

ПРОБЛЕМА НАУЧНОГО ИЗМЕРЕНИЯ В ПСИХОДИАГНОСТИКЕ

Ассанович М.А.

УО «Гродненский государственный медицинский университет», Гродно, Беларусь

Классическая теория тестов не содержит положений и разработок методологии измерения латентных психологических конструктов. Психологические тесты нельзя рассматривать как психодиагностические измерительные инструменты. Необходимо применение новых вероятностных моделей конструирования психодиагностических методик с построением равноинтервальной измерительной шкалы.

Ключевые слова: психодиагностика, психометрия, классическая теория тестов, измерение, аддитивность

Психодиагностика является областью научных знаний о применении методов и методик оценки психологических конструктов или свойств. К психологическим конструктам (свойствам) относятся различные аспекты психологических проявлений человека, такие как интеллектуальные способности, когнитивные функции, личностные особенности. Оценка психологических конструктов у индивида подразумевает описание, с одной стороны, только ему присущих особенностей этих конструктов, а с другой стороны, сопоставление полученных данных с состоянием оцениваемого психологического свойства у других людей.

В медицине психодиагностика направлена на оценку клинически значимых психологических конструктов. Например, такими конструктами являются депрессия, тревога, когнитивные функции, межличностные отношения и т.д. В клинике значение имеет не только констатация наличия у пациента каких-либо индивидуальных психологических особенностей, но особую значимость приобретает оценка выраженности клинически важного психологического конструкта. Определение выраженности клинико-психологического конструкта позволяет определить степень имеющихся нарушений, ресурсов, сделать прогноз, провести планирование лечения.

Большинство психодиагностических методик содержат числовые шкалы, которые используются для определения уровня оцениваемого психологического свойства. Оценка выраженности всегда включает в себя измерительный процесс [2, 8, 9]. К сожалению, в плане измерения используемые в клинике психодиагностические методики имеют существенные проблемы. Эти проблемы во многом связаны с тем, что подавляющее большинство методик создано на основе так называемой классической теории тестов. Прежде чем перейти к рассмотрению данных проблем, следует затронуть базовые понятия измерения с научной точки зрения.

Основные положения измерения в науке были заложены много веков назад [10]. Становление концепции измерения и ее основных положений было связано с разработкой мер расстояния, веса, температуры и других физических величин. Применительно к психодиагностическим измерениям интерес представляет история формирования измерения температуры [10]. Это обусловлено тем, что температура, как и психологические конструкты, является латентной (ненаблюдаемой) переменной. Оценки тепла и холода носят выраженный индивидуальный характер. Интенсивность ощущений тепла может быть вербализована, однако ее трудно проверить прямым образом. До 16 века не существовало метода согласования различий в оценках температуры. Отсутствовала теория, объясняющая природу температуры как измеряемой

переменной. Существовали догадки того, что температура имеет отношение к каким-то физическим процессам, но не более того [11].

В конце 16 века Галилей представил эксперимент, показывающий возможность наглядного измерения температуры. Аппарат для измерения температуры представлял собой стеклянную трубку, один конец которой был открыт, а второй вставлен в колбу. Открытый конец трубки был погружен в сосуд со спиртовой жидкостью (вероятно вином) и жидкость могла частично затекать в трубку. При нагревании колбы объем жидкости в трубке уменьшался, а в процессе охлаждения – увеличивался. Теперь известно, что нагревание воздуха изменяет его давление и это влечет за собой изменение уровня жидкости в трубке. Галилей назвал свой аппарат «термоскопом», поскольку он позволял визуально оценивать изменения температуры посредством изменения уровня жидкости. Идея опосредованного измерения температуры, воплощенная Галилеем в термоскопе, легла в основу изобретения термометра. Разработка термометра представляет собой пример создания надежной и точной интервальной шкалы измерения. В плане надежности, валидности и точности измерения температуры важен факт обоснования эквивалентности между изменением температуры и изменением столбика ртути (или спирта). В основе обоснования лежит теоретическое описание конструкта температуры в терминах термокинетической энергии вещества. Изменение термокинетической энергии и обуславливает изменение столбика ртути при повышении температуры [9, 10].

