

Проверка измерительной инвариантности для тестов когнитивных и языковых способностей

Логвиненко Т.И.

*Университетская клиника, Мюнхенский университет имени Людвига-Максимилиана,
г. Мюнхен, Германия; Институт образования, Национальный исследовательский университет
«Высшая школа экономики» (ФГАОУ ВО «НИУ ВШЭ»), Москва, Российская Федерация
ORCID: <https://orcid.org/0000-0001-7430-1963>, e-mail: logvinenkota.spb@gmail.com*

В статье рассматривается феномен измерительной инвариантности психодиагностических инструментов. Проверка инвариантности помогает оценить, насколько измеряемый конструкт имеет сопоставимое значение для разных групп или точек измерения. Такая проверка требуется, чтобы убедиться, что сравнение этих групп или измерений между собой является адекватным. В статье приведен обзор современных и классических публикаций зарубежных исследователей, демонстрирующих различные содержательные и методологические аспекты проверки инвариантности. Особое внимание уделено имеющимся данным о проверке инвариантности для тестов когнитивных и языковых способностей, а также о значении инвариантности для оценки индивидуальных и межгрупповых различий. В завершение обсуждаются актуальные подходы к оценке инвариантности в рамках моделирования структурными уравнениями и современной теории тестирования. Данная работа может стать первым шагом для знакомства с понятием измерительной инвариантности и будет полезна для всех читателей, заинтересованных в улучшении качества измерительных инструментов для получения более надежных выводов о психологических феноменах.

Ключевые слова: измерительная инвариантность, психометрика, психодиагностика, межгрупповые сравнения, дифференциальное функционирование теста.

Финансирование. Исследование выполнено за счет гранта Российского научного фонда (проект № 23-28-01399).

Благодарности. Автор благодарит за помощь и супервизию научного руководителя Е.Л. Григоренко.

Для цитаты: Логвиненко Т.И. Проверка измерительной инвариантности для тестов когнитивных и языковых способностей [Электронный ресурс] // Современная зарубежная психология. 2024. Том 13. № 2. С. 62—70. DOI: <https://doi.org/10.17759/jmfp.2024130206>

Testing the Tests: Investigating Measurement Invariance in Assessments of Cognitive and Language Abilities

Tatiana I. Logvinenko

*University Hospital, Ludwig Maximilian University of Munich (LMU), Munich, Germany;
National Research University Higher School of Economics, Moscow, Russia
ORCID: <https://orcid.org/0000-0001-7430-1963>, e-mail: logvinenkota.spb@gmail.com*

This paper discusses the measurement invariance of assessment instruments. The purpose of measurement invariance testing is to assess the extent to which a measured construct has comparable meaning across different groups or measurement points, to ensure that comparisons of these groups or measures to each other are appropriate. The paper provides an overview of recent and classic papers discussing conceptual and methodological aspects of invariance testing. Specific focus is placed on the available data on invariance testing for the assessments of cognitive and language abilities, and on the significance of invariance for the accurate assessment of individual and between-group differences. Finally, current approaches to invariance testing within structural equation modelling and item response theory are discussed. This paper may serve as an introduction to the concept of measurement invariance. It will be useful to readers interested in improving the quality of measurement instruments and making more reliable inferences about psychological phenomena.

Keywords: measurement invariance, psychometrics, psychological assessment, between-group differences, differential test functioning.

Funding. This work is supported by the Russian Science Foundation (project № 23-28-01399).

Acknowledgements. The author is grateful for supervision to her doctoral advisor, Dr. E.L. Grigorenko.

For citation: Logvinenko T.I. Testing the Tests: Investigating Measurement Invariance in Assessments of Cognitive and Language Abilities. *Sovremennaya zarubezhnaya psikhologiya* = Journal of Modern Foreign Psychology, 2024. Vol. 13, no. 2, pp. 62—70. DOI: <https://doi.org/10.17759/jmfp.2024130206> (In Russ.).

Понятие измерительной инвариантности

Инвариантность психодиагностической методики — эквивалентность измерения для групп или точек измерения. Измерительная инвариантность демонстрирует, что измеряемый психологический конструкт имеет одинаковое значение для разных исследуемых групп или при разных измерениях (например, повторных). Инвариантность измерения принимает различные формы и является ключевым фактором в психологических исследованиях, поскольку она показывает, что измерения сопоставимы, что позволяет в дальнейшем сравнивать эти измерения между собой [21; 26]. Иными словами, при сравнении групп, при сравнении средних в разных точках измерения и в анализе отношений между конструктами по группам (например, взаимодействие с фактором группы) необходимо убедиться в инвариантности измерительного метода.

Для примера, перед проверкой различий в средних по конструкту (например, аспект когнитивных или поведения) исследователи должны проверить, что измерение работает одинаковым образом для мальчиков и девочек, для разных этнических и культурных групп, для людей разных возрастов [23]. Поскольку один и тот же конструкт может иметь разное значение, меняться со временем; исследователи также должны проверять инвариантность в разных точках измерения (например, инвариантность в лонгитюдных исследованиях, до и после интервенции) [26; 36]. Таким образом, оценка измерительной инвариантности нужна, чтобы убедиться, что мы не сопоставляем «теплое с мягким».

