

Вычислительная психометрика: ближайшее будущее или уже реальность

*Рецензия на книгу “Computational Psychometrics:
New Methodologies for a New Generation of Digital
Learning and Assessment”¹*

Ксения Тарасова, Дарья Грачева

Статья поступила в редакцию в мае 2023 г. Тарасова Ксения Вадимовна — кандидат педагогических наук, заместитель руководителя лаборатории измерения новых конструктов и дизайна тестов Центра психометрики и измерений в образовании Института образования, Национальный исследовательский университет «Высшая школа экономики». Адрес: 101000, Москва, Потаповский пер., 16, стр. 10. E-mail: ktarasova@hse.ru. ORCID: <https://orcid.org/0000-0002-3915-3165> (контактное лицо для переписки)

Грачева Дарья Александровна — младший научный сотрудник Центра психометрики и измерений в образовании Института образования, Национальный исследовательский университет «Высшая школа экономики». E-mail: dgracheva@hse.ru. ORCID: <https://orcid.org/0000-0002-4646-7349>

Аннотация Рецензируемое издание продолжает серию книг о методологии образовательного тестирования и оценивания, написанных ведущими психометриками и исследователями в области оценивания в образовании. Вычислительная психометрика рассматривается как сочетание методов вычислительных наук и психометрических принципов измерений для анализа данных, полученных в результате тестирования с использованием технологически усовершенствованных форматов теста. В первой части книги обсуждаются изменения, которые произошли в обучении и образовательном оценивании под влиянием цифровых технологий. Во второй части представлен обзор методов вычислительной психометрики: от традиционных психометрических моделей до технологий машинного обучения. Материалы книги могут быть полезны студентам и исследователям в области психометрики, которые занимаются разработкой, проектированием, анализом систем обучения и измерениями с использованием сложных тестовых форматов и данных. Сильной стороной книги является электронное приложение, содержащие код среды программирования *R* или *Python* для методологических глав.

Ключевые слова вычислительная психометрика, машинное обучение, образовательное оценивание, Evidence-Centered Design

¹ Davier von A.A., Mislevy R.J., Hao J. (eds) (2022) *Computational Psychometrics: New Methodologies for a New Generation of Digital Learning and Assessment. With Examples in R and Python*. Cham: Springer Nature.

Для цитирования Тарасова К.В., Грачева Д.А. (2023) Вычислительная психометрика: ближайшее будущее или уже реальность. Рецензия на книгу "Computational Psychometrics: New Methodologies for a New Generation of Digital Learning and Assessment". *Вопросы образования / Educational Studies Moscow*, № 3, сс. 221–230. <https://doi.org/10.17323/vo-2023-17938>

Computational Psychometrics: Near Future or Reality

Review of the book "Computational Psychometrics: New Methodologies for a New Generation of Digital Learning and Assessment"

Ksenia Tarasova, Daria Gracheva

Ksenia V. Tarasova — Candidate of Sciences in Education, Deputy Head of the Laboratory for Measuring New Constructs and Test Design, Centre for Psychometrics and Measurement in Education, National Research University Higher School of Economics. Address: Bld. 10, 16 Potapovsky Ln, 101000 Moscow, Russian Federation. E-mail: ktarasova@hse.ru. ORCID: <https://orcid.org/0000-0002-3915-3165> (corresponding author)

Daria A. Gracheva — Junior Research Fellow at the Center for Psychometrics and Measurements in Education, Institute of Education, National Research University Higher School of Economics. E-mail: dgracheva@hse.ru. ORCID: <https://orcid.org/0000-0002-4646-7349>

Abstract This book continues a series of books on the methodology of educational testing and assessment written by leading psychometricians and researchers in the field of educational assessment. Computational psychometrics is defined as the combination of computer science methods and psychometric measurement principles for analysing data obtained as a result of testing using technologically advanced test formats. The first part of the book discusses the changes that have occurred in teaching and educational assessment under the influence of digital technologies. The second part provides an overview of computational psychometric methods: from traditional psychometric models to machine learning technologies. The material in the book can be useful to students and researchers in the field of psychometrics who are involved in the development, design and analysis of learning systems and measurements using complex test formats and data. The strength of the book is an electronic application containing the code of the R or Python programming environment for the methodological chapters.

