Колачев Н.И.[1] Использовать ли показатель трудности при анализе психологических тестов?

Kolachev N.I.[1] Is the difficulty index to be used when analyzing psychological tests?

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

В предлагаемой статье рассматривается вопрос применимости показателя трудности в рамках психометрического анализа психологических тестов. Показывается, что в традиционных психометрических исследованиях авторы игнорируют этот показатель, поскольку парадигмой конфирматорного факторного анализа, на основе которой большинство психологов проводит свои исследования, его изучение не предусмотрено. На аналитическом и эмпирическом примерах продемонстрировано, что трудность – аналог энтропии (меры количества информации), не имеющий единиц измерения. Проведено сравнение обоих показателей на предмет преимуществ и недостатков и даны рекомендации по их использованию в исследованиях качества диагностических методик.

Ключевые слова: психодиагностика, энтропия, показатель трудности, дискриминативность, конфирматорный факторный анализ, классическая теория тестирования, современная теория тестирования

Введение

В психологии за время её существования в качестве научной дисциплины появилось достаточно большое число различных диагностических методик, направленных на измерение латентных характеристик человека [Бурлачук, 2011]. С развитием психодиагностики развивались и методы оценки качества разрабатываемых инструментов измерения (тестов). В зависимости от цели тестирования в психометрике существуют различные критерии и показатели качества тестовой методики, однако неизменными остаются трудность и дискриминативность каждого отдельного утверждения (вопроса, пункта шкалы) и теста в целом. Первая, как правило, показывает, насколько трудно респондентам выразить согласие с тем или иным утверждением методики, в то время как последняя – насколько утверждения диагностического инструмента успешно дифференцируют респондентов с разным уровнем выраженности изучаемого признака [Крокер, Алгина, 2012].

Различительная способность исследователями-психологами активно изучается, её присутствие в научных статьях является обязательной. Однако в стандартных статьях по адаптации или разработке новой диагностической методики нет упоминаний о трудности. Дело в том, что психодиагностические методики анализируются в парадигме конфирматорного факторного анализа (КФА), который описывает связь латентной и наблюдаемой характеристики следующим образом:

$$y_{ij} = \lambda_{ij} F + e_{ij}$$

где y_{ij} – ответ i-го респондента на j-е утверждение методики, λ_{ij} – факторная нагрузка (регрессионный коэффициент) j-го утверждения, F – латентный фактор, e_{ij} – доля необъяснённой дисперсии j-го утверждения (остаток). Факторные нагрузки, по сути, являются показателями корреляции наблюдаемой и латентной переменных. Следовательно, чем выше факторная нагрузка, тем сильнее связь утверждения с латентным фактором и тем выше его различающая способность [Stark et al., 2006]. На факторные нагрузки обращается пристальное внимание, на их основе рассчитывается средняя извлеченная дисперсия (average variance extracted), которая используется при исследовании валидности инструмента и должна быть более 50%, соответственно, факторные нагрузки, используемые в её расчете, должны превосходить значения 0,50–0,70 по модулю [DiStefano, Hess, 2005; Rönkkö, Cho, 2022].

По отношению к показателю трудности не всё так однозначно. Во-первых, коллеги-психологи в личных беседах всё чаще высказывают недоумение относительно использования показателя трудности. Действительно, он изначально появился для анализа тестов достижений, в которых

есть заведомо верный ответ. В психологических методиках нет правильных ответов, а результаты говорят лишь о степени выраженности той или иной психологической характеристики. Для тестов с правильным ответом показатель трудности (как доля респондентов, верно выполнивших задание) довольно понятен и вполне информативен. Во-вторых, в парадигме конфирматорного факторного анализа трудность наблюдаемых индикаторов не изучается, что не мешает исследователям говорить о качестве разработанных инструментов; как правило, обращается внимание на факторные нагрузки (показатель дискриминативности), индексы согласия с моделью и корреляции латентных факторов (в случае многомерных моделей).

В вопросе оценки трудности оппозицию парадигме КФА составляет современная теория тестирования (IRT) – совокупность моделей и методов, как и КФА, описывающих то, как ответы респондентов связаны с латентными характеристиками и свойствами утверждений теста [Embretson, Reise, 2013], но параметризующих эту связь немного иначе, чем КФА. Важным оцениваемым параметром в моделях IRT является трудность задания, а также трудность шагов [Маслак, 2016], которые влияют на вероятность выбора той или иной категории ответа. Наиболее распространённая[1] модель современной теории тестирования, применяющаяся для рейтинговых шкал, выглядит следующим образом:

$$ln\left(\frac{P_{ijk}}{P_{ij(k-1)}}\right) = \theta_i - \left(\beta_j - \tau_k\right)$$

где P_{ijk} – вероятность выбора i-м респондентом k-й категории в j-м утверждении, θ_i – уровень выраженности латентной характеристики i-го респондента, β_j – трудность j-го утверждения, τ_k – трудность k-й категории шкалы ответов. Таким образом, при шкалировании результатов тестирования модель учитывает трудность утверждений и шагов (переходов от одной категории к другой) шкалы ответа. Важно отметить, что при параметризации конфирматорных моделей методами, предназначенными для анализа категориальных данных, рассчитываются трудности шагов [см. Li, 2016], но не моделируется средняя трудность пунктов шкалы (в терминологии КФА – индикаторов/наблюдаемых переменных).

Трудность утверждения (пункта шкалы) можно рассчитать без применения IRT-моделей. Трудность отдельного пункта методики вычисляется следующим образом:

[1] Если зайти на сайт Института объективных измерений (Institute for Objective Measurement) и взглянуть на перечень журналов, которые симпатизируют Раш-парадигме, то можно увидеть журналы по медицине и образованию (https://www.rasch.org/friendly.htm). Именно в этих областях модель приобрела наибольшую популярность.

$$d_j = \frac{\frac{1}{n}\sum_{i=1}^{n} x_{ij}}{k_j}$$

где d_j – трудность j-того пункта методики, n – количество респондентов, $\frac{1}{n}\sum_{i=1}^{n} x_{ij}$ – среднее арифметическое результатов j-го пункта методики, k_j – максимальный балл j-го пункта методики; $d_j \in [0, 1]$.

