

ШКАЛЫ АТИПИЧНОСТИ ОТВЕТОВ КАК ИНСТРУМЕНТ ВЫЯВЛЕНИЯ НЕКООПЕРАТИВНОГО ТЕСТОВОГО ПОВЕДЕНИЯ

К.В. Сугоняев

Военная академия Генерального штаба ВС РФ, г. Москва

Некооперативное тестовое поведение (НКТП) представляет собой серьезную, но часто недооцениваемую угрозу достоверности тестовых оценок и валидности их интерпретации. Одним из методов выявления НКТП может быть включение в состав многошкальных опросников шкал атипичности ответов (АТ), построенных на учете частоты подтверждения тех или иных опций ответа. Учитывая практическую важность задачи различения реалистических данных от проявлений НКТП при проведении массовых обследований военнослужащих, осуществлена разработка и включение в компьютерные версии ряда многошкальных опросников шкал АТ в сочетании с настройкой алгоритмов выявления хаотического реагирования, как одного из наиболее частых и деструктивных стилей НКТП. Проверка дифференцирующей способности шкал АТ осуществлялась путем сопоставления эмпирических данных, полученных в результате обследования восьми независимых выборок ($N = 406-5689$) с помощью шести опросников с разным форматом ответов, и множествами программно сгенерированных протоколов со случайным выбором ответов. Показано, что в среднем около 95 % сгенерированных таким образом протоколов идентифицируются как недостоверные, в том числе, 65 % из них опознаны как хаотические. Практика применения оснащенных шкалами АТ опросников в сочетании с результатами программного моделирования позволяют предположить, что включение подобных шкал в известные и вновь разрабатываемые опросники может существенно снизить риски ошибочных диагностических выводов.

Ключевые слова: достоверность, некооперативное тестовое поведение, атипичные ответы, конструирование опросников.

Введение

Проблема обеспечения достоверности личностных измерений относится к числу остро дискуссионных и нередко поляризует позиции организационных психологов: в то время как одни ратуют за широкое использование личностных опросников (ЛО) при отборе кандидатов (Hogan, 2007; Hough, 2008), другие оспаривают их пригодность к применению в ситуациях с «высокими ставками» обосновывая это их потенциально высокой подверженностью намеренным искажениям (Morgeson, 2007). Однако в этих дискуссиях основное внимание приковано к такой угрозе достоверности, как смещение тестовых оценок вследствие преднамеренных попыток респондентов приукрасить свой имидж (данный феномен часто обозначается как *социальная желательность*), тогда как прочие источники контаминации данных самоотчета остаются в тени (Сугоняев, 2009; Johnson, 2005).

Наши исследования, выполненные на достаточно больших выборках с применением различных многошкальных ЛО, показывают, что влияние социальной желательности не является настолько деструктивным, как это принято считать, по крайней мере – в контексте отбора кандидатов на работу в массовых профессиях (Сугоняев, 2013, 2015). В то же время следует высказать наше убеждение в том, что более серьезной (и при этом существенно недооцененной) является проблема выявления и исключения из анализа протоколов обследования, являющихся продуктом невнимательного или намеренно некооперативного тестового поведения (НКТП), при котором респондент имитирует ответы на пункты опросника безотносительно к тексту инструкции и содержанию пунктов. В англоязычной литературе подобный стиль реагирования обозначается как *аберрантное* (Karabatsos, 2003; Conijn, 2015), *небрежное* (Meade, 2012; Kam,

2015), *реагирование с недостаточными усилиями* (Huang, 2012, 2015) и т.п.

Считается, что НКТП редко встречается при кадровом отборе, однако при проведении административных обследований уже зачисленного на работу или службу персонала доля лиц, прибегающих к подобной тактике, может достигать нескольких процентов (Meade, 2012; Maniaci, 2014; Huang, 2015). Вероятность НКТП возрастает при проведении исследований на обучающихся (студентах, курсантах и др.), особенно в случае проведения обследования в анонимном варианте. Между тем, присутствие в составе выборки даже относительно небольшой доли лиц, избирающих тактику НКТП, серьезно деформирует психометрические характеристики (ПМХ) ЛЮ, включая надежность, структурную и внешнюю валидность шкал (Crede, 2010; Kam, 2015). Поскольку для исследовательских публикаций академического характера довольно типично доминирование исследований, выполненных на студенческих выборках, отсутствие соответствующего контроля достоверности исходных данных может компрометировать валидность представленных в них результатов.

Показано, что НКТП в наибольшей степени искажает ПМХ шкал, в которых большинство ответов добросовестных респондентов группируется вблизи одного из полюсов этих шкал (Meade, 2012). Именно такого рода результаты демонстрируют клинические опросники, особенно в тех случаях, когда их пытаются применять на нормативных популяциях. Впрочем, как показывают наши наблюдения, вопреки распространенной апологизации нормального распределения, практически невозможно назвать какой-либо многошкальный ЛЮ (даже пользующийся репутацией нормативного), в составе которого не нашлось хотя бы нескольких шкал, характеризующихся смещенным и асимметричным распределением оценок, полученных в условиях административного обследования.

