

Тюменева Ю.А.¹ Что дают числа для измерения в психологии? Обзор критики в адрес нетестируемых допущений психометрики и возможные альтернативы «измерительной» психологии

Tyumeneva Y.¹ What do numbers provide for measurement in psychology? A Review of Critiques of the untested assumptions of psychometrics and possible alternatives to "measurement" in psychology

¹ Национальный исследовательский университет «Высшая школа экономики», Москва, Россия

На фоне увеличивающегося числа психометрических исследований в русскоязычной литературе практически никакого критического отношения эти работы не встречают. В то же время, в последние годы критика в адрес психометрики растет: обсуждают не только неправомерное использование количественных измерений в психологии, но часто дезориентирующую роль чисел в вопросе упорядочивания индивидов в континууме черты или способности. Цель этой работы – обеспечить русскоязычную аудиторию представлением о некоторых острых вопросах, поднимаемых в отношении психометрики, которые иначе могут казаться либо уже решенными, либо несущественными. Обзор сфокусирован на критическом анализе нескольких допущений, на которых строится психометрика: психологические признаки (черты и способности) количественны и измеримы; современная теория тестирования открывает доступ к интервальным измерениям; если данные соответствуют модели, то тест может точно упорядочить индивидов в континууме латентной черты (способности); агрегированные данные пригодны для выводов об индивиде. Предлагаются примеры альтернативных практик, свободных от непроверенных допущений и решающих научные (исследовательские), а не инструментальные (измерительные) задачи.

Ключевые слова: психометрика, измерение, современная теория тестирования, модель Раша, ошибка измерения

Введение в проблему

Типичный образ действия в психометрике заключается в следующем: на выборке респондентов с помощью специальных процедур (тестов или наблюдений) провоцируется заданное поведение, скажем, решение математических задач; по определенным параметрам этого поведения, например, по количеству правильно решенных задач, генерируются числовые данные; данные агрегируются; анализируются и сравниваются с модельным их распределением по выборке, и делается вывод о структуре теста. Затем, в соответствии с выявленной структурой, респондентам начисляются баллы, отражающие их позицию в континууме латентного «конструкта» – черты, предположительно влияющей на выполнение теста. Техники получения тестового балла и его интерпретации различаются в разных психометрических подходах, например, классической теории тестирования (СТТ, classical test theory), измерении Раша (Rasch measurement), современной теории тестирования (IRT, item response theory); но в целом этот *modus operandi* остается тем же.

Для проверки качества «измерения» обычно используются различные индексы согласия данных с моделью. Кроме того, могут привлекаться и дополнительные критерии, по корреляции которых с баллами текущего теста судят о качестве последнего. Однако эти методы не могут проверить ни одно из базовых допущений, на которых они основаны, просто в силу того, что сами эти методы функционируют при условии, что эти допущения верны. Допускается, например, что:

- ненаблюдаемая черта различима от других черт;
- структура черты строго идентична у всех респондентов, а величина (количество, интенсивность) черты меняется только экстенсивно (без сопутствующих качественных изменений самой черты);
- черта присутствует у людей в некотором количестве;
- люди могут быть упорядочены в континууме черты;
- люди обладают идентичной способностью воспринимать задания теста;

- люди, выполнившие одинаковое число заданий определенной трудности, обладают одинаковым количеством черты и пр.

Допущения должны соответствовать истинному положению вещей, чтобы измерение было возможным и чтобы оно информировало об этом положении вещей. Поэтому если специальных усилий по проверке этих допущений не предпринимается и психометрика полагается на непроверенные допущения, то ее выводы относительно психологической реальности находятся в весьма уязвимой позиции.

На сегодняшний день различные точки зрения относительно научной обоснованности психометрических практик достаточно ясно оформлены в литературе, хотя и разрозненны. В этой работе мы фокусируемся только на допущениях, связанных с измеримостью любого психологического признака. Здесь ставится задача представить систематическую критику этих допущений. Кроме того, мы указываем на альтернативные исследовательские возможности, которые не полагаются на непроверяемые или необоснованные допущения и которые показали свою продуктивность в исследовании психологических феноменов.

Проблема (не)измеримости психологических признаков

Измерение внутренне связано с числом и количеством. Это явствует из самого определения измерения, принятого в науке: измерение – это установление отношения одной величины признака, взятого за единицу измерения, к другой величине того же самого признака. Если такие отношения удастся установить, то это автоматически ведет к возможности получения констант и пропорций. Но измерить можно только признаки с количественной (аддитивной) структурой. Аксиоматические условия количественной структуры известны со времен Евклида, систематически были изложены математиком О. Гёльдером [Hölder, 1901] и в связи с психологическими измерениями обсуждались Дж. Мичелом [Michell, Ernst, 1996]. В качестве примера можно взять коммутативность и ассоциативность. Формально говоря, признак является количественным, если и только если любые его три уровня, a , b , и c , такие, что: $a + b = b + a$ (коммутативность); и $a + (b + c) = (a + b) + c$ (ассоциативность).

В физике, которая была и остается образцом измеряющей науки для психологии, открытие какого-то свойства сопровождается множеством экспериментов только для проверки его измеримости. Канонический *modus operandi* для доказательства измеримости признака в физике – это:

- 1) связывание изменения изучаемого признака с изменением какой-то хорошо изученной экстенсивной величины (чаще всего длины), и только с ее изменением;
- 2) установление пропорций в изменении зависимой и независимой величин при контроле посторонних влияний.

Например, Георг Ом в результате многолетних экспериментов, смог показать, что при постоянном сопротивлении отношение величин силы тока константно для любого напряжения в цепи:

$$\frac{I_1}{I_2} = \frac{I_1}{I_2} = \frac{I_1}{I_2} = \text{const. для любого } U.$$

Ясно, что в психологии такой образ действия невозможен в силу невозможности аппаратных экспериментов в отношении психологических феноменов [Trendler, 2013; 2019]. Собственно, поэтому после провала фехнеровского проекта по квантификации ощущений [Ferguson, 1940; Michell, 1999] было вынужденно изменено само определение измерения, теперь основанное на «приписывании чисел» [Stevens, 1946]. Основанием для такого кардинального сдвига послужила репрезентационная теория измерений (РТИ), на которую ссылается Стивенс, где измерение рассматривается как приписывание числа (а не как открытие числовых отношений). Но РТИ предполагает существование отношений между измеряемыми объектами и символическим множеством таким образом, чтобы числа (как символы) репрезентировали отношения реальных объектов, причем так, чтобы репрезентация была логически независима от операций идентификации этих отношений. Однако Стивенс не следует РТИ в этом вопросе. В согласии с операционализмом, он понимает измерение не как отсылку к реальным объектам и не как абстракцию эмпирических отношений этих объектов, а как результат произвольной операции: «измерение в шкале отношений возможно там, где «существуют операции для определения равенности рангов, равенности интервалов, равенности отношений» [Stevens, 1946, p. 679].

Таким образом, подход к измерению Стивенса, переопределяющий ключевой для науки термин, позволил изящно обойти проблему (не)измеримости психологических признаков и формально оставил за психологией статус измеряющей науки. Однако факт, что психометрика оперирует количественными шкалами (и, соответственно, количественными моделями), которые вводятся, по сути, произвольно¹ на любом психологическом признаке и без каких-

либо эмпирических на то оснований, хорошо осознан в современной литературе [Laming, 1997; Michell, 1999; Velleman, Wilkinson, 2011].

Наша задача – критически проанализировать аргументы, которые выдвигаются в защиту такого положения дел. Логика изложения аргументов исходит преимущественно из принципа от частного к общему: не разобравшись с локальными проблемами внутренней психометрической кухни, невозможно в полной мере понять более «видимые» глобальные проблемы измерений в психологии. Эта логика также неидеальна, но решает поставленную автором задачу.

