

Психометрические исследования: современные методы и новые возможности для образования

Елена Карданова, Алина Иванова

Статья поступила
в редакцию
в сентябре 2023 г.

Карданова Елена Юрьевна — кандидат физико-математических наук, научный руководитель центра психометрики и измерений в образовании Института образования, Национальный исследовательский университет «Высшая школа экономики». E-mail: ekardanova@hse.ru. ORCID: <https://orcid.org/0000-0003-2280-1258>

Иванова Алина Евгеньевна — кандидат наук об образовании, старший научный сотрудник центра психометрики и измерений в образовании Института образования, Национальный исследовательский университет «Высшая школа экономики». Адрес: 101000, Москва, Потаповский пер., 16, стр. 10. E-mail: aeivanova@hse.ru. ORCID: <https://orcid.org/0000-0003-3340-7651> (контактное лицо для переписки)

Аннотация

Качественные измерения — фундаментальное требование к исследовательской практике в сфере социальных наук. Качество измерений определяет валидность интерпретаций и выводов, которые мы можем сделать, решений, которые мы можем принять на основе полученных в результате измерений данных. Для качественных измерений в социальных науках необходимы инструменты оценки, а также методы анализа данных, позволяющие связать наблюдаемые результаты измерений с теоретическими атрибутами. Научную основу для их разработки дает психометрика.

Предваряя специальный выпуск журнала «Вопросы образования / Educational Studies Moscow», посвященный психометрике, приглашенные редакторы этого выпуска освещают основные вехи истории психометрики, выделяют несколько исключительно значимых публикаций, отмечают профессиональные институции и авторов, которые внесли весомый вклад в развитие данной отрасли науки. Особое внимание авторы уделяют истории психометрики в России. Оценивая возможности, перспективы и ограничения психометрики, авторы высказывают свою точку зрения на дискуссионные вопросы, и она не всегда совпадает с мнением авторов специального выпуска.

В этом выпуске представлены примеры использования современных психометрических методов для решения актуальных проблем в исследованиях образования, а также в исследованиях на стыке образования и психологии, образования и разных сфер бизнеса. Всех авторов представленных статей объединяет стремление совершенствовать исследовательскую практику в социальных науках за счет по-настоящему качественных измерений.

Ключевые слова

психометрика, измерения в социальных науках, история психометрики

Для цитирования

Карданова Е.Ю., Иванова А.Е. (2023) Психометрические исследования: современные методы и новые возможности для образования. *Вопросы образования / Educational Studies Moscow*, № 3, сс. 8–19. <https://doi.org/10.17323/vo-2023-17951>

Psychometric Research: Modern Methods and New Opportunities for Education

Elena Kardanova, Alina Ivanova

Elena Yu. Kardanova — Candidate of Sciences (PhD) in Differential Equations, Dynamic Systems and Optimal Control; Scientific Supervisor, Centre for Psychometrics and Measurement in Education, Institute of Education, National Research University, Higher School of Economics. E-mail: ekardanova@hse.ru. ORCID: <https://orcid.org/0000-0003-2280-1258>

Alina E. Ivanova — Senior Researcher at the Center for Psychometrics and Measurements in Education, Institute of Education, National Research University Higher School of Economics. Address: Bld. 10, 16 Potapovsky Ln, 101000 Moscow, Russian Federation. E-mail: aeivanova@hse.ru. ORCID: <https://orcid.org/0000-0003-3340-7651> (corresponding author)

Abstract Qualitative measurement is a fundamental requirement for research practice in the social sciences. The quality of measurements determines the validity of the interpretations, conclusions and decisions we can make based on the data obtained from the measurements. Qualitative measurement in the social sciences requires assessment tools as well as data analysis methods to link observed measurement results to theoretical attributes. The scientific basis for their development is provided by psychometrics.

