

ОБЩАЯ ПСИХОЛОГИЯ И ПСИХОЛОГИЯ ЛИЧНОСТИ

DOI: 10.17223/17267080/69/1

УДК 159.9.072.43

К.В. Сугоняев¹, Ю.И. Радченко², А.А. Соколов²

¹ Институт психологии Российской академии наук (Москва, Россия)

² Военная академия Генерального штаба ВС РФ (Москва, Россия)

Добровольное интернет-тестирование как источник валидных оценок группового психометрического интеллекта

Статья подготовлена при поддержке Российского научного фонда,
проект № 17-78-30035.

Представлены результаты психометрического анализа масштабного массива данных, полученного в условиях добровольного интернет-тестирования, который включал результаты выполнения методики КОТ-30, зарекомендовавшей себя в условиях контролируемого обследования как надежный индикатор общей когнитивной способности. Предполагалось, что результаты выполнения этого теста в режиме онлайн могут оказаться пригодными для получения оценок психометрического интеллекта больших групп (возрастных, образовательных, региональных и т.п.). Для проверки данного предположения была выполнена серия сопоставлений результатов выполнения ряда когнитивных тестов, включая КОТ-30, в различных форматах обследования. Предложена и апробирована методология оценки эквивалентности результатов выполнения когнитивных тестов, выходящая за рамки традиционного сопоставления мер центральной тенденции, вариативности и надежности. Гипотеза относительно возможности использования результатов онлайн-тестирования в качестве источника валидных оценок группового психометрического интеллекта подтвердилась.

Ключевые слова: неконтролируемое интернет-тестирование; интеллект; общая когнитивная способность; психометрические характеристики; измерительная эквивалентность; факторный анализ.

Введение

Широкое распространение персональных компьютеров в последние три десятилетия позволило перенести компьютерную психодиагностику в России из научных лабораторий в общественную практику, прежде всего в тех сферах, где психологическое тестирование в силу необходимости носит массовый характер [1]. Неудивительно, что одним из пионеров компьютеризации практической психодиагностики в России стало Министерство обороны, где широкое внедрение компьютерных технологий – первоначаль-

но в сфере психологического отбора, а затем и психологического сопровождения деятельности военнослужащих – началось уже в конце 1990-х гг. [2, 3]. Преимущества компьютеризации психодиагностики (равно как и возможные издержки, проистекающие из чрезмерного увлечения процессом в ущерб содержанию) уже неоднократно и детально обсуждались в литературе [1, 4], что избавляет нас от необходимости вновь возвращаться к обсуждению этих вопросов.

Развитие и широкое проникновение современных информационных технологий, прежде всего глобальной информационной сети Интернет, формирует новые вызовы, но и представляет новые, недоступные ранее возможности для расширения круга потенциальных респондентов, поскольку открывает доступ к тестовым процедурам в удобное для них время и в удобном месте при минимальных затратах для бенефициара. Благодаря этому интернет-тестирование не только быстро стало элементом мероприятий рекрутинга и начальных этапов отбора персонала [5], но со временем превратилось в важное средство получения научных знаний, в том числе относящихся к сфере общей и дифференциальной психологии [6]. В результате в психологической литературе в последнее десятилетие стали публиковаться результаты исследований, выполненных в относительно короткие сроки на выборках, насчитывающих от нескольких десятков тысяч до миллионов респондентов [7, 8]. Хотя в России также реализуются сходные по направленности проекты (например, «Клуб испытателей тестовых технологий», учрежденный компанией «Гуманитарные технологии» [1. С. 393]), отечественные публикации, основанные на интернет-опросах, пока еще уступают по охвату аудитории зарубежным аналогам. Следует отметить, что интерес к данным, полученным на больших выборках, дополнительно стимулируется обескураживающе низкой воспроизводимостью результатов, полученных на более «традиционных» для академической психологии немногочисленных студенческих выборках¹. Именно поэтому включение в научный анализ данных, полученных в рамках программ добровольного интернет-тестирования, рассматривается некоторыми авторами как одно из средств преодоления «кризиса воспроизводимости» в психологии [10].

Однако вопрос о валидности полученных с помощью неконтролируемого интернет-тестирования (НИТ) данных продолжает оставаться предметом дискуссий. Если в отношении личностных инструментов получены достаточно убедительные свидетельства высокой степени измерительной эквивалентности данных, полученных в разных форматах обследования (бумажно-карандашном, компьютерном диалоговом, НИТ) [11–13], то к когнитивным тестам отношение исследователей гораздо более настороженное. Практически единодушно считается, что результаты выполнения когнитивных тестов в условиях НИТ могут легко фальсифицироваться, вследствие чего их результа-

¹ За рубежом для обозначения доминирующей в научных исследованиях категории респондентов получила распространение аббревиатура WEIRD (western, educated, industrialized, rich, democratic) [9].

ты требуют перепроверки в контролируемых условиях [1, 5]; данная норма закреплена, в частности, в п. 45.3 стандарта Международной тестовой комиссии [14]. Следует, однако, подчеркнуть, что данное предостережение относится к *индивидуальным* результатам и представляется обоснованным в *аттестационных* (high stake) ситуациях при принятии персонализированных решений. В то же время сравнительно мало известно о том, насколько оправдана подобная настороженность по отношению к *групповым* результатам, полученным в условиях *самопознания*. Есть основания полагать, что собранные в подобных условиях данные могут быть вполне пригодными для использования в решении как практических, так и некоторых фундаментальных вопросов, стоящих перед современной психологией [6, 9, 15].

Немногочисленные исследования сопоставимости групповых результатов выполнения когнитивных тестов, полученных в условиях традиционного администрирования и НИТ, демонстрируют противоречивые результаты: одни тесты характеризуются высокой степенью сопоставимости результатов, другие нет [16–18]. Обращает на себя внимание ограниченность используемых при этом методов оценки сопоставимости результатов тестирования, которые в большинстве случаев сводятся к сравнению мер центральной тенденции, вариативности и одномоментной надежности интегральных тестовых оценок.

В 2012 г. при участии авторов статьи на официальном сайте Минобороны России www.mil.ru в разделе, посвященном военной службе по контракту, была размещена система интернет-тестирования, включающая три психометрических методики (когнитивную, личностную и мотивационную) и алгоритм оценки индивидуальных результатов, настроенный на скрининг лиц, имеющих явные противопоказания к подобной службе [19]. Цель размещения системы – привлечь внимание граждан, интересующихся перспективой военной службы по контракту, к тому факту, что Минобороны России нуждается в качественных человеческих ресурсах. Особенностью системы является то, что тестирование носит добровольный и анонимный характер, интерпретация его результатов доводится только до самого респондента и не учитывается в последующих отборочных процедурах, т.е. данная процедура была спланирована и реализована как элемент рекрутинга, а не отбора. К концу 2017 г. данный раздел сайта посетили несколько сот тысяч человек, большинство из которых выполнили хотя бы одну методику. В частности, было получено свыше 260 тыс. протоколов выполнения когнитивного теста (КТ), содержащих некоторые социально-демографические данные о респонденте, итоговую оценку успешности выполнения КТ и ответы на отдельные задания. Поскольку столь большой массив данных выполнения одного КТ в сопоставимых условиях интерактивного интернет-тестирования, охватывающего все регионы страны, был аккумулирован в России, возможно, впервые, представлялось весьма заманчивым оценить его пригодность для решения некоторых актуальных вопросов, в частности оценки качества региональных когнитивных ресурсов [20, 21]. Для достижения этой цели мы обратились к способам доказательства валидности оценок КТ, основанным на внутренней структуре теста и соотноше-

ниях с доступными внешними переменными. В качестве последних были использованы данные, которые сообщал о себе респондент в ходе регистрации, а именно дата рождения, место проживания (страна, регион, населенный пункт) и уровень образования.

Исследование выполнялось в два этапа.

На первом этапе были исследованы психометрические характеристики интегрального массива оценок выполнения КТ, их связь с возрастом и образовательным уровнем респондентов, предпринята попытка оценить долю повторных прохождений КТ и степень их влияния на групповые индексы. Также был осуществлен качественный анализ идентифицирующей респондента информации с целью определения наиболее вероятных угроз ее достоверности.

На втором этапе комплекс количественных индексов, характеризующих результаты выполнения КТ в условиях НИТ, был сопоставлен с аналогичными индексами, рассчитанными по результатам выполнения этого же и пяти других тестов в условиях контролируемого обследования в диалоговом или бланковом формате. Дополнительные тесты, для которых имелись достаточно большие массивы данных, были использованы в нашем исследовании для того, чтобы обеспечить интерпретируемость полученных оценок конгруэнтности. В сопоставление были включены результаты выполнения тестов как в общем формате (бланк–бланк или диалог–диалог), так и в разных (бланк–диалог).

Методики

В исследовании было использовано 7 когнитивных тестов, первый из которых выполнялся в формате НИТ и являлся центральным для данного исследования. Результаты выполнения этого же и шести других КТ в контролируемых условиях в аттестационном контексте были использованы в интересах сопоставления и верификации эффектов, выявленных на основе анализа интернет-данных.

1. Методика с условным обозначением КОТ-30 входит в состав тестовой батареи, реализуемой системой интернет-тестирования. Методика разработана нами в 2011 г. специально для данной системы и является укороченной и модифицированной версией Краткого ориентировочного теста [22]. По сравнению с прототипом КОТ-30 обладает улучшенными психометрическими характеристиками благодаря исключению менее дискриминативных пунктов и частичному редактированию оставшихся. Тест состоит из 30 заданий, на выполнение которых отводится 15 минут. Задания адресованы вербальному, счетному, пространственному и перцептивному факторам интеллекта (с преобладанием первых двух) и имеют от трех до пяти вариантов ответа, лишь один из которых является правильным. Пригодность теста для оценки общей когнитивной способности подтверждается коэффициентом корреляции со Стандартными прогрессивными матрицами Равена 0,598 без коррекции и 0,675 после коррекции с учетом неидеальной надежности обоих тестов ($N = 654$).

