

Д.А. Федерякин, И.Л. Углонова, М.А. Скрыбин

НОВЫЕ ИСТОЧНИКИ ИНФОРМАЦИИ В КОМПЬЮТЕРНОМ ТЕСТИРОВАНИИ

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 19-29-14110.

Проводится разграничение психометрической целевой и коллатеральной информации, получаемой во время компьютерного тестирования. Вводится классификация источников коллатеральной информации на основе того, какую информацию они описывают – респондентов, задания или их взаимодействие. Приводится обзор современных, преимущественно зарубежных, источников, где иллюстрируется использование коллатеральной информации.

Ключевые слова: коллатеральная информация; компьютерное тестирование; современная теория тестирования; модели с латентными переменными; психометрика; педагогические практики.

Введение

Повсеместная компьютеризация трудовой деятельности стремительно меняет жизнь современного человека. Сегодня компьютерные и вычислительные технологии внедряются в различные отрасли производственной деятельности – от автоматизированных станков до программ биржевых торгов.

Компьютеризация влияет в том числе и на индустрию образовательных измерений, и одним из главных проявлений этого является возрастающий интерес к характеристикам, которые сложно или практически невозможно измерить с помощью классических технологий оценки [1]. Так, симуляционные тренажеры применяются для подготовки и аттестации в некоторых сферах деятельности с высокими ставками еще с 1980-х гг. (например, пилотирование самолетов) [2]. Однако повсеместное использование подобных технологий в образовании наблюдается лишь в последние годы. В частности, межстрановое сравнительное исследование уровня компетентности 15-летних школьников PISA использует компьютерные симуляции для оценки навыков совместного решения проблем с 2015 г. [3]. Более того, сама цифровая грамотность становится объектом исследований, и для нее также создаются компьютеризированные инструменты измерения [4].

Компьютеризация тестирования не только расширяет область применения тестов, но и обладает рядом других преимуществ. В частности, нельзя проигнорировать такие преимущества компьютерного тестирования, как удобство доставки тестовых материалов респондентам, автоматизация начисления баллов за задания, предоставление автоматической обратной связи, автоматическая генерация сопоставимых заданий и пр. Отдельным направлением развития психометрики является компьютерное адаптивное тестирование, основанное на подборе следующего задания на основе предыдущих ответов респондента.

В силу перечисленных особенностей компьютерное тестирование обладает огромным потенциалом для изменения педагогических практик. Например, оно позволяет облегчить реализацию дифференцированного подхода в классно-урочной образовательной системе. Поскольку компьютерное тестирование позволяет радикально сократить цикл внедрения инстру-

мента от сбора данных до использования результатов, учителя имеют возможность быстро выделить относительно гомогенные группы внутри класса, оптимизируя свою работу с ними [5]. Более того, компьютерное тестирование позволяет проводить лонгитюдные измерения изменений способностей учеников по единой шкале (в случае применения соответствующего математического аппарата обработки результатов), что дает возможность учителям получать доказательную информацию о скорости развития учеников, выделять так называемый «отрицательный прогресс», позволяя индивидуально осуществлять педагогическое воздействие и взаимодействие, а также предсказывать развитие респондентов в будущем и превентивно диагностировать потенциальные образовательные дефициты на индивидуальном уровне [6].

Кроме того, компьютерное тестирование открыло доступ к новым источникам данных, которые не были доступны ранее. Так, одним из бурно развивающихся направлений психометрики является учет времени ответа на задания [7]. Обратим внимание, что не только время решения тестовых заданий, но и время освоения материала в фокусе внимания наук об образовании: методология интеллектуального анализа данных (data mining) позволила выявить специфику временных затрат студентов при освоении разных учебных дисциплин [8]. Однако предметом нашего исследования является использование дополнительной информации в первую очередь для образовательного тестирования. Например, в последние годы большое развитие получило направление анализа процесса ответов [9], а также применение data mining к таким данным [10].

Однако правомерное использование дополнительной информации, собираемой в компьютерном тестировании, становится вызовом для исследователей образовательного тестирования. На сегодняшний день как никогда активны дискуссии о валидности использования этих данных и их пригодности для проектирования педагогических воздействий и взаимодействий [11]. Поэтому в целом интерес исследователей к данной теме только растет [12].

Психометрические исследования в рассматриваемой области направлены на увеличение точности оценки за счет всевозможных дополнительных источников информации таким образом, чтобы не изменить интерпретацию оценки. Тем не менее на данный мо-

мент нет работ, которые содержат обзор методологических разработок из области психометрики, направленных на использование дополнительных источников информации.

Настоящая работа направлена на классификацию исследований по теме использования коллатеральной информации и особенностей ее анализа.

1. Целевая и коллатеральная информация в образовательном тестировании

Цель любого тестирования – получить валидные данные о таких характеристиках респондентов, которые явно не наблюдаются, но могут быть оценены через релевантное наблюдаемое поведение. Таким образом, при разработке тестов уделяется особое внимание тому, чтобы наблюдаемые действия респондентов отражали именно целевую характеристику (целевой конструкт), а не другие их черты. Например, если цель тестирования – проверить знания по математике, и формат заданий предполагает выбор одного ответа из нескольких, разработчик должен убедиться, что наблюдаемое поведение – выбор варианта ответа – отражает именно знание предмета. Другие характеристики респондентов (например, знания других предметных областей, уровень владения языком, скорость когнитивных процессов, понимание инструкции, гендер, культурная принадлежность и др.) не должны влиять на вероятность дать правильный ответ напрямую. Наблюдаемое поведение, которое отражает целевой конструкт, называется индикатором этого конструкта.

Такие требования сохраняются для любого инструмента измерения вне зависимости от его формата и содержания. Но специфика компьютерного тестирования позволяет использовать больший спектр наблюдаемого поведения [13]. В компьютерном тестировании индикатором может выступать не только выбор варианта ответа из предложенных, но и сложные действия

респондента в насыщенной медиасреде [14]. В некоторых ситуациях сам процесс решения задачи может рассматриваться как индикатор. Так, для сценарных заданий, направленных на измерение совместного решения проблем, важным источником информации является последовательность и виды действий, совершаемых респондентами [15]. При этом компьютерное тестирование не только расширяет набор индикаторов, но и позволяет фиксировать такое наблюдаемое поведение, которое не считается отражением целевого конструкта. Например, можно фиксировать обращение к инструкции, пропуск и возврат к заданию, точнее фиксировать время решения и т.п.

Дополнительная информация, которая не отражает целевую характеристику респондентов, не должна влиять на итоговую оценку и интерпретацию результатов. И именно этот аспект задает определение коллатеральной информации. В общем смысле коллатеральной информацией считается любая дополнительная информация, включение которой в статистическую модель обработки данных тестирования не меняет интерпретацию целевых тестовых баллов [16], но повышает точность измерения (т.е. понижает ошибку измерения и повышает надежность). При этом часто, выбирая, какую коллатеральную информацию анализировать, исследователи принимают во внимание легкость ее сбора из разных источников, что также делает этот аспект коллатеральной информации важным для её описания.