В связи со всем изложенным выше возникают вопросы: имеет ли смысл при анализе психодиагностических тестов рассчитывать показатель трудности утверждений, ведь в парадигме КФА можно обойтись без анализа трудности пунктов шкалы и шагов шкалы ответа? Какую информацию трудность даёт исследователям и пользователям инструментов измерения латентных характеристик?

Целью этой статьи является демонстрация того, что показатель трудности информативен при анализе качества психодиагностических методик, при этом он может быть дополнен или заменён показателем энтропии. То есть сделана попытка продемонстрировать важность показателя трудности через его связь с мерой количества информации (энтропией).

Рассуждения

Для того, чтобы ответить на вопросы о важности показателя трудности, предлагаю обратиться к теории информации. «Теория информации представляет собой математическую теорию, посвященную измерению информации, её потока, «размеров» канала связи и т.п.» [Лидовский, 2004, с. 5]. Одним из ключевых понятий теории информации является энтропия, введённая в научный обиход Клодом Шенноном [Shannon, 1948]. В математической статистике энтропия является количественной мерой измерения информации и характеризует неопределённость вероятностного состояния случайной величины [Лидовский, 2004]. Энтропия дискретной случайной величины рассчитывается по следующей формуле:

$$H(x) = -\sum_{i=1}^{k} p_i \cdot \log_2 p_i$$

где p_i – вероятность появления состояния из набора дискретных состояний, $\log_2 p_i$ – неопределённость одного состояния x_i, или, по-другому, частная энтропия [Каргин, 2013], $\sum p_i = 1$. Как правило, энтропия выражается в таких единицах измерения, как бит (от словосочетания *binary digit*), поскольку используется логарифм по основанию 2. Если для нахождения энтропии использовать натуральный логарифм, то единицей измерения становится нат (от словосочетания *natural logarithm*).

В психологической диагностике случайной величиной выступают как результаты измерения конструкта в целом, выраженные в первичных баллах, так и ответы на то или иное утверждение методики по заданной шкале ответов. Как ответам, так и первичным баллам можно поставить в соответствие вероятность. Если нам дана некоторая шкала ответов, например, 5-балльная шкала Ликерта от «Полностью не согласен» до «Полностью согласен», то выбор ответа на определенное утверждение диагностической методики является дискретной случайной величиной, принимающей значений 1, 2, 3, 4 или 5. В результате эмпирического исследования каждой ответной категории того или иного утверждения диагностической методики можно поставить в соответствие (в пределе) вероятность выбора этой категории (хотя на деле это – доля респондентов, выбравших эту категорию). Таким образом, используя формулу вычисления энтропии, возможно рассчитать количество информации для каждого утверждения и для теста в целом.

С точки зрения теории информации, максимальная неопределённость возникает при равномерном распределении вероятности по всем элементам дискретной случайной величины [Cover, Thomas, 2006]. Если рассмотреть тест достижений, то наиболее информативным дихотомически оцениваемым заданием (0 или 1 балл) является такое, трудность которого равна 0,50 (то есть 50% респондентов, выполнявших это задание, получили 1 балл). В этом случае энтропия (неопределённость) максимальна и равна 1 биту (по формуле Шеннона с использованием логарифма по основанию 2). Если трудность задания близка к единице или нулю, то его энтропия приближается к нулю, что означает низкую информативность такого задания, с точки зрения дифференцирования респондентов. Если мы рассмотрим психологические методики, то их результаты получаются в ходе применения политомических шкал, где, как правило, имеется более двух градаций ответа. К примеру, в методике, в которой используется 4-балльная шкала ответов, равномерное распределение предполагает по 25% ответов на каждую категорию.

Зная вероятность выбора той или иной категории шкалы ответа, можно описать трудность утверждения (пункта шкалы) через математическое ожидание случайной величины:

$$d_j = \frac{\sum_{m=1}^{k} m_j \cdot p_{jm}}{k_j}$$

где m_j – значение категории шкалы j-го утверждения (можно начинать с 0), p_{jk} – вероятность выбора m-ой категории j-го утверждения, k_j – максимальный балл j-го пункта методики. Необходимо отметить, что если обозначения категорий шкалы начинать с 0, а не 1, то при

любом количестве категорий оптимумом трудности будет выступать значение 0.5; если с 1 – то оптимум будет смещён к более высокому значению в связи со смещением начала отсчёта, хотя при увеличении градаций шкалы ответов оптимум трудности стремится к 0.5.

Теперь отметим, что если оба показателя – трудность и энтропия – зависят от вероятности выбора той или иной категории шкалы ответов, то это значит, что они тесно взаимосвязаны. Чтобы показать эту взаимосвязь, приведём аналитический пример и проверим предположения на реальных данных.

При этом может возникнуть закономерный вопрос: зачем исследовать показатели, которые зависят от особенностей выборки (распределение вероятностей зависит от выборки)? Ведь у семейства моделей Г. Раша есть свойство специфической объективности, которое означает, что результаты сравнения выполнения теста испытуемыми зависят только от их уровня подготовленности (выраженности черты) и не зависят от предъявляемых заданий/утверждений [Fischer, 1987; Rasch, 1977], а оценки трудности тестовых заданий/утверждений инвариантны относительно контингента испытуемых, по результатам тестирования которых они получены [Карданова, 2004]. Тем не менее, несмотря на описанное свойство, исследователи фиксируют, что на практике показатели классической теории тестирования сопоставимы с параметрами моделей современной теории тестирования, а также инвариантны относительно различных групп испытуемых [Macdonald, Paunonen, 2002; Kohli, Kora, Henn, 2015]. Кроме того, учёные подчёркивают, что, хотя обе теории измерений дают сравнимые результаты, пользователям этих результатов более понятны постулаты классической теории тестирования [Fan, 1998].