2. Вербальный тест (ВТ). Предназначен для оценки вербальных способностей. Состоит из 36 заданий с четырьмя вариантами ответов в каждом, на выполнение которых отводится 9 минут. Задания требуют умения анализировать соотношения между понятиями, устанавливать аналогии, находить синонимичные и антонимичные понятия. В исследовании использовался интегральный массив данных, полученных в диалоговом формате обследования.

3. Исключение слова (ИС). Методика предназначена для оценки способности к обобщению вербальной информации и нахождению семантически дистантного понятия. Включает 30 заданий с пятью вариантами ответов в каждом, на выполнение которых отводится 8 минут. В сопоставление было включено два бланковых массива.

4. Установление закономерности (УЗ). Методика предназначена для оценки скорости и точности нахождения лексического эквивалента для знаковой последовательности. Включает 30 заданий с пятью вариантами ответов в каждом, на выполнение которых отводится 8 минут. В сопоставление было включено два бланковых массива, полученных на тех же выборах, которые выполняли предыдущую методику.

5. Аналогии (Ан). Методика предназначена для оценки способности к пониманию характера логических связей и отношений между понятиями. Включает 30 заданий с пятью вариантами ответов в каждом, на выполнение которых отводится 5 минут. В исследование было включено два бланковых и два диалоговых массива данных.

6. Стандартные прогрессивные матрицы Равена с 20- и 30-минутным ограничением (СПМ₂₀; СПМ₃₀); для каждой из версий имелись данные, полученные как в бланковом, так и в диалоговом форматах.

7. Укороченная 20-пунктовая версия Продвинутых прогрессивных матриц Равена (ППМ-20), составленная из заданий 1-го и 2-го наборов, отобранных по параметрам трудности и дискриминативности. На выполнение методики отведено 12 минут. В исследование было включено два бланковых и один диалоговый массив данных.

Для всех тестов рассчитывался комплекс оценок, характеризующих различные аспекты выполнения КТ (продуктивность, скорость, точность и эффективность), однако в настоящем исследовании анализируется лишь один индекс успешности выполнения теста – сумма правильно решенных заданий (продуктивность).

Характеристика выборок

Исходный массив протоколов выполнения КОТ-30 в формате НИТ, собранных в период с сентября 2012 по декабрь 2017 г., составил 262 652. Часть протоколов (3 505), не имеющих региональной принадлежности или выполненных нерезидентами Российской Федерации, была исключена из анализа. Таким образом, до чистки от повторных прохождений теста количество включенных в анализ протоколов составило 259 147. Согласно заданным ограничениям респондентами были мужчины в возрасте 18–

40 лет² ($26,593 \pm 4,936$) и уровнем образования не ниже среднего общего. За редкими исключениями все участники НИТ являлись гражданскими лицами, представляющими 85 регионов России.

Прочие выборки, привлеченные для сравнительного анализа, представлены военнослужащими различных категорий либо кандидатами на военную службу по контракту или обучение в военных учебных заведениях. Их краткая характеристика представлена в табл. 1.

Т а б л и ц а 1

Характеристика респондентов из контрольных выборок

Обозначение методики	Формат обследования	Период сбора данных	Объем выборки	Контингент	Гендерный состав	Возраст: $M \pm SD$ (размах)
КОТ-30	Бланковый	2014–2015	3 182	ВСП № 1	Мужчины	$20,01 \pm 1,54$ (18–26)
ВТ	Диалоговый	2009–2014	27 388	Кандидаты на ВСК № 1	Мужчины	$25,22 \pm 4,43$ (18–40)
ИС	Бланковый	2013	11 338	Абитуриенты № 1	Мужчины	$18,42 \pm 1,29$ (16–28)
	Бланковый	2014	11 908	Абитуриенты № 2	Мужчины	$17,95 \pm 0,43$ (17–18)
УЗ	Бланковый	2013	11 338	Абитуриенты № 1	Мужчины	$18,42 \pm 1,29$ (16–28)
	Бланковый	2014	11 908	Абитуриенты № 2	Мужчины	$17,95 \pm 0,43$ (17–18)
Ан	Диалоговый	2007	1 162	ВСП № 2	Мужчины	$19,60 \pm 1,23$ (18–24)
	Диалоговый	2009	1 096	ВСП № 3	Мужчины	$20,30 \pm 1,71$ (18–25)
	Бланковый	2008–2012	2 226	ВСП № 4	Мужчины	$20,11 \pm 1,67$ (18–25)
	Бланковый	2010–2012	2 422	ВСП № 5	Мужчины	$20,26 \pm 1,84$ (18–26)
СПМ ₂₀	Диалоговый	2009–2012	3 647	ВСП № 6	Мужчины	$20,73 \pm 1,95$ (18–26)
	Бланковый	2011–2012	1 411	ВСП № 7	Мужчины	$20,06 \pm 1,78$ (18–26)
СПМ ₃₀	Диалоговый	2008–2012	1 085	ВСП № 8	Мужчины	$20,49 \pm 1,72$ (18–26)
	Бланковый	2010–2012	3 146	ВСП № 9	Мужчины	$20,54 \pm 2,14$ (18–28)
ППМ-20	Диалоговый	2011–2014	973	ВСК и кандидаты на ВСК № 2	4,6% женщин	$27,63 \pm 7,10$ (16–53)
	Бланковый	2010–2016	1 839	Курсанты	21% женщин	$22,06 \pm 1,17$ (17–30)
	Бланковый	2017–2018	1 681	ВСК № 3	13% женщин	$36,08 \pm 7,22$ (20–55)

Примечания. ВСП – военнослужащие по призыву; ВСК – военная служба по контракту.

² До октября 2013 г. возрастной диапазон был ограничен 19–30 годами; с сентября 2017 г. возрастной ценз снижен до 18 лет.

Все выборки – за исключением абитуриентских 2013 и 2014 гг. – были независимыми. В некоторых случаях выборки были составными, но их объединение происходило только при отсутствии существенных различий средних и гистограмм распределения тестовых оценок. Во всех случаях в анализ включались только результаты первичного прохождения теста.

Процедура исследования

Результаты НИТ, включая социально-демографические данные о респондентах и результаты их тестирования (на уровне шкальных оценок и ответов на отдельные пункты), предоставлялись в ВАГШ ВС РФ ежемесячно в формате Microsoft Excel. Поскольку интернет-обследование носит анонимный характер, выявление протоколов повторных прохождений КОТ-30 опиралось на такие идентифицирующие признаки, как повторение в пределах одного региона в течение 1–2 месяцев сочетаний одинаковой даты рождения, уровня образования и наименования населенного пункта. Таких протоколов выявлено 19 706 (7,6% от числа включенных в анализ), причем число повторов у отдельных респондентов варьировало от 1 до 26.

Обследование контрольных выборок проходило в аттестационном контексте. Во всех случаях, когда обследование проходило в диалоговом формате, предъявление заданий, регистрация ответных действий респондентов и оценивание результатов осуществлялись с помощью компьютерного психодиагностического комплекса (КПДК). В случае бланкового формата (paper & pencil) обследование осуществлялось в составе групп. Испытуемые получали стандартные тестовые материалы (стимульные буклеты и регистрационные бланки), начало и завершение тестирования осуществлялись по команде администратора. Заполненные испытуемыми регистрационные бланки обрабатывались автоматически путем сканирования с использованием аналогичных КПДК.

Накопленные данные анализировались и вычищались (от повторных обследований и outliers) с помощью встроенных средств КПДК, после чего экспортировались в формате табличного процессора MS Excel. С помощью Excel рассчитывались некоторые дескриптивные статистики и точный возраст респондентов. Корреляционный и факторный анализ данных выполнялся с использованием пакета прикладных программ StatSoft Statistics, v. 8.0. Статистический анализ сопоставимости результатов, полученных в разных форматах и выборках, включал в себя не только сравнение средних, мер вариативности и одномоментной надежности итоговых тестовых оценок, но и сравнения на пунктовом уровне, а именно коэффициентов трудности и дискриминативности заданий, а также их факторных нагрузок на генеральный фактор.

Результаты

Психометрические характеристики оценок КОТ-30 в формате НИТ оценивались как в интегральной выборке (до и после чистки от повторных прохождений теста), так и в некоторых ее фрагментах (табл. 2).

Таблица 2

Основные психометрические индексы, характеризующие выполнение КОТ-30

Выборка интегральная		N	M	SD	ОМН
		259 147	20,073	6,022	0,863
В том числе по годам выполнения	2012–2013	35 544	19,813	5,962	0,857
	2014	56 821	19,933	6,021	0,862
	2015	67 866	20,197	5,994	0,862
	2016	54 400	20,234	6,029	0,865
	2017	44 516	20,070	6,090	0,868

Примечания. ОМН – одномоментная надежность, оцениваемая с помощью коэффициента альфа Кронбаха.

Представленные в табл. 2 данные свидетельствуют о высокой степени устойчивости оценок КОТ-30, полученных в формате НИТ, а значения индексов ОМН указывают на их высокую измерительную точность. Сходные по значениям индексы ОМН были получены и при иных способах структурирования интегральной выборки (по уровню образования, возрасту, региональной принадлежности).

Оценки психометрического интеллекта связаны с продолжительностью обучения; например, в работе Э. Ханта (2012) приводятся данные о том, что каждый год формального образования ведет к росту оценок КТ, эквивалентному 2,7 единиц шкалы IQ (т.е. 0,18 стандартного отклонения) [23. С. 298]. Исходя из этого, можно было предполагать, что оценки лиц со средним и высшим образованием могут различаться в среднем на 0,7–0,9 SD. Усредненные оценки по трем основным образовательным категориям³ (до и после чистки от повторов) представлены в табл. 3.

