

УДК: 159.9.072

doi: 10.11621/vsp.2021.01.06

ПРОБЛЕМА ОДНОРОДНОСТИ ШКАЛ ИНТЕЛЛЕКТУАЛЬНЫХ СПОСОБНОСТЕЙ: ПСИХОМЕТРИЧЕСКАЯ ОЦЕНКА

А.Н. Кричевец¹, А.А. Корнеев^{1*}, К.В. Сугоняев²

¹ МГУ имени М.В. Ломоносова, Москва, Россия.

² Институт психологии РАН, Москва, Россия.

Для контактов*. E-mail: korneeff@gmail.com

Актуальность. При разработке психометрических инструментов и оценке их структуры часто используются стандартные процедуры, предусматривающие расчет ограниченного набора коэффициентов. Однако шаблонное использование таких процедур не позволяет учитывать возможные нюансы исследуемых шкал и может приводить к получению неадекватных результатов. В силу этого важным представляется подробный анализ структуры психометрических инструментов, который допускает различные нестандартные приемы.

Цель нашей работы — на основе данных выполнения двух интеллектуальных субтестов провести более детальный анализ их структуры на уровне отдельных заданий и оценить достаточность применения стандартных методов оценки надежности и структурной валидности для этих субтестов.

Методики и выборка. В качестве материала для анализа мы использовали данные, полученные в ходе тестирования интеллекта у большой выборки респондентов (11335 человек) с помощью батареи «КР-3». В работе подробно рассматривается структура субтестов «Силлогизмы» и «Аналогии»: анализируется надежность шкал, оцениваемая с помощью коэффициента альфа Кронбаха, и факторная структура на пунктовом уровне с помощью конфирматорного факторного анализа.

Результаты и выводы. Оценка надежности анализируемых шкал с помощью коэффициента альфа Кронбаха показала, что получаемые оценки зависят от того, учитывается ли при анализе и отборе результатов влияние ограничения времени, стандартно используемого в групповых тестах на интеллект. С другой стороны, подробный анализ отдельных заданий в каждом из субтестов позволил выделить дополнительный, не заложенный

изначально фактор способностей предположительно более общего типа, чем специальные способности, на оценку которых они направлены. Конфирматорный факторный анализ показал существенность этого фактора с точки зрения оценки структуры шкал. Проведенный анализ позволяет заключить, что без подробного анализа процедуры тестирования и структуры заданий исследователь может не заметить важных свойств анализируемых шкал и сделать неполные или неадекватные выводы об их психометрических качествах.

Ключевые слова: структура интеллекта, психометрический интеллект, надежность шкалы, конфирматорный факторный анализ.

Благодарности. Исследование выполнено при поддержке РФФИ, проект 17-29-07030.

Для цитирования: Кричевец А.Н., Корнеев А.А., Сугоняев К.В. Проблема однородности шкал интеллектуальных способностей: психометрическая оценка // Вестник Московского университета. Серия 14. Психология. 2021. № 1. С. 144–169. doi: 10.11621/vsp.2021.01.06

Поступила в редакцию: 07.11.2020 / Принята к публикации: 25.12.2020

THE HOMOGENEITY OF SCALES OF INTELLECTUAL ABILITIES: PSYCHOMETRIC ANALYSIS

Anatoly N. Krichevets¹, Alexey A. Korneev^{1*}, K.V. Sugonyaev²

¹ Lomonosov Moscow State University, Moscow, Russia.

² Institute of Psychology of Russian Academy of Sciences, Moscow, Russia.

Corresponding author*. E-mail: korneeff@gmail.com

Relevance. Nowadays the researchers commonly use a limited set of standard procedures and statistical coefficients when develop psychometric instruments and investigate their structure. The routine using of such procedures without taking into account the specific features of the psychometric scales can lead to incomplete or even inadequate results. In this context detailed consideration of the structure of psychometric instruments seems to be important and it may demand various non-standard ways of statistical analysis.

Objectives. To conduct detailed analysis of the results of two intelligent subtests at the item level and to assess the sufficiency and adequacy of using standard methods for estimation of reliability and structural validity for these subtests.

Methods. We analyze the data collected in intelligence testing of a large sample of respondents (11335 young adults). The respondents passed the KR-3 battery. In this study we examine in detail the structure of the subtests “Syllogisms” and “Analogies”. Specifically, we estimated the reliability of the scales by the Cronbach’s alpha coefficient, and the structure at the item level using the confirmatory factor analysis.

Results and conclusions. Estimation of the reliability of the scales by Cronbach’s alpha coefficient showed importance of taking into account the time limitation, which is commonly used in intelligence tests. On the other hand, a detailed analysis of each subtest items made it possible to find out an additional factor which was not originally proposed in the factor structure. This is factor of higher-order abilities of abstract analysis, whilst the subtest originally aimed at estimation of the special abilities. Confirmatory factor analysis showed improvement of fit when this factor was added. The results allow to conclude that the researcher may miss the important properties of scales if not making a detailed analysis of testing procedures and the structure of subtest at the item level, and so may draw incomplete or inadequate conclusions about their psychometric properties

Keywords: structure of intelligence, psychometric intelligence, reliability, confirmatory factor analysis.

For citation: Krichevets, Korneev, A.A., A.N., Sugonyaev, K.V. (2021) The homogeneity of scales of intellectual abilities: psychometric analysis. *Vestnik Moskovskogo Universiteta. Seriya 14. Psikhologiya [Moscow University Psychology Bulletin]*, 1, P. 144–169. doi: 10.11621/vsp.2021.01.06

Received: November 07, 2020 / Accepted: December 25, 2020

Введение

Проблема психометрических свойств и структуры инструментов, используемых в исследованиях различных психических феноменов — тема, активно обсуждаемая в современной научной литературе. При этом, однако, часто проверка и оценка методик происходит по стандартным алгоритмам, которые, как и любой шаблон, с одной стороны упрощают задачу анализа, но с другой, могут приводить к потере важной информации об исследуемой реальности и тех инструментах, которые для этого используются. В рамках настоящей работы мы обсуждаем структуру двух шкал психометрического интеллекта, и

анализируем, насколько оправдан стандартный, шаблонный подход к его оценке и что может дать более пристальный взгляд на особенности этих шкал и их состава.

Проблема оценки надежности и структуры психометрического инструмента

В настоящее время сложилась практически общепринятая процедура анализа структуры многомерного теста. Считается, что каждый конструкт, отражающий психологическое качество, должен оцениваться в тесте определенным набором заданий (задач, вопросов, утверждений), ответы на которые показывают выраженность этого качества у данного индивида. Такой набор заданий как правило называется шкалой или субтестом, а несколько шкал/субтестов составляют тест. Стандартная проверка структуры теста включает в себя, по крайней мере, две процедуры. Во-первых, структура каждого субтеста оценивается на предмет надежности (reliability) получаемого с его помощью итогового показателя, а во-вторых, оценивается соотношение пунктов субтеста между собой и с латентными переменными (психологическими конструктами), на измерение которых он направлен, а также соотношение этих конструктов между собой, для чего применяется факторный анализ (так называемая факторная или структурная валидность, см., например, Митина, 2013).