Концептуально, инвариантность измерения выражает идею о том, что измерительные свойства X по отношению к целевому латентному признаку Wt одинаковы в разных популяциях или в разных точках измерений. Альтернативно, можно сказать, что в условиях инвариантности измерения знание о принадлежности индивида к популяции ничего не должно нам говорить о X сверх знания о Wt . Эта идея формально выражается следующим образом. Предположим, что у нас есть K популяций, $k = 1, 2, \dots, K$, члены которых прошли измерение по X , и далее предположим, что Wt представляет один или более целевых латентных признаков, которые X призван измерять. Пусть $P_k(X|Wt)$ — функция ответа для X в k -й популяции. Тогда мы будем говорить, что инвариантность измерения имеет место для X относительно Wt и $k = 1, 2, \dots, K$, тогда и только тогда, когда для $k = 1, 2, \dots, K$ и для всех Wt справедливо:

$$P_k(X|Wt) = P(X|Wt).$$

То есть, согласно принципу инвариантности измерения, мы не должны обнаружить никаких популяционных различий в распределении баллов по измеряемому переменным X в группе испытуемых, которые совпадают по целевой латентной переменной с Wt .

Этот принцип лежит в основе всех существующих статистических методов проверки инвариантности, хотя конкретная реализация этого принципа различается в разных подходах [21].

Понятие «инвариантность измерений» имеет тесную связь с понятиями эквивалентности измерений (measurement equivalence), предвзятости измерений (также — смещение оценки; measurement bias), дифференциального функционирования тестов (differential test functioning), а также с более широким понятием справедливости тестирования (test fairness). Разграничение этих понятий является увлекательным и несколько спорным вопросом, однако не входит в задачи данной статьи, и потому интересующийся читатель может обратиться к работам [25; 32]. В данном тексте термин «инвариантность» используется как зонтичный термин.

Роль проверки инвариантности в общей психометрической оценке инструментов

Идеи о важности проверки инвариантности измерений появились в литературе более 60 лет назад [например: 18]. Однако с рубежа веков оценка инвариантности для различных психодиагностических инструментов и разработка предназначенных для этого математико-статистических инструментов переживают настоящий подъем. Методисты чаще стали обращать внимание на значимость инвариантности измерений, особенно в рамках моделирования структурными уравнениями. Вместе с этими возникли и громкие прецеденты нарушения инвариантности: например, была обнаружена несправедливость в отношении этнических групп у тестов SAT (Scholastic Assessment Test), используемых для приема и отбора в высшие учебные заведения США. Подробное изложение истории в области исследований инвариантности и ее математических оснований можно найти у Р. Миллсапа и В. Мередит [22].

В настоящий момент интерес к вопросу инвариантности продолжает нарастать. Это совпадает, с одной стороны, с общим стремительным развитием психометрики, с другой стороны, подкрепляется набирающей мощь концепцией оценки, основанной на доказательствах — составной части общего движения за доказательную практику в психологии [5; 20]. При этом инвариантность часто рассматривают несколько отдельно от вопросов валидности. Вероятно, это связано с тем, что изначально инвариантность является термином статистическим, ассоциирующимся прежде всего с моделированием структурными уравнениями. Вместе с тем инвариантность можно рассматривать как показатель генерализуемости валидности, так как она

демонстрирует, насколько психометрические свойства инструмента могут быть переносимы на другие группы респондентов или условия тестирования. Кроме того, у проблемы инвариантности измерения есть еще одна сторона: когда обнаруживается, что измерительный инструмент функционирует по-разному в зависимости от принадлежности человека к той или иной группе, возникает вопрос о том, что на самом деле измеряет этот инструмент. То есть речь идет также и о конструктивной валидности.

Важность оценки инвариантности не только рассматривается в методологической литературе, но и закреплена в общепризнанных стандартах. Так, в действующей версии стандартов тестирования в области образования и психологии, разработанных Американской ассоциацией исследований в области образования, Американской психологической ассоциацией и Национальным советом по измерениям в образовании [30], измерительная инвариантность и отсутствие предвзятости оценок рассматривается как одна из ключевых составляющих справедливого тестирования. Под справедливым прежде всего понимается тестирование, в котором интерпретации показателей являются сопоставимыми для всех участников тестирования. Справедливость тестирования, в свою очередь, даже стала темой отдельной секции-книги [6] наравне с уже традиционными для стандартов секциями по валидности и надежности.

Вместе с тем по вопросу измерения инвариантности существует и много противоречий. Так, в активной фазе находится полемика об инвариантности как обязательном условии для валидных групповых сравнений. Например, К. Вельцель и соавторы [24] в целом считают проблему не-инвариантности измерений несколько преувеличенной и не столь существенной, по крайней мере, в кросс-культурных исследованиях; их тезисы вызвали яркую реакцию представителей противоположного мнения [34]. Этот вопрос также поднимается в недавней статье А. Робича и О. Людтке [27]. Авторы указывают на существующие противоречия и лакуны в доминирующем подходе к инвариантности и призывают не отождествлять инвариантность и эквивалентность измерений. Содержательную эквивалентность авторы связывают с валидностью, однако утверждают, что отсутствие статистической инвариантности, как правило, не угрожает валидности выводов о групповых различиях. Авторы не склоняют отказываться от проверок инвариантности вовсе, однако предлагают смотреть на эти проверки только как на инструмент изучения данных, а также описывают большое количество техник совладения с не-инвариантностью.

Можно предположить, что часть критики вызвана тем, что из статей сторонников проверки измерительной инвариантности иногда можно сделать побочный вывод о том, что в случае нарушения инвариантности сравнения запрещены [8]. Однако возможна и менее категоричная интерпретация: оценка инвариантности

измерений дает большое количество ценной информации о данных, и эту информацию не стоит игнорировать при проведении межгрупповых сравнений. Вместо этого она должна быть тщательно проанализирована и учтена, что в ряде случаев требует достаточно продвинутых методов анализа и понимания данных. Так или иначе, можно утверждать, что мнения о конкретных методах проверки инвариантности, а также о «допустимом» уровне инвариантности для выводов о различиях между группами значительно различаются [26].