2. Классификация коллатеральной информации

В целом данные, полученные при измерениях в социальных науках, можно представить в виде схемы (рис. 1). На первом уровне данные можно разделить на две группы: данные, представляющие целевую информацию, и данные, представляющие коллатеральную информацию.

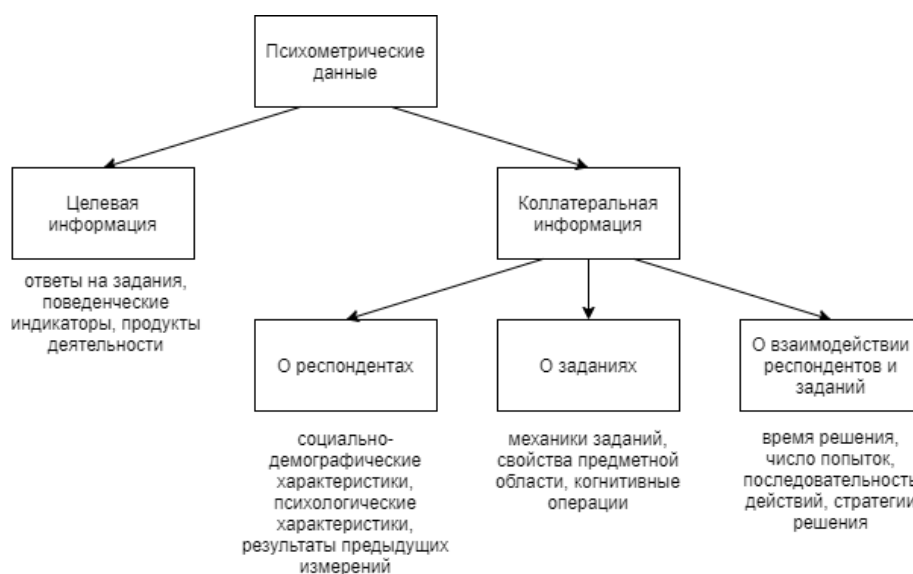


Рис. 1. Классификация психометрических данных

Хотя на данный момент не существует устоявшейся классификации источников коллатеральной информации, представляется возможным описать их систематически на основе того, какой тип данных они описывают. В качестве одного из подходов к классификации источников коллатеральной информации может использоваться способ классификации моделей современной теории тестирования [17]. Согласно этой систематизации выделяется три типа источников данных:

- 1) коллатеральная информация о респондентах;
- 2) коллатеральная информация о заданиях;
- 3) коллатеральная информация о взаимодействии респондентов и заданий.

В основном методы анализа данных тестирования связаны с интересом только к целевой информации. Интерес к построению моделей, использующих коллатеральную информацию, возник недавно и связан преимущественно с распространением компьютерного тестирования.

Для корректного учета коллатеральной информации требуются инновационные методы построения измерительных моделей, к которым относятся модели современной теории тестирования (Item Response Theory; IRT) [18]. Современная теория тестирования является фреймворком, который доминирует в области измерений в образовании и психологии, однако он также применяется и в других социальных науках – социологии, маркетинге, политологии и пр.

Далее мы более подробно представим особенности анализа данных для каждого из трех типов коллатеральной информации.

2.1. Коллатеральная информация о респондентах

Для учета этого типа коллатеральной информации используются одни из самых известных моделей IRT, например модели латентной регрессии [19]. Они позволяют выяснить, как наблюдаемые характеристики респондентов (например, социально-демографическая информация) связаны с целевым конструктом, и за счет этого уменьшают неопределенность итоговых оценок. Аналогичные модели известны в парадигме методов факторного анализа как MIMIC-модели [20]. Использование подобных техник анализа результатов измерений распространено в межстрановых сравнительных исследованиях образования для повышения надежности результатов [21].

Еще одним примером дополнительной информации о респонденте являются другие латентные характеристики респондента, которые также измерялись в процессе тестирования [22, 23]. Так, большинство крупномасштабных мониторинговых исследований измеряют одновременно несколько конструктов. Например, PISA кроме базовых областей языковой, математической и естественно-научной грамотности измеряет целую серию социологических и психологических характеристик учащихся [24]. В таком случае для анализа результатов применяются многомерные модели IRT, которые позволяют учесть информацию о распределении респондентов по нескольким латентным конструктам [25]. Таким образом, эти модели

позволяют конструктам «черпать» информацию друг из друга, повышая надежность за счет увеличения длины профиля ответов, который анализируется.

Отдельным аспектом использования многомерных моделей является измерение прогресса респондентов. Для измерения образовательного прогресса во времени требуется применение техник моделирования, которые позволяют расположить респондентов на одной шкале способности в разные точки времени, даже если содержание тестовых материалов было существенно изменено от одной волны измерения к другой [26]. Это возможно при использовании лонгитюдных моделей IRT, которые рассматривают каждую волну как отдельную размерность [27, 28]. Соответственно, эти модели также могут быть рассмотрены как специальные многомерные модели [29]. Кроме того, их применение ведет к повышению надежности измерений по каждой отдельной размерности благодаря использованию остальных размерностей как коллатеральной информации.

2.2. Коллатеральная информация о заданиях

Использование коллатеральной информации о заданиях описывается в первую очередь в контексте семейства моделей Latent Linear Test model (LLTM) [30–32]. Основной идеей этого семейства является положение о том, что некоторые задания могут иметь общие параметры, в отличие от исходного положения фреймворка IRT о том, что каждое задание имеет свой уникальный параметр. Как результат, это позволило исследователям изучать эффекты посторонних переменных на вероятность ответа. Первоначально LLTM использовалась для изучения того, как содержание заданий и когнитивные операции, необходимые для их решения, влияют на трудность заданий [30]. Такой способ применения этих моделей требует использования очень проработанной теоретической рамки: она должна содержать в себе информацию о детальной и корректной разметке теоретически ожидаемых когнитивных операций, используемых в процессе решения задания [33]. Эти модели применялись для изучения компонентов заданий, которые важны для их автоматической генерации [34], эффекта позиции задания в тесте [35, 36], влияния способа презентации информации на вероятность решения [37] и т.д. Подобное применение LLTM является ярким примером использования коллатеральной информации для увеличения точности оценки.