В литературе по рассматриваемой проблеме можно найти ряд рекомендаций, направленных на выявление недостоверных протоколов. Среди них – включение в состав опросника фиктивных пунктов (содержащих нереалистичные утверждения), пунктов с однозначными ответами, сравнение семантических и психометрических синонимов или антонимов, отслеживание длинных последова-

тельностью однотипных ответов испытуемых, учет времени реагирования на пункт и др. (Johnson, 2005; Meade, 2012; Desimone, 2015). К сожалению, некоторые из таких рекомендаций выполнимы только при интерактивном обследовании; другие предполагают вмешательство в содержание опросника и включение в него дополнительных пунктов, что не всегда бывает удобно.

Следует отметить, что многие авторы, обращаясь к теме НКТП, отмечают отсутствие идеального способа выявления подобного стиля реагирования и ратуют за применение комплекса мер для обеспечения чистоты данных (Karabatsos, 2003; Johnson, 2005; Desimone, 2015 и др.). Одним из самых ранних способов выявления недостоверных протоколов стало включение в состав ЛЮ шкал валидности, основанных на анализе частоты выбора тех или иных опций ответов. Именно этот принцип был положен в основу формирования таких шкал, как *F* в ММРІ, *Сm* в СРІ и т.п. Подобные шкалы присутствуют во многих зарубежных ЛЮ и продолжают разрабатываться для новых опросников (например, Suhr, 2011). В то же время в оригинальных отечественных разработках такие шкалы почти не встречаются. Реальное функционирование шкал валидности в отечественных адаптациях зарубежных ЛЮ также не привлекает к себе должного внимания, а заимствованная из первоисточника интерпретация воспринимается как нечто, не подлежащее сомнению. Между тем, проведенный нами ранее анализ больших массивов данных, полученных с применением отечественных версий методик ММРІ и СРІ, показал, что среди пунктов контрольных шкал *F* и *Сm* встречаются такие, в которых вероятность выбора ключевого ответа приближается к 50 % (Сугоняев, 2009). Присутствие пунктов с такими характеристиками противоречит сущности данных шкал (*infrequency / commonality*) и ведет к переоценке доли «невалидных» протоколов.

Практика психологического отбора и сопровождения профессиональной деятельности военнослужащих ограничивает возможности психолога уделять много времени каждому сотруднику. В этих условиях методы группового психометрического тестирования являются важнейшим, а иногда даже единственно возможным источником информации о профессионально релевантных индивидуальных особенностях обследуемых. Внедрение

средств компьютерной психодиагностики и методов психометрического анализа данных позволило еще десятилетие назад установить, что формируемые по результатам скрининговых обследований группы «повышенного психологического внимания» оказываются в той или иной степени инфильтрованными лицами, избравшими тактику НКТП. Таким образом, неумение дифференцировать «проблемные» личностные профили от тех, которые являются следствием НКТП, способно направить работу практического психолога по ложному пути, рассеивая его внимание между теми, кто действительно нуждается в психологической помощи, и теми, кто заслуживает не более чем воспитательных воздействий. Именно поэтому, начиная с 1998 года, мы начали оснащать компьютерные версии ЛО, широко применяемых в силовых структурах, шкалами, направленными на выявление НКТП и получившими общее наименование «*атипичность ответов*» (АТ).

Практическая реализация

В шкалу АТ рекрутируются наиболее «трудные» (наименее популярные) варианты ответов на пункты ЛО. Таким образом, основной вклад в шкалу АТ вносят пункты с асимметричным распределением вариантов ответа, у которых информационная функция смещена в сторону одного из крайних уровней выраженности измеряемого свойства. Практика применения шкал АТ показала, что для обеспечения приемлемой дискриминативности шкала должна быть достаточно мощной (состоять из нескольких десятков пунктов). Достижение таких параметров возможно для тех опросников, которые имеют относительно большое произведение длины теста на число опций ответа (более 200). Обычно в состав шкалы АТ включаются лишь те варианты ответов, вероятность выбора которых в выборке стандартизации не превышает 5–6 %, однако в случае дефицита таких пунктов этот лимит может быть увеличен до 8–10 %. В шкале АТ не должно быть преобладания пунктов, общих с какой-то одной из базисных шкал; следует избегать включения пунктов, дифференциально функционирующих в различных сегментах целевой популяции (например, гендерных); желательно стремиться к балансу полярностей включенных в шкалу опций ответов. Суммирование баллов по ключу в шкале АТ может происходить одним из двух спо-

собов: 1) за каждое «попадание» в редко выбираемую опцию ответа начисляется 1 балл; 2) начисляется взвешенный балл, учитывающий частоту выбора опции в популяции (чем она ниже, тем больше баллов начисляется). Опыт применения обоих способов показал превосходство первого в точности идентификации хаотического реагирования.

В ЛО с дихотомической шкалой ответов максимальный теоретический балл по шкале АТ равен числу включенных в шкалу опций ответа. В ЛО с полихотомической шкалой ответов число опций ответа может превышать число пунктов, из которых они рекрутированы, поскольку наиболее асимметричные пункты могут быть представлены в ключе шкалы АТ двумя–тремя и более опциями.