В самом общем виде, надежность в классической теории тестирования связывается со способностью психометрического инструмента измерять реальное состояние конструкта, на оценку которого он направлен, или, в математическом выражении, это отношение (оценки) дисперсии истинных баллов k (оценке) дисперсии наблюдаемых баллов. Вот одна из характеристик надежности: «Согласно классической теории тестов, надежность теста указывает на степень, в которой различия в тестовых баллах респондентов являются функцией от истинных различий между ними, а не ошибкой измерения» (Фер, Бакарак, 2010, стр. 91). Существуют различные подходы к оценке надежности, подробное их рассмотрение выходит за пределы данной работы, мы остановимся на понимании надежности как внутренней согласованности результатов, получаемых с помощью психометрического инструмента (Netemeyer, Bearden, Sharma, 2003).

Структура тестов на интеллект, как правило, включает в себя несколько субтестов или шкал, которые направлены на оценку отдельных частных способностей. При этом, традиционный психометрический подход подразумевает, что эти шкалы относительно

независимы, а задания, входящие в один субтест или шкалу однородны — они могут различаться по уровню сложности, но принципиально направлены на измерение одной способности. В силу этого, например, предполагается, что измеренная с помощью линейных коэффициентов корреляции или ковариации согласованность ответов на задания, относящихся к одному субтесту, будет достаточно высока, и при этом сильнее, чем с ответами на задания из других субтестов и шкал. На этих самых общих идеях основаны такие стандартные психометрические процедуры, как оценка надежности шкалы с помощью коэффициента альфа Кронбаха (Cronbach's α).

Достаточно очевидно, что такая «идеальная» структура психометрического инструмента (будь то тест на интеллект или какая-либо другая методика), в которой все пункты направлены на измерение только одного конструкта и четко дифференцируются по шкалам — ситуация достаточно редкая, а подчас и невозможная. По крайней мере, усредненные оценки ковариаций, к которым сводится коэффициент альфа Кронбаха — это часто слишком грубая оценка, которая в определенных обстоятельствах может дать недооценку или переоценку реальных психометрических свойств исследуемых методик. В литературе отмечается, что использование таких оценок надежности в качестве показателей однородности пунктов, входящих в одну шкалу, нежелательно; возможны достаточно высокие значения коэффициента альфа Кронбаха даже при неоднородности (т.е. неоднородности) пунктов, включенных в ту или иную шкалу (Фер, Бакарарк, 2010, гл. 6, Netemeyer, Bearden, Sharma, 2003). Для прояснения этого вопроса более целесообразно использование методов факторного анализа применительно к отдельным заданиям теста, и в рамках нашей работы мы показываем, что этот подход может прояснить некоторые неоднозначные вопросы, касающиеся «мерности» того или иного тестового задания.

Также при использовании стандартных методов оценки психометрических свойств теста нужно учитывать, что даже однородные по структуре задания могут быть связаны с использованием различных способностей. Так, решение заданий в одном из субтестов на интеллект, подчас, может требовать не только специфических способностей, на оценку которых направлен субтест или шкала, но и более общих интеллектуальных возможностей. Без подробного рассмотрения специфики пунктов с этой точки зрения, анализ структуры психометрического инструмента оказывается неполным и наша работа направлена на демонстрацию возможностей такого анализа.

Еще одна проблема, которая возникает при оценке психометрических свойств именно тестов на интеллект, связана с часто используемым ограничением времени на их выполнение. Введение ограничения на время выполнения теста естественным образом делает существенным фактор скорости реакции респондента (в самом широком смысле). Связь времени реакции и интеллекта обсуждается достаточно давно (Jensen, Munro, 1979; Barrett, Eysenck, Lucking, 1986 и др.), при этом интеллект связывается со скоростью переработки информации в более общем смысле (см. обзор Sheppard, Vernon, 2008). Оставляя за пределами данной работы подробное обсуждение этого вопроса, отметим, что в таком случае в структуру тестов имеет смысл включать и отдельный фактор скорости, имея в виду при этом, что успешность выполнения заданий с большим порядковым номером будут автоматически больше зависеть от этого фактора, чем выполнение заданий с меньшим номером. Это показано в работе (Estrada, Román, Abad, Colom, 2017), в которой последовательное включение в структурную модель фактора скорости для пунктов из второй половины интеллектуального теста, выполняемого в жестких временных ограничениях, приводило к улучшению оценок качества моделей, при этом факторные нагрузки по основному фактору измеряемой способности для этих пунктов ослабевали. Возможно также, что ограничение может по-разному влиять на успешность выполнения задач разного типа и сложности, что еще больше усложнит факторную структуру. Кроме того, есть данные, что введение временных ограничений ухудшает психометрические свойства тестов на интеллект и может «размывать» оценки структурной модели (Корнилова и др., 2019)

Метод

Выборка и методика

В рамках данного исследования мы рассматриваем результаты тестирования интеллекта у абитуриентов военно-учебных заведений Минобороны России (результаты были собраны в 21 вузе из 18 регионов Российской Федерации). Они включают в себя в общей сложности ответы 11355 респондентов мужского пола (средний возраст $18,4 \pm 1,29$ лет). В рамках данной работы мы анализируем результаты двух субтестов из батареи «КР-3» (подробное описание батареи — см. Методические рекомендации по организации и проведению профессионального психологического отбора в военно-учебных заведениях Министерства обороны Российской

Федерации, 2002) — «Аналогии» (далее AN) и «Силлогизмы» (далее SIL), каждое из которых содержит 30 заданий. Выше мы упоминали трудности, возникающие при анализе факторной структуры тестов, имеющих ограничения на время выполнения. Рассматриваемые субтесты имели такие ограничения (соответственно, 5 и 10 минут). Поскольку нас интересовали только те задания, на которые респонденты давали обдуманый ответ, мы решили включить в анализ подмножество заданий и только тех респондентов, которые успевали на них ответить. Для двух рассматриваемых субтестов мы остановились на начальном отрезке из 19 заданий, которые включали по три задания с правильным ответом «нет варианта» в каждом субтесте (в субтесте «аналогии» их всего три, в «силлогизмах» — пять, подробное обсуждение специфики этих пунктов — см. в первой части раздела «результаты»). Мы отбирали респондентов, которые успевали дать ответ на двадцатые задания в обоих субтестах, то есть ответ на девятнадцатое задание с большой вероятностью был дан ими обдуманно. Таких респондентов оказалось 8529.

Статическая обработка данных

Полученные в результате тестирования бинарные переменные — правильность ответов на каждое из заданий. Вероятность правильного ответа зависит от трудности задания. Два задания, измеряющие строго одну интеллектуальную функцию, но имеющие разную трудность, в психометрическом смысле должны характеризоваться как коллинеарные. Однако коэффициент корреляции Пирсона может для них принимать сколь угодно малое значение — тем меньшее, чем больше разница трудностей. Мы использовали тетрагорический коэффициент корреляции (также известный как тетрагорический коэффициент сопряженности Пирсона, Digvi, 1979), который дает в данном случае адекватный высказанным соображениям результат.

Для проверки гипотезы о наличии, помимо факторов частных способностей, обеспечивающих выполнение отдельных субтестов, более общего фактора, позволяющего успешнее решать более сложные задания с отсутствием правильного варианта ответа (об этом см. ниже, в разделе «Содержательный анализ субтестов и входящих в них пунктов»), мы использовали метод конфирматорного факторного анализа. При этом, так как исходные переменные представляли бинарные шкалы, мы использовали метод взвешенных наименьших квадратов с поправками среднего и дисперсии (WLSMV, Muthén,

Muthén, 2012), пригодный для построения и оценки структурных моделей на основе категориальных и бинарных данных. Расчет моделей проводились в пакете Mplus, ver. 8.4.