Таблица 3

Психометрические индексы, характеризующие выполнение КОТ-30 в образовательных категориях (без учета возраста) до и после чистки от протоколов ретестирования

Образование	До чистки				После чистки				Различие средних (d Коэна)
	N	M	SD	ОМН	N	M	SD	ОМН	
Высшее	123 184	21,370	5,677	0,856	114 403	21,317	5,652	0,856	–0,009
Среднее общее	44 795	19,741	6,136	0,866	41 234	19,763	6,098	0,866	0,004
Среднее профессиональное	77 752	18,084	5,952	0,846	70 459	18,027	5,905	0,845	–0,009

³ До сентября 2017 г. учитывалась также категория «незаконченное высшее образование», однако выбирающие ее респонденты по своим характеристикам незначительно отличалась от лиц, указывающих категорию «высшее образование».

Представленные в табл. 3 результаты свидетельствуют о том, что чистка данных от повторных прохождений КТ не оказала существенного влияния ни на один из индексов, рассчитываемых по большим массивам данных. В то же время различия средних, рассчитанных для трех образовательных категорий, оказались ниже зарубежных оценок и составили 0,26 стандартного отклонения в паре *высшее–среднее общее* и 0,55 в паре *высшее–среднее профессиональное*.

Более детальный анализ с учетом возраста показал, что недооценка различий в уровне психометрического интеллекта групп с высшим и средним образованием связана главным образом с более юными и многочисленными участниками тестирования (18–24 года), тогда как к 30 годам различие достигает уровня 0,5–0,6 стандартного отклонения. Также было выявлено присутствие среди респондентов 18–20 лет от 10 до 20% тех, кто выбрал в меню пункт «высшее образование», хотя это едва ли соответствует реальности. Можно предположить, что в эту категорию включают себя, в том числе, те участники НИТ, которые еще не закончили обучение в вузе. Возрастная динамика оценок в каждой из трех указанных образовательных категорий представлена на рис. 1.

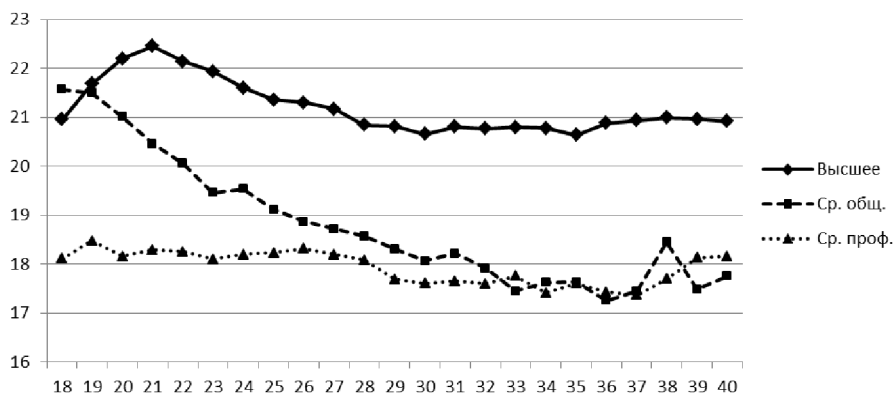


Рис. 1. Возрастные изменения оценок КОТ-30 в трех образовательных категориях

Из представленных на рис. 1 данных следует, что в двух образовательных категориях оценки в исследованном диапазоне возрастов остаются относительно стабильными, тогда как в группе со средним общим образованием наблюдается их заметное снижение. На уровне индивидуальных оценок теснота связи с возрастом в интегральной выборке слабая ($-0,048$), хотя и значимая ($p < 0,001$), однако в группе лиц со средним общим образованием она увеличивается до $-0,197$. Если возраст заменить годом рождения, нелинейный характер динамики оценок КОТ-30 во времени проявляется более четко, чем это видно на рис. 1: в когортах 1973–1987 гг. оценки колеблются около уровня 19,8 баллов, но затем они начинают линейно расти, в результате чего лица, рожденные в 1998 г., достигают средней оценки 20,9 баллов.

Для того чтобы уточнить, насколько выявленные различия специфичны для НИТ, аналогичный анализ был выполнен для контрольной выборки, выполнявшей в условиях аттестационного обследования, предполагающего точную фиксацию реального образовательного уровня, методику ВТ. В интегральной выборке средний балл составил $20,383 \pm 6,853$ при $OMH = 0,877$; в выборке лиц с высшим образованием ($N = 5\ 096$) – $22,671 \pm 6,515$; в выборке лиц со средним общим образованием ($N = 10\ 613$) – $19,127 \pm 6,790$; в выборке лиц со средним профессиональным образованием ($N = 10\ 040$) – $20,212 \pm 6,776$. Таким образом, в контрольной выборке стандартизированное различие средних в группах с высшим и средним общим образованием составило 0,52 стандартного отклонения. Другое отличие этой выборки от выборки НИТ состоит в том, что доля лиц с высшим образованием ни в одной из возрастных групп не превышала 30%, тогда как среди участников НИТ на наличие высшего образования указали почти 48% респондентов.

Анализ распределения возрастов среди посетителей сайта показал, что имеет место быстрый рост представленности от 18 до 22 лет с последующим постепенным, практически линейным спадом в диапазоне 22–40 лет. Сравнение с распределением возрастов среди кандидатов, выполнявших ВТ, показало высокую степень сходства двух кривых: корреляция рядов чисел, отражающих представленность различных возрастов в диапазоне 18–40 лет, составила 0,943 ($p < 0,0005$), при этом мода в обоих случаях приходилась на 22 года. Тот факт, что в двух независимых популяциях граждан, интересующихся перспективой военной службы по контракту, распределение возрастов оказалось практически идентичным, может рассматриваться как косвенное подтверждение довольно высокой степени достоверности данных о дате рождения в ситуации НИТ. Об этом же свидетельствуют равная частотность чисел в датах рождения и отсутствие выраженных пиков около дат с облегченным набором (вроде 11.11.1991).

Качественный анализ возможных угроз валидности групповых оценок в условиях НИТ. Хотя в условиях анонимного НИТ с целью самопознания у респондентов нет явных причин фальсифицировать данные о себе, исключить вероятность искажений априори едва ли возможно. Выше мы уже отмечали, что по ряду признаков есть основания с высокой степенью доверия относиться к данным о дате рождения респондента (которая в сопоставлении с датой обследования позволяет рассчитать точный возраст в момент тестирования). Более подверженной контаминации представляется информация, сообщаемая респондентом о своем образовании, причем здесь, помимо случайных ошибок, возможна и направленная тенденция к завышению имеющегося образовательного уровня. На присутствие первой составляющей указывает то, что в сериях протоколов, идентифицированных как ретестирование, иногда наблюдается варьирование образовательной категории. Вторая выявляется в таких признаках, как довольно высокая доля лиц с якобы высшим образованием в возрасте 18–20 лет и вызывающее сомнение преобладание лиц с высшим образованием в интеграль-

ной выборке. Тем не менее благодаря масштабам выборки даже такие смещенные оценки позволяют выделять «полезный сигнал» из информационного шума и выявлять некоторые общие тенденции.

Региональная принадлежность респондентов представляет большой практический и научный интерес (например, она может давать четкое представление о географическом распределении потенциальных кадровых ресурсов для военной службы по контракту и их динамике во времени), поэтому оценка достоверности сведений о географической локализации – несмотря на ее трудоемкость – представляется очень важной.

Подводя итог многомесячной работе по выявлению возможных искажений данной информации, мы можем сделать общий вывод: подавляющее большинство респондентов, скорее всего, указывает *реальную* региональную принадлежность. Вместе с тем выявлен ряд предпосылок к возможному искажению этой информации в относительно небольшом проценте случаев и предприняты шаги по ее исправлению, если признаки таких искажений удавалось обнаружить. Их перечень представлен ниже.

1. Хотя русскоговорящим гражданам иностранных государств предоставлена возможность регистрироваться в системе по месту реального проживания и проходить тестирование, некоторые из них предпочитают указывать произвольный регион России, чаще всего – граничащий с данным государством. К счастью, многие из них на следующем шаге указывают не только населенный пункт (находящийся за пределами России), но и реальную страну пребывания. Было выявлено свыше 500 таких протоколов.

2. Наиболее частые ошибки региональной локализации связаны с промахом при выборе региона проживания из алфавитного перечня субъектов Российской Федерации, вследствие чего случайно может оказаться выбран смежный (по списку) регион. Хотя в целом частота таких ошибок ничтожна (обычно не выходит за пределы десятых долей процента), она может повышаться в тех случаях, когда наименования регионов-смежников созвучны или содержат общие структурные элементы. К числу пар регионов «повышенного риска» неточной локализации относятся Республика Алтай и Алтайский край; Краснодарский и Красноярский края; Курская и Курганская области; Нижегородская и Новгородская области; Ненецкий и Ямало-Ненецкий автономные округа. В этих случаях вероятность неверной локализации может достигать нескольких процентов.

3. Наибольшая частота ошибок локализации, достигающая в процентах двузначных значений, связана с тремя парами субъектов: Москва и Московская область; Санкт-Петербург и Ленинградская область; Севастополь и Республика Крым. Распространенная ошибка жителей мегаполисов и Севастополя – указание в качестве места проживания *области* (республики), тогда как истинная локализация попадает в графу «населенный пункт». В данном случае ошибка носит однонаправленный характер и связана, с одной стороны, с нечеткостью инструкции, а с другой – с тем, что не все граждане вникают в тонкости административно-территориального устройства страны.

4. Среди участников НИТ имеется незначительная доля тех, кто по каким-то причинам не хочет указывать регион проживания, но поскольку этот пункт носит обязательный характер, выбирает один из пунктов в начале списка, в большинстве случаев – республики Адыгея и Алтай. Именно на эти два региона приходится наибольшее число протоколов с неясной локализацией, которые желательно исключить из анализа на региональном уровне.

В абсолютном большинстве случаев коллизий, подобных перечисленным выше, истинная локализация может быть уточнена по наименованию населенного пункта (что и было осуществлено).