Современные математические основы концепции измерения были заложены немецким математиком О. Хелдером (1901 г.). Он указывал, что измерение возможно только в том случае, если объект измерения является количественной величиной. Это означает, что данная величина имеет размеры и выраженность. Различные уровни выраженности величины состоят в отношениях друг с другом. Эти отношения могут быть выражены в цифрах, т.е. измерены [10, 11].

В современном виде научное определение измерения выглядит следующим образом. Измерение – это когнитивный процесс установления отношения выражаемой величины к единице измерения этой величины [8, 9, 10, 11, 12]. В данном определении можно выделить три важных момента. Измерение всегда предоставляет новую информацию, т.е. является познавательным процессом. Это означает, что до проведения измерения мы ничего не знаем о величине и уровнях выраженности количественной величины. В результате измерительного процесса мы получаем совершенно новую информацию о сущности данной величины. Вторым моментом заключается в том, что измерение всегда связано с расчетными

операциями, поскольку суть измерения состоит в установлении отношения. Третий аспект измерения включает обязательное наличие единицы измерения. Без этого аспекта научное измерение невозможно по своей сути. Во многих технических и коммерческих сферах жизни существуют юридически узаконенные единицы измерения. История становления ряда измерительных систем во многом связана с длительным поиском и формированием единиц измерения. Единица измерения – это фиксированная величина, которой обоснованно присвоено значение единицы. При этом значение, равное 1, эквивалентно определенному фиксированному количеству измеряемой величины [11,12]. Во многих случаях единицы измерения носят абстрактный характер, по сути, не имея никакого генетического отношения к измеряемой величине. Однако единица измерения должна иметь строгую и научно доказанную эквивалентность фиксированному количеству измеряемой величины.

Научное измерение возможно только при соблюдении определенных принципов и условий: аддитивности и конкатенация, объективность и независимость, воспроизводимость, изоморфизм, наличие равноинтервальной шкалы и единиц измерения [8, 9, 10, 11, 12].

Для понимания существующих проблем измерений в психодиагностике рассмотрим эти принципы подробнее.

Аддитивность означает, что состав величины полностью образуется за счет складывающихся ее частей. Сложение частей величины называется конкатенацией. Важно понимать, что с прибавлением каждой новой части количество или выраженность величины увеличивается на размер прибавляемой части. Величина ABC складывается как $A+B+C$. Если к расстоянию 4 метра прибавить расстояние 6 метров, то получится расстояние 10 метров. Расстояние является аддитивной величиной, доступной измерению. Аддитивные величины легче поддаются измерению по сравнению с неаддитивными. Для измерения неаддитивных переменных необходимо создание абстрактной аддитивной структуры, которая находилась бы в эквивалентных отношениях с неаддитивной измеряемой величиной. Например, температура не является аддитивной величиной. К тому же, как указывалось выше, она недоступна для прямого измерения. Говоря современным научным языком, температура является неаддитивной латентной переменной. Однако в процессе поиска возможности ее измерения было найдено, что температура может измеряться посредством оценки высоты столбика ртути или спирта. При этом высота столбика представляет собой уже аддитивную наблюдаемую переменную. Таким образом, через построение эквивалентных отношений становится возможным измерение неаддитивных переменных [10, 11, 12].

Следующими принципами научного измерения являются объективность и независимость измерения. Данные принципы означают, что на процесс измерения не должны оказывать влияние проводящий измерение, измерительный инструмент, а также объект, характеристика которого измеряется. Из этих принципов логически вытекают следующие два: принципы воспроизводимости и инвариантности измерения. Принцип воспроизводимости измерения состоит в том, что результаты измерения должны повторяться в последующих измерениях другими исследователями, с помощью других подобных измерительных инструментов и у других объектов, обладающих таким же

уровнем измеряемой величины. Например, если рост человека составляет 170 см, то это значение величины роста должно воспроизводиться любым исследователем, с помощью любой стандартной линейки. Кроме того, у всех людей такого же роста при измерении должен получаться один и тот же результат [10, 11, 12].