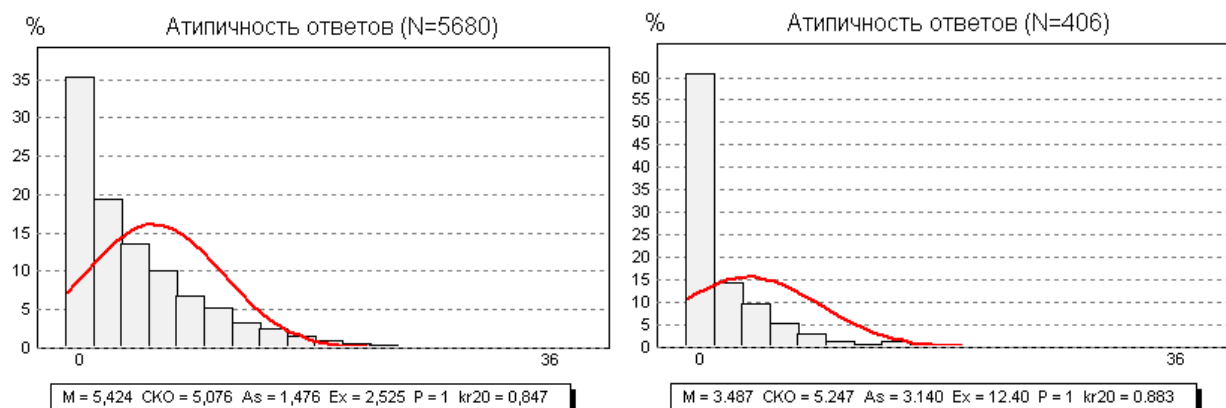
Хотя в работе Р. Тетта с соавт. (Tett, 2009) было показано, что для получения стабильных тестовых норм на обычные содержательные шкалы ЛО может быть достаточно выборки стандартизации 100–200 человек, представляющих целевую популяцию, для получения стабильной шкалы АТ – из-за того, что она формируется из редко подтверждаемых опций ответа – желательны более многочисленные выборки (500–1000 и более респондентов).

Распределение оценок по хорошо сконструированной шкале АТ близко к экспоненциальному: абсолютное большинство респондентов имеет нулевые или близкие к нулю оценки, и лишь незначительное меньшинство получает более высокие баллы, имеющие диагностическое значение. Примеры гистограмм распределения оценок по шкалам АТ представлены на рисунке.

Оценки по шкалам АТ могут выражаться в любой метрике, но для лучшего понимания особенностей их функционирования и эффективного использования шкалы в качестве инструмента оценки достоверности протокола более удобным является ее встраивание в график индивидуального личностного профиля, в котором все шкальные оценки представлены в процентах от теоретического максимума.

Наиболее частой причиной повышения оценок по шкале АТ является намеренное НКТП, которое может проявляться в тенденции давать быстрые случайные, либо однотипные, либо периодически чередующиеся ответы.

В случае *хаотического реагирования* обычно регистрируется оценка АТ, близкая к



Типичные гистограммы распределения оценок по шкале АТ

вероятности «случайного попадания» (50 % для тестов с дихотомическим форматом ответов, 33 % – для тестов с трихотомическим выбором и т.д.), в сочетании со сглаженным профилем оценок по содержательным шкалам. В ЛО, ориентированном на оценивание как социально одобряемых, так и менее желательных характеристик, подобный профиль может заметно отличаться от типичного рельефа, который в разработанных под нашим руководством компьютерных психодиагностических комплексах (КПДК) ряда «Мультипсихометр» (Сугоняев, 2000, 2014) отмечается двумя линиями «нормативного коридора» (от $M-\sigma$ до $M+\sigma$).

При использовании ЛО, направленных на диагностику пограничных психических состояний или признаков девиантного поведения, иногда можно зарегистрировать паттерны ответов, характерных для *аггравации*, т.е. стремления респондента преувеличить тяжесть имеющихся у него проблем. Хотя подобная тактика реагирования более характерна для ситуаций консультирования и может быть мотивирована стремлением испытуемого привлечь к себе внимание психолога с целью получения какой-либо помощи, аггравация возможна и при аттестационном тестировании. В последнем случае она может использоваться как способ выражения протеста против вынужденного обследования или как средство достижения более прагматических целей (например, уклонения от выполнения нежелательной миссии). В подобных случаях значения по шкале АТ могут превышать вероятность случайного попадания; отмечается заметное повышение оценок по шкалам, оценивающим нежелательные характеристики, и снижение уровня шкал, измеряющих социаль-

но одобряемые свойства. Как следствие, показатели корреляции индивидуального профиля с усредненным профилем популяции смещаются в область отрицательных значений.

Иногда можно зафиксировать результат, при котором индивидуальный профиль оказывается практически зеркальным отражением нормативного коридора, а значение по шкале АТ приближается к теоретическому максимуму. Подобный результат с высокой вероятностью отражает ненамеренное перепутывание полярности ответов респондентом, невнимательно изучившим инструкцию к тесту.

Наибольшие сложности для диагностики представляют случаи, когда регистрируется существенный подъем оценок по шкале АТ (до 9–10 стэнов), не достигающий уровней, характерных для «случайного попадания». В подобных ситуациях возникает проблема дифференциации наличия реальных личностных проблем от дефицита внимания при выполнении теста (например, развития скуки в ходе выполнения длинного опросника) или недостаточного уровня владения языком; иногда к сходному результату приводит уклончивый стиль ответов, проявляющийся, например, в предпочтении неопределенных вариантов в шкале Лайкерта с нечетным числом уровней. Оценку валидности такого протокола облегчил бы расчет дополнительных количественных индексов, характеризующих стилистические особенности ответов, например, индексы «соглашательства», «закрытости», поляризации и др. Полезная информация может быть извлечена также из сопоставления содержания ответов на конкретные пункты: например, обнаружение неоднократных смысловых противоречий в ответах может указывать на НКТП или неустойчивость внимания.