Аргумент 1. Переход к интервальной шкале обеспечивается современной теорией тестирования: измерение Раша – аналог фундаментального измерения

Аргумент отсылает к допущению о вероятностной природе ответов на тест и положениям измерительной модели Георга Раша. В ней вероятность выполнения задания (P) зависит от способности респондента (θ) и трудности задания (b), так что $P = f(\theta, b)$, где f – это логистическая функция [Rasch, 1960]. Логистическое преобразование нужно, чтобы приблизить распределение баллов к нормальному, так как в модели Раша предполагается, что изучаемая латентная характеристика (черта, свойство) нормально распределена в популяции.

Способность (θ) рассчитывается как логарифм от пропорции верных ответов (p) по тестуⁱⁱ: $\theta = \ln\left(\frac{p}{1-p}\right)$. Трудность задания (b) рассчитывается аналогично, только p – пропорция людей из выборки, выполнивших данное задание. В итоге и θ , и b располагаются на одной шкале, так что можно определить, какую вероятность справиться с заданием такой-то трудности имеет человек с такой-то способностью.

Эксплицитно модель утверждает только инвариантность порядка, а не отношений, то есть утверждается неизменный порядок: любой человек имеет большую вероятность ответить на легкие вопросы, чем на трудные. Аналогично, человек с высоко оцененной способностью имеет больше шансов справиться с любым заданием, чем человек с более низкой способностью [Rasch, 1960]. Но при этом и способность и трудность оцениваются в *единицах* измерения (логитах), которые, в отличие от сырых баллов, лежат на шкале равных интервалов [Hambleton et al., 1991; Write, 1993]. Логит – натуральный логарифм от шансов какого-то события: например, хотя и сильно упрощая, если человек выполнил 60 заданий теста из 90, то

Тюменева Ю.А. Что дают числа для измерения в психологии...

его шансы справиться с заданиями теста к шансам не справиться: 60/30 или 2/1, и оценка в логитах способности этого человека – натуральный логарифм от 2, то есть 0,69.

Желанную интервальную шкалу измерения обеспечивает именно логарифмирование. И логарифмирование же обеспечивает аддитивность оцененного признака. Это иллюстрируется по аналогии с производным измерением в физике, например, через второй закон Ньютона [Rasch, 1960]:

$a = \frac{F}{m}$, где a – ускорение тела, F – равнодействующая сил, приложенных к телу, m – масса тела.

Сравнение двух объектов по их массам при фиксированной силе не зависит от F :

$$\frac{a_1}{a_2} = \frac{\frac{F}{m_1}}{\frac{F}{m_2}} = \frac{m_2}{m_1}$$

В этом случае логарифмирование, будь оно применено, дало бы аддитивную структуру:

$$\ln\left(\frac{m_1}{m_2}\right) = \ln(m_2 - m_1)$$

Отсюда заключается, что точно так же логарифмическая трансформация отношения способности к трудности $P = \frac{\theta}{b}$ дает аддитивную структуру:

$$\frac{P_1}{P_2} = \frac{\frac{\theta}{b_1}}{\frac{\theta}{b_2}} = \frac{b_2}{b_1}$$

$$\ln\left(\frac{b_1}{b_2}\right) = \ln(b_2 - b_1).$$

Это же доказывает возможность констант и пропорций [Rasch, 1960; Write 1993; Schmidt, 2017].

Контраргумент

Проблема этой аргументации заключается в том, что она выстраивается на формальной аналогии, которая оказывается ложной, если не верны изначальные посылки. Физики

Тюменева Ю.А. Что дают числа для измерения в психологии...

опытным путем доказывают возможность устанавливать пропорции, причем не для каждого признака это оказывается возможным. Для модели Раша же, при всем формальном сходстве ее с физическими формулами и наличии формальных констант, декларируемые соотношения никогда не проверяются эмпирически [Barrett, 2003; Ferguson, 1940; Kyngdon, 2011; Michell, 1999, 2009; Trendler, 2013]. То есть никогда не было эмпирически показано, что отношение способностей двух индивидов, θ_1/θ_2 , остается константой для заданий с любой трудностью:

$$\frac{\theta_1 b_1}{\theta_2 b_1} = \frac{\theta_1 b_2}{\theta_2 b_2} = \frac{\theta_1 b_3}{\theta_2 b_3} = P (const.)$$

То есть никогда не было показано, что $\frac{\theta_1}{\theta_2} = const.$ для любого b [Trendler, 2019].

То же касается логарифмического преобразования, приводящего к аддитивности. Если бы константы были эмпирически обнаружены, то путем логарифмирования можно превратить их отношение в вычитание. Но для этого нужно убедиться экспериментально, что переход к аддитивности обоснован. Так, Ом варьировал длину проводника на фиксированные интервалы и затем устанавливал связь этих величин с интервалами в отклонении стрелки. Соответственно, для измерения Раша нужно тогда найти такой интервал от θ_1 к θ_2 , который требовал бы нескольких равных изменений в b , так чтобы $\theta_1 - \theta_2 = b_1 - b_2 = b_2 - b_3 = \dots = b_{n-1} - b_n$.

Поскольку этого никогда не было продемонстрировано эмпирически, то нельзя утверждать, что модель Раша дает интервальную аддитивную шкалу измерения. Логарифмирование как математическая операция всегда может быть представлено аддитивно, но такая аддитивность ничего не репрезентирует. В этом смысле данный аргумент продолжает логику операционализма Стивенса: «измерение в шкале отношений возможно там, где «существуют операции для определения равенности рангов, равенности интервалов, равенности отношений» [Stevens, 1946. P.679].

Аргумент 2. Измерение Раша – вероятностный аналог совместного измерения и потому может считаться аддитивным

Рассмотрим ту часть аргументации в пользу аддитивности измерения Раша, которая отсылает к вероятностной природе ответов на тест. Вообще, использование стохастического подхода к измерению в психологии решает множество проблем. Одно из главных преимуществ его над

детерминистским – смягчение требований к «строгим» результатам измерения – то есть фактическое отсутствие констант теперь допустимо в пределах так называемой «ошибки измерения». Сам переход к вероятностным моделям означает и переход к агрегированным данным. Причем оба допущения – и рассмотрение ответов как реализации случайности, и оценка индивида на основе исключительно групповых данных – не являются обоснованными, на чем мы остановимся ниже. А сейчас коротко опишем концепцию совместного измерения (СИ).

СИ – это способ обосновать аддитивность, когда прямой доступ к величинам невозможен, но когда возможно представить одну величину как функцию двух других не взаимодействующих между собой величин. Примером СИ для физических величин может служить скорость как функция расстояния и времени, а для психологических – достижение по тесту как функция уровня подготовки и трудности задания. Поскольку для всех трех величин достаточно быть порядковыми, чтобы их аддитивность была проверена, то СИ должно открывать доступ к доказательству возможности измерения (количественного, аддитивного) в психологии. Иными словами, СИ предлагает способ удостовериться в количественных свойствах признака, даже если нельзя провести операцию сложения на этом признаке ни прямо, ни косвенно [Michell, 2014].

В рамках РТИ Р. Льюс и Дж. Туки [Luce, Tukey, 1964] опубликовали условия совместного измерения (СИ), которые, если соблюдаются при комбинировании уровней любых трех величин, эмпирически подтверждают аддитивную структуру этих трех переменных. Предположим, что есть два множества упорядоченных элементов $A = \{a_1, a_2, a_3\}$ и $X = \{x_1, x_2, x_3\}$, так что $a_1 \geq a_2 \geq a_3$, и от $x_1 \geq x_2 \geq x_3$. Если эти множества не имеют общих элементов и если они связаны с третьей переменной P , то P является функцией их совместного влияния, $P = f(A, X)$.