Результаты

Содержательный анализ субтестов и входящих в них пунктов

Прежде чем оценивать психометрические свойства включенных в анализ субтестов, рассмотрим подробнее особенности входящих в них пунктов. Первый субтест, рассматриваемый в рамках нашей работы, как было указано выше, это «Силлогизмы», направленный на «оценку способности к логическому рассуждению» (Сугоняев, Радченко, 2018, с. 8). Выполнение этого субтеста подразумевает способность к логическому выводу на основе формальных правил. Подобные задания достаточно активно обсуждаются в исследованиях интеллекта (см, например, Johnson-Laird, 1983; Ушаков, 2005) и широко распространены в различных тестах на интеллект. Второй субтест «Аналогии» позволяет провести «оценку способности к пониманию характера логических связей и отношений между понятиями» (там же). Задания этого типа также являются достаточно известными и часто включаются в субтесты, направленные на измерение вербального интеллекта; например, они входят в состав IST-70 Р. Амтхауэра (Елисеев, 2003), Школьного теста умственного развития (ШТУР, Гуревич и др., 1987) и других). Таким образом, оба субтеста направлены на установление логических связей, но разных их аспектов. Опишем более подробно их структуру.

Субтест «Силлогизмы» состоит из 30 наборов посылок, из которых надо вывести верное умозаключение. Пример задания (здесь и далее полужирным выделены правильные ответы):

Посылки:

*Ни одна рыба не имеет четырехкамерного сердца
Все акулы — рыбы*

Варианты ответа:

- А) Все акулы имеют четырехкамерное сердце*
- Б) Ни одна акула не имеет четырехкамерного сердца***
- В) Некоторые акулы имеют четырехкамерное сердце*
- Г) Некоторые акулы имеют двухкамерное сердце*
- Д) Нет варианта*

Подробное рассмотрение отдельных пунктов показало, что среди заданий этого субтеста наиболее трудными для испытуемых оказались задания такого типа (пример придуман нами, чтобы не публиковать ответы на входящее в тест задание):

Щука — это рыба.

Ни один окунь не является щукой.

А) Всякий окунь — рыба

Б) Ни один окунь не является рыбой

В) Некоторые окуни являются рыбой

Г) Некоторые окуни не являются рыбой

Д) Нет ответа

В этом задании два начальных утверждения не являются посылками силлогизма, и из них невозможно вывести какое-либо умозаключение. В то же время среди ответов имеются содержательно правильные утверждения, и их оценка как неправильных требует более глубокого понимания формальной структуры силлогизма, в отличие от заданий типа предыдущего.

Второй субтест — «Аналогии» состоит из 30 наборов слов, в каждом из которых первые два составляют пару, а к третьему нужно из четырех вариантов ответа подобрать слово, находящееся с ним в таком же отношении, как второе слово с первым. Также в каждом задании есть вариант «нет ответа», который может оказаться правильным. В каждом из заданий есть только один верный ответ.

Пример задания с вариантами ответов:

Летчик — небо, моряк — ?

А) Парус

Б) Море

В) Корабль

Г) Шторм

Д) Нет варианта

Аналогично предыдущему, в данном субтесте также относительно более трудными были задания такого типа:

Стол — мебель; береза — ?

А) Ствол

Б) Трава

В) Лес

Г) Топор

Д) Нет ответа

Эти задания оказались все же проще, чем подобные задания в субтесте «Силлогизмы», вероятно потому, что четвертое понятие практически очевидно — в данном случае дерево (возможно, растение), и его отсутствие в списке устанавливается достаточно легко. В результате, если в «Силлогизмах» задания с правильным ответом «Д» существенно труднее остальных заданий, то в «Аналогиях» разрыв значительно меньше, и даже не все задания с ответом «Д» являются труднейшими. Тем не менее, мы решили проверить гипотезу о том, что для решения заданий этого типа в обоих субтестах требуется достаточный уровень развития особой функции, обеспечивающей работу с формальной стороной логических связей. В таком случае можно ожидать, что более высокий уровень развития этой функции приводит к успеху сначала в решении задач на аналогии, а затем и более трудных задач на силлогизмы. Так как эти трудные вопросы предположительно связаны со способностью прорабатывать формально-логические структуры, в дальнейшем мы их называем «формальными», так же как и факторы, соответствующие способностям такого типа.

Оценка надежности шкал

Мы рассчитали описательную статистику и провели анализ надежности двух шкал с помощью стандартного коэффициента надежности альфа Кронбаха, широко использующегося для этих целей. Мы рассчитали этот коэффициент для каждого из двух субтестов в различных условиях: в полной выборке для всех 30 заданий и для первых 19 заданий, а также отдельно для двух подвыборок: тех, кто успевал дать обдуманый ответ на двадцатое задание в обоих субтестах (далее — «быстрые»), и тех, кто в двадцатом пункте хотя бы одного из субтестов давал неверный ответ (далее — «медленные»), и отдельно для набора из 30 и 19 заданий. Результаты приведены в табл. 1.

Как видно из таблицы, исходная надежность обоих субтестов достаточно высокая ($\alpha > 0,8$). Однако при сокращении числа вопросов коэффициент уменьшается, что ожидаемо, так как он зависит от числа пунктов в шкале (см., например, подробный разбор в Фер, Бакарак, 2010). Другой фактор, влияющий на оценку надежности — это разнородность респондентов: при прочих равных условиях, чем выше вариативность уровней способности в выборке, тем выше оказывается коэффициент надежности (там же).

Таблица 1

**Описательная статистика и Альфа Кронбаха
 для двух анализируемых субтестов в различных условиях**

Выборка	Число заданий	AN		SIL	
		Среднее (ст. откл)	α Кронбаха	Среднее (ст. откл)	α Кронбаха
Полная (n = 11 335)	Полный набор (30 заданий)	18,78 (5,46)	0,844	15,6 (4,91)	0,813
	Сокращенный на- бор (19 заданий)	14,35 (3,15)	0,735	12,65 (3,38)	0,752
Быстрые (n = 8539)	Полный набор (30 заданий)	20,26 (4,79)	0,796	16,83 (4,41)	0,767
	Сокращенный на- бор (19 заданий)	14,86 (2,83)	0,689	13,23 (3,09)	0,721
Медленные (n = 2796)	Полный набор (30 заданий)	14,25 (4,84)	0,829	11,83 (4,38)	0,788
	Сокращенный на- бор (19 заданий)	12,78 (3,52)	0,765	10,88 (3,63)	0,752

Table 1

**Descriptive statistics and Cronbach's alpha
 for two analyzed subtests in different conditions**

Sample	Number of items	AN		SIL	
		Mean (standard deviations)	Cronbach's α	Mean (standard deviations)	Cronbach's α
Full sample (n = 11 335)	Full set of items (30 items)	18,78 (5,46)	0,844	15,6 (4,91)	0,813
	Reduced set of items (19 items)	14,35 (3,15)	0,735	12,65 (3,38)	0,752
"Fast" subsample (n = 8539)	Full set of items (30 items)	20,26 (4,79)	0,796	16,83 (4,41)	0,767
	Reduced set of items (19 items)	14,86 (2,83)	0,689	13,23 (3,09)	0,721
"Slow" subsample (n = 2796)	Full set of items (30 items)	14,25 (4,84)	0,829	11,83 (4,38)	0,788
	Reduced set of items (19 items)	12,78 (3,52)	0,765	10,88 (3,63)	0,752

Используемое нами разделение выборки приводит к снижению вариативности ответов (ср. стандартные отклонения в различных условиях), и также предсказуемо снижает оценки внутренней согласованности шкал.

