

УДК 303.21

СЕРГЕЙ ДЕМБИЦКИЙ,

кандидат социологических наук, научный сотрудник отдела методологии и методов социологии Института социологии НАН Украины

Относительно проблематичности статистической проверки теоретических конструкторов

Аннотация

В статье описаны результаты методических экспериментов, посвященных проблеме статистической проверки теоретических конструкторов. Продемонстрировано, что удовлетворительные результаты статистической проверки не являются достаточным условием валидации шкал. В качестве примера используется комплексная шкала, позволяющая определить тип студента в зависимости от усилий, прилагаемых им в процессе обучения (безразличный, декларативно заинтересованный, ответственный).

Ключевые слова: *шкалирование, теоретическая валидация, университетское образование*

Само по себе подтверждение гипотез мало о чем говорит, если опирается исключительно на кабинетные размышления и результаты измерения (пусть его и принято называть в социологии наблюдением). Даже использование “ломаных” теоретических конструкторов может привести к получению статистически значимых результатов, а также демонстрации высоких коэффициентов корреляции или значимых различий (соответствующий пример приведен ниже). Поэтому измерение, “замкнутое само на себя”, подходит для решения многих задач, но лучше всего — для получения наукообразных артефактов.

Я уже не говорю о повсеместно распространенных в социологии ситуациях, когда статистически значимые результаты, полученные на больших выборках (скажем, 1500 и более единиц), расцениваются в качестве ценных

с научной точки зрения (или подтверждающих что-либо) лишь по факту своей статистической значимости. А на больших выборках, как известно, практически все статистически значимо.

Такие размышления выдвигают на передний план проблему теоретической валидности, достижение которой призвано обеспечить исследователя валидными в отношении изучаемых социальных феноменов теоретическими конструктами. Следовательно, теоретическая валидизация является во многом автономной и определяющей для дальнейшей операционализации и эмпирической интерпретации сферой. Легко представить ситуацию, когда теоретический конструкт описан достаточно исчерпывающе, однако разработанная методика не обеспечивает удовлетворительного отображения данного конструкта. В соответствии с “лучшими” традициями количественного подхода в такой ситуации можно смело отказаться от теоретического конструкта, не прошедшего проверку. И противоположный случай — хотя методика позволяет вполне адекватно отобразить конструкт, но этот конструкт содержательно не валиден (например, охватывает только часть концептуального поля, которое находится в фокусе исследования). В данном случае будут приняты опять-таки ошибочные выводы — в этот раз уже о правильности невалидного конструкта.

Для более подробного описания последней ситуации обратимся к следующему примеру. В одной из своих недавних публикаций [Дембицкий, 2012] я описал специфику теоретической валидизации измерительных шкал. В качестве демонстрационного примера была взята шкала, предназначенная для определения типа студента, указывающего на его отношение к учебе (безразличный, декларативно заинтересованный или ответственный) на основании используемых студентом поведенческих стратегий.

Эта шкала имеет следующий вид:

Давайте предположим, что в самом общем виде студентов можно разделить на три типа, в зависимости от их отношения к учебе:

- А) наименее прилежные студенты, в основном не прилагающие усилий для учебы, надеющиеся на удачное стечение обстоятельств и на то, что все долги можно будет закрыть в конце семестра, а экзамены и зачеты сдать благодаря помощи одногруппников или даже практически ничего не делая;*
- В) балансирующие между подготовкой наиболее важных заданий и игнорированием заданий, не имеющих принципиального значения, использующие ранжирование преподавателей по строгости и требовательности с целью соответствующей концентрации усилий и предпочитающие посещать те пары, которые имеют больший вес в рамках дисциплины;*
- С) наиболее старательные, которые относятся к учебе максимально ответственно, стараются готовить все задания, предусмотренные учебными требованиями, получить максимум знаний и постоянно посещают пары.*

А теперь, используя приведенную ниже шкалу, УКАЖИТЕ МЕСТО, НА КОТОРОЕ ВЫ БЫ ПОСТАВИЛИ ПЕРЕЧИСЛЕННЫХ НИЖЕ СТУДЕНТОВ, сравнивая с этими тремя типами. При этом нельзя выбирать “чистый” тип (А, В или С), а необходимо выбрать ОДНО из значений между двумя “чистыми” типами (между А и В ИЛИ между В и С).

A	1	2	3	4	5	B	6	7	8	9	10	C
	A-90% B-10%	A-70% B-30%	A-50% B-50%	A-30% B-70%	A-10% B-90%		B-90% C-10%	B-70% C-30%	B-50% C-50%	B-30% C-70%	B-10% C-90%	

В ходе валидизации этой шкалы был проведен ряд опросов, в том числе и с помощью экспертов, в качестве которых выступили преподаватели и сами студенты (в этом случае они оценивали не только себя, но и своих одногруппников). Опросы в трех вузах, в рамках которых привлекались эксперты ($N_1 = 20$, $N_2 = 24$, $N_3 = 23$), я и использую для того, чтобы показать, как “ломаный” конструкт может привести к хорошим статистическим результатам.

В первом и втором опросе каждый из экспертов (два преподавателя и один студент) оценили каждого из студентов одной учебной группы по 10-балльной шкале. Не вдаваясь в подробности, отмечу лишь то, что значения от 1 до 5 соответствовали безразличному типу, от 6 до 7 — декларативно заинтересованному, от 8 до 10 — ответственному. Далее были рассчитаны: а) обобщенная оценка на основании среднего арифметического оценок всех экспертов; б) коэффициент корреляции между обобщенной оценкой каждого студента и его средним баллом успешности сдачи сессий за четыре года обучения (см. табл. 1).