Preceding the special issue of the journal “Voprosy obrazovaniya / Educational Studies Moscow” devoted to psychometrics, the guest editors of this issue cover the main milestones of the history of psychometrics, highlight some significant publications, note the professional institutions and authors who have made their valuable contribution to the development of this branch of science. The authors pay special attention to the history of psychometrics in Russia. Assessing the possibilities, prospects and limitations of psychometrics, the authors express their point of view on the debatable issues of psychometrics, and it does not always coincide with the opinion of the authors of the special issue. This issue presents examples of using modern psychometric methods to solve actual problems in education research, as well as in research at the intersection of education and psychology, education and different spheres of business. All the authors of the presented articles are united by the desire to improve research practice in the social sciences through truly qualitative measurements.

Keywords psychometrics, measurement in social sciences, history of psychometrics

For citing Kardanova E.Yu., Ivanova A.E. (2023) Psikhometricheskie issledovaniya: sovremennye metody i novye vozmozhnosti dlya obrazovaniya [Psychometric Research: Modern Methods and New Opportunities for Education]. *Voprosy obrazovaniya / Educational Studies Moscow*, no 3, pp. 8–19. <https://doi.org/10.17323/vo-2023-17951>

Психометрика создает теорию и методологию измерений в социальных науках, и поэтому область ее интересов обширна, а возможности применения очень широки. Психометрики разрабатывают инструменты оценки, методы измерения конструкторов, предлагают и применяют формализованные модели, которые могут служить для связи наблюдаемых явлений

с теоретическими атрибутами. Иначе говоря, психометрика определяет принципы разработки инструментов измерений в социальных науках, а также принципы работы с данными измерений. Она позволяет получить надежные и валидные данные измерений, которые можно использовать для принятия управленческих решений или проверки исследовательских гипотез.

Психометрика возникла в конце XIX — начале XX в. из социального запроса на объективные и надежные методы измерения психологических характеристик людей. Начало развития психометрики связывают с именами Ф. Гальтона, Ч. Спирмена, А. Бине, Р. Кэттелла, которые создали первые тесты и заложили основы статистического анализа данных тестирования. В дальнейшем большой вклад в развитие психометрики внесли исследователи компании *Educational Testing Service*¹ в Принстоне, среди которых можно выделить Ф. Лорда, Б. Грина, А. Бирнбаума. Своими работами они заложили основы современной теории тестирования и дали ее статистическое обоснование.

В 1968 г. вышла книга *Statistical Theories of Mental Test Scores* [Lord, Novick, 1968]. Она, во-первых, содержала тщательное обоснование классической теории тестирования. Во-вторых, четыре главы этой книги, написанные А. Бирнбаумом, посвящены основам современной теории тестирования — *item response theory* (IRT), активно развивавшейся в предыдущее десятилетие. В-третьих, в книге рассмотрены практические вопросы тестирования, в частности массового. Позднее Ф. Лорд опубликовал книгу *Applications of Item Response Theory to Practical Testing Problems* [Lord, 1980], в которой рассмотрел опыт применения IRT в практике тестирования: как оцениваются параметры заданий и испытуемых, чем полезны на практике характеристическая и информационная функции задания и теста, как можно выровнять разные формы теста. В этой книге Ф. Лорд также описывал базовые идеи компьютерного адаптивного тестирования. В дальнейшем идеи IRT получили развитие в работе Р. Хамблтона с соавторами *Fundamentals of Item Response Theory* [Hambleton, Swaminathan, Rogers, 1991], в книге С. Эмбретсон и С. Рейса *Item Response Theory for Psychologists* [Embretson, Reise, 2000] и многих других изданиях.