Важной вариацией LLTM для работы с коллатеральной информацией выступают многофасетные модели IRT (multifaceted IRT models) – модели для анализа деятельности реитеров, оценивающих задания открытого типа [38]. В данном подходе характеристики реитеров рассматриваются как специфическая характеристика процедуры измерения. В таких ситуациях фокус смещается с увеличения точности измерений на коррекцию индивидуальных искажений экспертных оценок и обеспечение справедливости измерений [39, 40]. В качестве коллатеральной информации выступают различные характеристики ситуации проверки открытых заданий (например, время

дня) или принадлежность респондентов и рейте-ров к каким-либо социально-демографическим группам. Показано, что эти характеристики могут влиять на уровень строгости экспертов во время проверки заданий или вызывать систематическую дискриминацию какой-либо группы [41, 42]. Даже несмотря на то, что такие источники информации, строго говоря, описывают не задания (по нашей классификации они относятся к типу информации о заданиях из-за параметризации моделей, которые описывают деятельность рейте-ров), они принадлежат к классу LLTM [19].

2.3. Коллатеральная информация о взаимодействии респондентов и заданий

Коллатеральная информация о взаимодействии респондентов и заданий обычно описывает процесс тестирования. Основные типы такой информации – это время тестирования [7], стратегии, которые используют респонденты при ответе на вопросы теста [43], последовательности взаимодействий респондентов с интерфейсом [44] и т.д.

Сбор подобной информации стал возможен на больших выборках именно с использованием компьютерного тестирования. Помимо этого, преимущество компьютерного тестирования заключается в возможности применять технологию айтрекинга – отслеживания направления взгляда респондента. Такая информация дает много возможностей исследователям для изучения поведения респондента при компьютерном тестировании [45]. Однако это инновационная область исследований, которая только начинает развиваться.

Время решения тестовых заданий как источник коллатеральной информации. Характеристики скорости выполнения заданий давно интересуют исследователей в социальных науках. Э. Торндайк и соавт. [46] еще в начале XX в. в исследованиях интеллекта обращали внимание на аспекты, связанные со скоростью выполнения заданий. Тем не менее только в последнее время это направление исследований приобретает массовую популярность, что связано с разработкой дружелюбного к пользователям программного обеспечения и статистических парадигм, позволяющих обобщить большое количество известных измерительных моделей [47].

В последние годы предлагается все больше теоретических фреймворков, направленных на осмысление различных способов использования информации о времени решения заданий [46–48]. При этом возможны случаи, когда включение времени ответа в психометрическую модель меняет интерпретацию результатов [49]. Так, Д. Molenaar и соавт. [50] обсуждают, что в зависимости от выбора измерительной модели для времени способность может интерпретироваться точно так же, как и в традиционных измерительных моделях без времени ответа, а может – как дисперсия способности, которая не описывается быстротой. Во втором случае основной фокус с измерения способности респондентов смещается на измерение их быстроты. Именно поэтому выбор корректной психометри-

ческой модели является критически важным для обработки коллатеральной информации.

Отдельного внимания заслуживают исследования по разработке компьютерных адаптивных тестов с учетом времени ответов на основе моделей процессов Маркова, развиваемые российскими специалистами [51, 52]. Однако интерпретация таких моделей остается относительно неизученной, и ее сравнение с устоявшейся традицией интерпретации моделей современной теории тестирования является перспективным направлением исследований.

Стратегии решения тестовых заданий как источник коллатеральной информации. Компьютеризация тестирования позволяет разрабатывать все более сложные системы оценки, которые построены на анализе действий, выполняемых респондентами. Исходя из этой информации выделяются различные стратегии выполнения заданий. При этом важно отметить, что иногда стратегии решения заданий могут анализироваться без использования данных о процессе тестирования. В частности, зачастую эксперты способны восстановить стратегию решения открытого задания, когда они видят ответ. Это позволяет выяснить, какая из стратегий делает вопрос более легким [53, 54]. Более подробная информация об используемой стратегии может быть получена из анализа времени, потраченного на тестовое задание или отдельные его блоки, совершенных действий при выполнении заданий [55, 56], нажатий клавиш, используемых для ввода ответа на вопрос [57], паттернов пропуска ответов [58] и пр.

Методология анализа процесса решения заданий. При анализе процесса решения заданий фокус смещается с классификации респондентов по стратегиям на выявление того, какие действия респондентов провоцируются какими-то другими. Для этого в основном используются методы машинного обучения, напрямую примененные к информации, полученной из журнала действий [59]. Наиболее распространенные методы – глубокий анализ процессов (process mining) и обнаружение последовательных шаблонов (sequence pattern mining). Эти методы позволяют строить модель процесса (или последовательности действий), что способствует лучшему пониманию типичных стратегий решения задач респондентами [60].

Полученная информация обычно направлена на выявление дополнительных свидетельств валидности для разработчиков инструментов и относительно редко используется в качестве коллатеральной информации при оценке результатов респондентов [58]. Однако внедрение этой информации в процесс шкалирования результатов измерений является крайне интересным и потенциально многообещающим направлением развития психометрики и наук о данных. Пока что основным направлением использования таких данных является анализ сопоставимости стратегий решения заданий людьми из разных социально-демографических или национальных групп [55].

Заключение

Одним из главных трендов развития индустрии тестирования является компьютеризация процесса те-

стирования, что сказывается и на других общественных институтах, в которых используется тестирование. Так, внедрение в образовательную практику компьютерных тестов позволяет существенно сократить цикл использования результатов, что способно существенно изменить педагогическую практику за счет быстрого доступа к высококачественной и достоверной информации в понятной форме.

Современные инструменты объективных измерений все чаще разрабатываются с самого начала как компьютерные: все больше психологических и образовательных конструктов измеряется на основе того, как их поведенческие индикаторы проявляются именно в компьютерной среде. Однако коррекции в определении конструктов – не единственный тренд в использовании компьютерных технологий в тестировании. Одно из самых популярных направлений психометрических исследований – использование дополнительных источников информации для увеличения точности измерений. Учет таких источников информации приводит к усложнению измерительных моделей, поскольку корректная измерительная модель должна не менять интерпретацию оценки способности при включении в нее дополнительных поведенческих индикаторов, но использовать информацию из них. При использовании таких измерительных моделей информацию из дополнительных источников называют коллатеральной информацией (например, время ответа на задание) в противовес традиционной целевой информации (правильность ответа на задание).

В данной статье предпринята попытка классификации источников коллатеральной информации на основании того типа данных, которые она описывает. Так, мы говорим о трех типах коллатеральной информации: о заданиях (различные характеристики заданий); о респондентах (такие характеристики респондентов, как социально-демографический статус или другие ненаблюдаемые характеристики, связанные с целевой); об их взаимодействии (информация о процессе решения задания – время ответа, стратегия решения, последовательность действий и т.п.). Достоинства и особенности использования коллатеральной информации зависят от конкретной ситуации измерения, в том числе от измеряемого конструкта, целевой аудитории инструмента, важности решения, принимаемого по результатам тестирования, и т.п. Однако основной аспект, определяющий коллатеральную информацию, связан именно с тем, как она используется в психометрической модели. При этом важно отметить, что сама по себе дополнительная информация может использоваться не только для увеличения точности измерений, но и в исследовательских целях – для обеспечения дополнительных свидетельств ва-

лидности выводов, которые делаются из результатов измерений.