Практики проверки инвариантности в тестах когнитивных и языковых способностей

Хотя для большого числа известных и зарекомендовавших себя тестов проверка инвариантности является регулярной практикой, далеко не все измерительные инструменты соответствуют принципам инвариантности [3]. Умозрительно наиболее распространены такие проверки в исследованиях кросс-культурных различий. Вместе с тем очевидно, что если рассматривать инвариантность как *условие* для групповых сравнений, то этот вопрос касается гораздо более широкого круга психологических и социальных дисциплин.

В отношении тестов способностей нельзя сказать, что на данный момент существует большое количество публикаций, позволяющих сделать вывод о распространенности проверок инвариантности. Й. Вихертс [35] в своей работе делает обзор исследований предвзятости батарей тестов интеллекта (в том числе популярных, например, 4-го издания теста Векслера для взрослых) и приходит к выводу, что в более чем в половине публикаций не подтверждается измерительная инвариантность методик в отношении этнических групп, пола, образования, или возраста. Также в недавней работе Э. Маассен [31] и коллег был проведен систематический обзор психологических статей из журналов *Psychological Science* и *PLOS ONE*, в которых использовалось сравнение средних между группами или временными срезами и которые имели открытые данные. Так, в 96 подходящих публикациях всего было 929 сравнений, из них лишь для 4% метрик были доступны результаты проверки инвариантности. При этом, при попытке провести независимую оценку инвариантности на опубликованных вместе со статьями данных, было обнаружено, что в 58% случаях не удовлетворяется даже эквивалентность структуры модели, т. е. самый базовый уровень инвариантности. Авторы отмечают, что при таком частом нарушении практик исследования инвариантности и их документации есть основания сомневаться в надежности полученных в исследованиях эффектов.

Среди тестов грамотности и языковых способностей, зарекомендовавшие себя инструменты регулярно проверяются на инвариантность как издательствами, так и

независимыми исследователями. Например, так происходит с пережившими уже несколько изданий тестами академических достижений Вудкока—Джонсона или тестами чтения DIBELS [например: 9]. Однако зарекомендовавшие себя тесты представлены прежде всего для английского языка. Для тестов чтения, письма, речевых способностей на других языках встречаются отдельные публикации, включающие проверку инвариантности [например: 15; 17], но такие публикации достаточно редки. Вместе с тем десятилетия попыток изменить статус-кво англоцентричности научных знаний о развитии языка, чтения и их нарушений привели к интенсивному увеличению числа кросс-лингвистических исследований, но, по всей видимости, не привели к существенному улучшению методологии этих исследований. Об этом говорит сравнительно недавняя обзорная публикация Т. Пападопулоса и коллег [19] — видных исследователей чтения в греческом, чешском, французском, финском и венгерском языках. Авторы обозначают актуальные методологические проблемы исследований грамотности на разных языках, отдельно выделяя вопросы об эквивалентности измеряемых конструктов и проверке инвариантности измерений. По всей видимости, если говорить о материалах для оценки чтения, понимания прочитанного, недостаточно тщательной адаптации, но необходимы и строгие проверки инвариантности относительно факторов, имеющих отношение к языку, культуре, когнитивному и экономическому статусу участников исследований. Кроме того, стоит вопрос не только о проверке инвариантности как таковой, но и об определении источников этой инвариантности на уровне всего теста и отдельных заданий. Так, в обзоре 110 публикаций, посвященных инвариантности в тестах языковых способностей и понимания прочитанного в период с 1985 по 2019 г., авторы пришли к выводу, что, несмотря на усложнение используемых статистических процедур, попыток определить содержательные и контекстуальные источники дифференциального функционирования заданий не стало больше по мере развития поля с течением времени [13].

Итак, несмотря на оживленные дискуссии вокруг темы инвариантности в сфере психометрики вообще, в исследованиях индивидуальных и межгрупповых различий в когнитивных и языковых способностях этот аспект не пользуется большой популярностью. В некоторой степени такое положение лишь подчеркивает разрыв между современными возможностями моделирования и анализа данных и их практическим применением в отношении инструментов, которые могут, например, иметь большое клиническое значение. Кроме того, при этом обсуждении уместно иметь в виду и более широкий контекст, а именно — кризис воспроизводимости в психологии [38]. Очевидно, что текущие реалии требуют более строгого внимания к различным аспектам качества психодиагностических инструментов и методологии исследований для надежных выводов о психологических феноменах [7].

Способы проверки измерительной инвариантности

Инвариантность измерения традиционно проверяется в рамках двух подходов: моделирования структурными уравнениями (Structural Equation Modelling — SEM) и современной теории тестирования (Item Response Theory — IRT). В данной секции будет дана общая информация об этих подходах.

Моделирование структурными уравнениями

Как было сказано ранее, само понятие измерительной инвариантности изначально укоренено в подходе моделирования структурными уравнениями. В рамках SEM, а точнее, конфирматорного факторного анализа (CFA) как измерительной модели в SEM, предполагается, что отношения между наблюдаемыми переменными и латентными конструктами исследуются на инвариантность между группами. Для этого традиционным является использование многоклассового конфирматорного факторного анализа (multiple group confirmatory factor analysis — MGCFA).