Отдельная линия развития IRT связана с именем Георга Раша, датского математика, который изучал свойства моделей измерения и разработал семейство моделей, обладающих так называемой специфической объективностью: в таких моделях параметры испытуемого и задания полностью отделены. Свойства разработанных им моделей измерения Раш описал в книге *Probabilistic Models for Some Intelligence and Attainment*

¹ <https://www.ets.org/>

Tests [Rasch, 1960]. В дальнейшем они получили название моделей семейства Раша и вдохновили многих ученых. Последователи Раша расширили это семейство моделей и придали развитие его теории. Самым известным из последователей Раша является Б. Райт, который создал в Университете Чикаго лабораторию объективных измерений, где вместе с учениками проводил исследования в области Раш-моделирования. Б. Райт с коллегами написал две книги с изложением основ измерений в рамках моделей Раша: *Best Test Design* [Wright, Stone, 1979] и *Rating Scale Analysis* [Wright, Masters, 1982]. Многие из его учеников внесли большой вклад в развитие психометрики — Д. Эндрич, М. Вилсон, Дж. Мастерс, М. Линакр, К. Майфорд [Myford, Wolfe, 2003; 2004].

В 1935 г. Л.Л. Терстоун основал в Анн Арборе Психометрическое общество, которое сегодня признано наиболее авторитетным институтом, занимающимся психометрической наукой. В 2019 г. группа исследователей попыталась представить академическую генеалогию Психометрического общества [Wijssen et al., 2019] по аналогии с традиционным семейным генеалогическим деревом. Отношения между научными руководителями и их учениками образовали пять отдельных ветвей, берущих начало от Джеймса Р. Энджелла, Вильгельма Вундта, Уильяма Джеймса, Карла Фридриха Гаусса или Альберта Мишотта.

В России начало развития психометрики датируют концом XIX в. и связывают с открытием первой Лаборатории психологических экспериментов под руководством В. Бехтерева в 1885 г. и первой Лаборатории экспериментальной педагогической психологии под руководством А. Нечаева в 1901 г. Оба ученых могут считаться учениками В. Вундта: Бехтерев в 1884–1885 гг. слушал в Лейпциге курс лекций Вундта по экспериментальной психологии и занимался в его лаборатории [Будилова, Кольцова, 1985], а Нечаев в 1898 г. во время стажировки в Европе работал в лабораториях Вундта в Лейпциге, а также у А. Бине в Париже [Сироткина, Смит, 2016]. В 1904 г. он открыл первые в России педологические курсы, на которых проводились занятия по технике психологического эксперимента и основам статистических методов. Таким образом, первые российские исследователи, применявшие психометрические методы в своей работе, тесно связаны с академическим деревом мировой психометрики.

К сожалению, печально известное постановление 1936 г. «О педологических извращениях в системе наркомпросов»² на полвека затормозило развитие психометрики в нашей стране. Но в 1990-х годах в ходе адаптации системы образования в

² Постановление ЦК ВКП(б) «О педологических извращениях в системе наркомпросов». 4 июля 1936 г. См.: КПСС в резолюциях... Т. 6. 1933–1937 гг. М., 1985. С. 364–367.

России к новым реалиям постсоветской эпохи и возвращения страны в мировое научное поле интерес к количественным методам в социальных науках, и к психометрике в частности, начал быстро расти. Стали возникать локальные сообщества разработчиков и пользователей тестов, появились первые курсы повышения квалификации в области разработки тестов и первые методологические разработки, начали укрепляться международные связи. В 1995 г. вышла работа М.Б. Чельшковой «Разработка педагогических тестов на основе современных математических моделей», в которой впервые на русском языке рассматриваются основные дихотомические модели современной теории тестирования. В 2002 г. появился Приказ Министерства образования РФ³, в котором впервые нашей стране заданы стандарты разработки и оценки качества педагогических тестов. В этот же период появляются первые русскоязычные руководства по разработке тестов и первичному анализу их результатов — это книги В.С. Аванесова [1996] и М.Б. Чельшковой [2002], а также переводы на русский язык известных за рубежом книг, таких как «Справочное руководство по конструированию тестов» [Клайн, 1994] и «Индивидуальные различия» [Купер, 2000].