Проверка аддитивности заключается в сравнении эмпирических значений в матрице СИ (рис. 1) с условиями (аксиомами) аддитивности, которые должны выполняться, если эмпирическая структура аддитивна. Например, условие *двойной отмены* (double cancellation) выполняется, если и только если показано, что значения функции P на пересечении соответствующих уровней двух независимых переменных A и X дают специфические отношения в матрице СИ (рис. 1) [Michell, 2014]. Стрелки на рисунке соответствуют отношению «больше или равно»: сплошные линии – для условия (*если*), а прерывистые – для заключения (*то*).

	a_1	a_2	a_3
x_1	$a_1 x_1$	$a_2 x_1$	$a_3 x_1$
x_2	$a_1 x_2$	$a_2 x_2$	$a_3 x_2$
x_3	$a_1 x_3$	$a_2 x_3$	$a_3 x_3$

Рис. 1. Матрица совместного измерения и проверка условия двойной отмены.

Если аксиомы аддитивности выполняются, то переменные A и X можно квантифицировать по их совместному эффекту. Это осуществляется через подбор равных шагов в упорядоченных значениях переменной X (аналогично и по переменной A), так чтобы $a_1 - a_2 = x_1 - x_2 = x_2 - x_3 = \dots = x_{n-1} - x_n$. На примере со скоростью мы можем подобрать такие интервалы в расстоянии, которые будут эквивалентны в отношении одного и того же временного интервала для какой-то постоянной скорости. Например, можно показать, что расстояния между 5 км и 10 км и между 10 и 15 км будут эквиваленты в том смысле, что в обоих случаях для их прохождения потребуется 1 час при постоянной скорости 5 км/ч. Таким образом, теоретически СИ делает возможным измерение подходящей эмпирической структуры в шкале интервалов (математические доказательства изложены у [Krantz et al., 1971, p. 264-266]).

Из многих публикаций психометриков следует, что измерение Раша является вероятностным вариантом СИ. Во-первых, существует формальное сходство модели Раша, $P = f(\Theta \cdot b, \geq)$, и эмпирической структуры СИ, $P = f(A \cdot X, \geq)$ (см., например, [Borsboom, Mellenbergh, 2004; Perline et al., 1979]). Действительно, было показано, что вероятности решения заданий, для которых было достигнуто согласие с моделью Раша, могут быть упорядочены в совместной матрице соответственно аксиомам СИ [Perline et al., 1979; Scheiblechner, 1999]. Так, в работе Л. Станкова и А. Крегана [Stankov, Cregan, 1993] интеллектуальная способность рассматривалась как функция трудности задания и мотивации к его выполнению. Независимость мотивации и трудности достигалась за счет того, что уровни мотивации задавались временем, отпущенным на задание, а трудность заданий – их вычислительной комплексностью. Авторы показали, что аксиомы СИ действительно выполняются на матрице, где в ячейках указывалась частота успешного выполнения теста (т.е. способность как функция совместного эффекта мотивации и трудности выполнения заданий). Таким образом, был сделан вывод о том, что все три признака (мотивация, трудность, способность) обладают аддитивной (т.е. количественной) структурой.

Контраргумент 1

Для того чтобы события в матрице СИ интерпретировались как элементы двух множеств (Θ и b), эти множества не должны влиять друг на друга. Хотя в модели Раша декларируется такая независимость, она – чисто техническая; на самом деле, нет способа оценить способность отдельно от заданий какой-то трудности, а трудность – отдельно от людей с какой-то способностью. Скажем, в уже упомянутом исследовании [Ibid. 1993] нельзя исключить, что мотивация на всем протяжении эксперимента не взаимодействовала с трудностью заданий, успешностью в предыдущих заданиях, усталостью и другими сторонними факторами. Кроме того, нельзя быть уверенным, что сокращение времени на выполнение задания не давало разнонаправленных эффектов на мотивацию у разных людей. То есть, хотя характеристики определяются как будто бы независимо, фундаментальная проблема с их неконтролируемым взаимовлиянием остается. Такая ситуация не годится для проверки количественности, и поэтому СИ не может быть использовано для доказательства аддитивности психологических шкал [Luce, 1997; Trendler, 2019].

Контраргумент 2

Как указывает Кингдон [Kynngdon, 2011], доказательство аддитивности через вероятностные модели некорректно в принципе. Дело в том, что вероятности – это множества вещественных чисел. Поскольку и способность, и трудность в модели Раша выражаются через пропорцию частот, то это тоже множество вещественных чисел. Тогда в уравнении Раша схватывается разность между двумя вещественными числами ($\Theta - b$), и, отсюда, эта разность сама является числом. Таким образом, модель Раша – это просто изоморфное отображение двух структурно идентичных математических множеств. Модель не содержит эмпирических объектов совсем. Безотносительно к тому, соответствует ли модель эмпирическим данным или нет, она всегда будет изоморфно отображать вероятности (которые суть вещественные числа) на вещественные числа (элементы Θ и b). Более того, было показано, что модель делает это даже для данных, за которыми нет никакой латентной черты, а именно для вероятности выпадения орла или решки при подбрасывании монеты [Wood, 1978].

Проблема здесь в том, что неизвестно, как связана вероятность события (количественный показатель) с эмпирической структурой (качественный показатель). Как заметил один из основателей СИ [Luce, 1997, p. 82]: «Факт в том, что мы не знаем, как это сделать [связать]. Все подходы к концепции случайности, с которыми я знаком, предполагают числовую репрезентацию; у нас просто нет *качественной* теории этой концепции».

В качестве подытога этой части аргументации констатируем, что на сегодняшний день у психометрики нет способов продемонстрировать какие-либо константы в «измеренных» величинах: ни в классическом понимании измерения, ни в РТИ, ни в детерминистском, ни в стохастическом подходе.

Аргумент 3. Пусть психометрика и не дает строгих измерений, но за счет использования вещественных чисел достигается более точное упорядочивание

В последнее время в психометрической литературе встречается утверждение, что нельзя требовать от психологии строгих измерений, подобных физическим [см., например, [Markus, Borsboom, 2012]]. Скорее речь идет о более точном (благодаря числам) упорядочивании индивидов в континууме черты. С одной стороны, замена «объективного измерения» [Rasch, 1977] на упорядочивание позволяет обойти критику в адрес необоснованных количественных притязаний [Uher, 2018; 2020], а с другой – сохранить существующие психометрические практики.

Контраргумент 1

Методологически нельзя создать с помощью вещественных чисел «более точное» упорядочивание [Michell, 2012]. Дело в том, что психологические признаки гетерогенны (или должны рассматриваться как гетерогенные, пока не будет доказано обратное). Хотя мы и способны воспринять относительную интенсивность или выраженность признака, сам признак при разных интенсивностях не является самоидентичным. Легкая боль может сопровождаться щекочущим ощущением или чувством инородного тела на участке боли; тогда как сильная боль может восприниматься как жжение, растяжение, онемение или сверление и пр. В отличие от расстояния или силы тока, которые, возрастая или уменьшаясь, остаются все еще длиной и силой тока, боль может не восприниматься как боль при изменении в силе. Для более комплексных чувств, состояний и признаков тем более бездоказательно утверждать, что их изменение исключительно экстенсивное, не затрагивающее их внутренний состав. Например, говоря об измерении позиции школьников в континууме их академической мотивации (даже с учетом ее «подвидов» – вроде внутренней, внешней и пр.), мы допускаем, что каждый школьник испытывает эту мотивацию абсолютно одинаково в разное время и абсолютно одинаково с другими школьниками, а также что к этой мотивации примешиваются другие чувства, цели, ожидания и пр. Ясно, что такое допущение не реалистично.

Интересно отметить, что во многих науках и научно-прикладных исследованиях упорядочивание используется как самодостаточный метод оценки – например, в шкале болевых ощущений, степени рака, старческой беспомощности или силы землетрясения. Переход к аддитивной шкале здесь невозможен именно в силу гетерогенности самого оцениваемого признака. Но что более важно, в этих случаях именно порядковая оценка оказывается самой адекватной, потому что она *структурно идентична самому оцениваемому признаку*. Введение вещественных чисел в эту оценку снизило бы точность оценки и дезориентировало бы оценивающих. Иными словами, числа могут предоставлять неточную или вовсе ложную информацию, если они структурно не идентичны феномену, к которому применяются.