В дальнейшей части работы мы рассматриваем только сокращенную выборку (8529 респондентов) и только первые 19 заданий из каждого субтеста.

Корреляция между пунктами

Далее, для анализа структуры анализируемых субтестов, мы рассчитали тетракорические корреляции между ответами на все задания, включенные в анализ. Мы обозначили GA (Аналогии общего вида) группу задач субтеста «Аналогии», у которых в списке был правильный ответ, FA (Формальные аналогии) — задачи с правильным выбором «нет ответа». Аналогично GS и FS — задания общего вида и формальные для субтеста «Силлогизмы». Всего рассчитано 10 средних корреляций между группами и внутри групп, они представлены на рис. 1. Отметим, что, поскольку наша выборка велика, отдельные корреляции оцениваются весьма точно. Усреднение здесь не имеет статистического смысла, а служит только для компактности представления данных.

Как видно из рис. 1, максимальные корреляции наблюдаются между формальными пунктами внутри субтестов, несколько ниже, но тоже относительно высокие — между неформальными внутри субтестов. Это может свидетельствовать о том, что «формальные» задания и задания общего вида обособляются в субтестах. При этом особенно интересно, что явно большие корреляции получены в группах формальных вопросов, то есть оказывается, что более согласованными являются ответы на выделенные нами вопросы, не имеющие правильного ответа и предположительно требующие особой способности. Также важно, что среди шести оставшихся корреляций максимальна корреляция $FS*FA$, что соответствует нашей гипотезе: «формальные» задания субтестов меньше отличаются друг от друга, чем от заданий общего вида «своих» субтестов ($FA*GA$ и $FS*GS$). Несколько странно, что последняя из них так мала — это означает, что преуспевающие в решении «формальных» силлогизмов могут терпеть неудачу в силлогизмах общего вида. Также трудно интерпретировать довольно большую корреляцию ($FA*GS$).

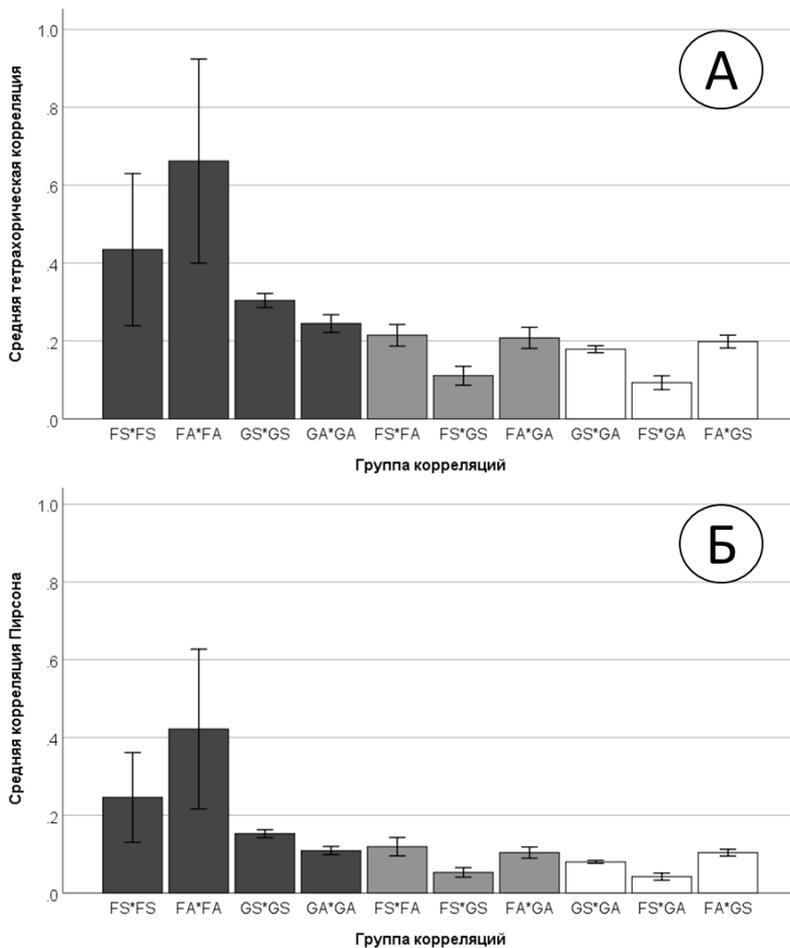


Рис. 1. Средние значения корреляций между пунктами разных типов. Черным обозначены средние корреляции между пунктами одного типа внутри субтестов; серым — средние корреляции между формальными пунктами разных субтестов и между формальными и неформальными пунктами в каждом из субтестов; белым — корреляции неформальных пунктов из разных субтестов и неформальных с формальными из разных субтестов. Столбики ошибок — 95% доверительные интервалы средних значений корреляций.

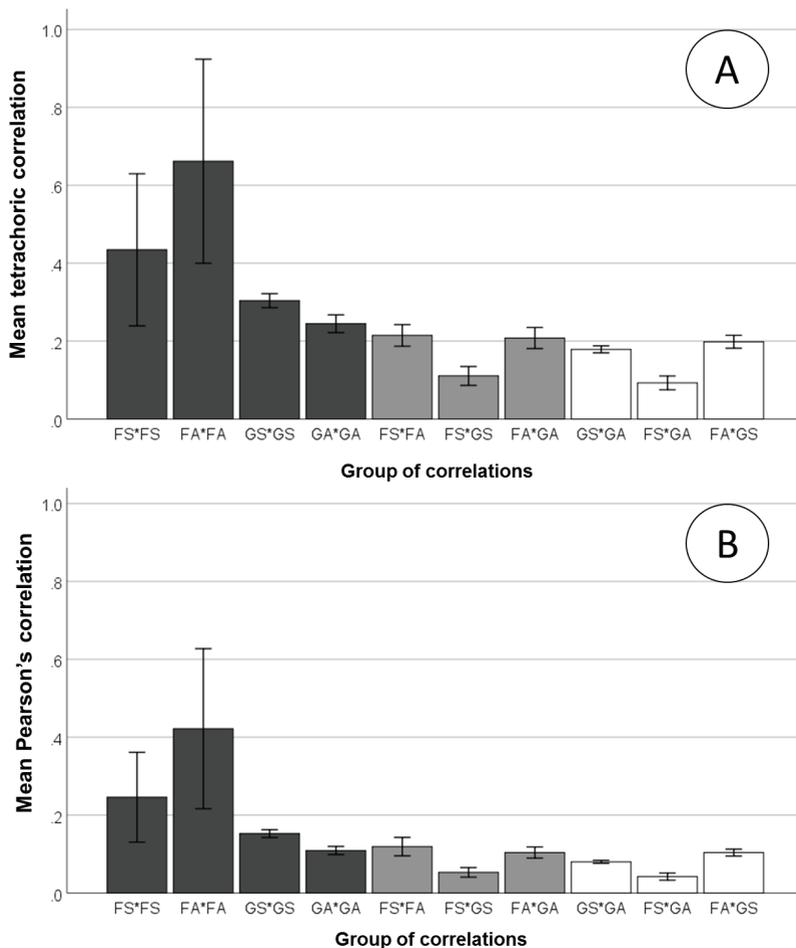


Fig. 1. Mean tetrachoric (A) and Pearson's (B) correlations between items of different types. Black color denotes the mean correlations between items of the same type within the subtests; gray — mean correlations between formal items of different subtests and between formal and informal items in each of the subtests; white — mean correlations between informal items and between informal and formal items from different subtests. Error bars — 95% confidence intervals of mean.