Keywords computational psychometrics, machine learning, educational assessment, evidence-centered design

For citing Tarasova K.V., Gracheva D.A. (2023) Vychislitel'naya psikhometrika: blizhajshee budushchee ili uzhe real'nost'. Retsenziya na knigu "Computational Psychometrics: New Methodologies for a New Generation of Digital Learning and Assessment" [Computational Psychometrics: Near Future or Reality. Review of the book "Computational Psychometrics: New Methodologies for a New Generation of Digital Learning and Assessment"]. *Voprosy obrazovaniya / Educational Studies Moscow*, no 3, pp. 221–230. <https://doi.org/10.17323/vo-2023-17938>

Психометрика возникла более века назад, когда для обоснования выводов о психологических свойствах и процессах стали применять количественные методы обработки данных, полученных в ходе наблюдения. В широком смысле содержание психометрики с тех пор не изменилось, несмотря на прогресс статистических теорий, моделей и методов.

Современные условия, в частности большой объем данных, получаемых в процессе тестирования, ставят перед исследователями в области психометрики новые задачи по их обработке, анализу и интерпретации. Среди актуальных вызовов — повышение точности и валидности измерений, выявление паттернов и стилей ответов, определение новых конструкторов и предоставление качественной обратной связи участникам тестирования и всем заинтересованным сторонам.

После нескольких лет обсуждений специфики новых методологий в тестировании исследователи обратили внимание на то, что они, несмотря на все различия, имеют одну общую черту — использование вычислительных моделей. Так родился лаконичный термин — вычислительная психометрика, введенный Алиной фон Давье на международной конференции по машинному обучению [Davier von, 2015].

Вычислительная психометрика стремится интегрировать методы науки о данных с методами машинного обучения в психометрике, придерживаясь установленных принципов измерений в психологии и образовании.

В книге представлены основы вычислительной психометрики: в первой части обсуждаются изменения в образовательном ландшафте, обусловившие переход к следующему поколению методов обучения и оценивания, а также отдельные технологии, способные трансформировать образовательный опыт; во второй части описаны конкретные методы как актуальные иллюстрации возможностей вычислительной психометрики.

**Обучение
и оценивание
в новых
условиях**

В написанном Р. Мислеви разделе книги, посвященном методологии вычислительной психометрики, представлен ситуативный, социокогнитивный взгляд на природу человеческих способностей, на их развитие и на способы их использования человеком для взаимодействия с социальным и физическим миром [Greeno, 1998]. Эта исследовательская перспектива создает запрос на широкий спектр типов оценки, обычно включающих взаимодействие со средой, часто в цифровых контекстах, а иногда требующих сотрудничества между людьми. Именно на основе оценки действий испытуемых в таких средах можно получить прямые свидетельства способностей, необходимых для формирования исследовательских навыков, умения решать проблемы, коммуникативных компетенций и т.п.

Вычислительная психометрика не просто предлагает совокупность методов, но может служить основой виртуальной среды обучения и оценки, такой как игры, симуляции, совместная работа и дополненные среды. В этих средах возможно получение большого количества данных сложной структуры.

В разделе, посвященном специфике вычислительной психометрики, А. фон Давье и ее соавторы подчеркивают, что сбор данных в цифровых средах должен быть целенаправленным в максимально возможной степени. Поэтому важно пересмотреть весь процесс разработки инструмента оценивания. Инструменты, ориентированные на доказательную базу или доказательный дизайн (*evidence-centered design, ECD*) [Mislevy, Almond, Lukas, 2003; Arieli-Attali et al., 2019], создаются с таким расчетом, чтобы обеспечить возможность фиксации и учета каждого элемента процесса выработки ответа или решения. В частности, дизайн таких инструментов включает концепции модели учащегося, модели задания и модели доказательств. Модель учащегося определяет конструкторы, о которых мы хотим сделать выводы. Модель задания предусматривает виды деятельности, которые позволят получить доказательства: это могут быть вопросы с множественным выбором, симуляция, игра или любая другая деятельность, в которой может участвовать испытуемый. В модели доказательств особое внимание уделяется их идентификации (извлечение ключевых элементов из продукта деятельности, созданного учащимся в результате тестирования) и агрегированию (аккумуляция информации в балл с помощью статистической модели). При анализе данных акцент смещается на то, как информация, полученная в результате выполнения задачи, обобщается для составления суждений о навыках учащихся, которые могут проявиться в разных контекстах. Эта перспектива особенно актуальна для оценивания высокоуровневых навыков.