Результаты анализа типичных ошибок, допускаемых респондентами, легли в основу подготовленных нами предложений по совершенствованию работы системы интернет-тестирования.

Исследование сопоставимости результатов тестирования в контролируемом и НИТ-формате осуществлялось путем сравнения интегральных тестовых оценок и оценок на пунктовом уровне. В первом случае основное внимание уделялось воспроизводимости средних, вариативности (стандартного отклонения) и ОМН. Основные психометрические индексы и результаты сравнения средних оценок представлены в табл. 4⁴.

Как видно из представленных в табл. 4 данных, все тесты имеют измерительную точность, отвечающую требованиям п. 4.4.2 Российского стандарта тестирования персонала [24]. Оценки КОТ-30 характеризуются весьма respectable значениями ОМН при любом формате проведения, их вариативность не превышает таковую других тестов в условиях контролируемого обследования и составляет от 1,5 до 3,7% (в контрольных выборках – от 0 до 7%). Колебания вариативности оценок КОТ-30 при транзите с контролируемого бланкового формата обследования на НИТ не превышают 9%; такой же порядок имеют межформатные колебания стандартного отклонения в методике «Аналогии», в случае СПМ они достигают 20%.

То же самое можно сказать о различиях средних оценок: хотя некоторые из них достигают уровней, которые Дж. Коэн предложил считать умеренными [25], они не выходят за пределы вариативности, характеризующей реальные различия в уровнях общей когнитивной способности разных популяций (в представленных в табл. 4 примерах – от 0,004 до 0,70).

Все же сравнение психометрических индексов, рассчитанных для теста в целом, является довольно грубой оценкой возможной измерительной эквивалентности разных форматов тестирования, к тому же ее затруднительно интерпретировать: вариации могут быть обусловлены не транзитом форматов, а идиосинкразическими особенностями использованных выборок, в частности их селективностью. Более точные данные, по нашему мнению, может предоставить изучение процессов выполнения теста на пунктовом уровне. Мы предположили, что высокая степень эквивалентности

⁴ Несмотря на формальные различия двух версий СПМ в части временного ограничения, мы объединили их в одну методику, поскольку результаты их выполнения различались несущественно.

тестовых результатов должна проявляться в сходстве распределения коэффициентов трудности и дискриминативности тестовых заданий, а также в сходных паттернах факторных нагрузок пунктов на генеральный фактор. Поскольку одной из наиболее популярных мер сходства двух профилей является коэффициент корреляции [26, 27], мы избрали в качестве меры конгруэнтности непараметрический коэффициент корреляции Спирмена (ρ).

Т а б л и ц а 4

Психометрические характеристики и сравнение средних в родственных интернет- и контрольных выборках

Обозначение методики	Формат обследования (условное обозначение выборки)	M	SD	OMH	Стандартизированное различие средних (d Коэна)		
					Б-Б	Д-Д	Д-Б
КОТ-30	НИТ (Д1 = высшее образование)	21,32	5,652	0,855	–	Д1–Д2: 0,26	Д1–Б: 0,56
	НИТ (Д2 = среднее общее)	19,76	6,098	0,866		Д1–Д3: 0,55	Д2–Б: 0,30
	НИТ (Д3 = среднее проф.)	18,03	5,905	0,846		Д2–Д3: 0,29	Д3–Б: 0,017
	Бланковый (Б)	17,93	6,192	0,878		–	–
ИС	Бланковый (Б1 = 2013)	20,19	4,418	0,756	0,08	–	–
	Бланковый (Б2 = 2014)	20,55	4,468	0,766			
УЗ	Бланковый (Б1 = 2013)	22,90	4,464	0,842	0,004	–	–
	Бланковый (Б2 = 2014)	22,92	4,587	0,851			
Ан	Диалоговый (Д1)	17,92	5,712	0,842	–	0,42	Д1–Б1: 0,06
	Диалоговый (Д2)	20,23	5,336	0,847			Д1–Б2: 0,28
	Бланковый (Б1)	18,29	5,782	0,903	0,34	–	Д2–Б1: 0,35
	Бланковый (Б2)	16,30	5,885	0,881			Д2–Б2: 0,70
СПМ	Диалоговый (Д1 = 20 мин)	43,37	8,692	0,909	–	0,05	Д1–Б1: 0,06
	Диалоговый (Д2 = 30 мин)	42,94	9,108	0,915			Д1–Б2: 0,32
	Бланковый (Б1 = 20 мин.)	43,87	8,885	0,913	0,27	–	Д2–Б1: 0,10
	Бланковый (Б2 = 30 мин)	45,95	7,383	0,887			Д2–Б2: 0,36
ППМ-20	Диалоговый (Д)	11,25	3,755	0,770	–	–	Д–Б1: 0,43
	Бланковый (Б1 = курсанты)	12,79	3,475	0,748			Д–Б2: 0,27
	Бланковый (Б2 = ВСК)	12,24	3,594	0,758			–

Учитывая дефицит данных, на которые можно опереться для интерпретации рассчитанных коэффициентов конгруэнтности (КК), мы рассчитали для каждого из трех параметров 23 подобных коэффициента для предположительно совместимых данных (относящихся к одному тесту). В группе тестов, имеющих одинаковую длину (30 заданий), для этих же параметров было рассчитано по 52 коэффициента межтестовой конгруэнт-

ности. По аналогии с одним из методов доказательства конструктивной валидности тестов с помощью расчета многопризнаковой-многометодной матрицы коэффициенты первой группы мы обозначили как конвергентные, второй – как дивергентные. Конвергентные КК могли быть классифицированы на внутриформатные (диалог / диалог и бланк / бланк) и межформатные (диалог / бланк). Поскольку интернет-тестирование по технологии взаимодействия с респондентом имеет максимальное сходство с диалоговым форматом тестирования на КПДК, ожидалось, что в благоприятном случае КК между результатами выполнения методики КОТ-30 разными образовательными группами будут близки к внутриформатным КК в контрольных выборках, а КК между результатами интернет- и бланкового тестирования будут близки к межформатным КК в этих выборках.

Матрицы КК, рассчитанные для 30-пунктовых тестов, включая КОТ-30, представлены в табл. 5.

Таблица 5

Матрицы КК, рассчитанные для индексов трудности, дискриминативности и факторных нагрузок заданий 30-пунктовых тестов

Методика	Выборка	КОТ-30				ИС		УЗ		Ан			
		Д1	Д2	Д3	Б	Б1	Б2	Б1	Б2	Д1	Д2	Б1	Б2
КОТ-30	Д1		0,996	0,99	0,90	0,33	0,36	0,65	0,66	0,46	0,46	0,67	0,69
	Д2	0,98		0,99	0,90	0,36	0,40	0,66	0,67	0,47	0,47	0,68	0,70
	Д3	0,93	0,95		0,86	0,37	0,41	0,64	0,64	0,48	0,48	0,68	0,70
	Б	0,86	0,85	0,85		0,28	0,32	0,73	0,74	0,51	0,50	0,75	0,76
ИС	Б1	0,46	0,45	0,47	0,30		0,99	0,41	0,40	0,41	0,37	0,36	0,37
	Б2	0,43	0,42	0,45	0,27	0,98		0,44	0,44	0,42	0,38	0,38	0,39
УЗ	Б1	0,46	0,46	0,54	0,67	0,28	0,29		0,998	0,47	0,45	0,78	0,79
	Б2	0,28	0,28	0,36	0,50	0,20	0,23	0,92		0,49	0,46	0,79	0,80
Ан	Д1	0,04	0,09	0,13	0,09	0,02	0,01	0,33	0,31		0,98	0,80	0,81
	Д2	0,19	0,22	0,24	0,19	0,18	0,20	0,40	0,38	0,87		0,81	0,81
	Б1	0,40	0,37	0,30	0,41	0,15	0,17	0,50	0,36	0,52	0,50		0,99
	Б2	0,33	0,29	0,20	0,31	0,13	0,13	0,37	0,25	0,59	0,52	0,95	
КОТ-30	Д1		0,99	0,95	0,86	0,48	0,46	0,07	–0,34	0,10	0,03	0,41	0,44
	Д2			0,95	0,86	0,44	0,42	0,04	–0,34	0,09	0,02	0,37	0,40
	Д3				0,88	0,47	0,46	0,05	–0,36	0,10	0,01	0,29	0,32
	Б					0,31	0,28	0,23	–0,22	0,11	0,02	0,37	0,41
ИС	Б1						0,98	0,09	–0,13	–0,02	0,06	0,14	0,17
	Б2							0,07	–0,16	0,00	0,09	0,16	0,17
УЗ	Б1								0,80	0,19	0,21	0,18	0,20
	Б2									–0,02	0,04	–0,13	–0,11
Ан	Д1										0,86	0,55	0,61
	Д2											0,45	0,51
	Б1												0,97

Примечание. В верхней части таблицы выше диагонали представлены КК, рассчитанные на основе коэффициентов трудности; ниже диагонали – на основе коэффициентов дискриминативности; в нижней части таблицы представлены коэффициенты КК, рассчитанные на основе факторных нагрузок. Жирным шрифтом выделены конвергентные КК. Все корреляции от 0,37 и выше значимы на уровне $p < 0,01$. Использованы те же обозначения выборок, что и в табл. 4.

Обобщая данные, представленные в табл. 6, можно констатировать следующее. Средняя конгруэнтность значений трудности заданий в разных популяциях, выполнявших один тест, очень высока (0,916), однако если внутриформатные КК приближаются к единице (0,99), то межформатные составили в среднем в исследованных методиках / популяциях 0,845. Межтестовая конгруэнтность существенно ниже (в среднем – 0,528), хотя заметно выше, чем при использовании двух других параметров. Возможная причина этого будет рассмотрена ниже. На фоне таких значений высокий уровень конгруэнтности данных, полученных в условиях НИТ, не вызывает сомнения: 0,99 между разными популяциями в условиях НИТ, 0,86–0,90 с результатами выполнения КОТ-30 в контролируемых условиях (в бланковом формате).