Контраргумент 2

Использование вещественных чисел для упорядочивания не всегда правомерно. Да, числа 1 и 0 можно использовать для кодировки отдельных наблюдений правильных и неправильных ответов. Но психометрики имеют дело с векторами ответов, суммируя эти коды. Однако сумма $1 + 1 + 0$ не соответствует никакому наблюдению, поэтому она не описывает поведения вообще. И числа не помогают упорядочить эти вектора. Математически, два разных вектора сопоставимы, если поэлементные сравнения векторов гомогенны [Алексеев, 2017; Harzheim, 2005; Ziegler, Vautier, 2014]. Отсюда следует, что, если появляются вектора (например, 110 и 001), то они будут лишь частично упорядоченными, и, соответственно, эти вектора несопоставимы (поскольку часть позаданных сравнений окажется гетерогенной: иногда – «более чем», иногда – «менее чем»).

Таким образом, набор логически возможных векторов не может служить порядковой шкалой, потому что порядковое измерение требует простого упорядочивания в противоположность частично упорядочиванию значений шкалы.

Аргумент 4. Оцененное на агрегированных данных распределение ответов позволяет оценить ошибку изменения, структуру и качество теста

При проверке согласия данных модели анализируются отклонения наблюдаемого ответа от математического ожидания согласно выбранной стохастической модели [Linacre, 1998]. Дисперсия по выборке, необъясненная данной моделью, разбивается на отдельные

компоненты. Если распределение дисперсии по этим компонентам равномерно, то принято считать, что никаких других «латентных» переменных, помимо модельных, нет. Поэтому полученная на агрегированных данных структура теста считается адекватной для всех индивидов в выборке, что открывает возможность для индивидуальной оценки.

Этот способ действия базируется на двух допущениях: во-первых, что ответы на вопросы теста – это реализация случайности, а во-вторых, что множественные наблюдения за единственным индивидом можно заменить на единичное наблюдение за множеством индивидов.

Оба допущения хорошо иллюстрируются высказыванием Георга Раша [Rasch, 1980, p. 11]: *«Что касается человеческих существ и их действий, представляется бесполезным делом конструировать модели для предсказания отдельных случаев. Наоборот, человеческое существо действительно кажется довольно случайным, не меньше, чем радиоактивная эмиссия».*

Действительно, в областях, где объекты по природе стохастичны (например, в термодинамике), естественные науки используют стохастические модели. Отсюда допускается, что стохастическое моделирование будет также успешно объяснять психологические феномены и индивидуальное поведение.

Контраргумент 1

Если учесть тот факт, что большинство людей каждый день успешно добираются до работы, готовят еду и заботятся о близких, едва ли допущение о непредсказуемости индивидуального поведения имеет серьезные основания. Более того, если разделять понятия случайности и изменчивости, и допустить изменчивость индивидуального поведения, то к этому можно относиться не как к «ошибке измерения», а как к новым исследовательским вопросам и разработке новых методов изучения индивида. Например, теория перцептивного контроля [Forssell, Powers, 2009] интерпретирует изменчивость индивидуальных реакций как ответ системы перцептивного контроля на меняющиеся стимулы, что с одной стороны делает изменчивость существенным компонентом адаптивной биологической системы, а с другой – открывает совершенно иные способы исследования индивидуального поведения. Или, например, подход Observation Oriented Modeling [Grice, 2011] позволяет получить информацию об индивидуальных паттернах поведения, которые пропускаются стандартными статистическими методами (поскольку последние основаны на усреднении агрегированных данных).

Контраргумент 2

Популяционные параметры являются характеристикой, возникающей на уровне популяции. Согласно научному реализмуⁱⁱⁱ, реальность организована по уровням, где новые характеристики появляются из процессов, происходящих на нижних уровнях [Psillos, 1999]: скажем, цвет глаз появляется на уровне фенотипа, но отсутствует на генетическом уровне, который и отвечает за цвет. Поэтому оценка психологических свойств, проведенная на уровне популяции, не может быть перенесена на уровень индивида [Cervone, 2005; Sixtl, 1980] без того, чтобы подтвердить эмпирически, что размерности черты одинаковы в меж-индивидуальной и внутри-индивидуальной моделях. Немногие исследования этого вопроса показали, однако, что структура черт, обнаруживаемых на меж-индивидуальном уровне, не воспроизводится на уровне отдельных индивидов [Mischel, Shoda, 1998; Feldman, 1995; Molenaar et al., 2003]. Так что наши знания, полученные на агрегированных данных (например, об общем интеллекте или Большой пятерке черт), могут являться лишь статистическим артефактом [Borsboom et al., 2003; Franić et al., 2014].

Другая проблема оценки индивида с опорой исключительно на агрегированные данные, заключается в том, что психометрическая точность моделей технически достигается только за счет дисперсии результатов. Поэтому «точность» измерения зависит не только от качества инструмента (как можно видеть в других измерительных науках), но и от численности и вариабельности результатов индивидуальных «замеров». Отсюда следует, что один и тот же тест неизбежно будет демонстрировать разную «точность» в разных ситуациях, что не может не противоречить самой идее инструмента измерения [Borsboom et al., 2009].

Более того, декларируемая «точность упорядочивания», которую якобы обеспечивают интервальные шкалы IRT и большие выборки, плохо согласуется с тем фактом, что сами интервалы технически выводятся из упорядоченных данных именно за счет неточности – ошибки измерения. Эта странная ситуация, когда более точную – интервальную – шкалу удается получить за счет ошибки измерения и когда при полной элиминации ошибки шкала превращается в менее точную – порядковую, была названа в литературе «парадоксом Раша» [Michell, 2008].

Альтернативные подходы к пониманию и оценке психологических феноменов

Из всего вышесказанного следует, что психометрические практики не могут проверить допущения, на которых сами же эти практики основаны. Более того, аргументы, которые обычно выдвигаются в поддержку психометрики, в свою очередь, базируются либо на столь же непроверенных, либо на ложных допущениях. Следует признать, однако, что в самих по себе психометрических измерительных моделях нет ничего неверного или вводящего в заблуждение. Источник проблемы – это их неизвестное репрезентационное качество. Иными словами, если мы не знаем природы феномена, который пытается репрезентировать, то вопросы о выборе подходящей шкалы, модели или «инструмента измерения» бессмысленны. Как указывал Мичелл, сначала должна решаться научная задача (онтологическая), а только потом – инструментальная (эпистемологическая) [Michell, 1999; 2003].

Возьмем, например, допущение о порядковой структуре психологического признака. Порядковая структура является необходимым, хотя и недостаточным, условием для количественной. И поэтому, если окажется, что признак невозможно даже упорядочить, то его автоматически невозможно измерить. Для такой проверки достаточно убедиться в транзитивности попарных сравнений уровней признака. Например, если три уровня А, В и С какого-то признака, будучи попарно сравнены, образуют транзитивные отношения: $A > B$, $B > C$, и $A > C$, то этот признак имеет порядковую структуру и, следовательно, может быть упорядочен. Однако при проверке транзитивности с утверждениями шкалы нейротизма одного из факторов «Большой пятерки черт» обнаружилось, что около четверти всех сделанных респондентами попарных сравнений нарушали транзитивность. Формально говоря, оказывалось, что если $A > B$ и если $B > C$, то $C > A$ [Morris et al., 2017]. Похожий результат был получен с утверждениями шкалы академической мотивации [Тюменева, Вергелес, 2021]. Это серьезно подрывает не только количественную оценку личностных черт, но и сами рейтинговые шкалы, которые обычно используются для оценки этой черты.