Одним из первых предложил общую стратегию проверки инвариантности К. Йореског [12] — она используется в настоящее время повсеместно. В этой стратегии вложенные модели организуются в иерархическом порядке с уменьшающимся числом параметров (или увеличивающимися степенями свободы), что влечет за собой добавление ограничений на параметры по одному. Все эти модели проверяются с точки зрения соответствия фита (качества подгонки) модели — в сторону улучшения или ухудшения в сопоставлении с другими. Вскоре были описаны четыре уровня инвариантности, которые соответствуют этапам ее проверки [37]: 1) конфигурационная или эквивалентность структуры модели; 2) метрическая (слабая факторная) или эквивалентность факторных нагрузок; 3) скалярная (сильная факторная) или эквивалентность констант (intercepts) или пороговых значений; 4) остаточная (строгая) или эквивалентность остатков. Также Р. Ванденберг и Ч. Лэнс [33] предложили восемь шагов для проверки инвариантности измерений: шаги со второго по пятый представляют уже указанные выше основные тесты инвариантности, а последние три этапа отражают структурную инвариантность самих полученных латентных факторов (например, эквивалентность факторных вариаций, ковариаций и средних). Первый же шаг предполагает оценку инвариантности ковариационных матриц, однако этот тест редко выполняется на практике и современные руководства теперь опускают этот первый шаг. В рамках моделирования структурными уравнениями измерительная инвариантность может дополнительно оцениваться с помощью моделей «множественных показателей — множественных причин» (multiple indicator multiple causes models — MIMIC), преимуществом которого в сравнении с многоклассовым конфирматорным факторным анализом являются меньшие требования к размеру выборки.

Для технической реализации анализа инвариантности в подходе моделирования структурными уравнениями чаще всего используется проприетарная программа MPlus или открытый пакет *lavaan* в R, реже — программы AMOS, LISREL, EQS. Так как многоклассовый конфирматорный факторный анализ и MIMIC-модели являются самым распространенным способом оценки инвариантности, для этих анализов доступны хорошие руководства [например: 21], в том числе — сопоставляющие традиционные методы и новые решения [например: 10; 14; 28].

Современная теория тестирования (IRT)

Современная теория тестирования или IRT является влиятельной парадигмой разработки, анализа и оценки психологических шкал и тестов. IRT-анализ определяет степень, в которой отдельные утверждения или вопросы шкалы оценивают индивидуальные различия по некоторому заданному конструкту. Параметры модели IRT предоставляют информацию о сложности задания отдельно от его дискриминативной способности (т. е. информативности для оценки индивидуальных различий между испытуемыми). Традиционно модели IRT для дихотомических баллов получают названия в зависимости от количества параметров, которые используются для моделирования характеристик вопросов. Однопараметрические модели (1PL) включают только сложность задания, двухпараметрические (2PL) включают сложность и дискриминативную способность, а трехпараметрические (3PL) дополнительно к сложности и дискриминативной способности включают моделирование случайно-угадывания [11].

В отличие от CFA, процедуры IRT были разработаны для оценки измерительной инвариантности (или, точнее, отсутствия измерительной инвариантности) для отдельных заданий, отсюда и терминология — дифференциальное функционирование заданий (DIF — differential item functioning). Для сравнения подходов SEM и IRT можно обратиться к работе Л. Тэй, А. Мида и М. Цао [29]. Однако, несмотря на общие моменты, нет однозначного соответствия между процедурами оценки инвариантности измерения и дифференциального функционирования заданий. Это объясняется тем, что IRT был разработан в контексте тестирования с акцентом на выявление конкретных тестовых заданий, которые могут «смещаться» в отношении определенной группы. Это не означает, что инвариантность всего теста не имеет значения: совокупный DIF по заданиям может привести к дифференциальному функционированию теста, которое также может быть оценено. Из-за прагматической направленности концепции для разработки тестов, в IRT также меньше внимания уделяется различиям в понимании/работе конструкта в разных группах тогда, когда обнаруживается неэквивалентность. Скорее, DIF часто используется для выявления потенциальных проблем в содержании и формулировке заданий.

Существует широкий спектр методов для исследования DIF. Прежде всего методы делятся по способу, которым определяется способность — через наблюдаемые переменные или через латентную способность. От этого зависит и способ «подбора» испытуемых с одинаковым уровнем латентной способности друг к другу. Так, анализ с помощью статистики Mantel-Haenszel, метод стандартизации и логистическая регрессия делают «подбор» испытуемых на основе общего балла, тогда как Lord's Wald -тест или Raju's тест определяют уровень испытуемого как латентную способность, уровень которой вычисляется вместе с другими параметрами IRT-модели. В настоящее время доступно большое количество обзоров по процедурам определения DIF [например: 11; 30], а сам анализ может быть реализован в R с помощью разнообразных пакетов [1; 4].

Непрерывно разрабатываются и новые статистические методы для проверки измерительной инвариантности [см. обзор: 2; 16]. Эти методы призваны преодолеть некоторые ограничения традиционных моделей и позволяют, например, накладывать меньшие ограничения на модели, облегчать вычислительный процесс, использовать большое число уровней переменной, относительно которой проверяется инвариантность, или использовать комбинации из континуальных и категориальных переменных. К ним относятся: многоклассовое байесовское моделирование структурными уравнениями (Bayesian approximate measurement invariance), метод выравнивания (alignment method), многоуровневый факторный анализ (multilevel factor analysis), смешанный многоклассовый факторный анализ (mixture multigroup factor analysis), эксплораторное моделирование структурными уравнениями (measurement invariance explorer), декомпозиция сдвига ответов и истинных изменений (response shift-true change decomposition approach), нелинейный модерационный факторный анализ (moderated nonlinear factor analysis).

Заключение

В заключение можно сказать, что проблема измерительной инвариантности является значимой и широко обсуждаемой в научной литературе. Интерес к вопросам инвариантности психодиагностических инструментов подогревается одновременно благодаря интенсивному развитию методов моделирования данных, кризису реплицируемости и общему движению в сторону к доказательным практикам и доказательной оценке в психологической науке.