Дальнейшее развитие психометрики в образовании в России связано с созданием Федерального центра тестирования (ФЦТ) и проведением им, начиная с 1995 г., централизованного тестирования — первого в нашей стране массового обследования на основе стандартизированных тестов и процедур тестирования, обработки, анализа и представления результатов. Осуществление такого тестирования потребовало разработки методологической базы, проведения научных исследований, объединения специалистов в области психометрики. ФЦТ проводил ежегодные конференции по тестированию в образовании, привлекавшие большое число участников со всей страны, выпускал журнал «Вопросы тестирования в образовании», в котором представлялись и методологические проблемы разработки тестов в образовании, и результаты первых в стране научных исследований в области психометрики. В 2000-х сотрудники ФЦТ издали две книги, посвященные теории и практике современной теории тестирования, — «Введение в теорию моделирования и параметризации педагогических тестов» [Нейман, Хлебников, 2000] и «Моделирование и параметризация тестов: основы теории и приложения» [Карданова, 2008].

Наконец, в 2002 г. в связи с подготовкой и введением Единого государственного экзамена в стране появился специализированный институт — Федеральный институт педагогических

³ Приказ Минобрнауки РФ от 17 апреля 2000 г. № 1122 «О сертификации качества педагогических тестовых материалов».

измерений (ФИПИ), основной задачей которого стало создание контрольно-измерительных материалов для государственной итоговой аттестации.

Основное препятствие на пути развития психометрики в России — это огромный дефицит профессиональных кадров. Как ответ на этот вызов, в 2010 г. в Институте образования НИУ ВШЭ создана магистерская программа «Измерения в психологии и образовании», которая в 2020 г. преобразована в программу «Обучение и оценивание как наука». До сих пор она является единственной на постсоветском пространстве программой высшего образования, готовящей специалистов в области разработки инструментов измерения в образовании и других социальных науках, психометрики, анализа данных с применением современных методов статистики, машинного обучения. За время существования программа подготовила большое число специалистов в области психометрики, и мы рады, что большинство авторов статей нашего спецвыпуска, прошедших строгий отбор и рецензирование, — выпускники нашей магистерской программы.

Цель предлагаемого вниманию читателей специального выпуска журнала «Вопросы образования / Educational Studies Moscow» — предоставить широкому кругу авторов из разных организаций, разных сфер и уровней образования (и не только) возможность ознакомить профессиональное сообщество с примерами использования современных психометрических методов для решения актуальных проблем в исследованиях, проводимых в образовании, а также на стыке образования и психологии, образования и разных сфер бизнеса. При этом, как нам кажется, всех авторов этого выпуска объединяет стремление совершенствовать исследовательскую практику в сфере социальных наук за счет по-настоящему качественных измерений. Именно качественные измерения определяют валидность интерпретаций и выводов, которые мы можем сделать, решений, которые мы можем принять, применяя тот или иной инструмент измерения. Сегодня валидными мы считаем те измерения, для которых собраны доказательства и теория вместе поддерживают интерпретацию результатов тестирования для заявленных целей использования этих тестов [American Educational Research Association et al., 2014]. Не будет преувеличением сказать, что именно соображения валидности составляют сердцевину разработки тестов и оценки их качества.

Методы психометрики, в частности современная теория тестирования и конфирматорный факторный анализ, имеют долгую историю применения в исследовательской практике, они многократно усовершенствованы и доказали свою полезность и высокое качество производимых ими данных. Оба метода

предоставляют уникальную информацию о функционировании отдельных заданий и о качестве тестов или опросников в целом и вносят важный вклад в состав свидетельств валидности результатов измерения, подтверждая возможность использования тех или иных инструментов измерения в научных и практических целях.

Авторы специального выпуска напрямую или косвенно затрагивают проблему валидности, дискутируя об истории психологического знания, полученного с помощью и без помощи психометрики, рассуждая о природе и тонкостях психометрического моделирования, применяя возможности психометрического моделирования для поиска ответов на широкий круг исследовательских вопросов.