Структурные модели

Еще одна процедура, используемая при анализе психометрических свойств методик — это конфирматорный факторный анализ, позволяющий проверить соответствие структуры данных заложенным в методику шкалам. Мы рассчитали и оценили шесть моделей (см. рис. 2). Базовая модель, включающая в себя два фактора специальных способностей, соответствующих двум анализируемым субтестам, схематически представлена на рис. 2А. Также мы построили модель, включающую в себя еще один фактор, который связан с решением заданий без правильного ответа. Так как отдельный интерес представляет то, насколько этот фактор может быть связанным с факторами специальных способностей, мы рассчитали и оценили два

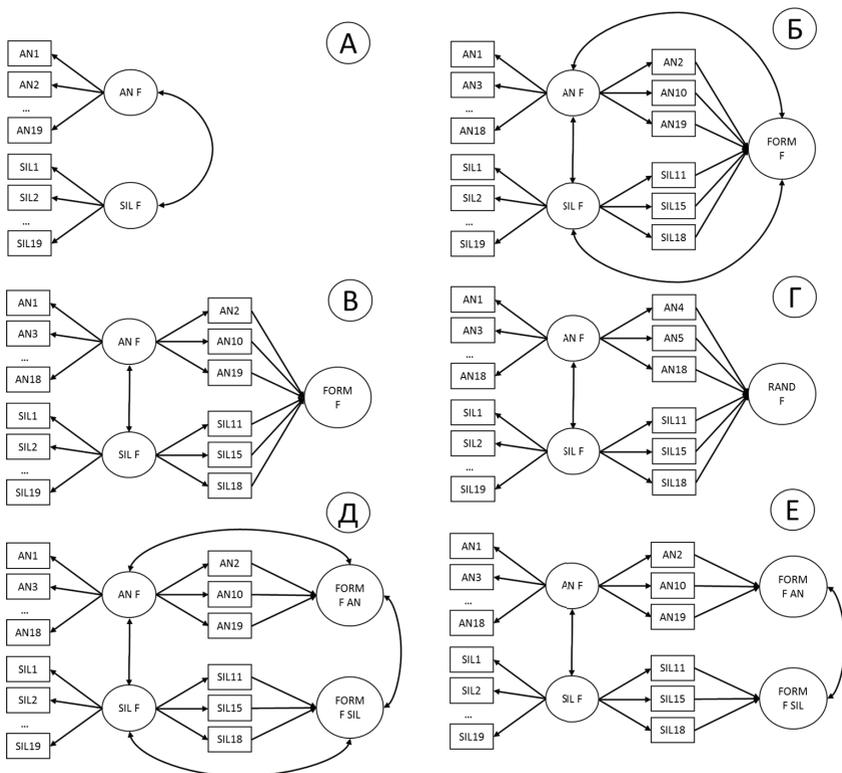


Рис. 2. Структурные модели, проверяемые в исследовании

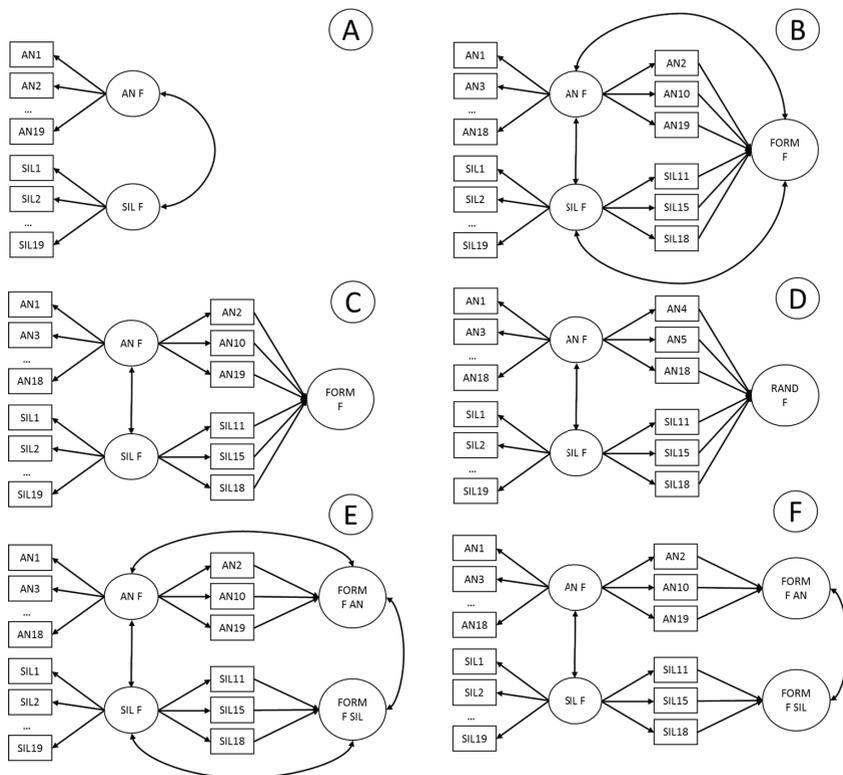


Fig. 2. Structural models in the study

варианта такой трехфакторной модели — с допущением корреляции между общим фактором и частными и без такого допущения (рис. 2Б и 2В соответственно). Для того, чтобы проверить, насколько изменение оценки модели при добавлении третьего фактора обусловлено именно спецификой пунктов без правильного ответа, а не просто добавлением новых связей, и не связан ли этот фактор с большей сложностью отобранных пунктов, была построена и проверена еще одна модель. По структуре она аналогична третьей модели, однако третий фактор в ней связывается с тремя сопоставимыми по сложности заданиями из субтеста «Аналогии» и тремя вопросами без ответов из субтеста «Силлогизмы» (так как аналогичных по сложности заданий в этом субтесте нет (см. рис. 2Г).

Наконец, исходя из результатов анализа корреляций между группами пунктов, была построена модель, в которой формальные пункты в каждом из субтестов связаны с отдельным фактором соответствующих способностей, то есть в модель включено 4 фактора. При этом рассмотрены две модели — допускающая корреляции между «формальными» и «неформальными» факторами для отдельных субтестов и без таких корреляций (рис. 2Д и 2Е соответственно).

Оценки качества моделей приведены в табл. 2.