Если обследование происходит в диалоговом формате, признаки НКТП могут быть замечены уже на ранних его этапах. Представляется, что лучше сократить вероятность попадания недостоверных протоколов в базу данных до минимума, нежели тратить время на безучастную регистрацию любых действий респондента с тем, чтобы по завершении процедуры тестирования вынести вердикт о недостоверности полученных данных. Поэтому в упомянутых выше КПДК в процедуру диалогового тестирования встроены предложенные нами алгоритмы, направленные на выявление таких признаков¹, за которым следует вывод на экран монитора транспарантов, предупреждающих респондента о нецелесообразности попыток «обмануть тест» и побуждающих его к добросовестному выполнению методики.

Однако практика эксплуатации КПДК в силовых структурах показывает, что наиболее популярным форматом обследования является бланковый. В этом случае временные характеристики ответов обследуемых неизвестны и основной способ выявления признаков НКТП сводится к апостериорному анализу массива ответов на пункты, и прежде всего – расчету оценок по шкале АТ. Реализованные в КПДК программные механизмы, обеспечивающие формирование интерпретации индивидуальных протоколов, позволяют сопоставлять различные признаки между собой и с заданными критериями. При обнаружении признаков НКТП в начало генерируемой интерпретации результатов выводится один из текстовых фрагментов, смысл которых может быть сведен к трем ситуациям: 1) хаотическое реагирование; 2) аггравация; 3) сомнительная достоверность (без явных признаков первых двух).

К настоящему времени шкалы АТ разработаны для нескольких десятков заимствованных и адаптированных методик, реализованных в КПДК. Среди них – все доступные версии 5- и 16-факторных ЛО, 6-факторный ЛО (на базе НЕХАСО-60), опросники межличностной диагностики Т. Лири, жизнестойкости С. Мадди, карьерных ориентаций Э. Шейна, ценностных портретов Ш. Шварца, черт характера для взрослых В. Русалова и

О. Маноловой, опросники «Склонность к отклоняющемуся поведению» А. Орла, «Шкала базисных убеждений» Р. Яноф-Бульман, «Прогноз», «Адаптивность» и др. Кроме того, подобные шкалы разрабатываются и включаются во все без исключения многошкальные ЛО, сконструированные нами за последнее десятилетие, включая все версии методик «Личностный профиль» (ЛП) и «Личностный опросник для скрининга кандидатов» (ЛОСК), а также «Личностный опросник руководителя» (ЛОР), «Личностный опросник контрактника» (ЛОК), «Комплексная оценка психологического состояния» (КОПС), «Субъективное восприятие ситуации» (СВС), «Диагностика межличностных диспозиций» (ДМД), «Мотивационный профиль обучающегося» (МПО) и др. Следует отметить, что состав шкал АТ является популяционно и контекстно специфическим, поэтому для некоторых тестов разработана более чем одна версия шкал.

Следующий после формирования шкалы АТ этап состоит в нахождении оптимальной разделительной оценки, превышение которой позволяет заподозрить факт НКТП. Для первых по времени разработки шкал АТ отсекающие оценки подбирались методом проб и ошибок. Однако в дальнейшем было установлено, что граничное значение АТ, позволяющее с наименьшей ошибкой дифференцировать достоверные данные от сомнительных, может быть определено на основе статистик шкалы АТ, рассчитанных для множества случайно сгенерированных протоколов путем вычитания двух стандартных отклонений из средней оценки AT_R .

Валидизация

С целью исследования критериальной валидности шкал АТ было проведено сопоставление оценок АТ, рассчитанных по результатам реальных административных обследований различных категорий военнослужащих, с аналогичными оценками, полученными из искусственно сгенерированных протоколов со случайным выбором ответов (имитация хаотического реагирования).

Для исследования были выбраны шесть опросников, различающихся по числу опций ответа (от двух до семи), для каждого из которых имелись электронные протоколы обследований относительно однородных выборок, численностью не менее 400 человек. В это число вошли (в порядке возрастания числа

¹ Среди них – сверхбыстрые ответы, длинные цепочки однотипных ответов, превышение текущего лимита по шкалам достоверности; содержание транспарантов учитывает эти различия.

опций ответа): модифицированная версия индивидуально-типологического опросника (ИТО+); «Личностный профиль» (ЛП); «Личностный опросник контрактника» (ЛОК); 6-факторный личностный опросник (6-ФЛО); шкала базисных убеждений (ШБУ); модифицированный опросник карьерных ориентаций (ОКО). Учитывая имеющиеся в литературе свидетельства о возможном влиянии лингвистической компетентности / образовательного уровня на оценки, получаемые с помощью ЛО (Johnson, 2005; Rammstedt, 2010), один из опросников (ИТО+) был представлен тремя выборками, различающимися по этим параметрам (военнослужащие срочной службы, курсанты военных вузов, офицеры).