Тем не менее потенциал применения компьютерных технологий в тестировании не ограничивается использованием только новых источников информации. Компьютерное тестирование позволяет оптимизировать процесс тестирования с помощью технологий компьютерного адаптивного тестирования (КАТ). В случае КАТ респонденту предъявляется не весь доступный банк заданий, а его часть, причем каждое следующее задание выбирается на основе предыдущих ответов. Это приводит к индивидуализации набора заданий и оптимизации их трудности для конкретного респондента при сохранении сопоставимости оценок.

Компьютерное тестирование может иметь вид сценарных инструментов измерения. Такие инструменты переворачивают само представление о тестировании за счет того, что больше напоминают компьютерные игры как в своем предъявлении респонденту, так и в процессе разработки. Эти инструменты измерения помещают респондента в контекст проблемной ситуации, для разрешения которой ему требуется проявить целевой измеряемый конструкт [61]. Подобные инструменты измерения хорошо подходят для измерения сложных конструктов, таких как критическое мышление и решение проблем [62] или навыки коммуникации и сотрудничества [63].

На фоне перечисленных возможностей компьютеризации тестирования такие его достоинства, как облегчение сбора и обработки информации, оптимизация работы с большими выборками, предоставление мгновенной обратной связи, выглядят тривиально.

Однако важно не забывать, что основным требованием для реализации компьютерного тестирования является использование компьютера. На сегодняшний день еще не все группы населения имеют дома доступ к компьютеру, что приводит к возможному неравенству при прохождении компьютерных тестов. Ожидается, что респонденты с доступом к компьютеру дома будут справляться с компьютерными инструментами измерения лучше, потому что они больше знакомы с контекстом ситуации тестирования. Таким образом, компьютерное тестирование открывает как новые горизонты в психометрике, так и новые трудности в их достижении.

Развитие компьютерного тестирования и применение коллатеральной информации повышают качество и достоверность результатов измерений психологических и образовательных конструктов респондентов. Это, в свою очередь, приводит к развитию доказательных педагогических практик и совершенствованию применяемых педагогических технологий.

ЛИТЕРАТУРА

1. Kyllonen P. New constructs, methods and directions for computer-based assessment // *The transition to computer-based assessment* / F. Scheuermann, J. Björnsson (eds.). Luxembourg : Office for Official Publications of the European Communities, 2009. P. 151–156. DOI: 10.2788/60083
2. Lee A.T. Flight simulation: virtual environments in aviation. London : Routledge, 2017. DOI: 10.4324/9781315255217
3. He Q., von Davier M., Greiff S., Steinhauer E.W., Borysewicz P.B. Collaborative Problem-Solving Measures in the Programme for International Student Assessment (PISA) // *Innovative assessment of collaboration* / A.A. von Davier, M. Zhu, P.C. Kyllonen (eds.). Cham : Springer, 2017. P. 95–111. DOI: 10.1007/978-3-319-33261-1_7