Средняя конгруэнтность значений дискриминативности заданий в разных популяциях, выполнявших один тест, также довольно высока (0,806), однако можно отметить существенный разброс значений между внутриформатными (в среднем 0,941) и межформатными (в среднем 0,736) КК. Межтестовая конгруэнтность характеризуется предсказуемо низкими значениями (в среднем 0,298). Здесь также интернет-данные демонстрируют высокую консистентность: 0,93–0,98 между интернет-выборками, 0,85–0,86 с результатами выполнения КОТ-30 в ином формате в контролируемых условиях.

Таблица 6
Межформатные КК, рассчитанные для тестов СПМ и ППМ-20

Параметр / выборки	Трудность	Дискриминативность	Факторные нагрузки
СПМ			
Д1/Б1	0,916	0,850	0,759
Д1/Б2	0,958	0,791	0,720
Д2/Б1	0,908	0,819	0,765
Д2/Б2	0,953	0,822	0,818
ППМ-20			
Д/Б1	0,953	0,759	0,717
Д/Б2	0,980	0,842	0,770

Наконец, средняя конгруэнтность факторных нагрузок составила 0,802 (от 0,730 в условиях смены формата до 0,930 при использовании одного формата). Средний КК межтестовых факторных нагрузок составил 0,137. И в этом случае результаты тестирования в условиях НИТ выглядят вполне респектабельно: 0,95–0,99 между интернет-выборками, 0,86–0,88 между НИТ и результатами выполнения КОТ-30 в ином формате в контролируемых условиях.

Для методик с другой длиной были получены следующие внутриформатные значения КК: 1) СПМ, трудность пунктов: 0,996 для пары Д1/Д2; 0,962 для пары Б1/Б2; дискриминативность: соответственно 0,945 и 0,841; факторные нагрузки: соответственно 0,906 и 0,842; 2) ППМ-20,

Б1/Б2: трудность пунктов – 0,989; дискриминативность – 0,871; факторные нагрузки – 0,917. Поскольку число межформатных КК существенно больше, они представлены в табл. 6.

Обсуждение результатов

Целью настоящей работы был анализ массива протоколов выполнения КТ в условиях НИТ с точки зрения оценки его пригодности для изучения психологических феноменов, например возрастной динамики и географии распределения оценок способностей, взаимосвязей когнитивных и личностных переменных и др. В качестве такого массива были использованы данные, собранные в период с 2012 по 2017 г. на сайте www.mil.ru, где размещена система самотестирования граждан, проявляющих интерес к военной службе по контракту в Вооруженных Силах Российской Федерации. Онлайн-тестирование обладает рядом преимуществ перед тестированием в контролируемых условиях – прежде всего благодаря дешевизне и возможности рекрутирования больших и диверсифицированных выборок – и потому становится все более популярным.

Поскольку массив данных НИТ подобного объема (около 260 тыс. протоколов) стал объектом анализа в России впервые, на начальном этапе исследования были выполнены некоторые расчеты, призванные продемонстрировать правдоподобие информации, извлекаемой из этих данных. Также были оценены потенциальные угрозы для валидности тестовых оценок и регистрационных данных респондентов и возможные способы их нейтрализации. На этом этапе были получены следующие основные результаты.

1. Психометрические характеристики оценок группового психометрического интеллекта, рассчитанные на основе как интегральных НИТ-данных, так и их отдельных фрагментов, не уступают таковым, рассчитанным по результатам применения этого же теста в контролируемых условиях.

2. Распределение возрастов среди респондентов НИТ посетителей тестовой системы, размещенной на сайте www.mil.ru, обнаружило высокую степень сходства с возрастным составом кандидатов, прошедших тестирование в контролируемых условиях ($r = 0,943$), что может свидетельствовать о высокой степени достоверности сведений о дате рождения.

3. Получены косвенные свидетельства того, что часть респондентов НИТ, возможно, завышает свой образовательный уровень. Отчасти это может происходить вследствие выбора пункта «высшее образование» не только обладателями дипломов о высшем образовании, но и студентами. Вместе с тем преобладание в популяции участников НИТ лиц с высшим образованием не обязательно может быть следствием намеренного искажения данных. Имеются признаки того, что среди участников НИТ было немало тех, кто проходил тестирование исключительно ради самопознания, удовольствия и удовлетворения любопытства. На это указывает наличие повторных серий прохождения КТ респондентами, которые уже в пер-

вой попытке показали отличный (максимальный или субмаксимальный) результат, а также комментарии, оставляемые некоторыми респондентами (вместо наименования населенного пункта), вроде *«просто интересно пройти тест»*. Недооценка различий в уровне когнитивной способности между лицами, выбравшими в меню «Образование» пункты «высшее» и «среднее общее», также может быть следствием не столько искажения информации респондентами, сколько различий в качестве высшего образования в России и США.

Тем не менее, даже при наличии определенной настороженности в отношении истинности сведений об образовании, полученные данные позволили обнаружить ряд эффектов, представляющих научный и практический интерес. Например, выявлено заметное снижение с возрастом уровня оценок в категории лиц со средним общим образованием (рис. 1), что может свидетельствовать о том, что наиболее способные представители этой категории со временем повышают свой образовательный уровень и переходят в другие категории либо успешно трудоустраиваются, вследствие чего их интерес к поиску альтернативных способов занятости снижается. Также обращает на себя внимание нарастающее расхождение оценок способностей между одногодками, относящимися к образовательным категориям «среднее общее» и «среднее профессиональное», родившимися после 1990 г. По всей видимости, это является следствием внедрения в конце нулевых годов ОГЭ, предопределяющего разные образовательные траектории лиц, различающихся по уровню интеллекта.

Может представлять интерес также линейный рост оценок КОТ-30 в когортах лиц, родившихся после 1987 г., в результате чего средняя оценка за 12 лет возросла на 1 балл (с 19,8 до 20,8). Является ли этот тренд отражением феномена, получившего наименование «эффект Флинна», требует дополнительного изучения.

4. Среди потенциальных угроз достоверности оценок КТ особое внимание было уделено анализу возможных влияний на агрегированные оценки повторных прохождений теста отдельными респондентами. Интерес к проблеме был обусловлен тем, что ретестирование, как хорошо известно, приводит к повышению оценок когнитивных тестов [28]. С другой стороны, можно было предположить, что более склонными к ретестированию окажутся респонденты с низким уровнем способностей, прежде всего те, кто получает по результатам выполнения КОТ-30 отрицательный вердикт и не допускается к дальнейшим этапам обследования. Несмотря на то, что доля протоколов, идентифицированных как ретестирование, оказалась неожиданно высока (в среднем 7,6% от общего числа протоколов), их влияние на агрегированные оценки оказалось ничтожным. Некоторые примеры величин эффекта от чистки представлены в табл. 3, хотя в общей сложности было сопоставлено несколько десятков индексов, рассчитанных до и после этой процедуры. Зафиксированные при этом сдвиги в числовых значениях показателей обычно не выходили за рамки десятых долей процента. Как показал анализ выявленных и изолированных повторных протоколов,

причина столь незначительного их воздействия на агрегированные оценки состоит в том, что среди респондентов, прибегающих к повторным прохождением КТ, примерно в равной пропорции присутствуют лица, обладающие уровнями способностей как выше, так и ниже среднего. Тем не менее присутствие повторных протоколов приводит к незначительному диффузному увеличению всех средних оценок; соответственно, после их удаления средние оценки группового интеллекта становятся немного ниже (см., например, табл. 3).

Предполагалось, что наибольший ущерб повторные протоколы могут нанести выборочным средним на региональном уровне. Но и в этом случае смещения средних оказались невелики, поскольку средние оценки, рассчитанные по повторным протоколам, довольно высоко коррелируют со средними оценками первичных протоколов ($r = 0,768$) и в большинстве случаев отличаются от них не более чем на 5%.

5. Региональная локализация респондентов в абсолютном большинстве случаев оценена как правдоподобная. Об этом свидетельствует тесная связь ($r = 0,872$) между числом респондентов, представляющих каждый из 85 субъектов Российской Федерации, и численностью их населения (по состоянию на 01.01.2013 г.). Более того, во всех случаях, когда имелось заметное отклонение координаты региона от линии регрессии, связывающей посещаемость сайта с численностью населения, этому можно было найти достаточно правдоподобное объяснение. Например, относительно меньшее число респондентов дали регионы, где добываются углеводороды и / или уровень доходов населения относительно высок⁵; противоположный вектор отклонений был характерен для приграничных регионов, имеющих развитую военную инфраструктуру. Интересно отметить, что на первое место по удельной посещаемости с большим отрывом вышел Севастополь: если учитывать только первичные протоколы – более 7,6 обращений на каждую тысячу жителей, что почти вдвое превышает аналогичный показатель занимающих следующие ранговые позиции Бурятия, Мурманской и Амурской областей; Республика Крым также получила довольно высокий ранг (14). Поскольку прохождение тестирования на сайте Минобороны России мотивировано в первую очередь интересом к военной службе по контракту, данный факт, на наш взгляд, является убедительным свидетельством тесной ментальной связи этих территорий с Российской Федерацией.

Потенциальные уязвимости достоверности региональной локализации респондентов НИТ были идентифицированы и нейтрализованы в интересах проведения дальнейших исследований.

Обнадеживающие результаты первого этапа позволили перейти к следующему, основному этапу исследования, направленному на оценку эквивалентности (или различий) индексов, характеризующих выполнение когнитивного теста в условиях НИТ, аналогичным индексам, рассчитан-

⁵ Это, очевидно, снижает интерес к поискам альтернативных способов трудоустройства.