Таблица 2

Оценки структурных моделей

Тип модели	(df) χ^2	RMSEA	CFI	Значимость различий оценок вложенных моделей от трехфакторной
Двухфакторная (модель А)	(664) 7222,5	0,034	0,872	<0,0001
Трехфакторная с корреляциями (модель Б)	(656) 3933,9	0,024	0,932	0,004
Трехфакторная без корреляций (модель В)	(658) 3925,8	0,024	0,936	
Трехфакторная с неформальными в AN (модель Г)	(658) 5596,4	0,030	0,904	
Четырехфакторная с корреляциями (модель Д)	(655) 2830,1	0,020	0,958	
Четырехфакторная без корреляций (модель Е)	(657) 2989,1	0,020	0,955	

Как видно из табл. 2, двухфакторная модель характеризуется самыми низкими оценками соответствия, при этом значение коэффициента CFI оказалось ниже самого либерального из используемых критериев равного 0,9 (Bentler, 1992). При добавлении третьего фактора, связанного с заданиями без правильного варианта, качество модели улучшается и становится уже вполне приемлемым. При этом наличие или отсутствие корреляций третьего фактора с двумя специальными практически не меняет оценки модели. Модель, в которой третий фактор связан с имеющими ответы пунктами из субтеста «Аналогии», имеет промежуточные оценки — ниже, чем трехфакторная модель, но лучше, чем двухфакторная. Добавим также, что при проверке дополнительной модели, в которой с третьим фактором связывались три «формальных» задания из Силлогизмов

и три случайно выбранных задания из Аналогий были получены результаты, минимально отличающиеся от двухфакторной модели ($\chi^2(658) = 7048,7$, RMSEA = 0,034, CFI = 0,875).

Table 2

Fit indexes of the structural models

Type of model	(df) χ^2	RMSEA	CFI	Results of chi-squares' difference testing for nested models (in comparison with three factors without correlations)
Two factors (model A)	(664) 7222,5	0,034	0,872	<0,0001
Three factors with correlations between formal and informal factors (model B)	(656) 3933,9	0,024	0,932	0,004
Three factors without correlations between formal and informal factors (model C)	(658) 3925,8	0,024	0,936	
Three factors with informal random items from subtest AN (model D)	(658) 5596,4	0,030	0,904	
Four factors with correlations between formal and informal factors (model E)	(655) 2830,1	0,020	0,958	
Four factors without correlations between formal and informal factors (model F)	(657) 2989,1	0,020	0,955	

Что касается четырехфакторных моделей, то оценки их качества оказываются еще более высокими, чем для трехфакторной модели, и, таким образом, оказывается, что введение разделения на специфические и формальные факторы в каждом из субтестов повышает соответствие структурной модели имеющимся данным. При этом наличие корреляций между формальными и неформальными факторами мало меняют модель, что указывает на их относительную независимость. Также можно отметить, что корреляция между двумя формальными факторами относительно небольшая (0,330 для последней модели), и при исключении возможности корреляций между двумя формальными факторами качество оценок модели снизилось незначительно ($\chi^2(658) = 3429,1$, RMSEA = 0,022, CFI = 0,946). Это свидетельствует в пользу относительной независимости

двух выделенных формальных факторов друг от друга. Таким образом, наша гипотеза о наличии дополнительного фактора способностей, связанных с решением формальных задач находит частичное подтверждение с помощью конфирматорного факторного анализа, однако, еще лучше данным соответствует модель, в которой два типа таких способностей при решении силлогизмов и аналогий оказываются разделены.

Обсуждение результатов

Полученные результаты показывают, что использование традиционных схем оценки структуры интеллектуальных тестов может привести к некоторым искажениям результатов.

При анализе надежности шкал, измеряющих специальные способности в исходном варианте субтестов, без специальной фильтрации испытуемых и заданий, которая учитывала бы влияние временных ограничений выполнения, коэффициент Кронбаха оказывается достаточно высоким, но при ограничении числа заданий и отборе испытуемых, которые заведомо успели обдуманно на них ответить, показатель заметно снижается. Снижение коэффициента при сокращении выборки не может объясняться чисто технически, поскольку альфа достаточно независима от размера выборки (Ercan et al., 2007). Скорее можно предположить, что в данном случае снижение надежности может быть связано с исключением относительно «слабой» части выборки, которая вносит важный вклад в дисперсию данных и, таким образом, повышает надежность шкалы, что хорошо видно в табл. 1: в группе «быстрых» респондентов стандартное отклонение заметно снижается по сравнению с полной выборкой.

Понижение альфы при исключении части пунктов, напротив, имеет техническое объяснение — из литературы известно, что альфа Кронбаха снижается при уменьшении числа пунктов, входящих в шкалу. Но это снижение может указывать также на то, что временные ограничения, используемые в процедуре тестирования, привносят влияние на результаты дополнительного фактора скорости. Это надо учитывать и на содержательном уровне, и при дальнейшей оценке структуры субтестов. Фактор скорости вносит свой вклад в корреляцию между ответами на отдельные задания и, судя по полученным нами данным, повышает оценки надежности шкалы, однако насколько это соответствует цели проведения методики и валидности интерпретации ее оценок — вопрос, который должен

задавать себе исследователь. Подробное рассмотрение этой проблемы выходит за пределы нашей работы, но можно указать, что в исследованиях интеллекта обсуждается роль временных ограничений. В частности, в работе (Estrada, Román, Abad, Colom, 2017) обсуждается разделение двух компонентов измерения интеллекта — «сила» (power) и «скорость» (speed). Сравнение выполнения тестов с разной жесткостью временных ограничений в этой работе показало, что надежность (альфа Кронбаха) в менее жестко ограниченном по времени тесте ниже, чем в более жестко ограниченных тестах. Структурная модель, включающая фактор скорости в этом исследовании оказалась заметно лучше, чем однофакторная модель, не имеющая такого фактора. Таким образом, важно иметь в виду, что такая деталь процедуры, как временное ограничение, может сама по себе вносить свой вклад в структуру получаемых данных и это надо учитывать при анализе получаемых результатов. В целом, мы хотим отметить, что если фактор скорости концептуально включен в состав измеряемых конструкторов, то это надо учитывать и при построении структуры методики. Если же действовать в предположении, что скорость решения задач не относится к собственно измеряемым интеллектуальным способностям, то это надо тем или иным способом учитывать при разработке и оценке психометрических свойств методик. В рамках данной работы мы просто указываем на существенность этого фактора, подробное обсуждение возможных способов его учета выходит за ее пределы, но проблема, судя по результатам нашего анализа, может иметь место. В рамках данной работы при анализе структуры субтестов мы предприняли попытку снизить влияние фактора скорости, искусственно ограничив размер выборки и число анализируемых пунктов.

Важный результат, который получен в нашей работе — это существенность учета связей между пунктами разного уровня сложности (названные нами «формальными» и «неформальными»), относящихся к одному специфическому субтесту. На уровне корреляций между отдельными пунктами получены явно большие корреляции между более сложными заданиями, которые, по нашему предположению, задействуют более общие способности к анализу логических структур, а не только специфические способности, на оценку которых направлен тот или иной субтест. Это вполне предсказуемо отражается в оценках факторных моделей, которые учитывают или не учитывают возможность присутствия таких дополнительных факторов. При

использовании более «традиционной» модели, включающей только факторы специальных способностей, оценки ее качества оказались не слишком высоки. Добавление дополнительного фактора «формальных» способностей заметно улучшило качество модели, что указывает на важность учета выделенных нами изначально заданий, не имеющих правильных ответов. Специальная проверка, в которой мы включили в дополнительный фактор не «формальные», а случайные «неформальные» пункты из одного из субтестов показала, что при таком изменении индексы соответствия модели неотличимы от получаемых на «простой» модели только с факторами специальных способностей. Это дает нам основания быть уверенными, что улучшения модели не вызваны просто добавлением новых связей между переменными.