Авторы статьи «Апробация шкалы установок к межкультурному обучению в вузе» М. Бульцева и С. Берриос Кальехас считают такие установки критически важным условием развития межкультурной компетентности студентов. Ввиду недостатка надежных средств измерения студенческих установок к межкультурному обучению авторы предлагают свой инструмент. Приведя подробный анализ концептов межкультурного обучения и межкультурной компетентности, авторы описывают процесс апробации Шкалы установок к межкультурному обучению в вузе и используют конфирматорный факторный анализ для ее валидации. Стоит отметить, что редакторы данного спецвыпуска придерживаются иного подхода в понимании валидности (мы уже привели его выше). Этот подход адресует скорее к работам С. Мессика [Messick, 1995], М. Кейна [Kane, 2001], Р. Мислеви [Mislevy, 2007], которые показали, что о валидности инструмента говорить некорректно. Вслед за ними мы полагаем, что доказательство валидности базируется на последовательном сборе ее свидетельств, на поиске аргументов, поддерживающих или опровергающих интерпретацию и использование результатов тестирования в заданных целях. Тем не менее многие исследователи, в том числе в смежных сферах социальных наук, например в менеджменте, придерживаются иной концепции валидности, применяют наборы статистических показателей валидности и оценивают дискриминантную, конвергентную и другие виды валидности инструментов. Надеемся, что читатели этого спецвыпуска выберут близкую для них оптику, в которой им удобнее будет работать самим и оценивать результаты исследований коллег.

В статье «Опыт использования бифакторных моделей для снижения эффектов социальной желательности на материале нормативного опросника универсальных компетенций» Е. Сагитов, И. Брун и С. Павлов используют метод конфирматорного факторного анализа для решения давней проблемы психо-

логического тестирования — искажения итоговых баллов по измеряемым конструктам, обусловленного социальной желательностью ответов респондентов. Авторы описывают метод внесения корректировок в итоговые баллы респондентов на примере своего нормативного опросника универсальных компетенций. Представленный ими подход к обработке и анализу данных позволяет минимизировать эффект социальной желательности при измерении психологических конструктов.

И. Антипкина и А. Иванова в статье «Декомпозиция трудности заданий в тесте читательской грамотности» выясняют, что именно составляет трудность заданий в тесте, проверяющем навык чтения у учеников начальной школы. Авторы показывают возможность с помощью современной теории тестирования не просто описать тестовое поведение испытуемых и их результаты по тесту, но и объяснить, какие характеристики заданий могут обуславливать степень их трудности. Применяемая в работе линейная логистическая тестовая модель с ошибкой (LLTM+e) использует характеристики задания как предикторы вероятности верного решения этого задания. С помощью предложенного методологического подхода авторы проверяют предположение о том, что при контроле «внешних» характеристик заданий (например, формата), параметры трудности, связанные с заложенными в задании группами читательских умений, будут образовывать иерархию — от поиска информации, данной в явном виде (наиболее простые задания) до оценивания текста в целом (наиболее трудные задания). Статья наглядно показывает, что гипотезы исследователей не всегда оказываются верны, но полученные результаты не становятся при этом менее интересными или полезными с точки зрения разработки тестовых материалов.

С. Тарасов, И. Зуева и Д. Федерякин в статье «Измерение образовательного прогресса на основе когнитивных операций» также используют модель LLTM, но значительно обогащают ее благодаря синтезу с одной из моделей для измерения образовательного прогресса — моделью Андерсена. Авторы отмечают, что, несмотря на все технологические и методологические усовершенствования в области исследований образования в последние годы, измерение динамики достижений учащихся, их образовательного прогресса остается нетривиальной методологической задачей. И чтобы эту задачу решить, авторы предлагают измерять образовательный прогресс с помощью когнитивных операций, освоение которых заранее закладывается и затем проверяется в тесте. Таким образом можно не просто измерить образовательный прогресс, но и существенно расширить возможности интерпретации тестовых баллов учеников как раз за счет когнитивных операций. Для иллюстрации

предлагаемого подхода использована линейка тестов, применявшихся для мониторинга образовательного прогресса в математике у учащихся 8–9-х классов средней школы.