Тестирование с использованием цифровых технологий позволяет разрабатывать насыщенные и технологически усовершенствованные типы заданий, которые лучше отражают ситуации из реальной жизни и позволяют измерять сложные навыки, обеспечивая тем самым более глубокое оценивание способностей учащихся. Эволюция от технологически усовершенствованных заданий к технологически усовершенствованному оцениванию происходит в несколько этапов [DiCerbo, Behrens, 2012]. На первом этапе технологии используются для создания новых типов элементов, механик и улучшения обратной связи; на втором заданию строится на основе центрального контекста — проблемы, которую необходимо решить (симуляционные задания); на третьем этапе создаются естественные цифровые среды, разрабатываемые с учетом аргумента оце-

нивания (использование скрытого оценивания в игре) [Shute, Ventura, 2013]; четвертый этап представляет собой экосистему, в которой накапливается информация из разных естественных цифровых сред, подобных созданным на третьем уровне.

Отдельный раздел в книге посвящен виртуальным тестовым средам (*virtual performance-based assessments*, VPBAs), в которых участники тестирования взаимодействуют с системами, иногда включающими других людей или агентов, чтобы представить доказательства своих знаний, умений или других качеств. Сбор цифровых данных позволяет получить подробную информацию о действиях испытуемых и развитии ситуаций, в которых эти действия происходят. При этом перед исследователем стоит задача сконструировать VPBA, способную инициировать проявление целевых способностей участников тестирования, дать им возможность продемонстрировать эти способности, зафиксировать соответствующие аспекты выполнения заданий, выявить значимые закономерности, которые являются доказательством целевых способностей, и предоставить основу для синтеза доказательств и описания их свойств.

К категории виртуальных тестовых сред современные исследователи относят: задания сценарного типа (*scenario-based assessments*), симуляционные задания (*simulation-based assessments*), оценивание с использованием заданий игрового типа (*game-based assessments*), оценивание совместной работы (*collaborative assessments*).

Оценка с помощью заданий сценарного типа актуализирует опыт испытуемого повествовательным контекстом с целью добавления слоя смысла к мотивам, мыслям и действиям. В самом простом варианте непрерывный контекст (развитие сюжета) вводится для последовательности заданий, которые часто имеют традиционный формат, например множественный выбор или краткий ответ. Более сложные задания сценарного типа могут содержать элементы симуляции и игр.

В симуляционных заданиях воспроизводятся обстоятельства, характерные для тех или иных видов реальной деятельности: например, проведение научного эксперимента, оказание медицинской помощи или полет на реактивном истребителе. Такой дизайн тестирования позволяет оценить способности индивидов применительно к конкретной сфере занятости [Davey et al., 2015]. Симуляционные задания выполняют очень важную функцию: они дают возможность людям продемонстрировать свои знания, умения и навыки в таких ситуациях, которые могут быть слишком опасными для того, чтобы проводить испытания в реальных условиях, или организация таких испытаний обходится слишком дорого либо затратна по времени [Snir, Smith, Grosslight, 1993].

Оценивание с использованием заданий игрового типа или основанное на игре объединяет принципы игрового дизайна с принципами дизайна заданий для измерения знаний, умений и навыков. Такие задания могут повысить мотивацию тестируемых за счет своей увлекательности. При разработке заданий игрового типа важно найти баланс между игровыми элементами и характеристиками, такими как увлекательность и вовлеченность, с одной стороны, и элементами оценивания — с другой, т.е. обеспечить соответствие игры психометрическим требованиям, такими как валидность, надежность и справедливость. Только так можно достичь заданной цели и предоставить испытуемому возможность продемонстрировать свои способности [Kim, Shute, 2015]. Создаваемые разработчиками игр сложные интерактивные среды подходят для исследования таких комплексных конструктов, как системное мышление и межкультурная компетентность.