Если классическая теория тестирования на практике во многом не уступает современной теории, то это значит, что первую нужно дальше изучать, развивать, расширять области применения. Наиболее ярким примером полезности постулатов классической теории и их применения за пределами анализа тестовых методик является формулировка и разработка моделей латентных состояний-черт (latent state-trait models) на основе модели классической теории тестов $X = T + E$. В рамках этих моделей дисперсия истинного балла[2] ($Var(T)$) декомпозируется на стабильную и изменчивую часть [Steyer, Mayer, Geiser, Cole, 2015], тем

[2] В психометрике под истинным баллом понимается средний результат бесконечного количества тестирований при условии, что респондент при каждом новом тестировании забывает содержание теста (то есть отсутствует эффект научения). При этом практическая реализация модели классической теории тестирования основана на работе с дисперсиями, поскольку в исследованиях мы имеем дело с несколькими респондентами, что позволяет работать с элементами модели классической теории не как с точечными оценками, а как с распределениями.

самым позволяя проверять гипотезы о структурных компонентах и динамических свойствах психологических конструктов.

Процедура и методика исследования

Предлагаемая работа основана на аналитическом применении и сравнении результатов формул расчёта трудности и энтропии. В качестве иллюстрации симулированы трудности утверждений, включающих два варианта ответа (допустим, «Нет» и «Да»), и поставлена в соответствие энтропийная мера. Трудность утверждений сгенерирована с шагом в 0.01.

В качестве эмпирической иллюстрации связи между трудностью и энтропией использованы результаты двух исследований. Первое посвящено измерению профессионального выгорания и жизнестойкости на выборке 620 респондентов. В инструменте выгорания используются 23 утверждения и 5-балльная шкала ответа, измеряющая частоту тех или иных симптомов, где 1 – «Никогда», 5 – «Постоянно». Этот опросник имеет приемлемые психометрические свойства и находится в открытом доступе [см. Schaufeli, De Witte, Desart, 2020]. Для измерения жизнестойкости использовалась скрининговая версия опросника жизнестойкости из 12 утверждений, разработанная Е.Н. Осиным и Е.И. Рассказовой [Осин, Рассказова, 2013]. Утверждения оцениваются по 4-балльной шкале, где 1 — «Нет», 4 — «Да». Выбор этих показателей обусловлен тем, что измерения профессионального выгорания на выборках нормотипичных людей носят смещённый характер (в сторону более низких значений), в то время как показатели жизнестойкости оказываются смещены в сторону более высоких значений (левосторонняя асимметрия).

Для полноты картины были проанализированы трудность и энтропия утверждений методики измерения рабочих требований и ресурсов [Иванова, 2016]. Данные получены от 239 респондентов. Опросник содержит 13 вопросов о ситуации на работе (6 вопросов о ресурсах и 7 вопросов о требованиях) и 5-балльную ответную шкалу, где 0 = Очень редко, никогда, 5 = Очень часто, постоянно. Трудность утверждений этой методики разнообразна, без явно выраженной асимметрии.

Для вычисления энтропии использован логарифм по основанию 2.

Результаты

Аналитический пример

Колачев Н.И. Использовать ли показатель трудности при анализе психологических тестов? Пусть имеется психодиагностический инструмент с 2-балльной шкалой ответа. Пусть категории именуются так: «Нет», «Да». Кроме того, пусть P_1 – вероятность выбора ответа «Нет», а P_2 – вероятность выбора ответа «Да». Закодируем ответ «Нет» как 0, ответ «Да» – как 1. Тогда трудность такого утверждения будет рассчитываться следующим образом: $0 \cdot P_1 + 1 \cdot P_2 = P_2$, а энтропия будет равняться $-(P_1 \cdot \log_2 P_1 + P_2 \cdot \log_2 P_2)$.

На рисунке 1 представлена взаимосвязь между показателями трудности и энтропии. График построен на основе узловых точек равномерной сетки значений трудности (с шагом 0.01) с применением интерполяции[3]. Заметно, что связь нелинейная, она описывается полиномом (многочленом) второй степени. Максимума функция достигает при $x = 0.5$. При этом в интервале $[0.01; 0.50]$ функция монотонно возрастает, а в интервале $[0.50; 0.99]$ монотонно убывает. Это связано с симметричностью энтропийной меры.

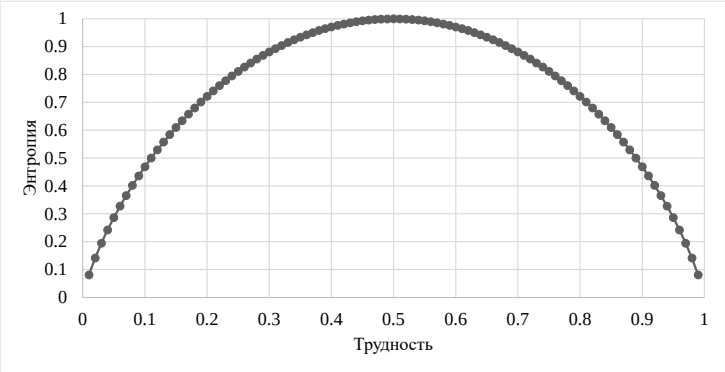

Рис. 1. Взаимосвязь трудности и энтропии.

Таким образом, можно сделать вывод, что в тесте, в котором наблюдается правосторонняя асимметрия распределения трудности пунктов (например, трудность всех утверждений находится в интервале $[0.10; 0.40]$), ожидается положительная связь трудности с показателем энтропии, при левосторонней асимметрии – отрицательная, при отсутствии заметной асимметрии – параболическая.