Или, например, было показано, что оценка с использованием ликертовских шкал связана не с «силой» оцениваемого «конструкта», а с семантикой самих утверждений. Так что, во-первых, «факторы», которые обычно получают психометрики в результате анализа агрегированных эмпирических данных, хорошо предсказываются семантической близостью утверждений в опроснике [Arnulf et al., 2019]. Во-вторых, отклонения индивидуальных интерпретаций от семантически ожидаемых расценивались традиционными методами факторного анализа как

Тюменева Ю.А. Что дают числа для измерения в психологии... «необъясненная дисперсия» и часто игнорировались как ошибка. Но оказалось, что сами эти отклоняющиеся интерпретации не являются случайными, а связаны с социальным контекстом жизни респондента, например, типом работы, которую он выполняет [Arnulf et al., 2020]. Показано также, как влияет особенность формулировок и контекст утверждений на ответы респондентов [Johnson, 2001; McGrane, 2009] и как легко списать эти влияния на «ошибку измерения», полагаясь на традиционные статические модели.

Надо сказать, что в целом вся вариативность, которую не удастся объяснить латентными переменными, обычно списывается на ошибку измерения; источники такой вариативности и индивидуальные процессы, отвечающие за них, почти не исследуются. Замечательны поэтому работы, которые проверяют существование вариативности на уровне индивида и ищут ее объяснения экспериментально. Так, например, интересна непараметрическая альтернатива логистической регрессии – «пороговый анализ» [Grice et al., 2016] – метод, который не «отменяет ошибки» через вычисление групповых средних, но представляет ответы респондентов как специфические паттерны ответов. При этом каждый последующий ответ оценивается как дискретное состояние, связанное не только с текущим утверждением, но и с предыдущим^{iv}. Этот подход позволяет вывести логику индивидуального поведения так, чтобы группа респондентов была как можно точнее разнесена по подгруппам с разными паттернами ответов [Grice et al., 2011; 2017]. Такого рода методы не только дают больше информации, чем стандартные психометрические измерения, но эта информация еще и индивидуализирована и не основана на непроверяемых допущениях вроде аддитивности и континуальности «психологического конструкта».

Иными словами, «измеряемый» феномен должен быть сначала изучен, и важно, что указанные примеры демонстрируют совершенно иной подход к изучению феноменов и их измерению, начинающийся с проверки самих его основ.

Все вышесказанное ни в коем случае не означает, что методы, предполагающие аддитивность и континуальность признака – такие как IRT и структурное моделирование – не обоснованы в принципе. Они могут обоснованно применяться там, где феномен хорош изучен, где его структура действительно количественна, а выводы касаются не индивидуальных, а популяционных характеристик. Вероятностное моделирование приносит большую пользу, например, в экологии [Arhonditsis et al., 2006], эволюционной биологии [Pugesek et al., 2003], генетике [Franić et al., 2012]. В науке о человеческом социальном поведении структурное моделирование и байесовский анализ довольно успешно применяются в исследованиях

Тюменева Ю.А. Что дают числа для измерения в психологии... принятия решений и экономического поведения [Birnbauм, 2008; Cavagnaro, Davis-Stober, 2014]. Здесь модели используют данные, полученные независимо от самих этих моделей, поэтому и результаты моделирования могут быть использованы для проверки модели в отношении реальности, которую модель должна репрезентировать.

Вообще же говоря, в противоположность популярному мнению, сохранившемуся с начала 20 века, науке необязательно иметь дело с количеством и измерением, чтобы оставаться наукой [Michell, 1999; 2003] – взять хотя бы структурную лингвистику или топологию. Более того, история самой психологии показывает, что наиболее сильные, получившие эмпирическую поддержку и сформировавшие «лицо» психологического знания, теории были разработаны вне измерительной парадигмы (например, представление о психологических защитных механизмах, процессах развития речи и мышления в онтогенезе, путях формирования навыков, волевого поведения, социальной идентичности и др.). Так что «психология без измерений» не означает «психология без научного метода», а всего лишь означает, что имеет смысл искать другие, адекватные психологической реальности, методы.

Конечно, когда потребности статуса, публикационная и академическая институциональная политика в психологии, а также связанное с этим финансирование, заставляют исследователей следовать типовым шагам, это не оставляет ни места, ни времени на внимательное и критическое к ним отношение. Однако аргументированная и развернутая критика вокруг необоснованного применения «измерений» должна быть как минимум знакома работающим в этой области исследователям. И здесь была сделана попытка представить коротко некоторые линии этой критики, а также указать на возможность альтернативных исследовательских подходов.

Литература

Алексеев В.Е. Дискретная математика: Учебное пособие. Н.Новгород: Нижегородский госуниверситет, 2017.

Тюменева Ю.А., Вергелес К.П. Исследование транзитивности самооценок мотивации. Социологический журнал, 2021, №2, 8-24.

Arhonditsis G.B., Stow C.A., Steinberg L.J., Kenney M.A., Lathrop R.C., McBride S.J., Reckhow K.H. Exploring ecological patterns with structural equation modeling and Bayesian analysis. *Ecological Modelling*, 2006, 192(3-4), 385-409. DOI:10.1016/j.ecolmodel.2005.07.028.

Arnulf J.K., Dysvik A., Larsen K.R. Measuring semantic components in training and motivation: a methodological introduction to the semantic theory of survey response. *Human Resource Development Quarterly*, 2019, 30(1), 17-38. DOI:10.1002/hrdq.21324.

Arnulf J.K., Nimon K., Larsen K.R., Hovland C.V., Arnesen M. The priest, the sex worker, and the CEO: measuring motivation by job type. *Frontiers in Psychology*, 2020, 11, 1321. DOI:10.3389/fpsyg.2020.01321.

Barrett P. Beyond psychometrics Measurement, non-quantitative structure, and applied numerics. *Journal of Managerial Psychology*, 2003, 18(3), 421–439. DOI:10.1108/02683940310484026.

Birnbaum M.H. New paradoxes of risky decision making. *Psychological Review*, 2008, 115, 463–501. DOI:10.1037/0033-295X.115.2.463.

Borsboom D., Mellenbergh G.J., Van Heerden J. The theoretical status of latent variables. *Psychological Review*, 2003, 110(2), 203. DOI:10.1037/0033-295X.110.2.203.

Borsboom D., Mellenbergh G., Heerden J. The Concept of Validity. *Psychological Review*, 2004, 111, 1061-71. DOI:10.1037/0033-295X.111.4.1061.

Borsboom D., Cramer A.O., Kievit R.A., Scholten A.Z., Franić S. The end of construct validity. In: *The Concept of Validity: Revisions, New Directions and Applications*. IAP Information Age Publishing, 2009. pp. 135-170.

Cavagnaro D.R., Davis-Stober C.P. Transitive in our preferences, but transitive in different ways: an analysis of choice variability. *Decision*, 1, 102–122. DOI:10.1037/dec0000011.

Cervone D. Personality architecture: Within-person structures and processes. *Annual Review of Psychology*, 2005, 56, 423–452. DOI:10.1146/annurev.psych.56.091103.070133.

Ferguson, A., Myers, C.S., Bartlett, R.J., Banister, H., Bartlett, F.C., Brown, W., Campbell, N.R., Craik, K.J.W., Drever, J., Guild, J., Houstoun, R.A., Irwin, J.O., Kaye, G.W.C., Philpott, S.J.F., Richardson, L.F., Shaxby, J.H., Smith, T., Thouless, R.H., & Tucker, W.S. Quantitative estimates of sensory events: Final report of the committee appointed to consider and report upon the possibility of quantitative estimates of sensory events. British Association for the Advancement of Science, 1940, 1, 331–349.

Feldman L.A. Variations in the circumplex structure of mood. *Personality and Social Psychology Bulletin*, 1995, 21(8), 806–817. DOI:10.1177/0146167295218003.

Forsell D., Powers W. Perceptual Control Theory. *Living Control Systems Pub*, 2009, 72, 74–75.

Franić S., Borsboom D., Dolan C.V., Boomsma D.I. The big five personality traits: psychological entities or statistical constructs? *Behavior Genetics*, 2014, 44(6), 591–604. DOI:10.1007/s10519-013-9625-7.

Franić S., Dolan C.V., Borsboom D., Boomsma D.I. Structural equation modeling in genetics. In: R.H. Hoyle (Ed.), *Handbook of structural equation modelling*. The Guilford Press, 2012. pp. 617–635.