4. Авдеева С.М., Руднев М.Г., Васин Г.М., Тарасова К.В., Панова Д.М. Оценка информационно-коммуникационной компетентности учащихся: подходы, инструмент, валидность и надежность результатов // Вопросы образования. 2017. № 4. С. 104–132. DOI: 10.17323/1814-9545-2017-4-104-132
5. Унт И.Э. Индивидуализация и дифференциация обучения. М. : Педагогика, 1990. 190 с.
6. Vlug K.F.M. Because every pupil counts: the success of the pupil monitoring system in The Netherlands // Education and Information Technologies. 1997. Vol. 2, № 4. P. 287–306.
7. De Boeck P., Jeon M. An overview of models for response times and processes in cognitive tests // Frontiers in psychology. 2019. № 10. P. 1–11. DOI: 10.3389/fpsyg.2019.00102
8. Bylieva D., Lobatyuk V., Safonova A., Rubtsova A. Correlation between the Practical Aspect of the Course and the E-Learning Progress // Education Sciences. 2019. № 9 (3):167. P. 1–14. DOI: 10.3390/educsci9030167
9. Ulinskas M., Damaševičius R., Maskeliūnas R., Woźniak M. Recognition of human daytime fatigue using keystroke data // Procedia computer science. 2018. № 130. P. 947–952.
10. Qiao X., Jiao H. Data mining techniques in analyzing process data: a didactic // Frontiers in psychology. 2018. № 9. P. 1–11. DOI: 10.3389/fpsyg.2018.02231
11. DiCerbo K., Shute V., Kim Y.J. The future of assessment in technology rich environments: Psychometric considerations // Learning, design, and technology: An international compendium of theory, research, practice, and policy / M. Spector, B.B. Lockee, M.D. Childress (eds.). Switzerland AG : Springer Nature, 2017. P. 1–21. DOI: 10.1007/978-3-319-17727-4_66-1
12. Goldhammer F., Zehner F. What to make of and how to interpret process data // Measurement: Interdisciplinary Research and Perspectives. 2017. № 15 (3-4). P. 128–132. DOI: 10.1080/15366367.2017.1411651
13. Csapó B., Ainley J., Bennett R.E., Latour T., Law N. Technological issues for computer-based assessment // Assessment and teaching of 21st century skills / P. Griffin, B. McGaw, E. Care (eds.). Dordrecht : Springer, 2012. P. 143–230. DOI: 10.1007/978-3-319-65368-6
14. Bennett R.E., Goodman M., Hessinger J., Kahn H., Liggett J., Marshall G., Zack J. Using multimedia in large-scale computer-based testing programs // Computers in Human Behavior. 1999. № 15 (3-4). P. 283–294. DOI: 10.1016/S0747-5632(99)00024-2
15. Polyak S.T., von Davier A.A., Peterschmidt K. Computational psychometrics for the measurement of collaborative problem solving skills // Frontiers in psychology. 2017. № 8. P. 1–16. DOI: 10.3389/fpsyg.2017.02029
16. Wang W., Chen P., Cheng Y. Improving measurement precision of test batteries using multidimensional item response models // Psychological Methods. 2004. № 9 (1). P. 116–136. DOI: 10.1037/1082-989X.9.1.116
17. De Boeck P., Bakker M., Zwitser R., Nivard M., Hofman A., Tuerlinckx F., Partchev I. The estimation of item response models with the lmer function from the lme4 package in R // Journal of Statistical Software. 2011. № 39 (12). C. 1–28. DOI: 10.18637/jss.v039.i12
18. Handbook of item response theory / W.J. van der Linden (ed.). CRC Press. 2016. Vol. 1: Models. DOI: 10.1201/9781315374512
19. Explanatory item response models: a generalized linear and nonlinear approach / P. De Boeck, M. Wilson (eds.). New York : Springer Science & Business Media, 2004. DOI: 10.1007/978-1-4757-3990-9
20. Markus K.A., Borsboom D. Frontiers of test validity theory: measurement, causation, and meaning. New York : Routledge, 2013. DOI: 10.4324/9780203501207
21. Foy P., Yin L. Scaling the TIMSS 2015 Achievement Data // Methods and procedures in TIMSS 2015 / M.O. Martin, I.V. Mullis, M. Hooper (eds.); TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College and International Association for the Evaluation of Educational Achievement (IEA). Boston, 2016. P. 13.1–13.62.
22. Wilson M., Gochyyev P. Having your cake and eating it too: Multiple dimensions and a composite // Measurement. 2020. № 151. P. 107–247. DOI: 10.1016/j.measurement.2019.107247
23. Wu M., Tam H.P., Jen T.H. Multidimensional IRT Models // Educational measurement for applied researchers. Theory into practice. Singapore : Springer, 2016. P. 283–296. DOI: 10.1007/978-981-10-3302-5
24. OECD. PISA 2018 Results (Vol. I): What Students Know and Can Do. Organisation for Economic Co-operation and Development (OECD) Publishing, 2020. DOI: 10.1787/19963777
25. Bock R.D., Mislevy R.J. Adaptive EAP estimation of ability in a microcomputer environment // Applied psychological measurement. 1982. № 6(4). P. 431–444. DOI: 10.1177/014662168200600405
26. González J., Wiberg M. Applying test equating methods using R. New York : Springer, 2017. 196 p. DOI: 10.1007/978-3-319-51824-4
27. Andersen E.B. Estimating latent correlations between repeated testings // Psychometrika. 1985. № 50. P. 3–16. DOI: 10.1007/BF02294143
28. Embretson S.E. A multidimensional latent trait model for measuring learning and change // Psychometrika. 1991. № 56. P. 495–515. DOI: 10.1007/BF02294487
29. Wilson M., Zheng X., McGuire L. Formulating latent growth using an explanatory item response model approach // Journal of applied measurement. 2012. № 13 (1). P. 1–22.
30. Fischer G.H. Logistic latent trait models with linear constraints // Psychometrika. 1983. № 48 (1). P. 3–26. DOI: 10.1007/BF02314674
31. Fischer G.H., Formann A.K. Some applications of logistic latent trait models with linear constraints on the parameters // Applied Psychological Measurement. 1982. № 6 (4). P. 397–416. DOI: 10.1177/014662168200600403
32. De Boeck P. Random item IRT models // Psychometrika. 2008. № 73 (4). P. 533–559. DOI: 10.1007/s11336-008-9092-x
33. Baker F.B. EQUATE 2.0: a computer program for the characteristic curve method of IRT equating // Applied Psychological Measurement. 1993. № 17 (1). DOI: 10.1177/014662169301700105
34. Sonnleitner P. Using the LLTM to evaluate an item-generating system for reading comprehension // Psychology Science Quarterly. 2008. № 50 (3). P. 345–362.
35. Baghaei P., Kubinger K.D. Linear Logistic Test Modeling with R // Practical Assessment, Research & Evaluation. 2015. № 20. Article 1. DOI: 10.7275/8f33-hz58
36. Hahne J. Analyzing position effects within reasoning items using the LLTM for structurally incomplete data // Psychology Science Quarterly. 2008. № 50. P. 379–390.
37. Rolfes T., Roth J., Schnotz W. Effects of tables, bar charts, and graphs on solving function tasks // Journal Für Mathematik-Didaktik. 2018. № 39 (1). P. 97–125. DOI: 10.1007/s13138-017-0124-x
38. Eckes T. Introduction to Many-Facet Rasch Measurement. Analyzing and Evaluating Rater-Mediated Assessments. Berlin : Peter Lang GmbH, Internationaler Verlag der Wissenschaften, 2015. 241 p. DOI: 10.3726/978-3-653-04844-5
39. Myford C.M., Wolfe E.W. Detecting and measuring rater effects using many-facet Rasch measurement: Part I // Journal of Applied Measurement. 2003. № 4 (4). P. 386–422.
40. Myford C.M., Wolfe E.W. Detecting and measuring rater effects using many-facet Rasch measurement: Part II // Journal of Applied Measurement. 2004. № 5 (2). P. 189–227.
41. Dobria L. Longitudinal Rater Modeling with Splines : Doctoral dissertation. Chicago, IL : University of Illinois, 2011.
42. Schaefer E. Rater bias patterns in an EFL writing assessment // Language Testing. 2008. № 25 (4). P. 465–493. DOI: 10.1177/0265532208094273
43. Goldhammer F., Naumann J., Kessel Y. Assessing Individual Differences in Basic Computer Skills: Psychometric characteristics of an interactive performance measure // European Journal of Psychological Assessment. 2013. № 29. P. 263–275. DOI: 10.1027/1015-5759/a000153
44. He Q., von Davier M., Han Z. Exploring Process Data in Computer-based International Large-scale Assessments // Data analytics and psychometrics: informing assessment practices / H. Jiao, R. Lissitz, A. van Wie (eds.). Charlotte, NC : Information Age Publishing, 2018. P. 53–76.

