующих за республиканцев, будет выше, чем у американцев, голосующих за демократов. В случае получения отрицательного результата можно сделать два вывода — либо теория не соответствует действительности, либо измерительный инструмент не является валидным. Вообще, результаты можно интерпретировать только в рамках хорошо разработанной теории” [Survey Methodology, 2004: p. 256].

Конструктивная валидность — это мера того, насколько измерительный инструмент измеряет конструкт, который должен быть измерен [Nunnally, Richard, Kubany, 1995: p. 239]. Понятие конструкта ввели Л.Кронбах и П.Мел в 1955 году. По их определению “конструкт — это некое характерное свойство человека, которое должно быть отражено в результатах теста. В рамках валидации теста характеристика, по поводу которой делают вывод в процессе интерпретации теста, представляет собой конструкт” [Cronbach, Meehl, 1955: p. 283]. В современной литературе существуют разные определения конструкта и разные соотношения между понятиями “конструкт” и “концепт”. Не буду углубляться в этот вопрос, поскольку это тема

отдельной статьи. Здесь будем понимать понятие конструкта как определенную описанную теоретически характеристику изучаемого явления, которую нужно измерить, поскольку это соответствует содержанию конструктивной валидации.

Конструктивная валидность базируется на теории измеряемого явления. Исследователь должен построить номологическую сеть — набор конструктов и связей между ними — и “вписать” в эту сеть явление, измеряемое инструментом, подлежащим валидации. На основе существующей литературы, результатов других исследований, логики, суждений экспертов и предварительных интервью с респондентами исследователь принимает гипотезы о том, с измерениями каких явлений должны коррелировать измерения исследуемого явления, а с какими не должны коррелировать, а затем тестирует свои гипотезы на собранных данных. Д.Кемпбелл и Д.Фиске разработали стратегию проверки конструктивной валидности. Для этого они ввели два подвида конструктивной валидности — конвергентную и дискриминантную валидности [Campbell, Fiske, 1959]:

- *конвергентная валидность* является оценкой того, насколько измерения, которые должны коррелировать, исходя из теоретических соображений, коррелируют между собой, согласно собранным данным;
- *дискриминантная валидность* является оценкой того, насколько измерения разных конструктов, которые, согласно теории, не должны коррелировать друг с другом, не коррелируют между собой исходя из собранных данных.

Конструктивную валидность любого измерительного инструмента можно оценить, если изучаемый конструкт находит место в теоретическом контексте. Если результаты измерения не расходятся с теоретическими ожиданиями, то измерительный инструмент можно считать конструктивно валидным [Carmines, Zeller, 1980].

Но что делать в случае, если исследователь не обнаруживает ожидаемой связи?

На этот вопрос ответили сами авторы понятия конструктивной валидности Л.Кронбах и П.Мел [Cronbach, Meehl, 1955]. Остальные исследователи поддержали эту стратегию и придерживаются ее поныне [28].

Причины расхождений результатов измерения с теоретическими ожиданиями могут быть следующими [Campbell, Fiske, 1959, p. 295]:

1. Инструменты, которыми измеряются другие конструкты номологической сети, в которую помещен изучаемый конструкт, не являются валидными.
2. Экспериментальный дизайн ошибочен.
3. Валидируемый измерительный инструмент не измеряет нужного нам конструкта.
4. Номологическая сеть (набор конструктов и гипотезы о связях между этими конструктами) ошибочна.

Когда мы не уверены в правильности экспериментального дизайна, мы должны повторить исследование с учетом предыдущих ошибок, чтобы убедиться в правильности нашего дизайна.

В случае сомнений в валидности измерительных инструментов, которые измеряют другие конструкты номологической сети, мы должны, по мере возможности, использовать другие, валидные, измерительные инстру-

менты для этих конструкторов и проверить связь на новых данных. Лучше всего проверять конструктивную валидность с использованием индикаторов, валидность и надежность которых достаточно высока. Таким образом, теоретически релевантные, валидные и надежные внешние индикаторы являются решающим моментом для оценки конструктивной валидности измерительного инструмента.

Если мы уверены в валидности измерительных инструментов, измеряющих конструкты номологической сети, и в экспериментальном дизайне, мы можем сделать вывод о том, что измерительный инструмент неадекватно измеряет конструкт. В таком случае нужно разработать и валидизировать другой тест. Валидизацию нового инструмента следует осуществлять на новых данных.

Можно подвергнуть сомнению номологическую сеть, лежащую в основе измерения конструкта. Тогда исследователь должен разработать новую номологическую сеть. Модифицируя сеть, он должен собрать новые данные для тестирования альтернативной гипотезы.