При этом, однако, дополнительный анализ показал, что выделение общего «формального» фактора с точки зрения оценок факторной структуры меньше соответствует структуре данных, чем введение двух дополнительных «формальных» факторов, отдельно для каждого субтеста. Возможно, дело здесь в том, что формальные задания в сравниваемых субтестах существенно различаются по трудности.

Увеличение мощности средств обработки данных все чаще побуждает исследователей отходить от классической организации тестовых данных (факторная структура на уровне субтестов) и работать непосредственно с результатами респондентов по отдельным тестовым заданиям (Molenaar, Kő, Rózsa, Mészáros, 2017; Shwabe, 2016). В нашей работе мы показали один возможный эффект, который можно обнаружить при переходе к схемам обработки на уровне заданий. Мы обнаружили, что респонденты, переходящие на более высокий уровень владения формальными аспектами логических отношений, относительно синхронно (поскольку четырехфакторная модель оказалась все же лучше трехфакторной, мы высказываем этот тезис с осторожностью) получают возможность решать соответствующие этому уровню задания в двух субтестах, которые вполне основательно рассматриваются как отражающие различные интеллектуальные способности. Таким образом, не только у детей (как это описано Ж. Пиаже, например, относительно освоения принципов сохранения в разных областях знаний), но и у взрослых наблюдаются переходы на качественно новые уровни мышления, хотя и в более узких областях. В работах, посвященных оценкам

психометрических свойств шкал интеллекта, в качестве альтернативы традиционной оценке надежности обсуждаются различные подходы. Так, указывается, что использование методов item-response theory (IRT, Bortolotti, Tezza, de Andrade, Bornia, de Sousa Júnior, 2013) может позволять анализировать и сопоставлять свойства отдельных пунктов на различных неоднородных выборках (см., например, Lazarević, Knežević, Mitić, Đurić-Jočić, 2018). Другой обсуждаемой альтернативой является использование так называемой композитной надежности для однородных измерительных моделей (composite reliability for congeneric measurement model, Raykov, 1997). В рамках этого подхода может оцениваться надежность пунктов, которые связаны с различными латентными переменными, а также с композитной латентной переменной, связанной с более общим латентным фактором. В приложении к тесту интеллекта это может быть фактор g , оценка которого связана как с отдельными специальными способностями, так и с более общими «формальными» (в терминах нашей работы) способностями, а также и с фактором скорости. Также на учет многомерности психометрических инструментов при оценке внутренней надежности направлен иерархический коэффициент ω Макдональда (McDonald's omega, Dunn et al., 2014), который базируется на предположении о возможном наличии нескольких факторов, с которыми связан каждый пункт методики, что позволяет оценить не только общую надежность (в этом качестве он мало отличается от альфы Кронбаха), но и разложить ее на составляющие, которые относятся к разным факторам (например, специальным и общим факторам интеллекта). На данном этапе наших исследований мы изучаем эти альтернативы, однако их адекватность в ситуации, подобной описанной в данной статье, требует отдельного исследования.

Вывод

Проведенный нами анализ двух шкал интеллекта позволяет указать на слабые места принятых в настоящее время, в качестве стандартных, процедур проверки психометрических свойств тестов на интеллект (и более широкого круга психометрических инструментов). Без детального рассмотрения самой процедуры тестирования и структуры заданий или вопросов исследователь может не заметить важных свойств анализируемых шкал и сделать неполные или неадекватные выводы об их психометрических качествах.

СПИСОК ЛИТЕРАТУРЫ

Корнилова Т.В., Корнилов С.А., Зиренко М.С., Чумакова М.А. Психометрические свойства модифицированной батареи Интернационального Ресурса Когнитивных Способностей (ICAR) // Национальный психологический журнал. 2019. № 3 (35). С. 32–45.

Методические рекомендации по организации и проведению профессионального психологического отбора в военно-учебных заведениях Министерства обороны Российской Федерации. М., 2002.

Митина О.В. Разработка и адаптация психологических опросников. М.: Смысл, 2013.

Гуревич К.М., Акимова М.К., Борисова Е.М., Козлова В.Т., Логинова Г.П. Школьный тест умственного развития (ШТУР). Методические рекомендации по работе с тестом (для школьных психологов). М., 1987.

Сугonyaев К.В., Радченко Ю.И. «Закон уменьшения отдачи» Спирмена: исследование на масштабных российских выборках // Вестник Южно-Уральского государственного университета. Серия: Психология. 2018. Т. 11. №. 1. С. 5–21.

Елисеев О.П. Тест структуры интеллекта (TSI) Р. Амтхауэра / Практикум по психологии личности. СПб.: Питер, 2001. С. 342–370.

Ушаков Д.В. Системность в психологии интеллекта: теория, подход, методология // Идея системности в современной психологии / под ред. ВА Барабанщикова. М.: Ин-т психологии РАН. 2005. С. 236–261.

Фер М.Р., Бакарак В.Р. Психометрика: Введение. Челябинск: Издательский центр ЮУрГУ, 2010.

Barrett, P., Eysenck, H.J., & Lucking, S. (1986). Reaction time and intelligence: A replicated study. *Intelligence*, 10 (1), 9–40.

Bentler, P.M. (1992). On the fit of models to covariances and methodology to the Bulletin. *Psychological bulletin*, 112 (3), 400–404.

Bortolotti S.L.V., Tezza R., de Andrade D.F., Bornaia A.C., de Sousa Júnior A.F. (2013). Relevance and advantages of using the item response theory. *Quality & Quantity*, 47 (4), 2341–2360.

Divgi, D.R. (1979). Calculation of the tetrachoric correlation coefficient. *Psychometrika*, 44 (2), 169–172.

Dunn, T.J., Baguley, T., & Brunsdon, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British journal of psychology*, 105 (3), 399–412.

Ercan, I., Yazici, B., Sigirli, D., Ediz, B., & Kan, I. (2007). Examining Cronbach alpha, theta, omega reliability coefficients according to sample size. *Journal of modern applied statistical methods*, 6 (1), 291–303.

Estrada, E., Román, F.J., Abad, F.J., & Colom, R. (2017). Separating power and speed components of standardized intelligence measures. *Intelligence*, 61, 159–168.

Jensen, A.R., & Munro, E. (1979). Reaction time, movement time, and intelligence. *Intelligence*, 3 (2), 121–126.

Johnson-Laird P.N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Harvard University Press.

Lazarević, L.B., Knežević, G., Mitić, M., & Đurić-Jočić, D. (2018). Psychometric properties of the Serbian version of the Wechsler adult intelligence scale: (WAIS-IV). *Psihologija*, 51 (3), 333–349.

McDonald R.R. (1999). *Test Theory: A Unified Treatment*. Mahwah, NJ: Lawrence Erlbaum Associates.