45. Klein P., Küchemann S., Brückner S., Zlatkin-Troitschanskaia O., Kuhn J. Student understanding of graph slope and area under a curve: a replication study comparing first-year physics and economics students // *Physical Review Physics Education Research*. 2019. № 15. P. 1–17. DOI: 10.1103/PhysRevPhysEducRes.15.020116
46. Thorndike E.L., Bregman E.O., Cobb M.V., Woodyard E. The measurement of intelligence. New York : Teachers College Bureau of Publications, 1926. 616 p. DOI: 10.1007/978-3-319-93846-2_64
47. Molenaar D., Tuerlinckx F., van der Maas H.L. A generalized linear factor model approach to the hierarchical framework for responses and response times // *British Journal of Mathematical and Statistical Psychology*. 2015. № 68. P. 197–219. DOI: 10.1111/bmsp.12042
48. van der Linden W.J. A hierarchical framework for modeling speed and accuracy on test items // *Psychometrika*. 2007. № 72 (3). P. 287–308. DOI: 10.1016/j.procs.2018.04.094
49. Goldhammer F. Measuring ability, speed, or both? Challenges, psychometric solutions, and what can be gained from experimental control // *Measurement: interdisciplinary research and perspectives*. 2015. № 13 (3–4). P. 133–164. DOI: 10.1080/15366367.2015.1100020
50. Molenaar D., Tuerlinckx F., van der Maas H.L. A bivariate generalized linear item response theory modeling framework to the analysis of responses and response times // *Multivariate Behavioral Research*. 2015. № 50 (1). P. 56–74. DOI: 10.1080/00273171.2014.962684
51. Куравский Л.С., Артеменков С.Л., Юрьев Г.А., Григоренко Е.И. Новый подход к компьютеризированному адаптивному тестированию // *Экспериментальная психология*. 2017. № 10 (3). С. 33–45. DOI: 10.17759/exppsy.2017100303
52. Куравский Л.С., Юрьев Г.А., Ушаков Д.В., Юрьева Н.Е., Валуева Е.А., Лаптева, Е.М. Диагностика по тестовым траекториям: метод паттернов // *Экспериментальная психология*. 2018. № 11 (2). С. 77–94. DOI: 10.17759/exppsy.2018110206
53. Wilson M. The ordered partition model: an extension of the partial credit model // *Applied Psychological Measurement*. 1992. № 16 (4). P. 309–325. DOI: 10.1177/014662169201600401
54. Lee Y. How to Make an Assessment More Informative and Interpretable Using the Ordered Partition Model // *Journal of Curriculum and Evaluation*. 2011. № 14. P. 333–361. DOI: 10.29221/jce.2011.14.3.333
55. Liao D., He Q., Jiao H. Mapping Background Variables with Sequential Patterns in Problem-Solving Environments: an Investigation of United States Adults' Employment Status in PIAAC // *Frontiers in Psychology*. 2019. № 10. P. 1–32. DOI: 10.3389/fpsyg.2019.00646
56. Chen Y., Li X., Liu J., Ying Z. Statistical Analysis of Complex Problem-Solving Process Data: an Event History Analysis Approach // *Frontiers in Psychology*. 2019. № 10. P. 1–10. DOI: 10.3389/fpsyg.2019.00486
57. Guo H., Deane P.D., van Rijn P.W., Zhang M., Bennett R.E. Modeling Basic Writing Processes from Keystroke Logs // *Journal of Educational Measurement Summer*. 2018. № 55/2. P. 194–216. DOI: 10.1111/jedm.12172
58. He Q., von Davier M. Analyzing Process Data from Problem-Solving Items with N-Grams. Insights from a computer-based large-scale assessment // *Handbook of Research on Technology Tools for Real-World Skill Development, Information Science Reference / Y. Rosen, S. Ferrara, M. Mosharraf (eds.). Hershey, PA : IGI Global, 2016. P. 750–777. DOI: 10.4018/978-1-4666-9441-5.ch029*
59. Begicheva A.A., Lomazov I.A. Discovering high-level process models from event logs // *Моделирование и анализ информационных систем*. 2017. № 24 (2). С. 125–140. DOI: 10.18255/1818-1015-2017-2-125-140
60. Tóth K., Rölke H., Goldhammer F., Barkow I. Educational process mining: New possibilities for understanding students' problem-solving skills // *The Nature of Problem Solving: Using Research to Inspire 21st Century Learning / B. Csapó, J. Funke (eds.). Paris : OECD Publishing, 2017. P. 193–209. DOI: 10.1787/9789264273955-14-en*
61. Shute V.J. Stealth assessment in computer-based games to support learning // *Computer games and instruction*. 2011. № 55 (2). P. 503–524.
62. Shute V.J., Wang L., Greiff S., Zhao W., Moore G. Measuring problem solving skills via stealth assessment in an engaging video game // *Computers in Human Behavior*. 2016. № 63. P. 106–117. DOI: 10.1016/j.chb.2016.05.047
63. Stoeffler K., Rosen Y., Bolsinova M., von Davier A.A. Gamified performance assessment of collaborative problem solving skills // *Computers in Human Behavior*. 2020. № 104. P. 106–136.

Статья представлена научной редакцией «Педагогика» 10 декабря 2020 г.

New Sources of Information in Computerized Testing

Vestnik Tomskogo gosudarstvennogo universiteta – Tomsk State University Journal, 2021, 465, 179–187.

DOI: 10.17223/15617793/465/24

Denis A. Federiakin, Higher School of Economics (Moscow, Russian Federation). E-mail: dafederiakin@hse.ru

Irina L. Uglanova, Higher School of Economics (Moscow, Russian Federation). E-mail: iuglanova@hse.ru

Maksim A. Skryabin, Higher School of Economics (Moscow, Russian Federation). E-mail: maxim.skryabin@gmail.com

Keywords: collateral information; computerized testing; item response theory; models with latent variables; psychometrics.

Traditionally, psychometrics is concerned with theory-based information about human behavior – indicators of the targeted construct, like item responses, performance assessment products, etc. However, over the past forty years, advances in psychometric modeling and the development of information technologies allowed for the analysis of the so-called collateral information. This information is not theory-based and easy to collect in computerized testing. However, most importantly, collateral information is intended solely to increase the reliability of measurements preserving the construct's original interpretation. This article distinguishes between target and collateral information gathered during computerized testing. A carefully crafted measurement model is required to properly process collateral information along with target information. Social scientists usually choose Item Response Theory (IRT) models as such measurement models due to their clear interpretation, facilitating the discussion of the results of measurements in terms of social sciences. Since the choice of the correct IRT-model is crucial for preserving the original interpretation of the parameter estimates, it is possible to use the classification of such models to describe sources of collateral information systematically. This article introduces a classification of sources of collateral information based on the type of data they describe: (i) collateral information about respondents, (ii) collateral information about items, (iii) collateral information about interactions between respondents and items. The latter type of collateral information is particularly intriguing. Typically, it includes such types of data as item response times, response strategies, actions log data, gaze data, and other types of process data. Additionally to IRT modeling, examples of process mining and sequence pattern mining are also provided as examples of collateral information. The article illustrates the use of collateral information in educational psychometrics with a recent literature review. We describe cases where the measurement model's choice changes the interpretation of the IRT parameter estimates, which causes the breaking of the conditions defining collateral information. There is large- and small-scale educational and psychological research among cases. We also highlight the most illustrative cases of using collateral information in modern psychometric practice with regard to its source and the IRT-model used to process it. Moreover, we demonstrate that using the new sources of information in computerized testing contributes to developing evidence-based pedagogical practices and makes their application more manageable. The directions for future research in the area of collateral information in psychometrics are provided.