Оценивание совместной работы — востребованная тестовая среда. Поскольку в современном мире планирование деятельности, принятие решений и урегулирование проблем по большей части осуществляется в составе команд, навыки, необходимые для достижения успеха в групповой работе, становятся важным критерием оценки персонала. В рамках оценивания совместной работы тестируемый получает возможность проявить свои умения эффективно работать или общаться с другими людьми. Для этого создается среда, которая обеспечивает его общение и сотрудничество с одним или несколькими лицами. Это общение может быть асинхронным, как, например, потоковое онлайн-обсуждение в системе управления обучением или на массовых открытых онлайн-курсах (MOOC) [Rosé, Ferschte, 2016; Wise, Chiu, 2011], или синхронным, когда люди общаются в режиме реального времени через чат или аудио- и видеоканалы [Andrews-Todd, Forsyth, 2020; Bower, Lee, Dalgarno, 2017]. Возможны среды, поддерживающие общение между человеком и одним или несколькими компьютерными агентами [Biswas et al., 2010; Graesser et al., 2012], взаимодействие двух людей с искусственными агентами [Liu et al., 2015]. В виртуальной среде можно фиксировать все действия и рассуждения членов команды, а не только ответы, которые они дают. По результатам такого оценивания можно судить не только о стратегиях, которые члены команды используют во время работы для решения проблемы, но и о характере их взаимодействия.

В последние годы новые возможности проектирования, сбора и анализа данных используются не только для создания цифровых сред оценки, но и для персонализации обучения на основе данных об учебном опыте студентов. Один из разделов книги посвящен обзору моделей адаптивного обу-

чения и методов интеллектуального анализа данных. Системы онлайн-обучения, использующие данные о характеристиках и успеваемости учащихся для подстраивания учебных программ к их потребностям, становятся адаптивными образовательными средами. Они учитывают широкий спектр характеристик студентов и конструкторов, влияющих на процесс обучения, — их знания, эмоции, поведенческую активность и мотивацию. Успешные адаптивные системы обучения могут использовать множество данных о деятельности учащегося в режиме реального времени, чтобы делать выводы, на основе которых учащемуся автоматически предоставляется необходимая для достижения успеха индивидуализированная поддержка.

Методы вычислительной психометрики

Во второй части книги рассматриваются методы вычислительной психометрики и возможности их применения в новой реальности: для сложных тестовых форматов и данных.

Вводный раздел посвящен фундаментальным идеям и моделям психометрики. В широком смысле модели психометрики связывают наблюдаемые данные тестирования с ненаблюдаемыми (латентными) характеристиками респондентов. В качестве наиболее распространенных сегодня психометрических моделей Р. Мислеви и М. Большова упоминают классическую теорию тестирования и ее расширения, в частности теорию генерализации, непрерывные и дискретные модели современной теории тестирования, а также эксплораторный и конфирматорный факторный анализ.

Развитие цифровых технологий и появление интерактивных заданий, реализуемых в насыщенной тестовой среде, предъявляет дополнительные требования к устоявшимся психометрическим практикам. В частности, на первый план выходят проблемы локальной зависимости заданий, многомерности оцениваемых конструкторов, больших объемов данных о процессе выполнения заданий (*process data*) и оценки навыков высокого порядка.

Перспективным для оценки параметров моделей в новых реалиях авторам представляется байесовский подход (*Bayesian inference*) с применением алгоритма Монте-Карло по схеме цепи Маркова (*Markov chain Monte Carlo*, MCMC), ему посвящен отдельный раздел книги. Авторы также высоко оценивают возможности в современных условиях таких психометрических практик, как моделирование иерархических структур данных (*hierarchical models*), компьютерное адаптивное тестирование на основе IRT (*computer adaptive testing*, CAT IRT), внедрение коллатеральной информации о респондентах и заданиях с опорой на когнитивные теории.

Одним из активно развивающихся направлений сегодня является сетевая психометрика [Marsman et al., 2018; Epskamp et al., 2018], которая предлагает альтернативное понимание психологических конструктов. В сетевой психометрике связи между наблюдаемыми переменными, входящими в сеть (конструкт), считаются корреляционными и при отсутствии общей причины для их значений, в то время как в традиционных психометрических моделях корреляционные отношения принято объяснять наличием общей причины — латентной переменной. Сетевые модели представляют структуру психологических конструктов в виде сложных взаимосвязей между психологическими, биологическими, социологическими и другими переменными без дополнительных допущений о причинности в отношениях между наблюдаемыми переменными, которыми оперирует традиционная психометрика.