В качестве дискуссионных вопросов можно обозначить, во-первых, содержательное понимание инвариантности и ее отношений с понятиями эквивалентности, предвзятости и справедливости; во-вторых, представления о значении показателей инвариантности для валидных выводов о групповых различиях; в-третьих —

большое число методологических вопросов о конкретных способах оценки инвариантности. К этим вопросам можно отнести типы, количество и порядок тестов для оценки инвариантности, допустимый уровень не-инвариантности психометрических инструментов, оценку инвариантности при специфических данных и задачах, например, при малой выборке, большом количестве сопоставляемых групп, моделей с малым и большим количеством вычисляемых параметров и т. д. Важно также отметить недостаточное внимание к этой проблеме в исследованиях индивидуальных и межгрупповых различий в когнитивных и языковых способностях, что свидетельствует о большом разрыве между современными методами анализа психометри-

ческих данных и их применением в отношении тестов, используемых в практике.

Вместе с тем продолжается разработка новых статистических методов для проверки инвариантности, которые призваны преодолеть ограничения традиционных моделей. Появляется все больше руководств, которые помогают исследователям внедрять проверку инвариантности в свой арсенал. Однозначно, остается много методологических вопросов, требующих дальнейших исследований и обсуждений, но активные дискуссии и разработка новых методов стимулируют прогресс в этой области и, вероятно, будут также в целом способствовать улучшению качества психодиагностических инструментов и методологии исследований.

Литература

1. An R toolbox for score-based measurement invariance tests in IRT models / L. Schneider, C. Strobl, A. Zeileis, R. Debelak // *Behavior Research Methods*. 2021. Vol. 54. P. 2101—2113. DOI:10.3758/s13428-021-01689-0
2. *Bauer D.J., Belzak W.C.M., Cole V.T.* Simplifying the Assessment of Measurement Invariance over Multiple Background Variables: Using Regularized Moderated Nonlinear Factor Analysis to Detect Differential Item Functioning // *Structural Equation Modeling: A Multidisciplinary Journal*. 2020. Vol. 27. № 1. P. 43—55. DOI:10.1080/10705511.2019.1642754
3. *Boer D., Hanke K., He J.* On Detecting Systematic Measurement Error in Cross-Cultural Research: A Review and Critical Reflection on Equivalence and Invariance Tests // *Journal of Cross-Cultural Psychology*. 2018. Vol. 49. № 5. P. 713—734. DOI:10.1177/0022022117749042
4. *Choi Y.-J., Asilkalkan A.R.* Packages for Item Response Theory Analysis: Descriptions and Features // *Measurement: Interdisciplinary Research and Perspectives*. 2019. Vol. 17. № 3. P. 168—175. DOI:10.1080/15366367.2019.1586404
5. Evidence-Based Assessment in Special Education Research: Advancing the Use of Evidence in Assessment Tools and Empirical Processes / E. Talbott, A. De Los Reyes, D.M. Kearns, J. Mancilla-Martinez, M. Wang // *Exceptional Children*. 2023. Vol. 89. № 4. P. 467—487. DOI:10.1177/00144029231171092
6. Fairness in educational and psychological testing: examining theoretical, research, practice, and policy implications of the 2014 standards / Eds. J.L. Jonson, K.F. Geisinger. Washington: American Educational Research Association, 2022. 466 p. DOI:10.2307/j.ctv2kzv0fw
7. *Flake J.K., Fried E.I.* Measurement Schmeasurement: Questionable Measurement Practices and How to Avoid Them // *Advances in Methods and Practices in Psychological Science*. 2020. Vol. 3. № 4. P. 456—465. DOI:10.1177/2515245920952393
8. *Funder D.C., Gardiner G.* Misgivings about measurement invariance // *European Journal of Personality*. 2024. 7 p. Ahead of Print. DOI:10.1177/08902070241228338
9. *Hajovsky D.B., Chesnut S.R.* Examination of differential effects of cognitive abilities on reading and mathematics achievement across race and ethnicity: Evidence with the WJ IV // *Journal of School Psychology*. 2022. Vol. 93. P. 1—27. DOI:10.1016/j.jsp.2022.05.001
10. *Han K., Colarelli S.M., Weed N.C.* Methodological and statistical advances in the consideration of cultural diversity in assessment: A critical review of group classification and measurement invariance testing // *Psychological Assessment*. 2019. Vol. 31(12). P. 1481—1496. DOI:10.1037/pas0000731
11. *Handbook of Item Response Theory: Three Volume Set* / Ed. W.J. Van Der Linden. Boca Raton: CRC Press, 2018. 1500 p. DOI:10.1201/9781315119144
12. *Jöreskog K.G.* Simultaneous factor analysis in several populations // *Psychometrika*. 1971. Vol. 36. P. 409—426. DOI:10.1007/BF02291366
13. *Li H., Hunter C.V., Bialo J.A.* A Revisit of Zumbo's Third Generation DIF: How Are We Doing in Language Testing? // *Language Assessment Quarterly*. 2022. Vol. 19. № 1. P. 27—53. DOI:10.1080/15434303.2021.1963253
14. *Luong R., Flake J.K.* Measurement invariance testing using confirmatory factor analysis and alignment optimization: A tutorial for transparent analysis planning and reporting // *Psychological Methods*. 2023. Vol. 28(4). P. 905—924. DOI:10.1037/met0000441
15. Measurement Invariance in Relation to First Language: An Evaluation of German Reading and Spelling Tests / L. Visser, F. Cartschau, A. Von Goldammer, J. Brandenburg, M. Timmerman, M. Hasselhorn, C. Mähler // *Applied Measurement in Education*. 2023. Vol. 36. № 2. P. 115—131. DOI:10.1080/08957347.2023.2201701
16. Measurement invariance in the social sciences: Historical development, methodological challenges, state of the art, and future perspectives / H. Leitgöb, D. Seddig, T. Asparouhov (et al.) // *Social Science Research*. 2023. Vol. 110. Article ID 102805. 30 p. DOI:10.1016/j.ssresearch.2022.102805