Для каждого из этих опросников с помощью встроенного в КПДК программного модуля было сгенерировано по 100 протоколов со случайным выбором ответов, которые были обработаны по тем же ключам, которые применяются для реальных данных.

Сопоставление эмпирических и сгенерированных данных включало оценивание степени перекрытия распределения баллов по шкале АТ между ними и способности механизмов автоматизированной оценки качества данных выявлять явно недостоверные протоколы.

Данные, представленные в табл. 1, свидетельствуют о том, что стандартизированные различия между средними оценками АТ в реальных и сгенерированных данных весьма велики: за исключением выборки военнослужащих срочной службы, выполнявших ИТО+, значения коэффициента d Коэна превышают 3,7. Это достигается благодаря тому, что

область совмещения гистограмм распределения оценок АТ между эмпирическими и сгенерированными данными составляет всего несколько процентов.

Обращает на себя внимание то, что значения коэффициента d варьируют от выборки к выборке и от теста к тесту. Анализ этих вариаций позволяет предположить, что они зависят как от технических характеристик шкалы АТ, так и от характеристик выборки. Среди первых наиболее вероятными факторами, влияющими на различительную способность оценок АТ, могут быть относительная мощность² шкалы АТ и величина порога включения опции в состав шкалы АТ. Рост различительной способности шкал АТ при переходе от выборки солдат к офицерской (на примере ИТО+) дает основания предположить ее прямую связь с лингвистической компетентностью популяции. Для более определенных выводов необходимы дополнительные исследования, включающие большее количество выборок, обследованных с помощью ЛО, оснащенных шкалами АТ.

Способность формализованных механизмов интерпретации результатов распознавать искусственно сгенерированные протоколы как недостоверные (в том числе – хаотические) иллюстрируется данными, представленными в табл. 2. Они подтверждают зависимость дифференцирующей способности шкал АТ от их относительной мощности. Что же касается идентификации хаотического реагирования, то, поскольку она осуществляется на основе учета формы профиля шкальных оценок, ее точность падает со снижением мощности содержательных шкал ЛО. Это связано с

Таблица 1

Сравнение дескриптивных статистик реальных и сгенерированных данных

Характеристики опросника			Реальные данные		Сгенерированные данные	Различие (d Коэна)
Обозначение	Кол-во опций ответа	Мощность шкалы АТ	Выборка	$M \pm \sigma$	$M \pm \sigma$	
			ИТО+	133×2	40	5689*
593**	2,578±4,187	4,374				
1022***	1,931±2,847	6,659				
ЛП-1	118×	54	2395***	4,387±4,78	25,27±3,29	4,369
ЛОК	180×	114	406**	3,487±5,247	37,31±4,885	6,446
6-ФЛО	75×5	48	957**	3,547±3,368	16,18±3,128	3,751
ШБУ	37×	37	721**	1,918±2,906	13,34±2,924	3,93
ОКО	50×	44	4320***	2,231±2,808	15,19±2,98	4,615

Примечание. * военнослужащие срочной службы; ** офицеры; *** курсанты военных вузов.

Таблица 2

Способность шкал АТ и сопутствующих настроек программного интерпретатора идентифицировать недостоверные данные

Опросник	Заключения (в % от объема сгенерированных данных)			
	Хаотическое реагирование	Достоверность сомнительна	Аггравация	Признаки НКТП не определены
ИТО+	67	32	–	1
ЛП-1	72	23	1	4
ЛОК	79	20	–	1
6-ФЛО	54	40	–	6
ШБУ	74	20	–	6
ОКО	45	47	–	8

тем, что при маломощных шкалах возрастает вероятность формирования – даже при хаотическом реагировании – рельефного профиля оценок, неотличимого от полученного в результате добросовестного выполнения методики (как это имеет место, например, в случае ОКО).

Представленные в табл. 2 данные показывают, что более 90 % искусственно сгенерированных протоколов опознаны как имеющие признаки недостоверности, в том числе, от 45 до 79 % идентифицированы как хаотические. Хотя в отношении 20–47 % сгенерированных протоколов вынесено более осторожное заключение («достоверность сомнительна»), предполагающее вовлечение психолога в принятие окончательного решения, подобные подсказки со стороны программы являются полезным подспорьем в его работе, так как в массивах исследованных нами реальных данных, полученных в ходе административного тестирования, доля проблемных протоколов обычно не превышает 1–2 %.

Привлекательность предлагаемого способа выявления НКТП состоит в том, что шкала АТ формируется из наименее популярных вариантов ответов на пункты, уже имеющиеся в составе ЛО, т. е. никакие изменения в содержание методики при этом не вносятся. Другим его достоинством является то, что, в отличие от методов выявления person-misfit, базирующихся на современной тестовой теории (IRT), настройка таких шкал посильна любому психологу (в том числе не обладающему познаниями в методологии IRT), располагающему достаточным количеством протоколов с детализацией до ответов на отдельные пункты.

Ограничением данного подхода к выявлению НКТП является то, что работоспособные шкалы АТ могут разрабатываться только на базе результатов проведения масштабных

административных обследований, и лишь после этого могут быть использованы для скрининга протоколов сомнительной достоверности в ситуациях, характеризующихся меньшей заинтересованностью респондентов в результатах тестирования.