ным на основе результатов выполнения этого же теста в контролируемых условиях. Сопоставление проводилось как на уровне интегральных индексов, характеризующих результаты выполнения теста в целом, так и на пунктовом уровне, когда в качестве объекта сравнения использовались серии индексов, характеризующих процесс выполнения теста (трудность, дискриминативность и факторные нагрузки частных заданий на генеральный фактор). Полученные при этом оценки конгруэнтности сопоставлялись с аналогичными оценками, рассчитанными для ряда других когнитивных тестов, выполнявшихся в контролируемых условиях.

Хотя сравнение уровневых оценок КОТ-30 в условиях группового бланкового обследования и НИТ выявило заметные различия ($d = 0,35$ с интегральной выборкой НИТ), анализ особенностей выполнения теста в условиях НИТ респондентами, относящимися к разным образовательным категориям, позволяет связать данное расхождение с особенностями выборок и контекста, в котором осуществлялось тестирование (см. табл. 4). Показано, что если ограничить сравнение выборки ВСП с НИТ-респондентами, имеющими среднее профессиональное образование, стандартизированное различие средних падает до исчезающе малой величины (0,017), а корреляция между графиками распределения оценок этих двух групп достигает величины 0,98 (аналогичный показатель для респондентов НИТ со средним общим образованием равен 0,85, для лиц с высшим образованием – 0,67). Контингент ВСП, выполнявший КОТ-30, во-первых, был представлен преимущественно (свыше 90%) лицами со средним и средним профессиональным образованием, во-вторых, тестировался в начальный период адаптации к военной службе, когда функциональное состояние участников едва ли может считаться оптимальным. Таким образом, зафиксированные различия уровней оценок (при высокой степени сходства их вариативности и надежности), скорее всего, отражают реальные различия в эффективности решения задач между двумя сравниваемыми популяциями и едва ли могут ставить под сомнение валидность результатов выполнения КОТ-30 в условиях НИТ.

Наиболее убедительные, как нам представляется, оценки измерительной эквивалентности результатов выполнения КОТ-30 в ситуации НИТ и контролируемых условиях были получены с помощью сопоставления трех групп коэффициентов конгруэнтности, характеризующих процессы выполнения теста на пунктовом уровне, а именно – паттерны трудности, дискриминативности и факторных нагрузок частных тестовых заданий на генеральный фактор. Во всех случаях, когда результаты выполнения КОТ-30 в условиях НИТ сравнивались с результатами выполнения этого теста в контролируемых условиях (и к тому же в другом формате – бланковом), конгруэнтность данных оказывалась весьма высокой и превосходила по этому параметру некоторые из контрольных тестов.

Хотя этот тезис справедлив для всех трех групп КК, некоторые из них оказались более эффективными, чем другие. В использованных нами выборках максимальную дистанцию между конвергентными и дивергент-

ными КК продемонстрировали те из них, которые рассчитывались на основе факторных нагрузок; наименее эффективными оказались КК, рассчитанные на основе индексов трудности пунктов. Причина последнего результата представляется достаточно ясной: многие неадаптивные КТ конструируются таким образом, чтобы в начале теста располагались менее трудные задания, а в конце – более трудные (принцип прогрессии). В этом случае ряды чисел, отражающих изменение индексов трудности от начала теста к концу, могут обнаруживать довольно высокую степень сходства даже в тестах разного содержания и направленности.

Поскольку при сравнении КК контрольных тестов были выявлены некоторые тесты-аутсайдеры, анализ возможных причин таких отклонений может представлять интерес как для пользователей тестов, так и для тех, кто занимается их конструированием. В частности, наибольшее ослабление КК зафиксировано для методики «Аналогии» в условиях транзита форматов. Анализ возможной причины этого показал, что данный тест отличается от всех прочих тем, что проводится в условиях чрезмерно жесткого (для данной популяции) пятиминутного временного ограничения, при котором от 36 до 86% респондентов не успевают дать ответ на последний пункт (в то время как рекомендуемый лимит для тестов уровня – не более 20–30% [29. Р. 2]). Парадоксальным, на первый взгляд, кажется тот факт, что в условиях бланкового обследования число не справившихся с последними заданиями оказалось вдвое больше, чем при диалоговом. Однако он может быть объяснен тем, что в условиях диалогового обследования респондент с помощью линейной шкалы в нижней части экрана информируется о соотношении доли выполненных заданий и доли затраченного времени и благодаря этому имеет возможность оптимизировать выполнение теста, действуя в более быстром темпе ценой некоторого снижения точности. В бланковом формате респонденты не имеют возможности контролировать затраты времени и из-за этого больше времени тратят на ранние задания, не успевая решить поздние (т.е. демонстрируют более высокую точность при меньшей скорости). В итоге примерное равенство выборочных средних, дисперсий и ОМН обеспечивается фактически разными тактиками выполнения теста, что наиболее ярко отразилось в низкой межформатной конгруентности факторных нагрузок. В данном случае, несмотря на примерное равенство средних оценок, эквивалентность двух форматов может быть поставлена под сомнение.

Еще одним «побочным» продуктом выполненного нами исследования стало получение убедительных свидетельств того, что между 20- и 30-минутными версиями СПМ нет сколько-нибудь заметных различий; это согласуется с выводами Д. Давыдова и Е. Чмыховой [30. С. 137].

Заслуживает обсуждения вопрос о репрезентативности данных, полученных в условиях НИТ в контексте привлечения потенциальных кандидатов на военную службу по контракту. Наиболее очевидная уязвимость исследованных данных состоит в том, что респондентами были исключительно (или, по крайней мере, в абсолютном большинстве) мужчины. Од-

нако содержание использованного теста таково, что он содержит как вербальные задания (дающие некоторое преимущество женщинам), так и задания, адресованные счетным и пространственным способностям (где незначительное преимущество имеют мужчины). Таким образом, благодаря сбалансированности тестового контента нет оснований ожидать, что при ином гендерном составе участников обследования распределение оценок и выявленные взаимосвязи могли бы оказаться существенно иными. Тем не менее включение в состав участников НИТ женщин, несомненно, повысило бы степень убедительности любых выводов, которые могут быть сделаны на основе этих или подобных им данных. В связи с недавними изменениями кадровой политики Минобороны России есть основания рассчитывать на то, что в ближайшем будущем такая возможность представится.

Говоря о репрезентативности данных, нельзя не упомянуть о довольно широком возрастном, образовательном и, тем более, географическом диапазоне участников НИТ, с которым едва ли сможет сравниться какое-либо из опубликованных до настоящего времени отечественных исследований в области психологии. Если еще 10–15 лет назад о пользователях сети Интернет можно было говорить как об особой популяции, то в период сбора данных для нынешнего исследования доля населения, имеющего в России доступ к Интернету, превысила 2/3 и к 2016 г. достигла 76,4% [31. С. 1084]. Включение в число респондентов максимально широких (по географии, национальному составу, образовательному уровню и т.п.) слоев населения и тесно связанная с этим высокая статистическая мощность подобных исследований традиционно отмечаются как важнейшие преимущества НИТ, которые способны с лихвой перевесить отдельные присущие ему недостатки [5, 6, 9, 15]. Подтверждением справедливости этого тезиса стал уже упоминавшийся выше коэффициент корреляции (0,872) между численностью населения региона и численностью респондентов, представляющих данный регион в массиве данных НИТ. Столь тесная связь позволяет считать полученные данные репрезентативными, в том числе на региональном уровне.

Заключение

Значительный по объему массив данных, полученный в условиях неконтролируемого интернет-тестирования, оценивался с точки зрения возможности извлечения из подобных данных валидных оценок группового психометрического интеллекта. Одним из основных способов валидации данных, полученных в условиях НИТ, является демонстрация измерительной эквивалентности оценок, полученных в контролируемых и неконтролируемых условиях. Поскольку в известной литературе описывается довольно ограниченный набор способов оценки эквивалентности результатов выполнения когнитивных тестов, нами был предложен и апробирован ряд дополнительных приемов для проведения такого сопоставления на пунктовом уровне.

Для проверки работоспособности предложенных коэффициентов конгруентности были привлечены архивные данные, содержащие результаты выполнения различных когнитивных тестов в контролируемых условиях и разных форматах обследования.

Анализ НИТ-данных показал, что психометрические характеристики оценок выполнения когнитивного теста в этих условиях – и на шкальном, и на пунктовом уровнях – не уступают таковым, рассчитанным в условиях контролируемого тестирования в аттестационном контексте. Показано, что, даже несмотря на наличие потенциальных уязвимостей, рассчитанные на основе интернет-данных групповые оценки демонстрируют высокий уровень робастности и иммунитета по отношению к контаминации. Установлено, что коэффициенты конгруэнтности между результатами выполнения методики КОТ-30 в условиях НИТ и в контролируемых условиях имеют тот же уровень значений, что и КК, рассчитанные для широко применяемых когнитивных тестов в условиях их проведения в разных форматах в аттестационном контексте. Среди апробированных КК наилучшую дифференциацию между конвергентными и дивергентными индексами продемонстрировали те, которые рассчитывались на основе факторных нагрузок.

Результаты исследования позволяют сделать вывод о возможности использования данного массива НИТ-данных для изучения групповых психологических феноменов, в частности на популяционном и географическом уровнях.