Принцип изоморфизма соответствует правилам эквивалентности отношений уровней выраженности величины и цифровых индикаторов этих уровней. Прежде всего, принцип изоморфизма касается эквивалентности распределений уровней выраженности величины и индикаторов измерительной шкалы. Если распределения не эквивалентны, то цифровые индикаторы не будут правильно отражать уровни выраженности измеряемой величины. В этих случаях требуются специальные методы установления правил эквивалентности на основе вероятностных моделей [10, 11, 12].

Последние два принципа научного измерения нами уже затрагивались выше. Измерение возможно только на равноинтервальной шкале. Равноинтервальная шкала представляет собой измерительную шкалу, в которой расстояние от предыдущей метки до данной метки равно расстоянию от данной метки до последующей метки. Именно такая шкала позволяет выполнять математические операции над полученными значениями выраженности измеренной величины. Результаты измерения представляются в единицах измерения [10, 11, 12].

Далее проведем анализ возможности и соблюдения принципов научного измерения в психометрике и психодиагностике.

Еще в 1860 году известный немецкий психофизик Г. Фехнер отмечал невозможность измерения психологических конструктов, поскольку они не являются количественными величинами [10]. Однако в начале 20 века британский психолог и математик Ч. Спирмен указал на необходимость и возможность измерения в психометрии [13]. В качестве довода он привел не научное обоснование измерения психологических свойств, а практическую целесообразность и статус психологии как науки. Спирмен выдвинул так называемый «числовой императив». Его логика заключалась в том, что во всех науках проводятся измерения. Психология является наукой, значит, в ней также возможны измерения. Далее он обратил внимание на прикладное значение использования психометрических тестов. По Спирмену, если психологические тесты применяются для принятия решений на практике, то это позволяет трактовать полученные результаты как измерения даже при отсутствии глубокого научного обоснования. В качестве теоретической основы психометрических измерений Спирмен выдвинул теорию истинной оценки и случайной ошибки, которая впоследствии получила название «классическая теория тестов» (КТТ) [6, 13].

КТТ содержит значительное количество теоретических допущений, которые используются в математических выводах [3, 4]. Основной постулат КТТ заключается в том, что наблюдаемая оценка по тесту состоит из двух компонентов: истинной оценки и случайной ошибки. Отсюда следует, что дисперсия наблюдаемой оценки складывается из дисперсии истинной оценки и дисперсии ошибки [1, 3, 5]. Истинная оценка по тесту является сугубо теоретическим конструктом, который не может быть реально наблюдаемым. Для получения истинной оценки необходимо было бы проводить исследова-

ние у одних и тех же испытуемых одним и тем же тестом бесконечное количество раз. Тогда истинная оценка равнялась бы среднему значению бесконечного количества тестирований. Так как истинная оценка по тесту является недостижимой величиной, то основная цель классического подхода в психометрии состоит в приближении наблюдаемой оценки к теоретической истинной оценке. Это может быть достигнуто посредством максимизации дисперсии наблюдаемой оценки и уменьшением дисперсии случайной ошибки измерения. Показателем способности приближения наблюдаемой оценки к истинной является надежность. В психометрическом смысле надежность теста представляет собой отношение дисперсии истинной оценки к дисперсии наблюдаемой оценки. Дисперсия истинной оценки всегда ниже дисперсии наблюдаемой оценки из-за доли дисперсии ошибки. Таким образом, надежность, по сути, эквивалентна доле истинной дисперсии, объясняемой за счет дисперсии наблюдаемой оценки [1, 3, 4, 5, 6].