Выводы

1. Наш опыт психометрического анализа достоверности разнообразных и значительных по объему массивов данных свидетельствует о том, что практически все они в той или иной степени контаминированы проявлениями НКТП. Их своевременное и надежное выявление позволяет пользователям ЛО верифицировать валидность оценок по содержательным шкалам, предотвратить потери времени на анализ недостоверных данных, повысить доверие к психологическому тестированию и его результатам со стороны респондентов и заказчиков.

2. Учет оценок по шкалам АТ выгодно отличается от альтернативных методов выявления НКТП тем, что он не требует введения в содержание ЛО дополнительных пунктов и использования сложных математических расчетов. Шкалы АТ обладают достаточно высокой дифференцирующей способностью, сопоставимой с гораздо более сложными методами, основанными на IRT, и способны выполнять возложенные на них функции как при диалоговом компьютерном, так и бланковом форматах обследования.

3. Эффективность шкал АТ является тем большей, чем более мощным является ЛО и входящие в его состав содержательные шкалы. Короткие и ультракороткие ЛО требуют разработки иных подходов к различению профилей лиц с проявлениями НКТП от профиля добросовестного респондента с признаками дезадаптации.

Литература

1. Сугоняев, К.В. Аппаратно-программный комплекс «Мультиспихометр-03» как потенциальное средство стандартизации измерений в профессиональной психодиагностике / К.В. Сугоняев // *Двойные технологии: спецвыпуск*. – 2000. – С. 12–14.
2. Сугоняев, К.В. Проблемы повышения достоверности личностных измерений: существующая практика и возможные усовершенствования / К.В. Сугоняев // *Здоровый образ жизни и физическое воспитание студентов и слушателей вузов: материалы науч.-практ. конф.*, 9 апр. 2009 г., г. Москва / под ред. А.В. Карасева, Е.А. Разумовского, В.А. Собины. – М.: ИНЭП, 2009. – С. 199–210.
3. Сугоняев, К.В. Немного правды о шкалах «лжи» / К.В. Сугоняев // *Здоровый образ жизни и физическое воспитание студентов и слушателей вузов: материалы XI межвузовской науч.-практ. конф.*, 4 апр. 2013 г., г. Москва / под ред. А.В. Карасева, В.А. Собины. – М.: ИНЭП, 2013. – С. 190–199.
4. Сугоняев, К.В. Разработка средств автоматизации профессиональной психодиагностики в РВСН: исторический очерк / К.В. Сугоняев, Н.В. Лазарев // *20 лет функционирования системы профессионального психологического отбора в Вооруженных Силах Российской Федерации, итоги и перспективы совершенствования: сб. материалов воен.-науч. конф. (Москва, 22 апр. 2014 г.) / под общ. ред. С.В. Чваркова*. – М.: ВАГШ ВС РФ, 2014. – С. 242–251.
5. Сугоняев, К.В. Что отражают шкалы социальной желательности? / К.В. Сугоняев // *Современная психодиагностика России. Преодоление кризиса: сб. материалов III Всерос. конф. по психол. диагностике: в 2 т. / редкол.: Н.А. Батулин (отв. ред.) и др.* – Челябинск: ЮУрГУ, 2015. – Т. 2. – С. 169–173.
6. Conijn, J.M. Detecting and explaining aberrant responding to the Outcome Questionnaire–45 / J.M. Conijn, W.H.M. Emons, K.D. Jong, K. Sijtsma // *Assessment*. – 2015. – Vol. 22, no. 4. – P. 513–524.
7. Credé, M. Random responding as a threat to the validity of effect size estimates in correlational research // *Educational and Psychological Measurement*. – 2010. – Vol. 70, no. 4. – P. 596–612.
8. Desimone, J.A. Best practice recommendations for data screening / J.A. Desimone, P.D. Harms, A.J. Desimone // *Journal of Organizational Behavior*. – 2015. – Vol. 36, no. 2. – P. 171–181.
9. Hogan, J. Personality measurement, faking, and employment selection / J. Hogan, P. Barrett, R. Hogan // *Journal of Applied Psychology*. – 2007. – Vol. 92, no. 5. – P. 1270–1285.
10. Hough, L.M. Personality testing and Industrial-organizational psychology: Reflections, progress, and prospects / L.M. Hough, F.L. Oswald // *Industrial and Organizational Psychology*. – 2008. – Vol. 1, no. 3. – P. 272–290.
11. Huang, J.L. Detecting and deterring insufficient effort responding to surveys / J.L. Huang, P.G. Curran, J. Keeney et al. // *Journal of Business and Psychology*. – 2012. – Vol. 27, no. 1. – P. 99–114.
12. Huang, J.L. Detecting insufficient effort responding with an infrequency scale: Evaluating validity and participant reactions / J.L. Huang, N.A. Bowling, M. Liu, Y. Li // *Journal of Business and Psychology*. – 2015. – Vol. 30, no. 2. – P. 299–311.
13. Johnson, J.A. Ascertaining the validity of individual protocols from Web-based personality inventories / J.A. Johnson // *Journal of Research in Personality*. – 2005. – Vol. 39, no. 1. – P. 103–129.
14. Kam, C.C.S. How careless responding and acquiescence response bias can influence construct dimensionality: The case of job satisfaction / C.C.S. Kam, J.P. Meyer // *Organizational Research Methods*. – 2015. – Vol. 18, no. 3. – P. 512–541.
15. Karabatsos, G. Comparing the aberrant response detection performance of thirty-six person-fit statistics / G. Karabatsos // *Applied Measurement in Education*. – 2003. – Vol. 16, no. 4. – P. 277–298.
16. Maniaci, M.R. Caring about carelessness: Participant inattention and its effects on research / M.R. Maniaci, R.D. Rogge // *Journal of Research in Personality*. – 2014. – Vol. 48. – P. 61–83.
17. Meade, A.W. Identifying careless responses in survey data / A.W. Meade, S.B. Craig // *Psychological Methods*. – 2012. – Vol. 17, no. 3. – P. 437–455.
18. Morgeson, F.P. Reconsidering the use of personality tests in personnel selection contexts / F.P. Morgeson, M.A. Campion, R.L. Dipboye, J.R. Hollenbeck et al. // *Personnel Psychology*. – 2007. – Vol. 60, no. 3. – P. 683–729.
19. Rammstedt B., Goldberg L.R., Borg I. The measurement equivalence of Big-Five factor markers for persons with different levels of education // *Journal of Research in Personality*. – 2010. – Vol. 44, no. 1. – P. 53–61.
20. Suhr, J.A. Development of an infrequency index for the CAARS / J.A. Suhr, M. Buelow, T. Riddle // *Journal of Psychoeducational Assessment*. – 2011. – Vol. 29, no. 2. – P. 160–170.
21. Tett, R.R. The use of personality test norms in work settings: Effects of sample size and relevance / R.R. Tett, J.R. Fitzke, P.L. Wadlington et al. // *Journal of Occupational and Organizational Psychology*. – 2009. – Vol. 82, no. 3. – P. 639–659.