Эмпирические примеры

На рисунке 2 представлена диаграмма рассеяния для показателей трудности и энтропии на данных опросника профессионального выгорания. Трудность пунктов этой методики характеризуется правосторонней асимметрией, то есть респонденты в основном демонстрировали невысокую выраженность этого синдрома, что естественно для

[3] Интерполяция – это метод нахождения неизвестных промежуточных значений некоторой функции по имеющемуся дискретному набору ее известных значений. Известные значения называются узловыми точками.

Колачев Н.И. Использовать ли показатель трудности при анализе психологических тестов? нормативной выборки. Показатель доли объяснённой дисперсии R^2 свидетельствует о высоком качестве линейной подгонки (приближении линейной функцией), тем самым мы можем говорить о линейной положительной связи между показателями в случае правосторонней асимметрии трудности пунктов диагностической методики. Конкретные значения трудности и энтропии для каждого пункта представлены в Приложении 1.

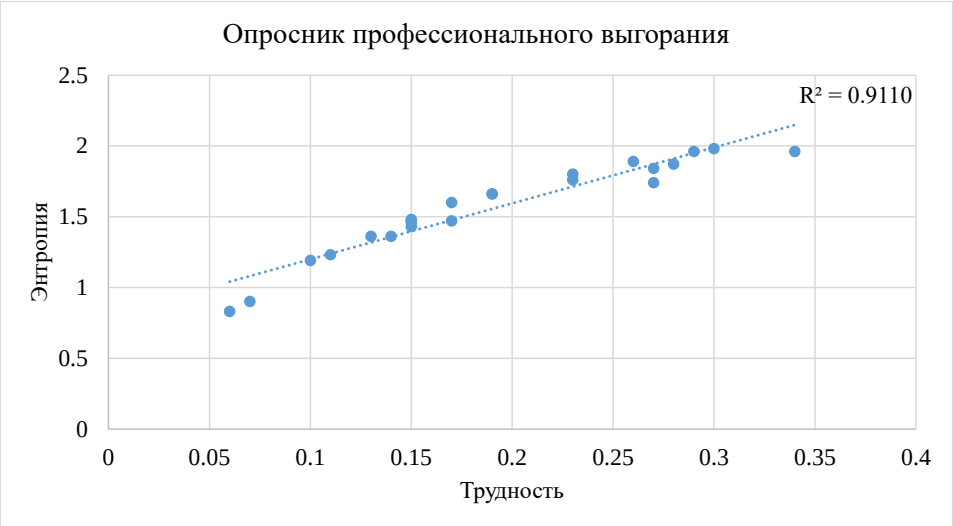

Рис. 2. Взаимосвязь трудности и энтропии на примере опросника профессионального выгорания.

На рисунке 3 представлена диаграмма рассеяния для показателей трудности и энтропии на данных скрининговой версии теста жизнестойкости. Трудность пунктов этой методики характеризуется левосторонней асимметрией, то есть респонденты в основном демонстрировали высокую самоотчетную жизнестойкость. Показатель доли объяснённой дисперсии R^2 свидетельствует о достаточно высоком качестве линейной подгонки (приближении линейной функцией), тем самым мы можем говорить о линейной отрицательной связи между показателями в случае левосторонней асимметрии трудности пунктов диагностической методики. Конкретные значения трудности и энтропии для каждого пункта представлены в Приложении 2.

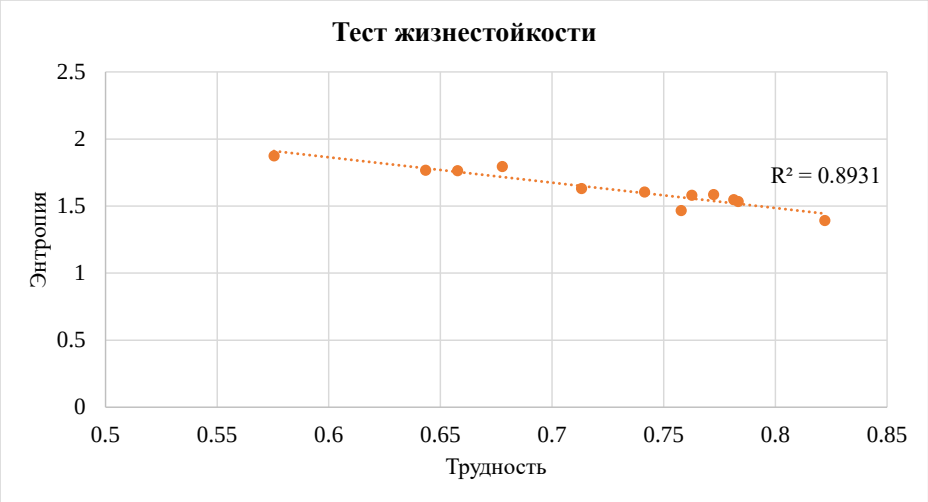

Рис. 3. Взаимосвязь трудности и энтропии на примере теста жизнестойкости.

На рисунке 4 представлена диаграмма рассеяния для показателей трудности и энтропии на данных опросника рабочих требований и ресурсов. Трудность пунктов этой методики характеризуется относительным разнообразием – от 0.25 до 0.80. Можно заметить, что, как и в случае аналитического примера, связь аппроксимируется полиномиальной функцией второй степени (многочленом второй степени). Конкретные значения трудности и энтропии для каждого пункта представлены в Приложении 3.