Статья Д. Грачевой «Роль контекста в заданиях сценарного типа при измерении универсальных навыков: применение теории генерализации» также иллюстрирует методологические достижения психометрики в решении практических задач оценивания в образовании. В современных быстро меняющихся и все более «цифровизирующихся» условиях большое внимание в образовании уделяется развитию и оцениванию универсальных навыков у школьников. Для такого оценивания необходимы новые тестовые форматы, основанные на наблюдаемых действиях учащегося в цифровой среде. Однако эти новые и, как правило, технологически сложные контекстуализированные форматы тестов несут с собой новые методологические вызовы. Каков вклад контекста в результаты оценивания универсальных компетенций? Какое количество контекстов сценарных заданий необходимо для надежного измерения универсальных навыков? Автор ищет ответы на эти вопросы, опираясь на методы теории генерализации. В статье содержится детальное описание теории генерализации и дизайна проведения анализа, она может стать методическим пособием для всех заинтересованных в изучении и применении психометрики в образовательных исследованиях.

Авторы еще двух статей этого специального выпуска ставят важные теоретические вопросы, открывая дискуссию о перспективах и ограничениях психометрических исследований. В статье «Психометрика и когнитивные исследования: противоречия и возможности кооперации» Ю. Кузьмина рассматривает историю взаимоотношений между экспериментальными исследованиями и психометрикой с конца XIX в. до настоящего времени. Когнитивная психология в основном развивалась в рамках экспериментальной парадигмы, в отличие от психометрики, занимающейся оценкой индивидуальных различий и корреляционными исследованиями. Автор статьи показывает, как возник и ширился разрыв между когнитивными исследованиями и психометрикой и как он связан с разной исследовательской логикой в этих двух подходах. Важно, что этот разрыв не предопределен — а следовательно, может быть преодолен. Опираясь на огромный пласт научных работ, автор показывает, что на всем протяжении развития психологии многие исследователи подчеркивали возможность сближения рассматриваемых подходов, которое способно обогатить психологию в целом.

В статье «Так ли полезна психометрика для академической психологии?» Ю. Тюменева анализирует логику психометрического моделирования как способа репрезентировать латент-

ный конструкт или обнаружить его структуру. Психологические теории относительно способностей и личностных черт часто полагаются на результаты психометрического моделирования. Ю. Тюменева задается вопросом: а действительно ли психометрическое моделирование является моделированием в общенаучном значении этого термина? Цепь логических рассуждений приводит автора к отрицательному ответу на этот вопрос. Поэтому, утверждает Ю. Тюменева, на основании психометрического моделирования можно делать выводы только о структуре тестовых данных, но не о структуре латентного конструкта.

Начиная обзор статей специального выпуска, мы отмечали, что наше видение как редакторов этого выпуска предложенных в статье идей и перспектив совсем не обязательно совпадает с видением авторов. Но мы верим, что именно различия во взглядах исследователей и возможность открыто обсуждать эти различия формируют академическую свободу.

Наконец, завершается специальный выпуск рецензиями на две выдающиеся, с нашей точки зрения, книги в области психометрики, изданные в течение последних десяти лет. И. Угланова знакомит читателей с книгой Р. Леви и Р. Мислеви *Bayesian Psychometric Modeling* (2016), посвященной байесовскому подходу к психометрике. К. Тарасова и Д. Грачева делятся своим видением психометрических перспектив, представленных А. фон Давьер, Р. Мислеви и Дж. Хао в книге *Computational Psychometrics: New Methodologies for a New Generation of Digital Learning and Assessment. With Examples in R and Python* (2022).

Мы надеемся, что этот специальный выпуск, в котором обсуждаются возможности, перспективы и ограничения применения психометрики в сфере образования, подарит читателям интеллектуальное удовольствие — такое же, какое он подарил нам во время его планирования, подготовки и публикации. Мы благодарим команду редакции журнала «Вопросы образования / Educational Studies Moscow» за предоставленную возможность провести интересную научную дискуссию на страницах журнала. И конечно же, мы благодарим авторов и рецензентов за проделанную работу и неоценимый вклад в эту дискуссию.