Главным недостатком КТТ является допущение о том, что в результате достижения адекватной надежности теста может быть осуществлено измерение психологического свойства. При этом понятие «тестирования» синонимично понятию «измерение». Однако в классической теории тестов отсутствует определение связи измеряемого психологического свойства с полученными в результате тестирования баллами [2, 7, 9, 10, 11]. Проблема измерения латентных переменных, к разряду которых относятся психологические конструкты, полностью проигнорирована. Для более глубокого понимания остановимся подробнее на этой проблеме.

Между измерением наблюдаемых переменных и психологических конструктов существует огромная разница. Наблюдаемые свойства в большинстве своем подлежат прямому измерению. Психологические конструкты прямо измерить невозможно. В случае прямого и наглядного измерения можно установить и тут же проверить пропорциональность связи между измеренной величиной свойства и самим свойством. Наблюдаемые переменные, так или иначе, в процессе измерения могут быть разделены на единицы и подсчитаны. Такой подход совершенно невозможен к измерению психологических конструктов. Основная причина состоит в том, что психологические конструкты относятся к латентным переменным, которые не могут быть непосредственно разделены на какие-либо единицы и подсчитаны числом этих единиц. Кроме того, крайне затруднительно найти четкое обоснование эквивалентности количества единиц (баллов), подсчитанных в результате психодиагностического исследования, оцениваемому уровню выраженности свойства [9, 10, 11, 12]. Данные проблемы уже встречались в истории, в частности, при измерении температуры, как описано нами выше. Исходя из того, что латентный психологический конструкт в ходе тестирования остается «за кадром», основное внимание в КТТ уделяется расчетам надежности.

В 1930 г. по заказу комитета Фергюссона известный британский физик Н. Кэмпбелл провел исследование, в котором показал невозможность научного измерения психологических конструктов. В качестве основной причины он отметил несоблюдение условия конкатенации [10]. Однако в 1946 г. американский психофизик С.С. Стивенс, известный открытием им степенным законом зависимости интенсивности ощущения от силы стимула, выдвинул свое опреде-

ление измерения, которое стало очень популярным среди психометристов. По определению Стивенса, измерение - «это присвоение чисел объектам или явлениям в соответствии с правилом, - любым правилом» [14]. Популярность данного определения в психодиагностике объясняется очень просто. Оно не имеет никаких спецификаций относительно конкретных правил научного измерения. Это открывает дорогу для различных допущений и умозрительных положений относительно измерения латентных переменных в психометрике и психодиагностике. Необходимо ввести какое-нибудь правило, и измерение будет иметь место. В КТТ таким правилом является получение общей суммарной тестовой оценки при достаточном уровне надежности [1, 3, 5, 6]. К сожалению, такой взгляд на процесс измерения психологических конструктов имеет негативные последствия для получаемых в ходе психодиагностического исследования результатов. Дело в том, что получаемые в рамках КТТ психодиагностические оценки не являются мерами психологического конструкта, но интерпретируются, как будто они таковыми являются [9, 10, 11, 12]. Полученный суммарный тестовый балл может иметь отношение к выраженности конструкта, но чтобы трактовать его как меру этого конструкта, во-первых, необходимо установить правила эквивалентности суммарного балла и уровня выраженности, а, во-вторых, представить полученные меры на равноинтервальной шкале. Тестовые баллы не могут быть представлены на равноинтервальной шкале по той простой причине, что они являются ранговыми оценками. Ранговые оценки могут быть локализованы только на порядковой шкале. Как известно, классические математические операции не могут производиться с данными на порядковой шкале [2, 6, 9, 10, 11, 12].