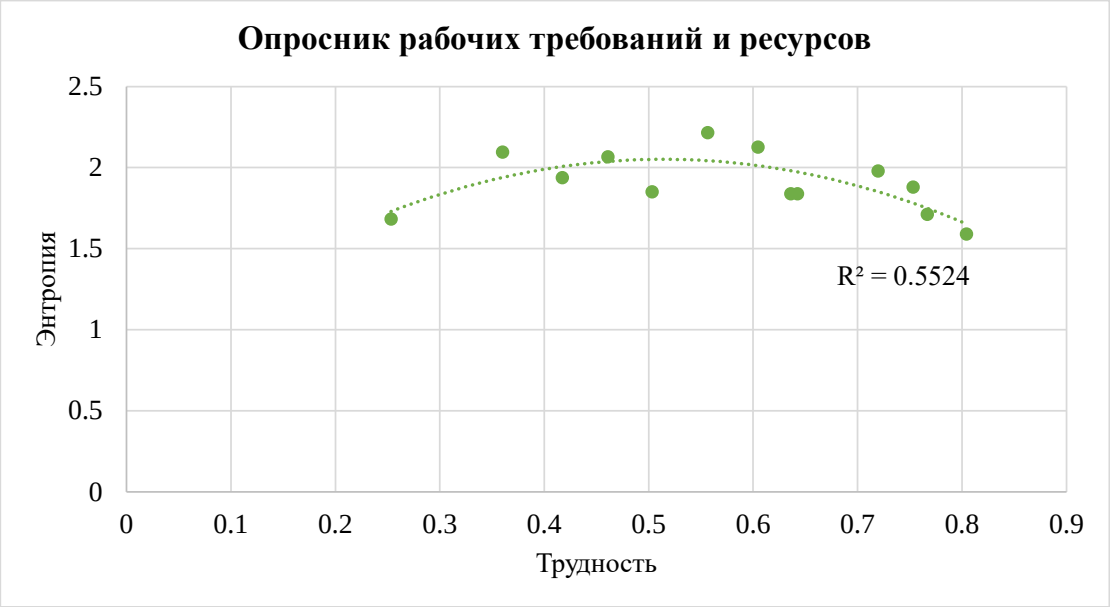

Рис. 4. Взаимосвязь трудности и энтропии на примере опросника рабочих требований и ресурсов.

В ходе анализа эмпирических примеров была замечена следующая закономерность: при одинаковой трудности (с учётом погрешности округления) утверждения могут давать разное количество информации о респондентах. К примеру, утверждения №1 и №16 опросника

выгорания имеют одинаковую трудность (0.17), но их энтропия различна – 1.84 и 1.74 соответственно. Другой пример: в опроснике жизнестойкости утверждения №1 и №6 имеют равную трудность (0.76), но значение энтропии отличается – 1.58 и 1.47 соответственно. Можно предположить, что это связано с распределением вероятностей выбора категорий ответа. И действительно, тест хи-квадрат на сходство распределений показал, что, несмотря на значимое отклонение от равномерного распределения, в утверждении №1 распределение вероятностей ближе к равномерному, чем в утверждении №6: $\chi^2(16) = 328.22$, $p = 8.80 \cdot 10^{-70}$ и $\chi^2(16) = 390.22$, $p = 1.97 \cdot 10^{-132}$ соответственно. То же самое характерно для упомянутых утверждений методики жизнестойкости: $\chi^2(9) = 316.76$, $p = 2.35 \cdot 10^{-68}$ и $\chi^2(9) = 374.11$, $p = 8.95 \cdot 10^{-81}$ соответственно.

Выявленная закономерность может свидетельствовать о том, что при создании альтернативных (эквивалентных) форм теста важно опираться не только на трудность входящих в них заданий (утверждений), как указано, к примеру, в энциклопедии по клинической нейропсихологии [Ferguson, Iverson, 2011], но и на количество получаемой с их помощью информации об испытуемых. Таким образом, опираясь на классическую теорию тестов, мы нашли доказательства постулата, который был сформулирован в рамках современной теории тестирования: параллельными являются те тестовые формы, которые имеют совпадающие информационные кривые [Lord, Novik, 1968; Sun et al., 2008].

Выводы

Из приведённых выше рассуждений можно сделать вывод, что классический психометрический показатель трудности является мерой количества информации. Однако, хоть он и отражает усреднённую вероятность согласия с утверждением, является безразмерной величиной, что затрудняет его интерпретацию. К тому же для информационного его толкования необходимо знать оптимальную трудность, что сопряжено с дополнительными расчётами. Классический показатель энтропии лишён этих недостатков: он имеет общепризнанные единицы измерения (например, бит, трит, нат и др.), при дифференциальной (нормо-ориентированной) диагностике легко интерпретируется – чем больше значение, тем лучше. Более того, на основе меры неопределённости можно рассчитать показатель эффективности информационного содержания теста через отношение эмпирически полученной энтропии к максимально возможному значению энтропии, которое получается в случае, когда вероятности равномерно распределяются по шкале ответов [Каргин, 2013]: $\frac{H_{\text{эмпир.}}}{H_{max.}}$. Благодаря этому показателю можно в процентном отношении узнать, насколько наблюдаемое информационное содержание методики близко к идеальному. В таблице 1

Колачев Н.И. Использовать ли показатель трудности при анализе психологических тестов? представлены эмпирическая и максимальная энтропии и показатель эффективности содержания трёх описанных диагностических методик. Наибольшая эффективность информационного содержания представлена у теста жизнестойкости и шкалы рабочих требований и ресурсов, что неудивительно, поскольку у последней трудность входящих в неё утверждений близка к оптимуму. Провести абсолютную оценку эффективности информационного содержания пока не представляется возможным, поскольку не определены критические значения эффективности; это требует дополнительных исследований.

Кроме того, при конструировании параллельных (альтернативных) форм теста опираться только на трудность входящих в них заданий (утверждений) некорректно, поскольку при одинаковой трудности задания могут давать различающееся количество информации о респондентах; количество получаемой информации зависит от распределения вероятностей выбора ответных категорий каждого задания. В такой ситуации важно изучать трудность совместно с энтропией.