REFERENCES

1. Kyllonen, P. (2009) New constructs, methods and directions for computer-based assessment. In: Scheuermann, F. & Björnsson, J. (eds) *The transition to computer-based assessment*. Luxembourg: Office for Official Publications of the European Communities. pp. 151–156. DOI: 10.2788/60083
2. Lee, A.T (2017). *Flight simulation: virtual environments in aviation*. London: Routledge. DOI: 10.4324/9781315255217
3. He, Q. et al. (2017) Collaborative Problem-Solving Measures in the Programme for International Student Assessment (PISA). In: von Davier, A.A. Zhu, M. & Kyllonen, P.C. (eds) *Innovative assessment of collaboration*. Cham: Springer. pp. 95–111. DOI: 10.1007/978-3-319-33261-1_7
4. Avdeeva, S.M. et al. (2017), Assessing Information and Communication Technology Competence of Students: Approaches, Tools, Validity and Reliability of Results. *Voprosy obrazovaniya – Educational Studies*. 4. pp. 104–132. (In Russian). DOI: 10.17323/1814-9545-2017-4-104-132
5. Unt, I.E. (1990) *Individualizatsiya i differentsiatsiya obucheniya* [Individualization and differentiation of teaching]. Moscow: Pedagogika.
6. Vlug, K.F.M. (1997) Because every pupil counts: the success of the pupil monitoring system in the Netherlands. *Education and Information Technologies*. 2 (4). pp. 287–306.
7. De Boeck, P. & Jeon, M. (2019) An overview of models for response times and processes in cognitive tests. *Frontiers in Psychology*. 10. pp. 1–11. DOI: 10.3389/fpsyg.2019.00102
8. Bylieva, D. et al. (2019) Correlation between the Practical Aspect of the Course and the E-Learning Progress. *Education Sciences*. 9 (3):167. pp. 1–14. DOI: 10.3390/educsci9030167
9. Ulinskas, M. et al. (2018) Recognition of human daytime fatigue using keystroke data. *Procedia Computer Science*. 130. pp. 947–952.
10. Qiao, X. & Jiao, H. (2018) Data mining techniques in analyzing process data: a didactic. *Frontiers in Psychology*. 9. pp. 1–11. DOI: 10.3389/fpsyg.2018.02231
11. DiCerbo, K., Shute, V. & Kim, Y.J. (2017) The future of assessment in technology rich environments: Psychometric considerations. In: Spector, M. Lockee, B.B. & Childress, M.D. (eds) *Learning, design, and technology: An international compendium of theory, research, practice, and policy*. Switzerland AG: Springer Nature. pp. 1–21. DOI: 10.1007/978-3-319-17727-4_66-1
12. Goldhammer, F. & Zehner, F. (2017) What to make of and how to interpret process data. *Measurement: Interdisciplinary Research and Perspectives*. 15 (3-4). pp. 128–132. DOI: 10.1080/15366367.2017.1411651
13. Csapó, B. et al. (2012) Technological issues for computer-based assessment. In: Griffin, P., McGaw, B. & Care, E. (eds) *Assessment and teaching of 21st century skills*. Dordrecht: Springer. pp. 143–230. DOI: 10.1007/978-3-319-65368-6
14. Bennett, R.E. et al. (1999) Using multimedia in large-scale computer-based testing programs. *Computers in Human Behavior*. 15 (3-4). pp. 283–294. DOI: 10.1016/S0747-5632(99)00024-2
15. Polyak, S.T., von Davier, A.A. & Peterschmidt, K. (2017) Computational psychometrics for the measurement of collaborative problem solving skills. *Frontiers in Psychology*. 8. pp. 1–16. DOI: 10.3389/fpsyg.2017.02029
16. Wang, W., Chen, P. & Cheng, Y. (2004) Improving measurement precision of test batteries using multidimensional item response models. *Psychological Methods*. 9 (1). pp. 116–136. DOI: 10.1037/1082-989X.9.1.116
17. De Boeck, P. et al. (2011) The estimation of item response models with the lmer function from the lme4 package in R. *Journal of Statistical Software*. 39 (12). pp. 1–28. DOI: 10.18637/jss.v039.i12
18. van der Linden (ed.) (2016) *Handbook of item response theory*. Vol. 1: Models. CRC Press. DOI: 10.1201/9781315374512
19. Boeck, P. & Wilson, M. (eds) (2014) *Explanatory item response models: a generalized linear and nonlinear approach*. New York: Springer Science & Business Media. DOI: 10.1007/978-1-4757-3990-9
20. Markus, K.A. & Borsboom, D. (2013) *Frontiers of test validity theory: measurement, causation, and meaning*. New York: Routledge. DOI: 10.4324/9780203501207
21. Foy, P. & Yin, L. (2016) Scaling the TIMSS 2015 Achievement Data. In: Martin, M.O., Mullis, I.V. & Hooper, M. (eds) *Methods and procedures in TIMSS 2015*. TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College and International Association for the Evaluation of Educational Achievement (IEA). Boston. pp. 13.1–13.62.
22. Wilson, M. & Gochyyev, P. (2020) Having your cake and eating it too: Multiple dimensions and a composite. *Measurement*. 151. pp. 107–247. DOI: 10.1016/j.measurement.2019.107247
23. Wu, M., Tam, H.P. & Jen, T.H. (2016) Multidimensional IRT Models. In: *Educational Measurement for Applied Researchers. Theory into Practice*. Singapore: Springer. pp. 283–296. DOI: 10.1007/978-981-10-3302-5
24. OECD. (2020) *OECD. PISA 2018 Results (Vol. 1): What Students Know and Can Do*. Organisation for Economic Co-operation and Development (OECD) Publishing. DOI: 10.1787/19963777
25. Bock, R.D. & Mislevy, R.J. (1982) Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement*. 6(4). pp. 431–444. DOI: 10.1177/014662168200600405
26. González, J. & Wiberg, M. (2017) *Applying test equating methods using R*. New York: Springer. DOI: 10.1007/978-3-319-51824-4
27. Andersen, E.B. (1985) Estimating latent correlations between repeated testings. *Psychometrika*. 50. pp. 3–16. DOI: 10.1007/BF02294143
28. Embretson, S.E. (1991) A multidimensional latent trait model for measuring learning and change. *Psychometrika*. 56. pp. 495–515. DOI: 10.1007/BF02294487
29. Wilson, M., Zheng, X. & McGuire, L. (2012) Formulating latent growth using an explanatory item response model approach. *Journal of Applied Measurement*. 13 (1). pp. 1–22.
30. Fischer, G.H. (1983) Logistic latent trait models with linear constraints. *Psychometrika*. 48 (1). pp. 3–26. DOI: 10.1007/BF02314674
31. Fischer, G.H. & Formann, A.K. (1982) Some applications of logistic latent trait models with linear constraints on the parameters. *Applied Psychological Measurement*. 6 (4). pp. 397–416. DOI: 10.1177/014662168200600403
32. De Boeck, P. (2008) Random item IRT models. *Psychometrika*. 73 (4). pp. 533–559. DOI: 10.1007/s11336-008-9092-x
33. Baker, F.B. (1993) EQUATE 2.0: a computer program for the characteristic curve method of IRT equating. *Applied Psychological Measurement*. 17 (1). DOI: 10.1177/014662169301700105
34. Sonnleitner, P. (2008) Using the LLTM to evaluate an item-generating system for reading comprehension. *Psychology Science Quarterly*. 50 (3). pp. 345–362.
35. Baghaei, P. & Kubinger, K.D. (2015) Linear Logistic Test Modeling with R. *Practical Assessment, Research & Evaluation*. 20. Article 1. DOI: 10.7275/8f33-hz58
36. Hahne, J. (2008) Analyzing position effects within reasoning items using the LLTM for structurally incomplete data. *Psychology Science Quarterly*. 50. pp. 379–390.
37. Rolfes, T., Roth, J. & Schnotz, W. (2018) Effects of tables, bar charts, and graphs on solving function tasks. *Journal Für Mathematik-Didaktik*. 39 (1). pp. 97–125. DOI: 10.1007/s13138-017-0124-x
38. Eckes, T. (2015) *Introduction to Many-Facet Rasch Measurement. Analyzing and Evaluating Rater-Mediated Assessments*. Berlin: Peter Lang GmbH, Internationaler Verlag der Wissenschaften. DOI: 10.3726/978-3-653-04844-5
39. Myford, C.M. & Wolfe, E.W. (2003) Detecting and measuring rater effects using many-facet Rasch measurement: Part I. *Journal of Applied Measurement*. 4 (4). P. 386–422.