Переход к цифровым инструментам оценивания сопровождается увеличением объема данных о респонденте. Традиционная психометрика основывается на конкретных и структурированных данных, например на баллах за выполнение заданий, в которых респондент поставлен перед ограниченным выбором из нескольких вариантов ответа. Вычислительная психометрика часто имеет дело с данными «на микроуровне» — с детализированными и неструктурированными сведениями о действиях тестируемых в процессе выполнения задания, например о движениях мыши, нажатии на клавиши, кликах. В состав этих данных могут входить нейрофизиологические показатели, например результаты фиксации взгляда.

Задача вычислительной психометрики состоит в том, чтобы на основе микроданных сформулировать содержательный вывод о респонденте. Эффективными инструментами для обнаружения закономерностей в микроданных служат подходы, сложившиеся в рамках интеллектуального анализа данных (*data mining*), учебной аналитики (*learning analytics*), машинного обучения. В рецензируемом издании описаны два основных направления машинного обучения — обучение с учителем (*supervised machine learning*) и обучение без учителя (*unsupervised machine learning*). Данные о процессе, т.е. о последовательности действий внутри задания, также могут быть проанализированы с использованием моделей временных рядов или социальных сетей (*social network analysis*).

Книга завершается обзором использования моделей обработки естественного языка (*natural language processing*, NLP) в области оценивания. Авторы рассматривают задачи интеллектуального анализа текстов (*text mining*) и автоматического скоринга заданий открытого типа (*automatic scoring*) с коротким и длинным (эссе) конструированным ответом.

Заключение Результатом использования цифровых форматов оценивания становится большой объем данных о поведении респондента в тестовой среде. Для анализа таких данных требуются специальные методы, расширяющие возможности традиционных психометрических моделей. Следуя этому тренду, *Ed-Tech*-компании начали применять к результатам тестирований продвинутое методы наук о данных, однако бизнес-сообщество уделяет мало внимания фундаментальным принципам измерений — валидности, надежности, справедливости оценивания, что отражается на качестве выводов о тестируемых. Вычислительная психометрика сочетает применение продвинутых методов вычислительных наук с соблюдением психометрических принципов измерений при анализе сложных данных. Рецензируемая книга дает представление об основных принципах измерений, которые должны соблюдаться на протяжении всего тестирования — от разработки заданий до представления результатов. Один из разделов книги посвящен обзору методов вычислительной психометрики с примерами кода среды программирования *R* или *Python*. Многие из этих методов еще не вошли в повседневный инструментарий психометрика, а навыки, необходимые для моделирования сложных данных, как правило, недостаточно освещаются в большинстве образовательных программ по измерениям. Таким образом, книга знакомит широкий круг читателей с идеями и методами вычислительной психометрики, которые способны обогатить образовательное оценивание и сделать более точными выводы в отношении каждого респондента и системы обучения и образования в целом.

- References**
- Andrews-Todd J., Forsyth C.M. (2020) Exploring Social and Cognitive Dimensions of Collaborative Problem Solving in an Open Online Simulation-Based Task. *Computers in Human Behavior*, vol. 104, January, Article no 105759. <https://doi.org/10.1016/j.chb.2018.10.025>
- Arieli-Attali M., Ward S., Thomas J., Deonovic B., von Davier A.A. (2019) The Expanded Evidence-Centered Design (E-ECD) for Learning and Assessment Systems: A Framework for Incorporating Learning Goals and Processes within Assessment Design. *Frontiers in Psychology*, vol. 10, April, Article no 853. <https://doi.org/10.3389/fpsyg.2019.00853>
- Biswas G., Jeong H., Kinnebrew J.S., Sulcer B., Roscoe R. (2010) Measuring Self-Regulated Learning Skills through Social Interactions in a Teachable Agent Environment. *Research and Practice in Technology Enhanced Learning*, vol. 5, no 2, pp. 123–152. <http://dx.doi.org/10.1142/S1793206810000839>
- Bower M., Lee M.J., Dalgarno B. (2017) Collaborative Learning across Physical and Virtual Worlds: Factors Supporting and Constraining Learners in a Blended Reality Environment. *British Journal of Educational Technology*, vol. 48, no 2, pp. 407–430. <http://dx.doi.org/10.1111/bjet.12435>
- Davey T., Ferrara S., Shavelson R., Holland P., Webb N., Wise L. (2015) *Psychometric Considerations for the Next Generation of Performance Assessment. Report of the Center for K-12 Assessment & Performance Management*. Available at: <https://>