Таблица 1

Эмпирическая, максимальная энтропии и показатель эффективности содержания трёх диагностических методик

Методика	$\boldsymbol{H}_{\text{эмпир.}}$	$\boldsymbol{H_{max.}}$	**Эффективность содержания** $\left(\frac{\boldsymbol{H}_{\text{эмпир.}}}{\boldsymbol{H_{max.}}}\right)$
Опросник профессионального выгорания	1.56	2.32	0.67
Тест жизнестойкости	1.63	2.00	0.82
Опросник рабочих требований и ресурсов	1.91	2.32	0.82

Необходимо отметить, что показатель неопределённости (энтропии) имеет недостаток – не указывает направление ответов респондентов на утверждение (из-за свойства симметричности логарифма). Если по трудности мы можем сказать, к каким ответам скорее тяготеют участники исследования (чем ниже значение, тем больше респонденты не соглашаются с утверждением

или выбирают меньшие значения на шкале ответов), то на основе энтропии это сделать без обращения к распределению ответов нельзя.

На основании вышеизложенного можно выдвинуть два практических соображения при анализе психодиагностических тестов:

1. Дополнять показатель трудности расчётом значения энтропии. По трудности специалисты смогут понять, в какую сторону смещены ответы респондентов; по энтропии – насколько далеко от оптимального находится трудность определённого утверждения и сколько информации об испытуемом оно даёт.
2. Использовать только показатель энтропии, а для исследования смещения ответов обращаться к распределению вероятностей выбора вариантов ответа (в виде таблиц или графиков).

Литература

Бурлачук Л.Ф. Психодиагностика: учебник для вузов. СПб.: Питер, 2011.

Иванова, Т.Ю. Функциональная роль личностных ресурсов в обеспечении психологического благополучия [канд. диссер.]. М.: Изд-во Моск. ун-та, 2016.

Карданова Е.Ю. Преимущества современной теории тестирования по сравнению с классической теорией тестирования. Вопросы тестирования в образовании, 2004, 10, 7-34.

Каргин Ю. Исследование взаимосвязи теории информации и теории педагогических измерений. Педагогические измерения, 2013, 2, 3-22.

Крокер Л., Алгина Д. Введение в классическую и современную теорию тестов. М.: Логос, 2012.

Лидовский В. В. Теория информации. М.: Компания Спутник+, 2004.

Маслак А.А. Теория и практика измерения латентных переменных в образовании. М.: Юрайт, 2016.

Колачев Н.И. Использовать ли показатель трудности при анализе психологических тестов? Осин Е.Н., Рассказова Е.И. Краткая версия теста жизнестойкости: психометрические характеристики и применение в организационном контексте. Вестник Московского университета. Серия 14. Психология, 2013, (2), 147-165.

Cover T.M., Thomas, J.A. Elements of Information Theory. New Jersey: Wiley-Interscience, 2006.

DiStefano C., Hess B. Using confirmatory factor analysis for construct validation: An empirical review. Journal of Psychoeducational Assessment, 2005, 23(3), 225-241. DOI:10.1177/073428290502300303

Embretson S.E., Reise S.P. Item Response Theory for Psychologists. Psychology Press, 2013. DOI:10.4324/9781410605269

Fan X. Item response theory and classical test theory: An empirical comparison of their item/person statistics. Educational and psychological measurement, 1998, 58(3), 357-381. DOI:10.1177/0013164498058003001

Ferguson K.E., Iverson G.L. Alternate Test Forms. In: Kreutzer, J.S., DeLuca, J., Caplan, B. (eds) Encyclopedia of Clinical Neuropsychology. NY: Springer, 2011. DOI:10.1007/978-0-387-79948-3_1170

Fischer G.H. Applying the principles of specific objectivity and of generalizability to the measurement of change. Psychometrika, 1987, 52(4), 565-587. DOI:10.1007/BF02294820

Kohli N., Koran J., Henn L. Relationships among classical test theory and item response theory frameworks via factor analytic models. Educational and Psychological Measurement, 2015, 75(3), 389-405. DOI:10.1177/0013164414559071

Li C.H. Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. Behavior Research Methods, 2016, 48(3), 936-949. DOI:10.3758/s13428-015-0619-7

Lord F. M., Novick M.R. Statistical theories of mental test scores. Addison-Wesley, 1968.

Колачев Н.И. Использовать ли показатель трудности при анализе психологических тестов?
Macdonald P., Paunonen S.V. A Monte Carlo comparison of item and person statistics based on item response theory versus classical test theory. Educational and psychological measurement, 2002, 62(6), 921-943. DOI:10.1177/0013164402238082

Rasch G. On specific objectivity: An attempt at formalizing the request for generality and validity of scientific statements. Danish Yearbook of Philosophy, 1977, 14, 58-94. DOI:10.1163/24689300-01401006

Rönkkö M., Cho E. An updated guideline for assessing discriminant validity. Organizational Research Methods, 2022, 25(1), 6-14. DOI:10.1177/1094428120968614

Schaufeli W.B., De Witte H., Desart S. (2020). Manual Burnout Assessment Tool (BAT) – Version 2.0. KU Leuven, Belgium: Unpublished internal report. Retrieved from https://burnoutassessmenttool.be/wp-content/uploads/2020/08/Test-Manual-BAT-English-version-2.0-1.pdf

Shannon C.E. A mathematical theory of communication. The Bell system technical journal, 1948, 27(3), 379–423. DOI:10.1002/j.1538-7305.1948.tb01338.x

Stark S., Chernyshenko O.S., Drasgow F. Detecting differential item functioning with confirmatory factor analysis and item response theory: toward a unified strategy. Journal of Applied Psychology, 2006, 91(6), 1292. DOI:10.1037/0021-9010.91.6.1292

Steyer R., Mayer A., Geiser C., Cole D.A. A theory of states and traits—Revised. Annual review of clinical psychology, 2015, 11, 71-98. DOI:10.1146/annurev-clinpsy-032813-153719

Sun K.T., Chen Y.J., Tsai S.Y., Cheng C.F. Creating IRT-based parallel test forms using the genetic algorithm method. Applied Measurement in Education, 2008, 21(2), 141–161. DOI:10.1080/08957340801926151

Приложения

Приложение 1. Распределение вероятностей выбора категории шкалы ответов, трудность и энтропия по каждому утверждению инструмента измерения профессионального выгорания