Сугоняев Константин Владимирович, кандидат технических наук, доцент, старший научный сотрудник Научно-практического центра, Военная академия Генерального штаба ВС РФ (Москва), skv-354@yandex.ru

Поступила в редакцию 29 января 2016 г.

ATYPICAL RESPONSES SCALES AS THE TOOL OF NON-COOPERATIVE TESTING BEHAVIOR DETECTION

K.V. Sugonyaev, *skv-354@yandex.ru*

Military Academy of General Staff, Moscow, Russian Federation

Non-cooperative testing behavior (NCTB) during questionnaire completion is serious threat to the valid interpretations of test scores. There are a number of methods recommended for use to identify protocols with NCTB features in order to enhance the credibility of personality testing results (Karabatsos, 2003; Desimone, 2015), but not all of them are proper and accessible for practitioners. One of the earliest methods of aberrant responding detection is based on incorporation of validity scales, including those composed of response options with rare endorsement. Assumptions and considerations regarding design (without adding supplementary items) and appropriate use of “atypical responses” (AR) scales are discussed. Such scales have been designed and implemented in some known and newly designed questionnaires. Findings suggest that the AR scale cutoff that is optimally efficient for detecting NCTB in high-stake testing may be calculated on randomly generated data as $MR - 2 * SDR$. Validation of AR cutoff scores was accomplished by means of comparison of eight empirical ($N = 406-5689$) and randomly generated data for six questionnaires with different number of response options. Established sensitivity (0.92–0.99) is provided evidence for acceptable discriminating power of AR scores. This figure is comparable with person-fit methods based on IRT. Overall, the results provide support for the AR scores as indexes of NCTB. Military personnel testing practice suggest that taking into account AR scores enables to enhance the credibility of testing results and to diminish risks of diagnostic failures.

Keywords: non-cooperative testing behavior, careless responding, atypical (aberrant) responses, credibility, validity, personality questionnaires design.

References

1. Sugonyaev K.V. [Hardware-Software Complex “Multipsihometr-03” as a Potential Way of Measurement Standards in Professional Psychodiagnostics] *Dvoynye tekhnologii: spetsvypusk* [Dual Technology: a Special Issue], 2000, pp. 12–14. (in Russ.)
2. Sugonyaev K.V. [Problems of Increasing the Reliability of Personality Dimensions: Current Practices and Possible Improvements]. *Zdorovyy obraz zhizni i fizicheskoe vospitanie studentov i slushateley vuzov: Materialy nauchno-prakticheskoy konferentsii* [A Healthy Lifestyle and Physical Education Students and University Students: Materials of Scientifically-Practical Conference]. Moscow, EEI Publ., 2009, vol. 7, pp. 199–210. (in Russ.)
3. Sugonyaev K.V. [A Little Truth about the Scales “Lie”]. *Zdorovyy obraz zhizni i fizicheskoe vospitanie studentov i slushateley vuzov* [A Healthy Lifestyle and Physical Education Students and University Students: Proceedings of the XI Inter-University Scientific and Practical Conference]. Moscow, EEI Publ., 2013, vol. 11, pp. 190–199. (in Russ.)
4. Sugonyaev K.V. [The Development of Automation Professional Psychodiagnostics in the Strategic Missile Forces: Historical Essay]. *Sovremennaya psikhodiagnostika Rossii. Preodolenie krizisa* [Modern Psychodiagnostics Russia. Overcoming the Crisis]. Chelyabinsk, SUSU Publ., 2015, vol. 2, pp. 169–173. (in Russ.)
5. Sugonyaev K.V., Lazarev N.V. [The Development of Automation Professional Psychodiagnostics in the Strategic Missile Forces: Historical Essay]. *20 let funktsionirovaniya sistemy professional'nogo psikhologicheskogo otbora v Vooruzhennykh Silakh Rossiyskoy Federatsii, itogi i perspektivy sovershenstvovaniya* [20 Years of the Professional Psychological Selection System in the Russian Federation Armed Forces, Results and Prospects of Improvement]. Moscow, MAGS Publ., 2014, pp. 242–251. (in Russ.)
6. Conijn J.M., Emons W.H.M., Jong K.D., Sijtsma K. Detecting and explaining aberrant responding to the Outcome Questionnaire–45. *Assessment*, 2015, vol. 22, no. 4, pp. 513–524. DOI: 10.1177/1073191114560882.
7. Credé M. Random responding as a threat to the validity of effect size estimates in correlational research. *Educational and Psychological Measurement*, 2010, vol. 70, no. 4, pp. 596–612. DOI: 10.1177/0013164410366686.
8. Desimone J.A., Harms P.D., Desimone A.J. Best practice recommendations for data screening. *Journal of Organizational Behavior*, 2015, vol. 36, no. 2, pp. 171–181. DOI: 10.1002/job.1962.
9. Hogan J., Barrett P., Hogan R. Personality measurement, faking, and employment selection. *Journal of Applied Psychology*, 2007, vol. 92, no. 5, pp. 1270–1285. DOI: 10.1037/0021-9010.92.5.1270.