Пока не существует четкой процедуры, позволяющей определить, какой из приведенных вариантов является основанием отрицательного результата проверки конструктивной валидности измерительного инструмента. Для выбора нужно выяснить, насколько хорошо теоретическая схема подтверждалась предыдущими данными, взвесить доказательства в пользу теоретической схемы и сопоставить с последним неудачным экспериментом.

После проведения процесса валидизации может понадобиться изменение измерительного инструмента для интересующего нас конструкта, или пересмотр теоретической базы конструкта.

Результаты валидизации могут обусловить потребность в очищении определения конструкта либо пересмотре измерительного инструмента или процесса проведения исследования [Standards, 1999].

Очевидная валидность. Некоторые исследователи выделяют также очевидную, или внешнюю валидность (*face validity*). Тест (в случае социологического измерения — измерительный инструмент) является очевидно валидным, если явно измеряет именно то, что имеют в виду респонденты. Очевидная валидность измерительного инструмента помогает добиться сотрудничества с респондентами, будучи связана только с тем, как воспринимают этот измерительный инструмент респонденты [Клайн, 1994: с. 26]. Очевидная валидность, по сути, не имеет отношения к валидности в классическом понимании; здесь не идет речь о методе проверки валидности.

Обобщенная схема проверки валидности

На основании изложенного обзора существующих методов проверки валидности попробуем наметить обобщенную схему проверки валидности конкретного измерительного инструмента.

1. Проверка адекватности модели измерительного явления с использованием результатов известных исследований, экспертных оценок и предварительных интервью с респондентами (в случае опроса). На этом этапе проверки валидности еще не проводится сбор данных, и проверку осуществляют логическим путем. Формализовать такой процесс не-

просто. Этот этап проверки валидности соответствует описанной выше содержательной валидации.

2. Проверка соответствия измерительного инструмента разработанной теоретической модели исследуемого явления. Этот этап осуществляется как с использованием предыдущего опыта и на основе исследования восприятия разработанных индикаторов респондентами (содержательная валидизация), так и на основании собранных данных (критериальная и конструктивная валидизация). Распишем подробнее процесс проверки соответствия измерительного инструмента разработанной теоретической модели:
 - a) разработка номологической сети — набора конструктов и связей между ними;
 - b) проведение эмпирического исследования для проверки гипотез о связях между конструктами номологической сети;
 - c) отбор критериев валидности (если они существуют) для критериальной валидизации инструмента и изучения связей между данными, полученными с использованием валидизируемого по критерию валидности измерительного инструмента;
 - d) выделение групп, которые должны отличаться по исследуемому признаку (если такие группы можно выделить) и проведение исследования в этих двух группах с применением валидизируемого инструмента;
 - e) в случае:
 - i) подтверждения гипотез о связях между конструктами номологической сети,
 - ii) выявления связи между данными, полученными с использованием измерительного инструмента, который валидизируется, и критерием валидности (если таковой существует),
 - iii) получения разных результатов измерения в выделенных группах (если таковые можно выделить)разработанный инструмент можно считать валидным.
 - f) в случае, когда:
 - i) гипотеза о связях между конструктами номологической сети не подтвердилась,
 - ii) связь между измерением и критерием валидности не обнаружена,
 - iii) разные результаты в разных группах не получены,нужно пересматривать разработанный измерительный инструмент и проводить валидизацию новых данных;
 - g) если после проведения повторной валидации пересмотренного измерительного инструмента на новых данных теоретические гипотезы не подтвердились и не обнаружена связь между измерением и критерием валидности, нужно вернуться к разработанной модели измерительного явления и номологической сети, в которую “встроено” это явление, чтобы пересмотреть их, а потом проводить валидизацию измерительного инструмента, скорректированного на основе обновленной модели и обновленной номологической сети на новых данных.

Валидизация измерительных инструментов — это постоянный, коллективный процесс [Cronbach, 1987: р. 163–164]. В нем должны участвовать все члены научного сообщества, занимающиеся измерением определенного явления. Ответственность за валидизацию лежит как на разработчике измерительного инструмента, так и на исследователях, использующих этот инструмент. Разработчик измерительного инструмента отвечает за доказательство валидности измерительного инструмента и за инструкции по правильному его использованию. Пользователи этого инструмента отвечают за оценку доказательств его валидности в конкретной ситуации применения [Standards, 1999]. Валидность имеющихся инструментов следует периодически снова проверять. Такую проверку валидности могут проводить другие исследователи, а не только автор измерительного инструмента. Это может обуславливаться либо значительным промежутком времени с момента разработки определенного инструмента, либо намерением использовать его в другом культурном контексте. Только постоянно подтверждая в разных исследованиях валидность измерительных инструментов, мы можем быть уверены в адекватности наших выводов об изучаемых явлениях.