Molenaar D., Kő N., Rózsa S., Mészáros A. (2017). Differentiation of cognitive abilities in the WAIS-IV at the item level. *Intelligence*, 65, 48–59.

Muthén L. K., Muthén B. (2017). *Mplus. The comprehensive modelling program for applied researchers: user's guide*. Los Angeles, CA: Muthén & Muthén.

Netemeyer R.G., Bearden W.O., Sharma S. (2003). *Scaling procedures: Issues and applications*. Sage Publications.

Raykov, T. (1997). Estimation of composite reliability for congeneric measures. *Applied Psychological Measurement*, 21 (2), 173–184.

Sheppard, L.D., & Vernon, P.A. (2008). Intelligence and speed of information-processing: A review of 50 years of research. *Personality and individual differences*, 44 (3), 535–551.

Shwabe I. *Nature, Nurture and Item Response Theory — A Psychometric Approach to Behaviour Genetics*. Thesis for: PhD, 2016.

REFERENCES

Kornilova T.V., Kornilov A.S., Zirenko M.S., Chumakova M.A. (2019). Psychometric properties of the modified International Cognitive Ability Resource (ICAR) test battery. *Natsional'nyi psikhologicheskii zhurnal (National psychological journal)*, 3 (35), 32–45. (in Russ.).

Guidelines for organizing and conducting professional psychological selection in military educational institutions of the Ministry of Defense of the Russian Federation (2002). Moscow. (in Russ.).

Mitina O.V. (2013). *Development and adaptation of psychological questionnaires*. Moscow: Smysl. (in Russ.).

Gurevich K.M., Akimova M.K., Borisova E.M., Kozlova V.T., Loginova G.P. (1987). School test of mental development. Methodological recommendations for working with the test (for school psychologists). Moscow. (in Russ.).

Sugonyaev K.V., Radchenko Yu.I. (2018). Spearman's law of diminishing returns: investigation on large-scales russian samples. *Vestnik Yuzhno-Ural'skogo gosudarstvennogo universiteta. Seriya: Psikhologiya (Bulletin of the South Ural State University. Series "Psychology")*, 11 (1), 5–21. (in Russ.).

Eliseev O.P. (2003). R. Amthauer's intelligence structure test. In *Workshop on Personality Psychology* (pp. 342–370). Saint-Petersburg: Piter. (in Russ.).

Ushakov D.V. Systematic approach in the psychology of intelligence: theory, approach, methodology. Barabanschikov V.A. (ed.) *The idea of the systems in modern psychology* (pp. 236–261). Moscow: Institut psikhologii RAS. (in Russ.).

Fer M.R., Bakarak V.R. (2010). *Psychometrics: Introduction*. Chelyabinsk: Челябинск: Publishing Centre of YurGU. (in Russ.).

Barrett, P., Eysenck, H.J., & Lucking, S. (1986). Reaction time and intelligence: A replicated study. *Intelligence*, 10 (1), 9–40.

Bentler, P.M. (1992). On the fit of models to covariances and methodology to the Bulletin. *Psychological bulletin*, 112 (3), 400–404.

Bortolotti S.L.V., Tezza R., de Andrade D.F., Bornaia A.C., de Sousa Júnior A.F. (2013). Relevance and advantages of using the item response theory. *Quality & Quantity*, 47 (4), 2341–2360.

Divgi, D.R. (1979). Calculation of the tetrachoric correlation coefficient. *Psychometrika*, 44 (2), 169–172.

Dunn, T.J., Baguley, T., & Brunsdon, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British journal of psychology*, 105 (3), 399–412.

Ercan, I., Yazici, B., Sigirli, D., Ediz, B., & Kan, I. (2007). Examining Cronbach alpha, theta, omega reliability coefficients according to sample size. *Journal of modern applied statistical methods*, 6 (1), 291–303.

Estrada, E., Román, F.J., Abad, F.J., & Colom, R. (2017). Separating power and speed components of standardized intelligence measures. *Intelligence*, 61, 159–168.

Jensen, A.R., & Munro, E. (1979). Reaction time, movement time, and intelligence. *Intelligence*, 3 (2), 121–126.

Johnson-Laird P.N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Harvard University Press.

Lazarević, L.B., Knežević, G., Mitić, M., & Đurić-Jočić, D. (2018). Psychometric properties of the Serbian version of the Wechsler adult intelligence scale: (WAIS-IV). *Psihologija*, 51 (3), 333–349.

McDonald R.R. (1999). *Test Theory: A Unified Treatment*. Mahwah, NJ: Lawrence Erlbaum Associates.

Molenaar D., Kó N., Rózsa S., Mészáros A. (2017). Differentiation of cognitive abilities in the WAIS-IV at the item level. *Intelligence*, 65, 48–59.

Muthén L.K., Muthén B. (2017). *Mplus. The comprehensive modelling program for applied researchers: user's guide*. Los Angeles, CA: Muthén & Muthén.

Netemeyer R.G., Bearden W.O., Sharma S. (2003). *Scaling procedures: Issues and applications*. Sage Publications.

Raykov, T. (1997). Estimation of composite reliability for congeneric measures. *Applied Psychological Measurement*, 21 (2), 173–184.

Sheppard, L.D., & Vernon, P.A. (2008). Intelligence and speed of information-processing: A review of 50 years of research. *Personality and individual differences*, 44 (3), 535–551.

Shwabe I. *Nature, Nurture and Item Response Theory — A Psychometric Approach to Behaviour Genetics*. Thesis for: PhD, 2016.

ИНФОРМАЦИЯ ОБ АВТОРАХ

Кричевец Анатолий Николаевич — кандидат физико-математических наук, доктор философских наук, профессор кафедры методологии психологии факультета психологии МГУ имени М.В. Ломоносова, Москва, Россия. ORCID: <https://orcid.org/0000-0002-4064-3858>. E-mail: ankrich@mail.ru

Корнеев Алексей Андреевич — кандидат психологических наук, старший научный сотрудник лаборатории нейропсихологии факультета психологии МГУ имени М.В. Ломоносова, Москва, Россия. ORCID: <https://orcid.org/0000-0001-6389-8215>. E-mail: korneeff@gmail.com

Сугоняев Константин Владимирович — кандидат технических наук, ассоциированный сотрудник лаборатории психологии и психофизиологии творчества Института психологии РАН, Москва, Россия. ORCID: <https://orcid.org/0000-0002-6207-7228>. E-mail: skv-354@yandex.ru

ABOUT THE AUTHORS

Anatoly N. Krichevets — ScD (Philosophy), Professor, Department of Psychological Methodology, Faculty of Psychology, Lomonosov Moscow State University, Moscow, Russia. ORCID: <https://orcid.org/0000-0002-4064-3858>. E-mail: ankrich@mail.ru

Aleksei A. Korneev — PhD in psychology, senior research associate, Laboratory of Neuropsychology, Faculty of Psychology, Lomonosov Moscow State University, Moscow, Russia. ORCID: <https://orcid.org/0000-0001-6389-8215>. E-mail: korneeff@gmail.com

Konstantin V. Sugonyaev — PhD in technical sciences, research associate, Laboratory of Psychology and Psychophysiology of creative activity, Institute of Psychology of Russian Academy of Sciences, Moscow, Russia. ORCID: <https://orcid.org/0000-0002-6207-7228>. E-mail: skv-354@yandex.ru