Категория шкалы ответов	Утв. 1	Утв. 2	Утв. 3	Утв. 4	Утв. 5	Утв. 6	Утв. 7	Утв. 8	Утв. 9	Утв. 10	Утв. 11	Утв. 12	Утв. 13	Утв. 14	Утв. 15	Утв. 16	Утв. 17	Утв. 18	Утв. 19	Утв. 20	Утв. 21	Утв. 22	Утв. 23
0 – «Никогда»	0.38	0.22	0.29	0.31	0.36	0.39	0.42	0.32	0.47	0.58	0.83	0.83	0.65	0.55	0.71	0.28	0.59	0.49	0.52	0.67	0.57	0.57	0.47
1 – «Редко»	0.23	0.35	0.34	0.33	0.33	0.35	0.32	0.32	0.33	0.25	0.10	0.09	0.18	0.30	0.18	0.43	0.26	0.28	0.36	0.26	0.33	0.33	0.41
2 – «Иногда»	0.32	0.32	0.26	0.29	0.23	0.22	0.21	0.26	0.16	0.11	0.06	0.05	0.13	0.14	0.10	0.25	0.13	0.19	0.10	0.07	0.09	0.10	0.12
3 – «Часто»	0.06	0.11	0.10	0.06	0.07	0.04	0.04	0.08	0.03	0.05	0.01	0.02	0.03	0.01	0.01	0.04	0.02	0.03	0.01	0.00	0.01	0.01	0.01
4 – «Постоянно»	0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.02	0.01	0.01	0.00	0.01	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Трудность	0.27	0.34	0.30	0.28	0.26	0.23	0.23	0.29	0.19	0.17	0.06	0.07	0.15	0.15	0.11	0.27	0.15	0.19	0.15	0.10	0.13	0.14	0.17
Энтропия	1.84	1.96	1.98	1.87	1.89	1.76	1.80	1.96	1.66	1.60	0.83	0.90	1.47	1.46	1.23	1.74	1.48	1.66	1.43	1.19	1.36	1.36	1.47

Приложение 2. Распределение вероятностей выбора категории шкалы ответов, трудность и энтропия по каждому утверждению теста жизнестойкости

Категория шкалы ответов	Утв. 1	Утв. 2	Утв. 3	Утв. 4	Утв. 5	Утв. 6	Утв. 7	Утв. 8	Утв. 9	Утв. 10	Утв. 11	Утв. 12
0 – «Нет»	0.04	0.06	0.03	0.03	0.04	0.01	0.04	0.06	0.06	0.03	0.05	0.16
1 – «Скорее нет»	0.12	0.20	0.13	0.08	0.15	0.11	0.12	0.23	0.23	0.13	0.12	0.18
2 – «Скорее да»	0.36	0.43	0.41	0.29	0.45	0.47	0.30	0.32	0.43	0.30	0.28	0.43
3 – «Да»	0.48	0.30	0.43	0.60	0.36	0.41	0.54	0.39	0.28	0.54	0.54	0.22
Трудность	0.76	0.66	0.74	0.82	0.71	0.76	0.78	0.68	0.64	0.78	0.77	0.58
Энтропия	1.58	1.76	1.60	1.39	1.63	1.47	1.55	1.80	1.77	1.53	1.59	1.88

Приложение 3. Распределение вероятностей выбора категории шкалы ответов, трудность и энтропия по каждому утверждению опросника рабочих требований и ресурсов

Категория шкалы ответов	Утв. 1	Утв. 2	Утв. 3	Утв. 4	Утв. 5	Утв. 6	Утв. 7	Утв. 8	Утв. 9	Утв. 10	Утв. 11	Утв. 12	Утв. 13
0 – «Очень редко, никогда»	0.03	0.02	0.01	0.00	0.10	0.12	0.02	0.06	0.15	0.28	0.29	0.16	0.02
1 – «Довольно редко»	0.08	0.07	0.01	0.03	0.07	0.13	0.08	0.17	0.17	0.22	0.43	0.20	0.06
2 – «Иногда»	0.20	0.18	0.14	0.23	0.35	0.33	0.35	0.53	0.44	0.34	0.26	0.46	0.41
3 – «Довольно часто»	0.35	0.34	0.43	0.39	0.26	0.25	0.43	0.19	0.17	0.11	0.02	0.15	0.35
4 – «Очень часто, постоянно»	0.34	0.39	0.41	0.36	0.22	0.17	0.13	0.06	0.07	0.06	0.00	0.02	0.15
Трудность	0.72	0.75	0.80	0.77	0.60	0.56	0.64	0.50	0.46	0.36	0.25	0.42	0.64
Энтропия	1.98	1.88	1.59	1.71	2.13	2.21	1.84	1.85	2.07	2.09	1.68	1.94	1.84

Поступила в редакцию: 26 мая 2022 г. Дата публикации: 23 июня 2023 г.

Сведения об авторах

Колачев Никита Игоревич. Психолог, стажёр-исследователь Международной лаборатории позитивной психологии личности и мотивации, Национальный исследовательский университет «Высшая школа экономики», ул. Славянская площадь, д. 4 стр. 2, 109240 Москва, Россия.

E-mail: nkolachev@hse.ru

Ссылка для цитирования

Колачев Н.И. Использовать ли показатель трудности при анализе психологических тестов? Психологические исследования. 2023. Т. 16, № 88. С. 1. URL: https://psystudy.ru

Адрес статьи: https://doi.org/10.54359/ps.v16i88.1374

Kolachev N.I.[1] Is the difficulty index to be used when analyzing psychological tests?

[1] National Research University Higher School of Economics, Moscow, Russia

The present article is aimed to examine the applicability of the difficulty index within the framework of psychometric analysis of psychological tests. It is demonstrated that this index is often disregarded in traditional psychometric research due to its absence in the paradigm of confirmatory factor analysis, which is commonly employed by most psychologists. Analytical and empirical examples are provided to illustrate that difficulty is analogous to entropy, a measure of information quantity, and does not possess a unit of measurement. A comparison of both indices is conducted to assess their respective advantages and disadvantages. Recommendations are provided regarding their utilization in studies on the quality of diagnostic methods.