Таблица 1

Сила и направление связи результатов экспертного оценивания со средними значениями оценок, полученных на всех сессиях (первый и второй вуз)

Эксперты	Сила связи	
	Первый вуз	Второй вуз
Первый преподаватель	0,83**	0,93**
Второй преподаватель	0,72**	0,91**
Студент	0,80**	0,93**
Среднее значение трех оценок	0,85**	0,95**

** $p < 0,01$.

В лучших традициях количественного подхода я бы мог решить, что я получил подтверждение (по крайней мере частичное) правильности применяемого теоретического конструкта. Но я и так знал, что он правильный, ведь я использовал валидный теоретический конструкт (валидность конструкта была достигнута на предыдущих этапах исследования посредством качественных методов).

Давайте теперь посмотрим, что произойдет, если из содержания используемого теоретического конструкта убрать значительную часть, а именно информацию — о безразличном типе (часть шкалы от 1 до 5). В этом случае каждому студенту, который получил обобщенную оценку от 1 до 5,499 (таких было восемь человек, или 36% от общего числа), необходимо присвоить 6, а остальные оценки оставить без изменений. Насколько сильно изменится коэффициент корреляции? После преобразования шкалы он стал равен 0,70 (вероятность ошибки по-прежнему не превышала 0,01%).

Вместе с тем это всего лишь мысленный эксперимент. Поэтому я провел дополнительный методический эксперимент в учебной группе другого вуза. В качестве экспертов выступили два преподавателя, хорошо знакомые с группой, и двое старост (каждый из старост давал оценку только по студентам своей подгруппы). На первом этапе они оценили всех студентов с помощью шкалы, которая была специально искажена. В ней отсутствовала вторая часть (от 6 до 10), то есть оценка осуществлялась на основании только двух типов студентов — безразличного и декларативно заинтересованного, представленных пунктами шкалы от 1 до 5 (см. приложение). Лишь один из экспертов (староста) поставил валидность шкалы под сомнение, но после объяснения, что это методический эксперимент и так надо, заполнил анкету. Были получены следующие результаты: индивидуальные и усредненные результаты оценок экспертов показали высокую взаимосвязь со средним баллом успешности сдачи сессий за четыре года обучения. Через месяц я попросил этих экспертов повторить оценку, но уже с использованием полноценной 10-балльной шкалы. В результате индивидуальные коэффициенты в первом случае увеличились, во втором — несколько уменьшились, в третьем — остались практически неизменными. Что касается обобщенной оценки, то она также практически не изменилась (см. табл. 2).

Таблица 2

Сила и направление связи результатов экспертного оценивания со средними значениями оценок, полученных на всех сессиях (третий вуз)

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**

* $p < 0,05$;

** $p < 0,01$.

Отталкиваясь исключительно от статистической информации, можно сделать вывод, что валидность двух вариантов шкалы (по крайней мере, при использовании в отношении этой конкретной группы студентов) является одинаковой. Вместе с тем, очень вероятно, что в случае использования неполноценной шкалы, эксперты ориентировались в первую очередь на числовые пункты, а не на содержание двух описанных в вопросе типов. По моему мнению, такая ситуация может иметь место в том случае, когда эксперты хорошо знакомы с изучаемой проблемой. Соответственно, неучтенная в тексте вопроса информация все равно учитывается ими в силу их знаний и опыта. Таким образом, статистическая валидность шкалы компенсируется посредством когнитивных возможностей экспертов. А вот содержательная (а более широко — теоретическая) валидность шкалы является сомнительной, поскольку она не позволяет перейти к трем типам студентов. В свою

очередь, полноценный вариант шкалы является валидным во всех отношениях — и в статистическом, и в теоретическом.

На этом примере я хочу показать лишь одно — искаженные конструкты вполне могут получить “подтверждение” на основании результатов измерения с дальнейшей статистической проверкой. И эта проблема, на мой взгляд, не может быть решена в рамках количественного подхода.

Первоочередное значение в разработке валидных теоретических конструктов должно занимать изучение окружающей социальной действительности посредством тщательного наблюдения происходящих в ней процессов, а также личного взаимодействия с участниками таких процессов.

Вместе с тем эмпирические свидетельства, представленные в статье, являются достаточно фрагментарными и, таким образом, скорее очерчивают методологическую проблему, чем доказывают положения, выдвинутые в статье. Полноценное же решение проблемы требует проведения дальнейших методических экспериментов.

ПРИЛОЖЕНИЕ

Предположим, что в самом общем виде студентов можно разделить на два типа в зависимости от их отношения к учебе:

- А) балансирующие между подготовкой наиболее важных заданий и игнорированием заданий, не имеющих принципиального значения, использующие ранжирование преподавателей по строгости и требовательности с целью соответствующей концентрации усилий и предпочитающие посещать те пары, которые имеют больший вес в рамках дисциплины;*
- В) наиболее старательные, которые относятся к учебе максимально ответственно, стараются готовить все задания, предусмотренные учебными требованиями, получить максимум знаний и постоянно посещать пары.*

А теперь, используя приведенную ниже шкалу, УКАЖИТЕ МЕСТО, НА КОТОРОЕ ВЫ БЫ ПОСТАВИЛИ ПЕРЕЧИСЛЕННЫХ НИЖЕ СТУДЕНТОВ, сравнивая с этими двумя типами. При этом нельзя выбирать “чистый” тип (А или В), а необходимо выбрать ОДНО из значений между двумя “чистыми” типами.

А	1	2	3	4	5	В
	A-90% B-10%	A-70% B-30%	A-50% B-50%	A-30% B-70%	A-10% B-90%	

Источники

Дембицкий С. Теоретическая валидизация измерительных шкал / Сергей Дембицкий // Социология: теория, методы, маркетинг. — 2012. — №2. — С. 53–65.