Литература

- Волович В.И.* Надежность информации в социологическом исследовании: Проблемы методологии и методики. — К., 1974.
- Головаха Е., Горбачик А., Любимая Т., Панина Н., Серета В., Урсулenco К.* Субъективная надежность: теория и методы измерения (ИСН) // Социология: теория, методы, маркетинг. — 2008. — №1. — С. 166–188.
- Девятко И.Ф.* Методы социологического исследования. — М., 2002.
- Дембицкий С.* Теоретическая валидность измерительной процедуры и смещение данных в социологическом исследовании // Социология: теория, методы, маркетинг. — 2008. — № 3. — С. 99–118.
- Докторов Б.З.* О надежности измерения в социологическом исследовании. — Л., 1979.
- Здравомыслов А.Г.* Методология и процедура социологических исследований. — М., 1969.
- Клайн П.* Справочное руководство по конструированию тестов. Введение в психометрическое проектирование. — К., 1994.
- Панина Н.В.* Технологія соціологічного дослідження. — К., 1996.
- Паниотто В.И.* Качество социологической информации (методы оценки и процедуры обеспечения). — К., 1986.
- Саганенко Г.И.* Надежность результатов социологического исследования. — Л., 1983.
- Ядов В.А.* Социологическое исследование: Методология, программа, методы. — М., 1972.
- American Psychologist.* — 1975. — Vol. 30. — P. 955–966.
- Blalock H.M.* The measurement problem // Methodology in Social Research / Ed. by H.M.Blalock, A.Blalock. — N.Y, 1969.
- Blalock H.M.* Theory Construction: From verbal to Mathematical Formulation. Englewood Cliffs, N.J., 1969.
- Bohrsted G.W.* Validity // Encyclopedia of Sociology, Second Edition / 2nd ed. — Vol. 5. — N.Y., 2000.
- Campbell D.T., Fiske D.W.* Convergent and discriminant validation by the multitrait-multimethod matrix // Psychological Bulletin. — 1959. — № 58. — P.81–105.
- Carmines E.G., Zeller R.A.* Measurement in the social sciences. — Cambridge, 1980.

- Carmines E.G., Zeller R.A.* Reliability and Validity Assessment. — Beverley Hills, 1979.
- Cronbach L.J., Meehl P.E.* Construct Validity in Psychological Tests // Psychological Bulletin. — 1955. — № 52. — P. 281–302.
- Cronbach L.J., Meehl P.E.* Construct Validity in Psychological Tests // Psychological Bulletin. — 1955. — № 52. — P. 281–302.
- Cronbach L.J.* Construct validity after thirty years // Linn R.E. (ed.), Intelligence: Measurement, Theory, and Public Policy. — Urbana, 1987.
- Cureton E.* Validity // Educational Measurement. — Washington, D.C, 1951. — Ch. 16. — P. 621–695.
- Guilford P.* Psychometric Methods. — S.I., 1936.
- Haynes S., Richard D., Kubany E.* Content Validity in Psychological Assessment: A Functional Approach to Concepts and Methods // Psychological Assessment. — 1995. — № 3. — P. 238–247.
- Marsella A.J., Kameoka V.* Ethnocultural issues in the assessment of psychopathology // Wetzler S. (ed.). Measurement mental illness: Psychometric assessment for clinicians. — Washington, DC, 1995. — P. 231–256.
- Messick S.* The standard problem: Meaning and values in measurement and evaluation // American Psychologist. — 1975. — October. — P. 955–966.
- Myrphy K.R., Davidshofer C.O.* Psychological testing: Principles and applications (3rd ed.). — Englewood Cliffs, N.J., 1994.
- Northrop F.S.C.* The Logic of the Sciences and the Humanities. — N.Y., 1947.
- Sage Handbook of social research methods. — S.I., 2008
- Technical recommendations for psychological tests and diagnostic techniques // Psychological Bulletin [Supplement]. — 1954. — № 51 (2, part 2).
- Suen H.K.* Principles of test theories. — Hillsdale, N.J., 1990.
- Survey Methodology / Series W., Groves R.M., Fowler F.G., Jr, Couper M.P., Lepkowski J.M., Singer E., Tourangeau R. — S.I., 2004.
- Standards for educational and psychological tests and manuals / American Psychological Association. — Washington, DC, 1966.
- Standards for educational and psychological tests and manuals / American Psychological Association. — Washington, DC, 1974.
- Standards for educational and psychological testing / American Educational Research Association. — Washington, DC, 1985.
- Standards for educational and psychological testing / American Educational Research Association, American Psychological Association and National Council on Measurement in Education. — Washington, DC, 1999.
- Thurstone L.L.* The Reliability and Validity of the Test. — Ann Arbor, MI, 1931.