Литература

1. Шмелев А.Г. Практическая тестология. Тестирование в образовании, прикладной психологии и управлении персоналом. М. : Маска, 2013. 688 с.
2. Использование новейших информационных технологий в мероприятиях по профессиональному психологическому отбору в Вооруженных Силах Российской Федерации : материалы научно-практической конференции (Москва, 2–3.10.2002 г.). М. : ГИШ ВС РФ, 2003. 248 с.
3. Компьютерная психодиагностика в профессиональном психологическом отборе военных специалистов РВСН : учеб.-метод. пособие / под ред. К.В. Сугоняева. М. : ЦИПК РВСН, 2005. 184 с.
4. Соколов А.А. Перспективы автоматизации профессионального психологического отбора в Вооруженных Силах Российской Федерации // 20 лет функционирования системы профессионального психологического отбора в Вооруженных Силах Российской Федерации, итоги и перспективы совершенствования : сб. мат. военно-науч. конф. (Москва, 22 апреля 2014 г.) / под общ. ред. С.В. Чваркова. М. : ВАГШ ВС РФ, 2014. С. 87–95.
5. Tippins N.T., Beaty J., Drasgow F., Gibson W.M., Pearlman K., Segall D.O., Shepherd W. Unproctored Internet testing in employment settings // *Personnel Psychology*. 2006. Vol. 59, № 1. P. 189–225. DOI: 10.1111/j.1744-6570.2006.00909.x.
6. Gosling S.D., Mason W. Internet research in psychology // *Annual Review of Psychology*. 2015. Vol. 66. P. 877–902. DOI: 10.1146/annurevpsych-010814-015321.
7. Soto C.J., John O.P., Gosling S.D., Potter J. Age differences in personality traits from 10 to 65: Big five domains and facets in a large cross-sectional sample // *Journal of Personality and Social Psychology*. 2011. Vol. 100, № 2. P. 330–348. DOI: 10.1037/a0021717.

8. Gebauer J.E., Sedikides C., Schönbrodt F.D., Bleidorn W., Rentfrow P.J., Potter J., Gosling S.D. The religiosity as social value hypothesis: a multi-method replication and extension across 65 countries and three levels of spatial aggregation // *Journal of Personality and Social Psychology*. 2017. Vol. 113, № 3. P. e18–e39. DOI: 10.1037/pspp0000104.
9. Gosling S.D., Sandy C.J., John O.P., Potter J. Wired but not WEIRD: the promise of the Internet in reaching more diverse samples // *Behavioral and Brain Sciences*. 2010. Vol. 33, № 2–3. P. 94–95. DOI: 10.1017/S0140525X10000300.
10. Gleibs I.H. Are all ‘research fields’ equal? Rethinking practice for the use of data from crowdsourcing market places // *Behavior Research Methods*. 2017. Vol. 49, № 4. P. 1333–1342. DOI: 10.3758/s13428-016-0789-y.
11. Chuah S.C., Drasgow F., Roberts B.W. Personality assessment: Does the medium matter? No // *Journal of Research in Personality*. 2006. Vol. 40, № 4. P. 359–376. DOI: 10.1016/j.jrp.2005.01.006.
12. Vecchione M., Alessandri G., Barbaranelli C. Paper-and-pencil and web-based testing: The measurement invariance of the Big Five personality tests in applied settings // *Assessment*. 2012. Vol. 19, № 2. P. 243–246. DOI: 10.1177/1073191111419091.
13. Le Corff Y., Gingras V., Busque-Carrier M. Equivalence of unproctored internet testing and proctored paper-and-pencil testing of the Big Five // *International Journal of Selection and Assessment*. 2017. Vol. 25, № 2. P. 154–160. DOI: 10.1111/ijsa.12168.
14. International Testing Commission. International guidelines on computer-based and Internet-delivered testing: Version 2005 // *International Journal of Testing*. 2006. Vol. 6, № 2. P. 143–172. DOI:10.1207/s15327574ijt0602_4.
15. Crump M.J.C., McDonnell J.V., Gureckis T.M. Evaluating Amazon’s Mechanical Turk as a tool for experimental behavioral research // *PLoS ONE*. 2013. Vol. 8, № 3. e57410. DOI: 10.1371/journal.pone.0057410.
16. Potosky D., Bobko P. Selection testing via the Internet: Practical considerations and exploratory empirical findings // *Personnel Psychology*. 2004. Vol. 57, № 4. P. 1003–1034. DOI: 10.1111/j.1744-6570.2004.00013.x.
17. Arthur W., Glaze R.M., Villado A.J., Taylor J.E. The magnitude and extent of cheating and response distortion effects on unproctored Internet-based tests of cognitive ability and personality // *International Journal of Selection and Assessment*. 2010. Vol. 18, № 1. P. 1–16. DOI: 10.1111/j.1468-2389.2010.00476.x.
18. Lievens F., Burke E. Dealing with the threats inherent in unproctored Internet testing of cognitive ability: Results from a large-scale operational test program // *Journal of Occupational and Organizational Psychology*. 2011. Vol. 84, № 4. P. 817–824. DOI: 10.1348/096317910X522672.
19. Радченко Ю.И., Сугоняев К.В. Разработка тестового комплекса для Интернет-самотестирования потенциальных кандидатов на военную службу по контракту // *Актуальные проблемы психологического обеспечения практической деятельности силовых структур : сб. мат. 3-й Всерос. науч.-практ. конф. СПб. : С.-Петербург. им. В.Б. Бобкова филиал РТА, 2014. С. 317–324.*
20. Grigoriev A., Ushakov D., Valueva E., Zirenko M., Lynn R. Differences in educational attainment, socio-economic variables and geographical location across 79 provinces of the Russian Federation // *Intelligence*. 2016. Vol. 58. P. 14–17. DOI: 10.1016/j.intell.2016.05.008.
21. Lynn R., Fuerst J., Kirkegaard E.O.W. Regional differences in intelligence in 22 countries and their economic, social and demographic correlates : a review // *Intelligence*. 2018. Vol. 69. P. 24–36. DOI: 10.1016/j.intell.2018.04.004.
22. Практикум по психодиагностике: конкретные психодиагностические методики. М. : МГУ, 1989. С. 112–126.
23. Hunt E. What makes nations intelligent? // *Perspectives on Psychological Science*. 2013. Vol. 7, № 3. P. 284–306. DOI: 10.1177/1745691612442905.

24. Российский стандарт тестирования персонала / Н.А. Батурин и др. // Организационная психология. 2015. Т. 5, № 2. С. 67–138.
25. Cohen J. A power primer // Psychological Bulletin. 1992. Vol. 112, № 1. P. 155–159. DOI: 10.1037/0033-2909.112.1.155.
26. McCrae R.R. A note on some measures of profile agreement // Journal of Personality Assessment. 2008. Vol. 90, № 2. P. 105–109. DOI: 10.1080/00223890701845104.
27. Furr R.M. A framework for profile similarity: Integrating similarity, normativeness, and distinctiveness // Journal of Personality. 2008. Vol. 76, № 5. P. 1267–1316. DOI: 10.1111/j.1467-6494.2008.00521.x.
28. Hausknecht J.P., Halpert J.A., Di Paolo N.T., Moriarty G.M.O. Retesting in selection: a meta-analysis of coaching and practice effects for tests of cognitive ability // Journal of Applied Psychology. 2007. Vol. 92, № 2. P. 373–385. DOI: 10.1037/0021-9010.92.2.373.
29. Kyllonen P.C., Zu J. Use of response time for measuring cognitive ability // Journal of Intelligence. 2016. Vol. 4, № 4. 14. DOI: 10.3390/jintelligence4040014.
30. Давыдов Д.Г., Чмыхова Е.В. Применение теста Стандартные прогрессивные матрицы Равена в режиме ограничения времени // Вопросы психологии. 2016. № 4. С. 129–139.
31. Регионы России. Социально-экономические показатели. 2017 : стат. сб. / Росстат. М., 2017. 1402 с.

Поступила в редакцию 06.05.2018 г.; принята 06.06.2018 г.

Сведения об авторах:

Сугоняев Константин Владимирович, кандидат технических наук, доцент, ассоциированный сотрудник Института психологии Российской академии наук, лаборатория психологии и психофизиологии творчества (Москва, Россия). E-mail: skv-354@yandex.ru

Радченко Юрий Иванович, начальник Научно-практического центра Военной академии Генерального штаба Вооруженных Сил Российской Федерации (Москва, Россия). E-mail: savaur@mail.ru

Соколов Александр Анатольевич, начальник лаборатории Научно-практического центра Военной академии Генерального штаба Вооруженных Сил Российской Федерации (Москва, Россия). E-mail: as450@mail.ru

VOLUNTARY INTERNET-BASED TESTING AS A SOURCE OF VALID GROUP PSYCHOMETRIC INTELLIGENCE SCORES

Sibirskiy Psikhologicheskii Zhurnal – Siberian journal of psychology, 2018, 69, 6–32. DOI: 10.17223/17267080/69/1

Konstantin V. Sugonyaev, Institute of Psychology of Russian Academy of Sciences (Moscow, Russian Federation). E-mail: skv-354@yandex.ru

Yurij I. Radchenko., Alexander A. Sokolov, Military Academy of General Staff (Moscow, Russian Federation). E-mail: savaur@mail.ru; as450@mail.ru

Keywords: Unproctored; Web-based testing; intelligence; general mental ability; psychometric properties; factor analysis.

Unproctored Internet-based testing (UIT) is becoming more popular in recruitment and employment settings due to its cost-effectiveness, and because it makes it possible to reach diverse and geographically distant populations. Growing body of evidences is accumulated about equivalence of UIT to traditional proctored paper-and-pencil testing as applied to surveys and personality inventories, but there is little information available about suitability of UIT for collecting ability-test data. We hypothesized that such data may be quite useful for estimation of psychometric intelligence of large groups differed on age, education level, regions of living and so on. To check this supposition we used large-scale data gathered via the

site www.mil.ru from September 2012 to December 2017, which included 262652 protocols of carrying-out of cognitive screening test (CST). This UIT-data were compared to results of carrying-out this test by military conscripts sample ($N = 3182$) in controlled conditions in paper-and-pencil mode (CPPT-data). Tests of measurement equivalency included both traditional procedures (comparison of scores distributions, means, standard deviations and reliability indexes) and three more fine grained ones based on item-level responses (comparison of patterns of item's difficulties, item-scale correlations and item's loadings on g-factor). Although mean scores were slightly higher for integrated UIT-sample, this difference disappeared if education level of UIT- and CPPT -responders has been equated. Links of aggregated IQ-scores extracted from UIT-data with age and education had high level of similarity with analogous links calculated for sample of 27388 candidates equated for age and gender, which was tested by other cognitive test in controlled high-stake conditions. Indexes of congruency between different segments of UIT-data and between UIT-, CPPT- and computer-based data on scale- and item-level were compared with analogous indexes calculated for 15 control samples fulfilled five cognitive tests in controlled conditions in the same or different modes (total size more than 47000). Congruency between different segments of UIT-data were none the worse than between samples examined by the same tests in same-mode cases; congruency between UIT- and CPPT-data for the CST was not the worse than congruency between data obtained from samples examined by the same tests in cross-mode cases. Although percentage of repeated testing protocols in UIT-data was near 7.6%, retesting didn't bias the mean IQ-scores of educational groups ($d < 0.01$). These finding suggests that UIT-data can be a source of valid information about group psychometric intelligence.