Литература

1. Аванесов В.С. (1996) *Композиция тестовых заданий*. М.: Ассоциация инженеров-педагогов.
2. Будилова Е.А., Кольцова В.А. (1985) 100-летие первой русской экспериментальной психологической лаборатории. *Вопросы психологии*, № 6, сс. 96–102.
3. Карданова Е.Ю. (2008) *Моделирование и параметризация тестов: основы теории и приложения*. М.: Федеральный центр тестирования.
4. Клайн П. (1994) *Справочное руководство по конструированию тестов. Введение в психометрическое проектирование*. Киев: ПАН.

5. Купер К. (2000) *Индивидуальные различия*. М.: Аспект.
6. Нейман Ю.М., Хлебников В.А. (2000) *Введение в теорию моделирования и параметризации педагогических тестов*. М.: Прометей.
7. Сироткина И., Смит Р. (2016) *История психологии в России: краткий очерк с авторскими акцентами. Препринт WP6/2016/01*. М.: НИУ ВШЭ.
8. Чельшкова М.Б. (2002) *Теория и практика конструирования педагогических тестов*. М.: Логос.
9. Чельшкова М.Б. (1995) *Разработка педагогических тестов на основе современных математических моделей*. М.: МИСИС.
10. American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational and Psychological Testing (U.S.) (2014) *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.
11. Embretson S.E., Reise S.P. (2000) *Item Response Theory for Psychologists*. Mahwah, NJ: Erlbaum.
12. Hambleton R.K., Swaminathan H., Rogers H.J. (1991) *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage.
13. Kane M.T. (2001) Current Concerns in Validity Theory. *Journal of Educational Measurement*, vol. 38, no 4, pp. 319–342. <http://dx.doi.org/10.1111/j.1745-3984.2001.tb01130.x>
14. Lord F.M. (1980) *Applications of Item Response Theory to Practical Testing Problems*. New York, NY: Routledge. doi.org/10.4324/9780203056615
15. Lord F.M., Novick M.R. (1968) *Statistical Theories of Mental Test Scores*. Menlo Park, CA: Addison-Wesley.
16. Messick S. (1995) Standards of Validity and the Validity of Standards in Performance Assessment. *Educational Measurement: Issues and Practice*, vol. 14, no 4, pp. 5–8. <https://doi.org/10.1111/j.1745-3992.1995.tb00881.x>
17. Mislevy R.J. (2007) Validity by Design. *Educational Researcher*, vol. 36, no 8, pp. 463–469.
18. Myford C.M., Wolfe E.W. (2004) Detecting and Measuring Rater Effects Using Many-Facet Rasch Measurement: Part 2. *Journal of Applied Measurement*, vol. 5, no 2, pp. 189–227.
19. Myford C.M., Wolfe E.W. (2003) Detecting and Measuring Rater Effects Using Many-Facet Rasch Measurement: Part 1. *Journal of Applied Measurement*, vol. 4, no 4, pp. 386–422.
20. Rasch G. (1960) *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen: Danmarks Paedagogiske Institut.
21. Wijzen L.D., Borsboom D., Cabaço T., Heiser W.J. (2019) An Academic Genealogy of Psychometric Society Presidents. *Psychometrika*, vol. 84, no 2, pp. 562–588. <http://dx.doi.org/10.1007/s11336-018-09651-4>
22. Wright B.D., Masters G.N. (1982) *Rating Scale Analysis. Rasch Measurement*. Chicago, IL: Mesa.
23. Wright B.D., Stone M.N. (1979) *Best Test Design. Rasch Measurement*. Chicago, IL: Mesa.

- References** American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational and Psychological Testing (U.S.) (2014) *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.
- Avanesov V.S. (1996) *Kompozitsiya testovykh zadaniy* [Composition of Test Tasks]. Moscow: Association of Engineers-Teachers.