Grice J.W. *Observation oriented modeling: Analysis of cause in the behavioral sciences*. Academic Press, 2011.

Grice J.W., Barrett P.T., Cota L., Taylor Z., Felix C., Garner S., Medellin E., Vest A. Four bad habits of modern psychologists. *Behavioral Sciences*, 2017, 7(3), Article 53. DOI:10.3390/bs7030053

Grice J.W., Cota L.D., Barrett P.T., Wuensch K.L., Poteat M. A simple and transparent alternative to logistic regression. *Advances in Social Sciences Research Journal*, 2016, 3(7). DOI:10.1177/2158244015604192.

Hambleton R.K., Swaminathan H., Rogers H.J. *Fundamentals of item response theory*. Sage, 1991. Vol. 2.

Harzheim E. *Ordered Sets*. Springer Science Business Media, 2005. pp. 71–83.

Hölder O. Die Axiome der Quantitaai und die Lehre vom mass. Ber. Verh. Sächs. Akad. Wiss. Leipzig Math. Phys, 1901, Kl. 53, 1–46.

Johnson T. Controlling the effect of stimulus context change on attitude statements using Michell's binary tree procedure. Australian Journal of Psychology, 2001, 53(1), 23-28. DOI:10.1080/00049530108255118.

Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. Foundations of measurement: Additive and polynomial representations (Vol. 1). New York, NY: Academic Press. 1971.

Kyngdon A. Psychological measurement needs units, ratios, and real quantities: A commentary on Humphry. Measurement, 2011, 9(1), 55–58. DOI:10.1080/15366367.2011.558791.

Laming D.R.J. The measurement of sensation. Oxford University Press. 1997. DOI: 10.1093/acprof:oso/9780198523420.001.0001.

Linacre J.M. Structure in Rasch Residuals: Why Principal Components Analysis. Rasch Measurement Transactions, 1998, 12(2), 636. <https://www.rasch.org/rmt/rmt122m.htm>.

Luce R.D. Several unresolved conceptual problems of mathematical psychology. Journal of mathematical psychology, 1997, 41(1), 79–87. DOI:10.1006/jmps.1997.1150.

Luce R.D., Tukey J.W. Simultaneous conjoint measurement: A new type of fundamental measurement. Journal of mathematical psychology, 1964, 1(1), 1–27. DOI:10.1016/0022-2496(64)90015-X.

Markus K.A., Borsboom D. The cat came back: Evaluating arguments against psychological measurement. Theory & Psychology, 2012. 22(4), 452–466. DOI:10.1177/0959354310381155.

McGrane J.A. Unfolding the conceptualization and measurement of ambivalent attitudes. Doctoral dissertation. University of Sydney, Sydney, Australia, 2009.

Michell J. Measurement in Psychology: A Critical History of a Methodological Concept. Vol. 53. Cambridge, UK: Cambridge University Press. 1999.

Michell, J. Pragmatism, positivism and the quantitative imperative. *Theory and Psychology*. 2003. 13, 45–52. DOI:10.1177/0959354303013001761.

Michell J. Conjoint Measurement and the Rasch Paradox: A Response to Kyngdon. *Theory and Psychology*. 2008, 18(1), 119 – 124. DOI:10.1177/0959354307086926.

Michell, J. The psychometricians' fallacy: Too clever by half? *British Journal of Mathematical and Statistical Psychology*. 2009, 62(1), 41–55. DOI:10.1348/000711007X243582.

Michell, J. “The constantly recurring argument:” inferring quantity from order. *Theory and Psychology*. 2012, 22, 255–271. DOI:10.1177/0959354311434656.

Michell, J. An introduction to the logic of psychological measurement. Psychology Press. 2014.

Michell, J., and Ernst, C. The axioms of quantity and the theory of measurement. *Journal of Mathematical Psychology*. 1996, 40, 235–252. DOI:10.1006/jmps.1996.0023.

Michell, J. An introduction to the logic of psychological measurement. Psychology Press. 2014.

Mischel W., Shoda Y. Reconciling processing dynamics and personality dispositions. *Annual review of psychology*. 1998, 49(1), 229-258. DOI:10.1146/annurev.psych.49.1.229.

Molenaar P.C., Huizenga H.M., Nesselroade J.R. The relationship between the structure of interindividual and intraindividual variability: A theoretical and empirical vindication of developmental systems theory. In: *Understanding human development*. Boston, MA: Springer, 2003. pp. 339–360.

Morris S.D., Grice J.W., Cox R.A. Scale imposition as quantitative alchemy: Studies on the transitivity of neuroticism ratings. *Basic and Applied Social Psychology*, 2017, 39(1), 1–18. DOI:10.1080/01973533.2016.1256288.

Perline R., Wright B.D., Wainer H. The Rasch model as additive conjoint measurement. *Applied Psychological Measurement*, 1979, 3(2), 237–255. DOI:10.1177/014662167900300213.

Psillos S. *Scientific realism: How science tracks the truth*. London, UK: Routledge, 1999.

Pugesek B.H., Tomer A., Von Eye A. (Eds.). *Structural equation modeling: applications in ecological and evolutionary biology*. Cambridge University Press, 2003.

Rasch G. *Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests*. Nielsen Lydiche, 1960.

Rasch G. On Specific Objectivity: An attempt at formalizing the request for generality and validity of scientific statements. *The Danish Yearbook of Philosophy*, 1977, 14, 58–93.

Rasch G. *Probabilistic models for some intelligence and attainment tests*. University of Chicago Press, 1980.

Scheiblechner H. Additive conjoint isotonic probabilistic models (ADISOP). *Psychometrika*, 1999, 64(3), 295-316.

Schmidt F.L. Beyond questionable research methods: The role of omitted relevant research in the credibility of research. *Archives of Scientific Psychology*, 2017, 5(1), 32–41. DOI:10.1037/arc0000033.

Sixtl F. Generalized laws of reaction, the average person, and interindividual variation. In: E.D. Lantermann, H. Feger (Eds.), *Similarity and choice: Papers in honour of Clyde Coombs*. Bern, Switzerland: Huber, 1980. pp. 100–107.

Stankov L., Cregan A. Quantitative and qualitative properties of an intelligence test: Series completion. *Learning and Individual Differences*, 1993, 5(2), 137–169. DOI:10.1016/1041-6080(93)90009-H.

Stevens S. On the Theory of Scales of Measurement. *Science*, 1946, 103(2684), 677–680. <http://www.jstor.org/stable/1671815>.

Trendler G. Measurement in psychology: A case of ignoramus et ignorabimus? A rejoinder. *Theory & Psychology*, 2013, 23(5), 591–615. DOI:10.1177/0959354313490451.

Trendler G. Measurability, systematic error, and the replication crisis: A reply to Michell (2019) and Krantz and Wallsten (2019). *Theory and Psychology*, 2019, 29(1), 144–151.

DOI:10.1177/0959354318824414.

Uher J. Quantitative data from rating scales: An epistemological and methodological enquiry. *Frontiers in Psychology*, 2018, 9(DEC), 2599. DOI:10.3389/fpsyg.2018.02599.

Uher J. Psychometrics is not measurement: Unraveling a fundamental misconception in quantitative psychology and the complex network of its underlying fallacies. *Journal of Theoretical and Philosophical Psychology*, 2020, 41(1), 58. DOI:10.1037/teo0000176.

Velleman P., Wilkinson L. Nominal, ordinal, interval, and ratio typologies are misleading. *The American Statistician*, 1993, 47(1), 65-72. DOI:10.1080/00031305.1993.10475938.

Wood R. Fitting the Rasch model—A heady tale. *British Journal of Mathematical and Statistical Psychology*, 2011, 31(1), 27–32. DOI:10.1111/j.2044-8317.1978.tb00569.x.

Write B.D. Logits? *Rasch Measurement Transactions*, 1993, 7(2), 288.

Ziegler M., Vautier S. A farewell, a welcome, and an unusual exchange. *European Journal of Psychological Assessment*, 2014, 30(2), 81–85. DOI:10.1027/1015-5759/a000203.