40. Myford, C.M. & Wolfe, E.W. (2004) Detecting and measuring rater effects using many-facet Rasch measurement: Part II. *Journal of Applied Measurement*. 5 (2). pp. 189–227.
41. Dobria, L. (2011) *Longitudinal Rater Modeling with Splines*: Doctoral dissertation. Chicago, IL: University of Illinois.
42. Schaefer, E. (2008) Rater bias patterns in an EFL writing assessment. *Language Testing*. 25 (4). pp. 465–493. DOI: 10.1177/0265532208094273
43. Goldhammer, F., Naumann, J. & Kessel, Y. (2013) Assessing Individual differences in Basic Computer Skills: Psychometric characteristics of an interactive performance measure. *European Journal of Psychological Assessment*. 29. pp. 263–275. DOI: 10.1027/1015-5759/a000153
44. He, Q., von Davier, M. & Han, Z. (2018) Exploring Process Data in Computer-based International Large-scale Assessments. In: Jiao, H. Lissitz, R. & van Wie, A. (eds) *Data analytics and psychometrics: informing assessment practices*. Charlotte, NC: Information Age Publishing. pp. 53–76.
45. Klein, P. et al. (2019) Student understanding of graph slope and area under a curve: a replication study comparing first-year physics and economics students. *Physical Review Physics Education Research*. 15. pp. 1–17. DOI: 10.1103/PhysRevPhysEducRes.15.020116
46. Thorndike, E.L. et al. (1926) *The measurement of intelligence*. New York: Teachers College Bureau of Publications. DOI: 10.1007/978-3-319-93846-2_64
47. Molenaar, D., Tuerlinckx, F. & van der Maas, H.L. (2015) A generalized linear factor model approach to the hierarchical framework for responses and response times. *British Journal of Mathematical and Statistical Psychology*. 68. pp. 197–219. DOI: 10.1111/bmsp.12042
48. van der Linden, W.J. (2007) A hierarchical framework for modeling speed and accuracy on test items. *Psychometrika*. 72 (3). pp. 287–308. DOI: 10.1016/j.procs.2018.04.094
49. Goldhammer, F. (2015) Measuring ability, speed, or both? Challenges, psychometric solutions, and what can be gained from experimental control. *Measurement: Interdisciplinary Research and Perspectives*. 13 (3–4). pp. 133–164. DOI: 10.1080/15366367.2015.1100020
50. Molenaar, D., Tuerlinckx, F. & van der Maas, H.L. (2015) A bivariate generalized linear item response theory modeling framework to the analysis of responses and response times. *Multivariate Behavioral Research*. 50 (1). pp. 56–74. DOI: 10.1080/00273171.2014.962684
51. Kuravskiy, L.S. et al. (2017) A new approach to computerized adaptive testing. *Ekspertimel'naya psikhologiya – Experimental Psychology* 10 (3). pp. 33–45. (In Russian). DOI: 10.17759/expsy.2017100303
52. Kuravskiy, L.S. et al. (2018) Diagnostics basing on testing paths: the method of patterns. *Ekspertimel'naya psikhologiya – Experimental Psychology*. 11 (2). pp. 77–94. (In Russian). DOI: 10.17759/expsy.2018110206
53. Wilson, M. (1992) The ordered partition model: an extension of the partial credit model. *Applied Psychological Measurement*. 16 (4). pp. 309–325. DOI: 10.1177/014662169201600401
54. Lee, Y. (2011) How to Make an Assessment More Informative and Interpretable Using the Ordered Partition Model. *Journal of Curriculum and Evaluation*. 14. pp. 333–361. DOI: 10.29221/jce.2011.14.3.333
55. Liao, D., He, Q. & Jiao, H. (2019) Mapping Background Variables with Sequential Patterns in Problem-Solving Environments: an Investigation of United States Adults' Employment Status in PIAAC. *Frontiers in Psychology*. 10. pp. 1–32. DOI: 10.3389/fpsyg.2019.00646
56. Chen, Y. et al. (2019) Statistical Analysis of Complex Problem-Solving Process Data: an Event History Analysis Approach. *Frontiers in Psychology*. 10. pp. 1–10. DOI: 10.3389/fpsyg.2019.00486
57. Guo, H. et al. (2018) Modeling Basic Writing Processes from Keystroke Logs. *Journal of Educational Measurement*. Summer. 2018. 55/2. pp. 194–216. DOI: 10.1111/jedm.12172
58. He, Q. & von Davier, M. (2016) Analyzing Process Data from Problem-Solving Items with N-Grams. Insights from a computer-based large-scale assessment. In: Rosen, Y., Ferrara, S. & Mosharraf, M. (eds) *Handbook of Research on Technology Tools for Real-World Skill Development, Information Science Reference*. Hershey, PA: IGI Global. pp. 750–777. DOI: 10.4018/978-1-4666-9441-5.ch029
59. Begicheva, A.A. & Lomazov, I.A. (2017) Discovering high-level process models from event logs. *Modelirovanie i analiz informatsionnykh sistem*. 24 (2). pp. 125–140. DOI: 10.18255/1818-1015-2017-2-125-140
60. Tóth, K. et al. (2017) Educational process mining: New possibilities for understanding students' problem-solving skills. In: Csapó, B. & Funke, J. (eds) *The Nature of Problem Solving: Using Research to Inspire 21st Century Learning*. Paris: OECD Publishing. pp. 193–209. DOI: 10.1787/9789264273955-14-en
61. Shute, V.J. (2011) Stealth assessment in computer-based games to support learning. *Computer Games and Instruction*. 55 (2). pp. 503–524.
62. Shute, V.J. et al. (2016) Measuring problem solving skills via stealth assessment in an engaging video game. *Computers in Human Behavior*. 63. pp. 106–117. DOI: 10.1016/j.chb.2016.05.047
63. Stoeffler, K. et al. (2020) Gamified performance assessment of collaborative problem solving skills. *Computers in Human Behavior*. 104. pp. 106–136.

Received: 10 December 2020