- Budilova E.A., Koltsova V.A. (1985) 100-letie pervoy russkoy eksperimental'noy psikhologicheskoy laboratorii [100th Anniversary of the First Russian Experimental Psychological Laboratory]. *Voprosy Psichologii*, no 6, pp. 96–102.
- Chelyshkova M.B. (2002) *Teoriya i praktika konstruirovaniya pedagogicheskikh testov* [Theory and Practice of Designing Pedagogical Tests]. Moscow: Logos.
- Chelyshkova M.B. (1995) *Razrabotka pedagogicheskikh testov na osnove sovremennykh matematicheskikh modeley* [Development of Pedagogical Tests Based on Modern Mathematical Models]. Moscow: MISIS.
- Cooper C. (2000) *Individual'nye razlichiya* [Individual Differences]. Moscow: Aspekt.
- Embretson S.E., Reise S.P. (2000) *Item Response Theory for Psychologists*. Mahwah, NJ: Erlbaum.
- Hambleton R.K., Swaminathan H., Rogers H.J. (1991) *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage.
- Kane M.T. (2001) Current Concerns in Validity Theory. *Journal of Educational Measurement*, vol. 38, no 4, pp. 319–342. <http://dx.doi.org/10.1111/j.1745-3984.2001.tb01130.x>
- Kardanova E.Yu. (2008) *Modelirovanie i parametrizatsiya testov: osnovy teorii i prilozheniya* [Modeling and Parameterization of Tests: Fundamentals of Theory and Applications]. Moscow: Federal Testing Center.
- Klein P. (1994) *Spravochnoe rukovodstvo po konstruirovaniyu testov. Vvedenie v psichometricheskoe proektirovanie* [Reference Guide for Designing Tests. Introduction to Psychometric Design]. Kiev: PAN.
- Lord F.M. (1980) *Applications of Item Response Theory to Practical Testing Problems*. New York, NY: Routledge. doi.org/10.4324/9780203056615
- Lord F.M., Novick M.R. (1968) *Statistical Theories of Mental Test Scores*. Menlo Park, CA: Addison-Wesley.
- Messick S. (1995) Standards of Validity and the Validity of Standards in Performance Assessment. *Educational Measurement: Issues and Practice*, vol. 14, no 4, pp. 5–8. <https://doi.org/10.1111/j.1745-3992.1995.tb00881.x>
- Mislevy R.J. (2007) Validity by Design. *Educational Researcher*, vol. 36, no 8, pp. 463–469.
- Myford C.M., Wolfe E.W. (2004) Detecting and Measuring Rater Effects Using Many-Facet Rasch Measurement: Part 2. *Journal of Applied Measurement*, vol. 5, no 2, pp. 189–227.
- Myford C.M., Wolfe E.W. (2003) Detecting and Measuring Rater Effects Using Many-Facet Rasch Measurement: Part 1. *Journal of Applied Measurement*, vol. 4, no 4, pp. 386–422.
- Neiman Yu.M., Khlebnikov V.A. (2000) *Vvedenie v teoriyu modelirovaniya i parametrizatsii pedagogicheskikh testov* [Introduction to the Theory of Modeling and Parameterization of Pedagogical Tests]. Moscow: Prometei.
- Rasch G. (1960) *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen: Danmarks Paedagogiske Institut.
- Sirotkina I., Smith R. (2016) *Istoriya psichologii v Rossii: kratkiy ocherk s avtorskimi aktsentami. Preprint WP6/2016/01* [History of Psychology in Russia: Short Review with the Authors' Emphasis. Working paper no WP6/2016/01]. Moscow: HSE.
- Wijsen L.D., Borsboom D., Cabaço T., Heiser W.J. (2019) An Academic Genealogy of Psychometric Society Presidents. *Psychometrika*, vol. 84, no 2, pp. 562–588. <http://dx.doi.org/10.1007/s11336-018-09651-4>
- Wright B.D., Masters G.N. (1982) *Rating Scale Analysis. Rasch Measurement*. Chicago, IL: Mesa.
- Wright B.D., Stone M.N. (1979) *Best Test Design. Rasch Measurement*. Chicago, IL: Mesa.