17. Measuring the development of general language skills in English as a foreign language—Longitudinal invariance of the C-test / B. Schnoor, J. Hartig, T. Klinger, A. Naumann, I. Usanova // *Language Testing*. 2023. Vol. 40. № 3. P. 796—819. DOI:10.1177/02655322231159829
18. *Meredith W.* Notes on factorial invariance // *Psychometrika*. 1964. Vol. 29. P. 177—185. DOI:10.1007/BF02289699
19. Methodological Issues in Literacy Research Across Languages: Evidence From Alphabetic Orthographies / T.C. Papadopoulos, V. Csépe, M. Aro, M. Caravolas, I.-A. Diakidoy, T. Olive // *Reading Research Quarterly*. 2021. Vol. 56. № S1. P. S351—S370. DOI:10.1002/rrq.407
20. *Miller L.A., Lovler R.L.* Foundations of psychological testing: A practical approach. Thousand Oaks: SAGE Publications Inc, 2020. 592 p.
21. *Millsap R.E., Meredith W.* Factorial invariance: Historical perspectives and new problems // *Factor analysis at 100: Historical developments and future directions* / Eds. R. Cudeck, R.C. MacCallum. New York: Routledge, 2007. P. 131—152. DOI:10.4324/9780203936764
22. *Millsap R.E.* Statistical approaches to measurement invariance. New York: Routledge, 2011. 368 p. DOI:10.4324/9780203821961
23. *Molenaar D.* A flexible moderated factor analysis approach to test for measurement invariance across a continuous variable // *Psychological Methods*. 2020. Vol. 26(6). P. 660—679. DOI:10.1037/met0000360
24. Non-invariance? An Overstated Problem With Misconceived Causes / C. Welzel, L. Brunkert, S. Kruse, R.F. Inglehart // *Sociological Methods & Research*. 2023. Vol. 52. № 3. P. 1368—1400. DOI:10.1177/0049124121995521
25. *Pedraza O., van de Vijver F.J.R.* Bias, equivalence, and fairness // *Clinical cultural neuroscience: An integrative approach to cross-cultural neuropsychology* / Ed. O. Pedraza. Oxford: Oxford University Press, 2020. P. 252—275.
26. *Putnick D.L., Bornstein M.H.* Measurement invariance conventions and reporting: The state of the art and future directions for psychological research // *Developmental Review*. 2016. Vol. 41. P. 71—90. DOI:10.1016/j.dr.2016.06.004
27. *Robitzsch A., Lüdtke O.* Why Full, Partial, or Approximate Measurement Invariance Are Not a Prerequisite for Meaningful and Valid Group Comparisons // *Structural Equation Modeling: A Multidisciplinary Journal*. 2023. Vol. 30. № 6. P. 859—870. DOI:10.1080/10705511.2023.2191292
28. *Svetina D., Rutkowski L., Rutkowski D.* Multiple-Group Invariance with Categorical Outcomes Using Updated Guidelines: An Illustration Using *Mplus* and the *lavaan/semTools* Packages // *Structural Equation Modeling: A Multidisciplinary Journal*. 2020. Vol. 27. № 1. P. 111—130. DOI:10.1080/10705511.2019.1602776
29. *Tay L., Meade A.W., Cao M.* An Overview and Practical Guide to IRT Measurement Equivalence Analysis // *Organizational Research Methods*. 2015. Vol. 18. № 1. P. 3—46. DOI:10.1177/1094428114553062
30. The Standards for educational and psychological testing [Электронный ресурс] // American Psychological Association. Washington, 2009. URL: <https://www.apa.org/science/programs/testing/standards> (дата обращения: 25.06.2024).
31. The dire disregard of measurement invariance testing in psychological science / E. Maassen, E.D. D'Urso, M.A.L.M. Van Assen, M.B. Nuijten, K. De Roover, J.M. Wicherts // *Psychological Methods*. 2023. 14 p. DOI:10.1037/met0000624
32. *Van de Vijver F.J., Leung K.* Methods and data analysis for cross-cultural research. Cambridge: Cambridge University Press, 2021. 250 p. DOI:10.1017/9781107415188
33. *Vandenberg R.J., Lance C.E.* A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research // *Organizational Research Methods*. 2000. Vol. 3. № 1. P. 4—70. DOI:10.1177/109442810031002
34. Why Measurement Invariance is Important in Comparative Research. A Response to Welzel et al. (2021) / B. Meuleman, T. Żółtak, A. Pokropek, E. Davidov, B. Muth n, D.L. Oberski, J. Billiet, P. Schmidt // *Sociological Methods & Research*. 2023. Vol. 52. № 3. P. 1401—1419. DOI:10.1177/00491241221091755
35. *Wicherts J.M.* The importance of measurement invariance in neurocognitive ability testing // *The Clinical Neuropsychologist*. 2016. Vol. 30. № 7. P. 1006—1016. DOI:10.1080/13854046.2016.1205136
36. *Widaman K.F., Ferrer E., Conger R.D.* Factorial Invariance Within Longitudinal Structural Equation Models: Measuring the Same Construct Across Time // *Child Development Perspectives*. 2010. Vol. 4. № 1. P. 10—18. DOI:10.1111/j.1750-8606.2009.00110.x
37. *Widaman K.F., Reise S.P.* Exploring the measurement invariance of psychological instruments: Applications in the substance use domain // *The science of prevention: Methodological advances from alcohol and substance abuse research* / Eds. K.J. Bryant, M. Windle, S.G. West. Washington: American Psychological Association, 1997. P. 281—324. DOI:10.1037/10222-009
38. *Wingen T., Berkessel J.B., English B.* No Replication, No Trust? How Low Replicability Influences Trust in Psychology // *Social Psychological and Personality Science*. 2020. Vol. 11. № 4. P. 454—463. DOI:10.1177/1948550619877412