Поступила в редакцию: 23 июля 2021 г. Дата публикации: 31 августа 2022 г.

Сведения об авторах

Тюменева Юлия Алексеевна. Кандидат психологических наук, ст. научный сотрудник Института образования Национального исследовательского университета «Высшая школа экономики», Потаповский пер., д. 16, стр. 10, 101000 Москва, Россия.

E-mail: jtiumeneva@hse.ru, jutu@yandex.ru

Ссылка для цитирования

Тюменева Ю.А. Что дают числа для измерения в психологии? Обзор критики в адрес нетестируемых допущений психометрики и возможные альтернативы «измерительной» психологии. *Психологические исследования*. 2022. Т. 15, № 84. С. 3. URL: <https://psystudy.ru>

Адрес статьи: <https://doi.org/10.54359/ps.v15i84.1198>

Тюменева Ю.А.¹ Что дают числа для измерения в психологии? Обзор критики в адрес нетестируемых допущений психометрики и возможные альтернативы «измерительной» психологии

Tyumeneva Y.¹ What do numbers provide for measurement in psychology? A Review of Critiques of the untested assumptions of psychometrics and possible alternatives to "measurement" in psychology

¹ HSE University, Institute of Education, Moscow, Russia

Despite the increasing number of Russian-language psychometric studies, almost no criticism of these works is encountered. At the same time, global criticism of psychometrics has been growing – not only the inappropriate use of quantitative measurements in psychology, but also the misleading role of numbers in ordering individuals on a continuum of trait or ability is discussed. This article is aimed to introduce Russian-speaking audience to some of the thorny issues of psychometrics, which might otherwise seem either already resolved or irrelevant. Here I critically analyze several fundamental assumptions of psychometrics: psychological attributes (traits and abilities) are quantitative and measurable; modern test theory allows for interval measures; and, if the data fits the model, the test can accurately order individuals on a latent trait (ability) continuum; aggregated data are suitable for drawing conclusions about an individual. Alternative practices free from the aforementioned assumptions are proposed. Those are thought to solve scientific rather than instrumental (measurement) problems.

Keywords: psychometrics, measurement, modern test theory, Rasch model, error of measurement

References

Alekseev V.E. Diskretnaya matematika: Uchebnoe posobie. N.Novgorod: Nizhegorodskij gosuniversitet, 2017. (in Russian)

Arhonditsis G.B., Stow C.A., Steinberg L.J., Kenney M.A., Lathrop R.C., McBride S.J., Reckhow K.H. Exploring ecological patterns with structural equation modeling and Bayesian analysis. *Ecological Modelling*, 2006, 192(3-4), 385-409. DOI:10.1016/j.ecolmodel.2005.07.028.

Arnulf J.K., Dysvik A., Larsen K.R. Measuring semantic components in training and motivation: a methodological introduction to the semantic theory of survey response. *Human Resource Development Quarterly*, 2019, 30(1), 17-38. DOI:10.1002/hrdq.21324.

Arnulf J.K., Nimon K., Larsen K.R., Hovland C.V., Arnesen M. The priest, the sex worker, and the CEO: measuring motivation by job type. *Frontiers in Psychology*, 2020, 11, 1321. DOI:10.3389/fpsyg.2020.01321.

Barrett P. Beyond psychometrics Measurement, non-quantitative structure, and applied numerics. *Journal of Managerial Psychology*, 2003, 18(3), 421–439. DOI:10.1108/02683940310484026.

Birnbaum M.H. New paradoxes of risky decision making. *Psychological Review*, 2008, 115, 463–501. DOI:10.1037/0033-295X.115.2.463.

Borsboom D., Mellenbergh G.J., Van Heerden J. The theoretical status of latent variables. *Psychological Review*, 2003, 110(2), 203. DOI:10.1037/0033-295X.110.2.203.

Borsboom D., Mellenbergh G., Heerden J. The Concept of Validity. *Psychological Review*, 2004, 111, 1061-71. DOI:10.1037/0033-295X.111.4.1061.

Borsboom D., Cramer A.O., Kievit R.A., Scholten A.Z., Franić S. The end of construct validity. In: *The Concept of Validity: Revisions, New Directions and Applications*. IAP Information Age Publishing, 2009. pp. 135-170.

Cavagnaro D.R., Davis-Stober C.P. Transitive in our preferences, but transitive in different ways: an analysis of choice variability. *Decision*, 1, 102–122. DOI:10.1037/dec0000011.

Cervone D. Personality architecture: Within-person structures and processes. *Annual Review of Psychology*, 2005, 56, 423–452. DOI:10.1146/annurev.psych.56.091103.070133.

Ferguson, A., Myers, C.S., Bartlett, R.J., Banister, H., Bartlett, F.C., Brown, W., Campbell, N.R., Craik, K.J.W., Drever, J., Guild, J., Houstoun, R.A., Irwin, J.O., Kaye, G.W.C., Philpott, S.J.F., Richardson, L.F., Shaxby, J.H., Smith, T., Thouless, R.H., & Tucker, W.S. Quantitative estimates of sensory events: Final report of the committee appointed to consider and report upon the possibility of quantitative estimates of sensory events. *British Association for the Advancement of Science*, 1940, 1, 331–349.

Feldman L.A. Variations in the circumplex structure of mood. *Personality and Social Psychology Bulletin*, 1995, 21(8), 806–817. DOI:10.1177/0146167295218003.

Forsell D., Powers W. *Perceptual Control Theory*. Living Control Systems Pub, 2009, 72, 74–75.

Franić S., Borsboom D., Dolan C.V., Boomsma D.I. The big five personality traits: psychological entities or statistical constructs? *Behavior Genetics*, 2014, 44(6), 591–604. DOI:10.1007/s10519-013-9625-7.

Franić S., Dolan C.V., Borsboom D., Boomsma D.I. Structural equation modeling in genetics. In: R.H. Hoyle (Ed.), *Handbook of structural equation modelling*. The Guilford Press, 2012. pp.617–635.

Grice J.W. *Observation oriented modeling: Analysis of cause in the behavioral sciences*. Academic Press, 2011.

Grice J.W., Barrett P.T., Cota L., Taylor Z., Felix C., Garner S., Medellin E., Vest A. Four bad habits of modern psychologists. *Behavioral Sciences*, 2017, 7(3), Article 53. DOI:10.3390/bs7030053.

Grice J.W., Cota L.D., Barrett P.T., Wuensch K.L., Poteat M. A simple and transparent alternative to logistic regression. *Advances in Social Sciences Research Journal*, 2016, 3(7). DOI:10.1177/2158244015604192.

Hambleton R.K., Swaminathan H., Rogers H.J. *Fundamentals of item response theory*. Sage, 1991. Vol. 2.

Harzheim E. *Ordered Sets*. Springer Science Business Media, 2005. pp. 71–83.

Hölder O. Die Axiome der Quantitaa und die Lehre vom mass. *Ber. Verh. Sächs. Akad. Wiss. Leipzig Math. Phys.*, 1901, Kl. 53, 1–46.

Johnson T. Controlling the effect of stimulus context change on attitude statements using Michell's binary tree procedure. *Australian Journal of Psychology*, 2001, 53(1), 23-28. DOI:10.1080/00049530108255118.

Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. Foundations of measurement: Additive and polynomial representations (Vol. 1). New York, NY: Academic Press. 1971.

Kyngdon A. Psychological measurement needs units, ratios, and real quantities: A commentary on Humphry. *Measurement*, 2011, 9(1), 55–58. DOI:10.1080/15366367.2011.558791.

Laming D.R.J. The measurement of sensation. Oxford University Press. 1997. DOI: 10.1093/acprof:oso/9780198523420.001.0001.

Linacre J.M. Structure in Rasch Residuals: Why Principal Components Analysis. *Rasch Measurement Transactions*, 1998, 12(2), 636. <https://www.rasch.org/rmt/rmt122m.htm>.