Keywords: psychodiagnostics, entropy, difficulty index, discriminative power, confirmatory factor analysis, classical test theory, item response theory

References

Burlachuk L.F. Psihodiagnostika: uchebnik dlya vuzov. SPb.: Piter, 2011. (in Russian)

Cover T.M., Thomas J.A. Elements of Information Theory. New Jersey: Wiley-Interscience, 2006.

DiStefano C., Hess B. Using confirmatory factor analysis for construct validation: An empirical review. Journal of Psychoeducational Assessment. 2005, 23(3), 225-241. DOI:10.1177/073428290502300303

Embretson S.E., Reise S.P. Item Response Theory for Psychologists. Psychology Press, 2013. DOI:10.4324/9781410605269

Fan X. Item response theory and classical test theory: An empirical comparison of their item/person statistics. Educational and psychological measurement, 1998, 58(3), 357–381. DOI:10.1177/0013164498058003001

Ferguson K.E., Iverson G.L. Alternate Test Forms. In: Kreutzer, J.S., DeLuca, J., Caplan, B. (eds) Encyclopedia of Clinical Neuropsychology. 2011, Springer. DOI:10.1007/978-0-387-79948-3_1170

Колачев Н.И. Использовать ли показатель трудности при анализе психологических тестов?
Fischer G. H. Applying the principles of specific objectivity and of generalizability to the measurement of change. Psychometrika, 1987, *52*(4), 565-587. DOI:10.1007/BF02294820

Ivanova, T.YU. Funkcional'naya rol' lichnostnyh resursov v obespechenii psihologicheskogo blagopoluchiya [kand. disser.]. M.: Izd-vo Mosk. un-ta, 2016.

Kardanova E.YU. Preimushchestva sovremennoj teorii testirovaniya po sravneniyu s klassi-cheskoj teoriej testirovaniya. Voprosy testirovaniya v obrazovanii, 2004, 10, 7-34.

Kargin YU. Issledovanie vzaimosvyazi teorii informacii i teorii pedagogicheskih izmerenij. Pedagogicheskie izmereniya, 2013, 2, 3-22. (in Russian)

Kohli N., Koran J., Henn L. Relationships among classical test theory and item response theory frameworks via factor analytic models. Educational and Psychological Measurement, 2015, 75(3), 389-405. DOI:10.1177/0013164414559071

Kroker L., Algina D. Vvedenie v klassicheskuyu i sovremennuyu teoriyu testov. M.: Logos, 2012. (in Russian)

Li C. H. Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. Behavior Research Methods, 2016, 48(3), 936-949. DOI:10.3758/s13428-015-0619-7

Lidovskij V. V. Teoriya informacii. M.: Kompaniya Sputnik+, 2004. (in Russian)

Lord F. M., Novick M. R. Statistical theories of mental test scores. Addison-Wesley, 1968.

Macdonald P., Paunonen S. V. A Monte Carlo comparison of item and person statistics based on item response theory versus classical test theory. Educational and psychological measurement, 2002, 62(6), 921-943. DOI:10.1177/0013164402238082

Maslak A. A. Teoriya i praktika izmereniya latentnyh peremennyh v obrazovanii. M.: YUrajt, 2016. (in Russian)

Колачев Н.И. Использовать ли показатель трудности при анализе психологических тестов? Osin E. N., Rasskazova E. I. Kratkaya versiya testa zhiznestojkosti: psihometricheskie ha-rakteristiki i primenenie v organizacionnom kontekste. Vestnik Moskovskogo universi-teta. Seriya 14. Psihologiya, 2013, (2), 147-165.

Rasch G. On specific objectivity: An attempt at formalizing the request for generality and validity of scientific statements. Danish Yearbook of Philosophy, 1977, 14, 58-94. DOI:10.1163/24689300-01401006

Rönkkö M., Cho E. An updated guideline for assessing discriminant validity. Organizational Research Methods, 2022, 25(1), 6-14. DOI:10.1177/1094428120968614

Schaufeli W.B., De Witte H., Desart S. (2020). Manual Burnout Assessment Tool (BAT) – Version 2.0. KU Leuven, Belgium: Unpublished internal report. Retrieved from https://burnoutassessmenttool.be/wp-content/uploads/2020/08/Test-Manual-BAT-English-version-2.0-1.pdf

Shannon C.E. A mathematical theory of communication. The Bell system technical journal, 1948, 27(3), 379–423. DOI:10.1002/j.1538-7305.1948.tb01338.x

Stark S., Chernyshenko O. S., Drasgow F. Detecting differential item functioning with confirmatory factor analysis and item response theory: toward a unified strategy. Journal of Applied Psychology, 2006, 91(6), 1292. DOI:10.1037/0021-9010.91.6.1292

Steyer R., Mayer A., Geiser C., Cole D. A. A theory of states and traits—Revised. Annual review of clinical psychology, 2015, 11, 71-98. DOI:10.1146/annurev-clinpsy-032813-153719

Sun K. T., Chen Y. J., Tsai S. Y., Cheng C. F. Creating IRT-based parallel test forms using the genetic algorithm method. Applied Measurement in Education, 2008, 21(2), 141–161. DOI:10.1080/08957340801926151

Information about authors

Kolachev Nikita Igorevich. Psychologist, Research Assistant at the International Laboratory of Positive Psychology of Personality and Motivation, HSE University, Slavyanskaya ploshchad', 4, building 2, 109240 Moscow, Russia.

Колачев Н.И. Использовать ли показатель трудности при анализе психологических тестов?
E-mail: nkolachev@hse.ru

For citation: Kolachev N.I. Is the difficulty index to be used when analyzing psychological tests? Psikhologicheskie Issledovaniya, 2023, Vol. 16, No. 88, p. 1. https://psystudy.ru