- www.ets.org/Media/Research/pdf/psychometric_considerations_white_paper.pdf (accessed 20 August 2023).
- Davier von A.A. (2015) *Virtual and Collaborative Assessments: Examples, Implications, and Challenges for Educational Measurement*. Paper presented at the 32nd International Conference on Machine Learning (ICML 2015) (Lille, France, 2015, July 6–11).
- DiCerbo K.E., Behrens J.T. (2012) Implications of the Digital Ocean on Current and Future Assessment. *Computers and Their Impact on State Assessment: Recent History and Predictions for the Future* (eds R. Lissitz, H. Jiao), Charlotte, NC: Information Age Publishing, pp. 273–306.
- Epskamp S., Maris G., Waldorp L.J., Borsboom D. (2018) Network Psychometrics. *The Wiley Handbook of Psychometric Testing* (eds P. Irwing, D. Hughes, T. Booth), New York, NY: Elsevier.
- Graesser A.C., D'Mello S., Hu X., Cai Z., Olney A., Morgan B. (2012) AutoTutor. *Applied Natural Language Processing: Identification, Investigation and Resolution* (eds P. McCarthy, C. Boonthum-Denecke), Hershey, PA: IGI Global, pp. 169–187. <http://dx.doi.org/10.4018/978-1-60960-741-8>
- Greeno J.G. (1998) The Situativity of Knowing, Learning, and Research. *American Psychologist*, vol. 53, no 1, pp. 5–26. <https://doi.org/10.1037/0003-066X.53.1.5>
- Kim Y.J., Shute V.J. (2015) The Interplay of Game Elements with Psychometric Qualities, Learning, and Enjoyment in Game-Based Assessment. *Computers & Education*, vol. 87, no 2, pp. 340–356. <http://dx.doi.org/10.1016/j.compedu.2015.07.009>
- Liu L., von Davier A.A., Hao J., Kyllonen P., Zapata-Rivera J.-D. (2015) A Tough Nut to Crack: Measuring Collaborative Problem Solving. *Handbook of Research on Computational Tools for Real-World Skill Development* (eds Y. Rosen, S. Ferrara, M. Mosharraf), Hershey, PA: IGI-Global, pp. 344–359.
- Marsman M., Borsboom D., Kruis J., Epskamp S., van Bork R., Waldorp L. et al. (2018) An Introduction to Network Psychometrics: Relating Ising Network Models to Item Response Theory Models. *Multivariate Behavioral Research*, vol. 53, no 1, pp. 15–35. <http://dx.doi.org/10.1080/00273171.2017.1379379>
- Mislevy R.J., Almond R.G., Lukas J.F. (2003) *A Brief Introduction to Evidence-Centered Design*. Center for the Study of Evaluation Report no 632. Los Angeles, CA: University of California. <http://dx.doi.org/10.1002/j.2333-8504.2003.tb01908.x>
- Rosé C.P., Ferschke O. (2016) Technology Support for Discussion Based Learning: From Computer Supported Collaborative Learning to the Future of Massive Open Online Courses. *International Journal of Artificial Intelligence in Education*, vol. 26, no 2, pp. 660–678. <http://dx.doi.org/10.1007/s40593-016-0107-y>
- Shute V.J., Ventura M. (2013) *Stealth Assessment: Measuring and Supporting Learning in Video Games*. Cambridge, MA: MIT. <https://doi.org/10.7551/mit-press%2F9589.001.0001>
- Snir J., Smith C., Grosslight L. (1993) Conceptually Enhanced Simulations: A Computer Tool for Science Teaching. *Journal of Science Education and Technology*, vol. 2, no 2, pp. 373–388. <https://doi.org/10.1007/BF00694526>
- Wise A.F., Chiu M.M. (2011) Analyzing Temporal Patterns of Knowledge Construction in a Role-Based Online Discussion. *International Journal of Computer-Supported Collaborative Learning*, vol. 6, no 3, pp. 445–470. <http://dx.doi.org/10.1007/s11412-011-9120-1>