References

1. Schneider L., Strobl C., Zeileis A., Debelak R. An R toolbox for score-based measurement invariance tests in IRT models. *Behavior Research Methods*, 2021. Vol. 54, pp. 2101—2113. DOI:10.3758/s13428-021-01689-0

2. Bauer D.J., Belzak W.C.M., Cole V.T. Simplifying the Assessment of Measurement Invariance over Multiple Background Variables: Using Regularized Moderated Nonlinear Factor Analysis to Detect Differential Item Functioning. *Structural Equation Modeling: A Multidisciplinary Journal*, 2020. Vol. 27, no. 1, pp. 43—55. DOI:10.1080/10705511.2019.1642754
3. Boer D., Hanke K., He J. On Detecting Systematic Measurement Error in Cross-Cultural Research: A Review and Critical Reflection on Equivalence and Invariance Tests. *Journal of Cross-Cultural Psychology*, 2018. Vol. 49, no. 5, pp. 713—734. DOI:10.1177/0022022117749042
4. Choi Y.-J., Asilkalkan A.R. Packages for Item Response Theory Analysis: Descriptions and Features. *Measurement: Interdisciplinary Research and Perspectives*, 2019. Vol. 17, no. 3, pp. 168—175. DOI:10.1080/15366367.2019.1586404
5. Talbott E., De Los Reyes A., Kearns D.M., Mancilla-Martinez J., Wang M. Evidence-Based Assessment in Special Education Research: Advancing the Use of Evidence in Assessment Tools and Empirical Processes. *Exceptional Children*, 2023. Vol. 89, no. 4, pp. 467—487. DOI:10.1177/00144029231171092
6. Jonson J.L., Geisinger K.F. (eds.), *Fairness in educational and psychological testing: examining theoretical, research, practice, and policy implications of the 2014 standards*. Washington: American Educational Research Association, 2022. 466 p. DOI:10.2307/j.ctv2kzv0fw
7. Flake J.K., Fried E.I. Measurement Schmeasurement: Questionable Measurement Practices and How to Avoid Them. *Advances in Methods and Practices in Psychological Science*, 2020. Vol. 3, no. 4, pp. 456—465. DOI:10.1177/2515245920952393
8. Funder D.C., Gardiner G. Misgivings about measurement invariance. *European Journal of Personality*, 2024. 7 p. Ahead of Print. DOI:10.1177/08902070241228338
9. Hajovsky D.B., Chesnut S.R. Examination of differential effects of cognitive abilities on reading and mathematics achievement across race and ethnicity: Evidence with the WJ IV. *Journal of School Psychology*, 2022. Vol. 93, pp. 1—27. DOI:10.1016/j.jsp.2022.05.001
10. Han K., Colarelli S.M., Weed N.C. Methodological and statistical advances in the consideration of cultural diversity in assessment: A critical review of group classification and measurement invariance testing. *Psychological Assessment*, 2019. Vol. 31(12), pp. 1481—1496. DOI:10.1037/pas0000731
11. Van Der Linden W.J. (ed.), *Handbook of Item Response Theory: Three Volume Set*. Boca Raton: CRC Press, 2018. 1500 p. DOI:10.1201/9781315119144
12. Jöreskog K.G. Simultaneous factor analysis in several populations. *Psychometrika*, 1971. Vol. 36, pp. 409—426. DOI:10.1007/BF02291366
13. Li H., Hunter C.V., Bialo J.A. A Revisit of Zumbo's Third Generation DIF: How Are We Doing in Language Testing? *Language Assessment Quarterly*, 2022. Vol. 19, no. 1, pp. 27—53. DOI:10.1080/15434303.2021.1963253
14. Luong R., Flake J.K. Measurement invariance testing using confirmatory factor analysis and alignment optimization: A tutorial for transparent analysis planning and reporting. *Psychological Methods*, 2023. Vol. 28(4), pp. 905—924. DOI:10.1037/met0000441
15. Visser L., Carstachau F., Von Goldammer A., Brandenburg J., Timmerman M., Hasselhorn M., Mähler C. Measurement Invariance in Relation to First Language: An Evaluation of German Reading and Spelling Tests. *Applied Measurement in Education*, 2023. Vol. 36, no. 2, pp. 115—131. DOI:10.1080/08957347.2023.2201701
16. Leitgöb H., Seddig D., Asparouhov T. [et al.]. Measurement invariance in the social sciences: Historical development, methodological challenges, state of the art, and future perspectives. *Social Science Research*, 2023. Vol. 110, article ID 102805. 30 p. DOI:10.1016/j.ssresearch.2022.102805
17. Schnoor B., Hartig J., Klinger T., Naumann A., Usanova I. Measuring the development of general language skills in English as a foreign language—Longitudinal invariance of the C-test. *Language Testing*, 2023. Vol. 40, no. 3, pp. 796—819. DOI:10.1177/02655322231159829
18. Meredith W. Notes on factorial invariance. *Psychometrika*, 1964. Vol. 29, pp. 177—185. DOI:10.1007/BF02289699
19. Papadopoulos T.C., Csipe V., Aro M., Caravolas M., Diakidoy I.-A., Olive T. Methodological Issues in Literacy Research Across Languages: Evidence From Alphabetic Orthographies. *Reading Research Quarterly*, 2021. Vol. 56, no. S1, pp. S351—S370. DOI:10.1002/rq.407
20. Miller L.A., Lovler R.L. *Foundations of psychological testing: A practical approach*. Thousand Oaks: SAGE Publications Inc, 2020. 592 p.
21. Millsap R.E., Meredith W. Factorial invariance: Historical perspectives and new problems. In Cudeck R., MacCallum R.C. (eds.), *Factor analysis at 100: Historical developments and future directions*. New York: Routledge, 2007, pp. 131—152. DOI:10.4324/9780203936764
22. Millsap R.E. *Statistical approaches to measurement invariance*. New York: Routledge, 2011. 368 p. DOI:10.4324/9780203821961
23. Molenaar D. A flexible moderated factor analysis approach to test for measurement invariance across a continuous variable. *Psychological Methods*, 2020. Vol. 26(6), pp. 660—679. DOI:10.1037/met0000360
24. Welzel C., Brunkert L., Kruse S., Inglehart R.F. Non-invariance? An Overstated Problem With Misconceived Causes. *Sociological Methods & Research*, 2023. Vol. 52, no. 3, pp. 1368—1400. DOI:10.1177/0049124121995521