В рамках КТТ сырые тестовые оценки подвергаются дальнейшему преобразованию, которое называется нормализацией. Цель нормализации состоит в приведении тестовых баллов к шкале с нормально распределенными оценками. Многие специалисты считают, что таким образом создается измерительная шкала равных интервалов. Однако совершенно ясно, и это отмечено во многих современных психометрических исследованиях, что нормализованная шкала является искусственной и уж никак не содержит информацию о реальном распределении уровней выраженности психологического конструкта в популяции людей. Таким образом, в классической теории теста произошла подмена понятия «измерение». Вместо установления эксплицитных правил эквивалентности отношений «уровень латентного конструкта - тестовая оценка» измерением стало считаться простое получение суммарного тестового балла с последующим искусственным приведением к нормальному распределению. При этом проигнорированы все современные принципы научного измерения [9, 10, 11, 12]. Рассмотрим последствия такого подхода.

Главным негативным последствием отсутствия научно обоснованного измерения в классической психометрике явилась выраженная нестабильность психодиагностических оценок. Результаты часто не воспроизводимы на других выборках испытуемых. Надежность теста не может считаться устоявшейся характеристикой теста. Скорее, она характеризует психодиагностическое исследование, в котором применяется данный тест. Это означает необходимость расчета надежности в каждом психодиагностическом исследовании. Нестабильность психодиагностиче-

ских оценок обусловлена несоблюдением принципов независимости, объективности, инвариантности измерения. Тестовые оценки зависят от распределения конструкта в выборке испытуемых. В выборке испытуемых, которые имеют высокие уровни конструкта, будут получены заниженные оценки. В выборке с низкими уровнями психологического свойства полученные оценки будут завышенными. Эта зависимость особенно сказывается при создании так называемых нормативов. Нормативы – это тестовые оценки, которые приняты как эталонные. Считается, что они отражают истинное распределение конструкта в популяции. На деле это далеко не так, поскольку классическая теория тестов не занимается оценкой истинного распределения конструкта в популяции. Нормативные выборки зачастую нерепрезентативны, полученные нормативные оценки быстро устаревают, а главное, не соблюдается изоморфизм распределений конструкта и тестовых баллов [2, 4, 9, 10, 11, 12].

Другая проблемная ситуация связана с трудностью тестовых заданий. Трудность задания теста обуславливает тот уровень психологического конструкта, который требуется для выполнения данного задания. Проблема научного измерения приводит к зависимости полученных тестовых оценок не только от распределения свойства в нормативной выборке, но и от трудности заданий теста. Выполняя тесты, включающие преимущественно легкие задания, испытуемые будут получать завышенные оценки. Выполняя трудные задания теста, испытуемые получат заниженные показатели [9, 10, 11, 12].

Отметим еще раз, что нормативные данные теста никоим образом не отражают истинное распределение уровней психологического конструкта в популяции. Линейные преобразования тестовых оценок в КТТ не приводят к построению равноинтервальной шкалы. По сути, получается эквивалентная порядковая шкала. Сравнивая тестовый балл данного конкретного испытуемого с нормативной оценкой, мы не определяем уровень выраженности конструкта. Мы всего лишь сравниваем положение нашего испытуемого в ряду других испытуемых нормативной выборки. Во многих случаях

нормативные выборки нерепрезентативны по социально-демографическим и временным характеристикам. Это еще больше усугубляет проблемность психодиагностического исследования, выполненного в рамках классической теории тестов [9, 10, 11, 12].

Отягчающим обстоятельством для психодиагностики в Республике Беларусь является то, что большинство психологических тестов содержат нормативные данные, полученные в России или СССР. Такие нормативные данные не годятся для использования в РБ ни с научной, ни с этической точки зрения.

Таким образом, подытоживая материал настоящей статьи, необходимо сделать следующие выводы.