References

1. Shmelev, A.G. (2013) *Prakticheskaya testologiya. Testirovanie v obrazovanii, prikladnoy psikhologii i upravlenii personalom* [Practical test. Testing in education, applied psychology and personnel management]. Moscow: Maska.
2. General Staff of the Armed Forces of the Russian Federation. (2003) *Ispol'zovanie novykh informatsionnykh tekhnologiy v meropriyatiyakh po professi-onal'nomu psikhologicheskomu otboru v Vooruzhennykh Silakh Rossiyskoy Federatsii* [Use of the newest information technologies in activities on professional psychological selection in the Armed Forces of the Russian Federation]. Proc. of the Conference. Moscow. October 2–3, 2002. Moscow: General Staff of the Armed Forces of the Russian Federation.
3. Sugonyaev, K.V. (ed.) (2005) *Komp'yuternaya psikhodiagnostika v professional'nom psikhologicheskoy otbore voennykh spetsialistov RVSN* [Computer psychodiagnostics in the professional psychological selection of military specialists of the Strategic Missile Troops of the Russian Federation]. Moscow: TsIPK RVSN.
4. Sokolov, A.A. (2014) *Perspektivy avtomatizatsii professional'nogo psikhologicheskogo otbora v Vooruzhennykh Silakh Rossiyskoy Federatsii* [Prospects of automation of professional psychological selection in the Armed Forces of the Russian Federation]. In: Chvarkov, S.V. (ed.) *20 let funktsionirovaniya sistemy professional'nogo psikhologicheskogo otbora v Vooruzhennykh Silakh Rossiyskoy Federatsii, itogi i perspektivy sovershenstvovaniya* [20 years of the system of professional psychological selection in the Armed Forces of the Russian Federation, results and prospects for improvement]. Moscow: VAGSh VS RF. pp. 87–95.
5. Tippins, N.T., Beaty, J., Dragow, F., Gibson, W.M., Pearlman, K., Segall, D.O. & Shepherd, W. (2006) Unproctored Internet testing in employment settings. *Personnel Psychology*. 2006. 59(1). pp. 189–225. DOI: 10.1111/j.1744-6570.2006.00909.x
6. Gosling, S.D. & Mason, W. (2015) Internet research in psychology. *Annual Review of Psychology*. 66. pp. 877–902. DOI: 10.1146/annurevpsych-010814-015321
7. Soto, C.J., John, O.P., Gosling, S.D. & Potter, J. (2011) Age differences in personality traits from 10 to 65: Big five domains and facets in a large cross-sectional sample. *Journal of Personality and Social Psychology*. 100(2). pp. 330–348. DOI: 10.1037/a0021717

8. Gebauer, J.E., Sedikides, C., Schönbrodt, F.D., Bleidorn, W., Rentfrow, P.J., Potter, J. & Gosling, S.D. (2017) The religiosity as social value hypothesis: a multi-method replication and extension across 65 countries and three levels of spatial aggregation. *Journal of Personality and Social Psychology*. 113(3). pp. e18–e39. DOI: 10.1037/pspp0000104
9. Gosling, S.D., Sandy, C.J., John, O.P. & Potter, J. (2010) Wired but not WEIRD: the promise of the Internet in reaching more diverse samples. *Behavioral and Brain Sciences*. 33(2–3). pp. 94–95. DOI: 10.1017/S0140525X10000300
10. Gleibs, I.H. (2017) Are all ‘research fields’ equal? Rethinking practice for the use of data from crowdsourcing market places. *Behavior Research Methods*. 49(4). pp. 1333–1342. DOI: 10.3758/s13428-016-0789-y
11. Chuah, S.C., Drasgow, F. & Roberts, B.W. (2006) Personality assessment: Does the medium matter? No. *Journal of Research in Personality*. 40(4). pp. 359–376. DOI: 10.1016/j.jrp.2005.01.006
12. Vecchione, M., Alessandri, G. & Barbaranelli, C. (2012) Paper-and-pencil and web-based testing: The measurement invariance of the Big Five personality tests in applied settings. *Assessment*. 19(2). pp. 243–246. DOI: 10.1177/1073191111419091
13. Le Corff, Y., Gingras, V. & Busque-Carrier, M. (2017) Equivalence of unproctored internet testing and proctored paper-and-pencil testing of the Big Five. *International Journal of Selection and Assessment*. 25(2). pp. 154–160. DOI: 10.1111/ijsa.12168
14. International Testing Commission. (2006) International guidelines on computer-based and Internet-delivered testing: Version 2005. *International Journal of Testing*. 6(2). pp. 143–172. DOI: 10.1207/s15327574ijt0602_4
15. Crump, M.J.C., McDonnell, J.V. & Gureckis, T.M. (2013) Evaluating Amazon’s Mechanical Turk as a tool for experimental behavioral research. *PLoS ONE*. 8(3). e57410. DOI: 10.1371/journal.pone.0057410
16. Potosky, D. & Bobko, P. (2004) Selection testing via the Internet: Practical considerations and exploratory empirical findings. *Personnel Psychology*. 57(4). pp. 1003–1034. DOI: 10.1111/j.1744-6570.2004.00013.x
17. Arthur, W., Glaze, R.M., Villado, A.J. & Taylor, J.E. (2010) The magnitude and extent of cheating and response distortion effects on unproctored Internet-based tests of cognitive ability and personality. *International Journal of Selection and Assessment*. 18(1). pp. 1–16. DOI: 10.1111/j.1468-2389.2010.00476.x
18. Lievens, F. & Burke, E. (2011) Dealing with the threats inherent in unproctored Internet testing of cognitive ability: Results from a large-scale operational test program. *Journal of Occupational and Organizational Psychology*. 84(4). pp. 817–824. DOI: 10.1348/096317910X522672
19. Radchenko, Yu.I. & Sugonyaev, K.V. (2014) Razrabotka testovogo kompleksa dlya Internet-samotestirovaniya potentsial'nykh kandidatov na voennuyu sluzhbu po kontraktu [Development of a test complex for Internet self-testing of potential candidates for military service under the contract]. In: Agrashenkov, A.V., Vinogradov, P.N., Lagun, A.V., Posokhova, S.T. & Reshetnikov, M.M. (eds) *Aktual'nye problemy psikhologicheskogo obespecheniya prakticheskoy deyatel'nosti silovykh struktur* [Topical problems of practical activity psychological support in power structures]. St. Petersburg: St. Petersburg Branch of the Russian Customs Academy. pp. 317–324.
20. Grigoriev, A., Ushakov, D., Valueva, E., Zirenko, M. & Lynn, R. (2016) Differences in educational attainment, socio-economic variables and geographical location across 79 provinces of the Russian Federation. *Intelligence*. 58. pp. 14–17. DOI: 10.1016/j.intell.2016.05.008
21. Lynn, R., Fuerst, J. & Kirkegaard, E.O.W. (2018) Regional differences in intelligence in 22 countries and their economic, social and demographic correlates: a review. *Intelligence*. 69. pp. 24–36. DOI: 10.1016/j.intell.2018.04.004
22. Panteleev, S.R. (1989) *Praktikum po psikhodiagnostike: konkretne psikhodiagnosticheskie metodiki* [Workshop on psycho-diagnostics: specific psychodiagnostic techniques]. Moscow: Moscow State University. pp. 112–126.

23. Hunt, E. (2013) What makes nations intelligent? *Perspectives on Psychological Science*. 7(3). pp. 284–306. DOI: 10.1177/1745691612442905
24. Baturin, N.A. et al. (2015) Russian Standard for Personnel Testing. *Organizatsionnaya psikhologiya – Organizational Psychology*. 5(2). pp. 67–138. (In Russian).
25. Cohen, J. (1992) A power primer. *Psychological Bulletin*. 112(1). pp. 155–159. DOI: 10.1037/0033-2909.112.1.155
26. McCrae, R.R. (2008) A note on some measures of profile agreement. *Journal of Personality Assessment*. 90(2). pp. 105–109. DOI: 10.1080/00223890701845104
27. Furr, R.M. (2008) A framework for profile similarity: Integrating similarity, normativeness, and distinctiveness. *Journal of Personality*. 76(5). pp. 1267–1316. DOI: 10.1111/j.1467-6494.2008.00521.x
28. Hausknecht, J.P., Halpert, J.A., Di Paolo, N.T. & Moriarty, G.M.O. (2007) Retesting in selection: a meta-analysis of coaching and practice effects for tests of cognitive ability. *Journal of Applied Psychology*. 92(2). pp. 373–385. DOI: 10.1037/0021-9010.92.2.373
29. Kyllonen, P.C. & Zu, J. (2016) Use of response time for measuring cognitive ability. *Journal of Intelligence*. 4(14). DOI: 10.3390/jintelligence4040014
30. Davydov, D.G. & Chmykhova, E.V. (2016) Primenenie testa Standartnye progressivnye matritsy Ravena v rezhime ogranicheniya vremeni [Application of the test Standard progressive matrices of Raven in the time-limited mode]. *Voprosy psikhologii*. 4. pp. 129–139.
31. Egorenko, S.N. (ed.) (2017) *Regiony Rossii. Sotsial'no-ekonomicheskie pokazateli. 2017* [Regions of Russia. Socio-economic indicators. 2017]. Moscow: [s.n.].

Received 06.05.2018;

Accepted 06.06.2018