25. Pedraza O., van de Vijver F.J.R. Bias, equivalence, and fairness. In Pedraza O. (ed.), *Clinical cultural neuroscience: An integrative approach to cross-cultural neuropsychology*. Oxford: Oxford University Press, 2020, pp. 252—275.
26. Putnick D.L., Bornstein M.H. Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, 2016. Vol. 41, pp. 71—90. DOI:10.1016/j.dr.2016.06.004
27. Robitzsch A., Lüdtke O. Why Full, Partial, or Approximate Measurement Invariance Are Not a Prerequisite for Meaningful and Valid Group Comparisons. *Structural Equation Modeling: A Multidisciplinary Journal*, 2023. Vol. 30, no. 6, pp. 859—870. DOI:10.1080/10705511.2023.2191292
28. Svetina D., Rutkowski L., Rutkowski D. Multiple-Group Invariance with Categorical Outcomes Using Updated Guidelines: An Illustration Using Mplus and the lavaan/semTools Packages. *Structural Equation Modeling: A Multidisciplinary Journal*, 2020. Vol. 27, no. 1, pp. 111—130. DOI:10.1080/10705511.2019.1602776
29. Tay L., Meade A.W., Cao M. An Overview and Practical Guide to IRT Measurement Equivalence Analysis. *Organizational Research Methods*, 2015. Vol. 18, no. 1, pp. 3—46. DOI:10.1177/1094428114553062
30. The standards for educational and psychological testing [Electronic resource]. American Psychological Association. Washington, 2009. URL: <https://www.apa.org/science/programs/testing/standards> (Accessed 25.06.2024).
31. Maassen E., D'Urso E.D., Van Assen M.A.L.M., Nuijten M.B., De Roover K., Wicherts J.M. The dire disregard of measurement invariance testing in psychological science. *Psychological Methods*, 2023. 14 p. DOI:10.1037/met0000624
32. Van de Vijver F.J., Leung K. *Methods and data analysis for cross-cultural research*. Cambridge: Cambridge University Press, 2021. 250 p. DOI:10.1017/9781107415188
33. Vandenberg R.J., Lance C.E. A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research. *Organizational Research Methods*, 2000. Vol. 3, no. 1, pp. 4—70. DOI:10.1177/109442810031002
34. Meuleman B., Żółtak T., Pokropek A., Davidov E., Muthén B., Oberski D.L., Billiet J., Schmidt P. Why Measurement Invariance is Important in Comparative Research. A Response to Welzel et al (2021). *Sociological Methods & Research*, 2023. Vol. 52, no. 3, pp. 1401—1419. DOI:10.1177/00491241221091755
35. Wicherts J.M. The importance of measurement invariance in neurocognitive ability testing. *The Clinical Neuropsychologist*, 2016. Vol. 30, no. 7, pp. 1006—1016. DOI:10.1080/13854046.2016.1205136
36. Widaman K.F., Ferrer E., Conger R.D. Factorial Invariance Within Longitudinal Structural Equation Models: Measuring the Same Construct Across Time. *Child Development Perspectives*, 2010. Vol. 4, no. 1, pp. 10—18. DOI:10.1111/j.1750-8606.2009.00110.x
37. Widaman K.F., Reise S.P. Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In Bryant K.J., Windle M., West S.G. (eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research*. Washington: American Psychological Association, 1997, pp. 281—324. DOI:10.1037/10222-009
38. Wingen T., Berkessel J.B., Englich B. No Replication, No Trust? How Low Replicability Influences Trust in Psychology. *Social Psychological and Personality Science*, 2020. Vol. 11, no. 4, pp. 454—463. DOI:10.1177/1948550619877412

Информация об авторах

Логвиненко Татьяна Игоревна, клинический психолог, научный сотрудник, Университетская клиника, Мюнхенский университет имени Людвига-Максимилиана, г. Мюнхен, Германия; приглашенный исследователь, Институт образования, Национальный исследовательский университет «Высшая школа экономики» (ФГАОУ ВО «НИУ ВШЭ»), г. Москва, Российская Федерация, ORCID: <https://orcid.org/0000-0001-7430-1963>, e-mail: logvinenkota.spb@gmail.com

Information about the authors

Tatiana Logvinenko, Clinical Psychologist, Research Fellow, University Hospital, Ludwig Maximilian University of Munich (LMU), Munich, Germany; National Research University Higher School of Economics, Moscow, Russia, ORCID: <https://orcid.org/0000-0001-7430-1963>, e-mail: logvinenkota.spb@gmail.com

Получена 05.05.2024

Принята в печать 24.06.2024

Received 05.05.2024

Accepted 24.06.2024