10. Hough L.M., Oswald F.L. Personality testing and industrial-organizational psychology: Reflections, progress, and prospects. *Industrial and Organizational Psychology*, 2008, vol. 1, no. 3, pp. 272–290. DOI: 10.1111/j.1754-9434.2008.00048.x.
11. Huang J.L., Curran P.G., Keeney J., Poposki E.M., DeShon R.P. Detecting and deterring insufficient effort responding to surveys. *Journal of Business and Psychology*, 2012, vol. 27, no. 1, pp. 99–114. DOI: 10.1007/s10869-011-9231-8.
12. Huang J.L., Bowling N.A., Liu M., Li Y. Detecting insufficient effort responding with an infrequency scale: Evaluating validity and participant reactions. *Journal of Business and Psychology*, 2015, vol. 30, no. 2, pp. 299–311. DOI: 10.1007/s10869-014-9357-6.
13. Johnson J.A. Ascertaining the validity of individual protocols from Web-based personality inventories. *Journal of Research in Personality*, 2005, vol. 39, no. 1, pp. 103–129. DOI: 10.1016/j.jrp.2004.09.009.
14. Kam C.C.S., Meyer J.P. How careless responding and acquiescence response bias can influence construct dimensionality: The case of job satisfaction. *Organizational Research Methods*, 2015, vol. 18, no. 3, pp. 512–541. DOI: 10.1177/1094428115571894.
15. Karabatsos G. Comparing the aberrant response detection performance of thirty-six person-fit statistics. *Applied Measurement in Education*, 2003, vol. 16, no. 4, pp. 277–298. DOI: 10.1207/S15324818AME1604_2.
16. Maniaci M.R., Rogge R.D. Caring about carelessness: Participant inattention and its effects on research. *Journal of Research in Personality*, 2014, vol. 48, pp. 61–83. DOI: 10.1016/j.jrp.2013.09.008.
17. Meade A.W., Craig S.B. Identifying careless responses in survey data. *Psychological Methods*, 2012, vol. 17, no. 3, pp. 437–455. DOI: 10.1037/a0028085.
18. Morgeson F.P., Campion M.A., Dipboye R.L., Hollenbeck J.R. et al. Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology*, 2007, vol. 60, no. 3, pp. 683–729. DOI: 10.1111/j.1744-6570.2007.00089.x.
19. Rammstedt B., Goldberg L.R., Borg I. The measurement equivalence of Big-Five factor markers for persons with different levels of education. *Journal of Research in Personality*, 2010, vol. 44, no. 1, pp. 53–61. DOI: 10.1016/j.jrp.2009.10.005.
20. Suhr J.A., Buelow M., Riddle T. Development of an infrequency index for the CAARS. *Journal of Psychoeducational Assessment*, 2011, vol. 29, no. 2, pp. 160–170. DOI: 10.1177/0734282910380190.
21. Tett R.R., Fitzke J.R., Wadlington P.L., Davies S.A., Anderson M.G., Foster J. The use of personality test norms in work settings: Effects of sample size and relevance. *Journal of Occupational and Organizational Psychology*, 2009, vol. 82, no. 3, pp. 639–659. DOI: 10.1348/096317908X336159.

Received 29 January 2016

ОБРАЗЕЦ ЦИТИРОВАНИЯ

Сугоняев, К.В. Шкалы атипичности ответов как инструмент выявления некооперативного тестового поведения / К.В. Сугоняев // Вестник ЮУрГУ. Серия «Психология». – 2016. – Т. 9, № 1. – С. 17–26. DOI: 10.14529/psy160102

FOR CITATION

Sugonyaev K.V. Atypical Responses Scales as the Tool of Non-Cooperative Testing Behavior Detection. *Bulletin of the South Ural State University. Ser. Psychology*. 2016, vol. 9, no. 1, pp. 17–26. (in Russ.). DOI: 10.14529/psy160102
