Проект «Использование контекстной информации и информации из цифровой среды оценивания при измерении индивидуального прогресса учащихся начальной школы с помощью цифровых технологий» носит сугубо методологический характер. Он направлен на разработку методов увеличения надежности психометрических измерений образовательных достижений в школе. В ходе этого проекта было получено 4 основных научных результата, которые заслуживают особого внимания.

Первым из них является первое в психометрике формальное определение Коллатеральной Информации (КИ). Мы сформулировали его, разведя понимание КИ с целевой информацией – той, которая имеет ключевое значение для обеспечения валидности результатов измерений. Говоря простыми словами, мы определили КИ как ту информацию, которая не является ответами на задания теста, но сопутствует им: время решения, попытки решения заданий, социально-демографические характеристики респондентов, характеристики заданий, и т.д. Эту информацию очень удобно собирать в компьютерном тестировании автоматически.

Обработка КИ является, тем не менее, сложным процессом, который способен исказить изначально предполагаемую интерпретацию способности респондентов. Поэтому здесь важным является выбор психометрической модели обработки результатов, которая способна включить в себя как целевые, теоретически обоснованные, индикаторы способности, так и саму КИ. Соответственно, из-за такого высокого значения психометрической модели при обработке КИ, представляется возможным классифицировать ее источники согласно тому, как классифицируют такие модели. Поэтому мы выделили следующие типы КИ:

1. О респондентах (социально-демографические данные, другие латентные характеристики, предыдущие замеры в моделировании образовательного прогресса);

2. О заданиях (различные атрибуты заданий – от формата ответа до способа презентации задания);

3. О взаимодействии респондентов и заданий (время решения, попытки решения, стратегии решения, процессные данные из журнала событий, и т.д.).

Далее, использую эту классификацию КИ как рамку проекта, мы показали, как КИ может быть использована для повышения валидности выводов о респондентах: на этапах от разработки теоретической рамки теста до анализа сценарных заданий или правомерности психометрических моделей, используемых для обработки результатов измерения.

В частности, в направлении разработки методов анализа КИ о респондентах, мы достигли одного из важнейших научных результатов проекта: мы показали, что существует возможность оценивать все корреляции между всеми факторами в бифакторных моделях (которые одновременно оценивают тестовый балл как по общему фактору – композиту, так и по его частям) современной теории тестирования. Это позволяет решить мировую психометрическую проблему, существовавшую с 1937 года в бифакторном моделировании. Традиционно эти модели требовали фиксации всех корреляций между всеми факторами-способностями к нулю, чтобы модель была идентифицирована. Это запрещало интерпретацию этих специфических факторов как компонентов композита, заставляя исследователей концентрироваться только на общем факторе. Тем не менее, наши результаты показывают, что при простых модификациях факторной структуры теста, все корреляции могут быть оценены, открывая новые возможности для интерпретации результатов из этих моделей и повышая точность оценки способности в них.

В направлении использования КИ о заданиях, мы разработали новую методологию измерения образовательного прогресса учеников. Традиционно, чтобы говорить об образовательном прогрессе и измерять его доказательно, требовалось не только применение специальных методов психометрического моделирования, но и применение специальных методов композиции теста – повторение отдельных заданий, чтобы достичь сопоставимости интерпретации способностей, которые измеряют шкалы. Без этого не представлялось возможным достичь того, чтобы один и тот же балл по разным шкалам соответствовал одному и тому же уровню когнитивной способности. Тем не менее, используя психометрические модели, описывающие не задания, а когнитивные операции, необходимые для их решения, мы смогли изменить эти жесткие требования, необходимые для измерения прогресса. Мы показали, что если тесты измеряют хотя бы частично совпадающий набор когнитивных способностей, то используя эту информацию о заданиях, шкалы тестов можно выровнять с помощью параметров трудностей этих когнитивных операций. Это ослабляет требования к повторяемости заданий в дизайне тестов, повышая, однако, требования к согласию модели с данными.

В направлении использования КИ о взаимодействии респондентов и заданий, мы описали возможности использования искусственных нейросетевых моделей для начисления баллов за задания с открытым ответом – точнее, за задания, в которых ученику нужно нарисовать рисунок на компьютере в тесте на креативность. В этом задании ученику нужно из предложенных компонентов составить фоторобот монстра. Начисление баллов в таком тесте, согласно теоретической рамке, происходит согласно критериям «Оригинальность» и «Детальность». Тем не менее, начисление баллов в открытых заданиях требует сложного дизайна, участия большого количества экспертов, начисляющих баллы, и специального психометрического моделирования, гарантирующего, что баллы респондентов не зависят от индивидуальных искажений экспертов. В данном направлении мы выделили два класса изображений по каждому критерию, используя автоматизацию обработки теоретически обоснованных критериев, а после этого обучили нейросетевую модель классифицировать изображения в эти классы. Точность классификации изображений составила порядка 90% по каждому из критериев. При этом, нейросетевые модели анализируют все изображение целиком, а не только теоретически обоснованные его элементы, что позволяет говорить об использовании всей информации о взаимодействии респондента и задания.

Кроме того, этот проект также позволил получить большой перечень частных научных результатов, касающихся отдельных инструментов измерения или направленных на изучение отдельных аспектов образовательного процесса. В общей сложности, по результатам проекта было подготовлено 30 выступлений на российских и международных (включая наиболее ведущие и престижные в области образовательных измерений) конференциях и семинарах, а также 18 публикаций в российских и зарубежных научных рецензируемых журналах и сборниках конференций, включающих РИНЦ, ВАК, а также Web of Science и Scopus (хотя часть из них находится на рецензировании в данный момент).