Практически все известные на сегодняшний день русскоязычные психодиагностические тестовые методики выполнены в рамках классической теории тестов. Психологические тесты, выполненные на основе КТТ, лишены главного диагностического компонента – научного измерения психологического конструкта. Отсюда следует, что психологические тесты в рамках КТТ нельзя называть измерительными инструментами. Они не способны точно и научно обоснованно оценивать уровень выраженности психологических конструктов, для оценки которых они изначально предназначены. В данных методиках не соблюдены принципы построения научного измерения латентных переменных, к которым относятся психологические конструкты. Вследствие этого получаемые в психодиагностических исследованиях результаты отличаются высокой степенью нестабильности, зависят от трудности применяемых тестов и особенностей нормативных выборок. Никакие ссылки или приведенные в литературе сведения об адаптации используемой психодиагностической методики за пределами РБ не могут являться научным и этическим основанием для применения данной методики в Республике Беларусь. Учитывая большое количество психодиагностических тестов с непроверенными и недоказанными измерительными свойствами, в РБ остро назрела необходимость использования научно обоснованных психометрических моделей создания и адаптации реальных психодиагностических инструментов.

Литература

1. Bakarak, V.R., Fer, R.M., Psihometrika. Vvedenie /V.R. Bakarak, R.M. Fer; per. s angl. A.S. Naumenko, A.Ju. Popova; pod red. N.A. Baturina, E.V. Jejdmana. – 44eljabinsk: Izdatel'skij centr JuUrGU, 2010. – 445s.
2. Coombs, C.H. Psychological scaling without a unit of measurement / C.H. Coombs //Psychological Review. - 1950. - №57. – P. 145-158.
3. Crocher, L., Algina, J. Introduction to Classical and Modern Test Theory / L. Crocher, J. Algina. – Cengage Learning, 2008. – 527 p.
4. Gruijter, D.N., van der Kamp, L.J. Statistical test theory for education and psychology/ D.N. Gruijter, I.J. van der Kamp. – Chapman&Hall, 2005.- 186 p.
5. Gulliksen, H. Theory of mental tests / H. Gulliksen. – New York: Wiley, 1950. - 512 p.
6. Lord, F.M., Novick, M.R. Statistical theories of mental test scores/ F.M. Lord, M.R. Novick. – Information Age Pub., 2008 – 568 p.
7. McDonald, R.P. Test theory: A unified treatment / R.P. McDonald. – NJ: Lawrence Erlbaum Associates, 1999. – 487 p,
8. Mehrens, W.A., Lehmann, I.J. Measurement and evaluation in education and psychology/ W.A. Mehrens, L. J. Lehmann. - Holt, Rinehart and Winston, 1984. - 658 p.
9. Michell, J. An introduction to the logic of psychological measurement / J. Michell. – Hillsdale, NJ: Lawrence Elbraum, 1990. - 190 p.
10. Michell, J. Measurement in psychology: Critical history of a methodological concept/ J. Michell. – Cambridge: Cambridge University Press, 1999.- 265p.
11. Michell, J. Quantitative science and the definition of measurement in psychology / J. Michelle // British journal of Psychology. – 1997. – №68. – P. 355-383.
12. Nunnally, J.C., Bernstein, I.H. Psychometric theory(3rd ed.) / J.C. Nunnally, I.H. Bernstein. – New York: McGraw-Hill, 1994. - 736 p.
13. Spearman, C. The proof and measurement of association between two things / C. Spearman // American Journal of Psychology. – 1904. - № 18. – P. 161 -169.
14. Stevens, S.S. On the theory of scales of measurement / S.S. Stevens// Science. – 1946. - №1103. – P. 677-680.

THE PROBLEM OF SCIENTIFIC MEASUREMENT IN PSYCHODIAGNOSTICS

Assanovich M.A.

Educational Establishment "Grodno State Medical University", Belarus, Grodno

The classical test theory does not include any theses and elaborations of latent psychological constructs measurement methodology. So, psychological tests cannot be considered as measuring tools in psychodiagnostics. New probabilistic models of psychological test design with equal interval measuring scale constructing are necessary to apply.

Key words: *psychodiagnostics, psychometrics, classical test theory, measurement, additivity.*

Адрес для корреспонденции: e-mail: marat.as@tut.by

Поступила 28.10.2013