Luce R.D. Several unresolved conceptual problems of mathematical psychology. *Journal of mathematical psychology*, 1997, 41(1), 79–87. DOI:10.1006/jmps.1997.1150.

Luce R.D., Tukey J.W. Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of mathematical psychology*, 1964, 1(1), 1–27. DOI:10.1016/0022-2496(64)90015-X.

Markus K.A., Borsboom D. The cat came back: Evaluating arguments against psychological measurement. *Theory & Psychology*, 2012. 22(4), 452–466. DOI:10.1177/0959354310381155.

McGrane J.A. Unfolding the conceptualization and measurement of ambivalent attitudes. Doctoral dissertation. University of Sydney, Sydney, Australia, 2009.

Michell J. *Measurement in Psychology: A Critical History of a Methodological Concept*. Vol. 53. Cambridge, UK: Cambridge University Press. 1999.

Michell, J. Pragmatism, positivism and the quantitative imperative. *Theory and Psychology*. 2003. 13, 45–52. DOI:10.1177/0959354303013001761.

Michell J. Conjoint Measurement and the Rasch Paradox: A Response to Kyngdon. *Theory and Psychology*. 2008, 18(1), 119 – 124. DOI:10.1177/0959354307086926.

Michell, J. The psychometricians' fallacy: Too clever by half? *British Journal of Mathematical and Statistical Psychology*. 2009, 62(1), 41–55. DOI:10.1348/000711007X243582.

Michell, J. “The constantly recurring argument:” inferring quantity from order. *Theory and Psychology*. 2012, 22, 255–271. DOI:10.1177/0959354311434656.

Michell, J. *An introduction to the logic of psychological measurement*. Psychology Press. 2014.

Michell, J., and Ernst, C. The axioms of quantity and the theory of measurement. *Journal of Mathematical Psychology*. 1996, 40, 235–252. DOI:10.1006/jmps.1996.0023.

Mischel W., Shoda Y. Reconciling processing dynamics and personality dispositions. *Annual review of psychology*. 1998, 49(1), 229–258. DOI:10.1146/annurev.psych.49.1.229.

Molenaar P.C., Huizenga H.M., Nesselroade J.R. The relationship between the structure of interindividual and intraindividual variability: A theoretical and empirical vindication of

developmental systems theory. In: Understanding human development. Boston, MA: Springer, 2003. pp. 339–360.

Morris S.D., Grice J.W., Cox R.A. Scale imposition as quantitative alchemy: Studies on the transitivity of neuroticism ratings. *Basic and Applied Social Psychology*, 2017, 39(1), 1–18. DOI:10.1080/01973533.2016.1256288.

Perline R., Wright B.D., Wainer H. The Rasch model as additive conjoint measurement. *Applied Psychological Measurement*, 1979, 3(2), 237–255. DOI:10.1177/014662167900300213.

Psillos S. *Scientific realism: How science tracks the truth*. London, UK: Routledge, 1999.

Pugesek B.H., Tomer A., Von Eye A. (Eds.). *Structural equation modeling: applications in ecological and evolutionary biology*. Cambridge University Press, 2003.

Rasch G. *Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests*. Nielsen Lydiche, 1960.

Rasch G. On Specific Objectivity: An attempt at formalizing the request for generality and validity of scientific statements. *The Danish Yearbook of Philosophy*, 1977, 14, 58–93.

Rasch G. *Probabilistic models for some intelligence and attainment tests*. University of Chicago Press, 1980.

Scheiblechner H. Additive conjoint isotonic probabilistic models (ADISOP). *Psychometrika*, 1999, 64(3), 295–316.

Schmidt F.L. Beyond questionable research methods: The role of omitted relevant research in the credibility of research. *Archives of Scientific Psychology*, 2017, 5(1), 32–41. DOI:10.1037/arc0000033.

Sixtl F. Generalized laws of reaction, the average person, and interindividual variation. In: E.D. Lantermann, H. Feger (Eds.), *Similarity and choice: Papers in honour of Clyde Coombs*. Bern, Switzerland: Huber, 1980. pp. 100–107.

Stankov L., Cregan A. Quantitative and qualitative properties of an intelligence test: Series completion. *Learning and Individual Differences*, 1993, 5(2), 137–169. DOI:10.1016/1041-6080(93)90009-H.

Stevens S. On the Theory of Scales of Measurement. *Science*, 1946, 103(2684), 677–680. <http://www.jstor.org/stable/1671815>.

Trendler G. Measurement in psychology: A case of ignoramus et ignorabimus? A rejoinder. *Theory & Psychology*, 2013, 23(5), 591–615. DOI:10.1177/0959354313490451.

Trendler G. Measurability, systematic error, and the replication crisis: A reply to Michell (2019) and Krantz and Wallsten (2019). *Theory and Psychology*, 2019, 29(1), 144–151. DOI:10.1177/0959354318824414.

Tyumeneva Yu.A., Vergeles K.P. Issledovanie tranzitivnosti samoocenok motivacii. *Sociologicheskij zhurnal*, 2021, №2, 8–24. (in Russian)

Uher J. Quantitative data from rating scales: An epistemological and methodological enquiry. *Frontiers in Psychology*, 2018, 9(DEC), 2599. DOI:10.3389/fpsyg.2018.02599.

Uher J. Psychometrics is not measurement: Unraveling a fundamental misconception in quantitative psychology and the complex network of its underlying fallacies. *Journal of Theoretical and Philosophical Psychology*, 2020, 41(1), 58. DOI:10.1037/teo0000176.

Velleman P., Wilkinson L. Nominal, ordinal, interval, and ratio typologies are misleading. *The American Statistician*, 1993, 47(1), 65-72. DOI:10.1080/00031305.1993.10475938.

Wood R. Fitting the Rasch model—A heady tale. *British Journal of Mathematical and Statistical Psychology*, 2011, 31(1), 27–32. DOI:10.1111/j.2044-8317.1978.tb00569.x.

Write B.D. Logits? *Rasch Measurement Transactions*, 1993, 7(2), 288.

Ziegler M., Vautier S. A farewell, a welcome, and an unusual exchange. *European Journal of Psychological Assessment*, 2014, 30(2), 81–85. DOI:10.1027/1015-5759/a000203.

Information about authors

Tyumeneva Yulia. PhD, Senior Research Fellow, Institute of Education HSE Higher School of Economics (HSE University), Potapovsky Pereulok, 16 / 10, 101000 Moscow, Russia.

E-mail: jtiumeneva@hse.ru, jutu@yandex.ru

To cite this article

Tyumeneva Y. What do numbers provide for measurement in psychology? A Review of Critiques of the untested assumptions of psychometrics and possible alternatives to "measurement" in psychology. *Psikhologicheskie Issledovaniya*, 2022, Vol. 15, No. 84, p. 3. <https://psystudy.ru>

ⁱ Произвольность заключается в том, что природа шкалы определяется исследователем, а не отношениями, которые демонстрируют *эмпирические* феномены.

ⁱⁱ Для простоты сейчас рассматриваются только тестовые задания с правильными и неправильными ответами, доминирующие в измерениях когнитивных способностей и результатов обучения. Модели Раша широко применяются и для измерения личностных черт, что накладывает специфику на подсчеты, но принцип оценки остается тем же.

ⁱⁱⁱ Общий тезис реализма состоит в том, что мир существует независимо от наших интерпретаций, и теории (как инструменты познания) иногда бывают близки к тому, каков мир на самом деле. Одна из ключевых доктрин научного реализма признает существование теоретических ненаблюдаемых объектов, постулируемых современной наукой (например, электрон существует в том же смысле, как и любой другой наблюдаемый физический объект). В философии науки принято называть научным реализмом семейство доктрин, различающихся по своей позиции в отношении истины, метода познания ненаблюдаемых объектов, и пр., но обсуждение этого вопроса выходит за рамки статьи.

^{iv} Действительно, сложно представить, что человек, оценивая утверждения в серии, полностью забывает свои предыдущие ответы (замечание, сделанное еще в работе Lord, Novick, 1968).

