

согласно предложенному плану и теоретической модели. Поэтому причинно-обусловленные модели используются обычно при проведении так называемой «эвалюации процесса» один или несколько раз в период выполнения проекта или программы.

Разработка инструментария эвалюации осуществляется путем математического моделирования второго вида, которое развивается в рамках современной теории измерений, получившей название *Item Response Theory* (IRT). Математические модели этой теории используются для анализа результатов, обеспечивая получение надежных и валидных данных о воздействии управляемых переменных на качество результатов образования.

Обращаясь к моделям первого вида, можно выделить несколько категорий в зависимости от видов анализируемых переменных. В их число входят как управляемые переменные, так и переменные, не поддающиеся контролю и управлению.<sup>2</sup> В целом при установлении причинных связей между переменными принято выделять: прямые, косвенные, ложные, модерируемые и двунаправленные причинно-обусловленные связи, а также связи, не поддающиеся анализу.

Помимо анализа процесса эвалюации необходимо обеспечить деятельность, которая поможет выявить наблюдаемые результаты проекта, получить информацию для объяснения его результатов и получения ответа на вопрос, был ли достигнут прогресс в обучении и реализованы цели, заявленные при планировании нововведений по результатам выполнения образовательного проекта?

Для этого приходится обращаться к *моделям второго вида* и теории измерений, позволяющей перенести результаты проекта на количественную или качественную шкалу для их дальнейшего анализа и интерпретации. Перенос результатов эвалюации на шкалу позволяет абстрагироваться от самих объектов эвалюации, в роли которых обычно выступают обучаемые, переходя к анализу взаимного расположения точек на шкале, являющихся отображением оценок измеряемых переменных.

### Шкала оценок

Наиболее современный и эффективный подход к построению шкал переменных основан на параметрических моделях IRT, устанавливающих функциональную связь между наблюдаемыми результатами измерений и латентными (скрытыми, ненаблюдаемыми) переменными, являющимися объектами воздействия в программах и проектах. Обычно роль латентных переменных отводится параметрам обучаемых, в роли которых выступают оценки характеристик учебных достижений (знаний или умений), а в последнее время — оценки компетенций.

Поскольку речь идет о параметрах, а не об их приближенных оценках, то подразумевается величина, отображающая характеристики учебных достижений или уровень освоения

компетенций с нулевой ошибкой измерения («истинная оценка», как принято в теории педагогических измерений). IRT обладает высокоразвитым математическим аппаратом, позволяющим осуществить переход от наблюдаемых, приближенных оценок параметров интервального характера к точечным истинным оценкам параметров, в роли которых выступают оценки наибольшего правдоподобия, получаемые на основе построения специальной функции по методу максимального правдоподобия Фишера.

В целом теория IRT, нацеленная на оценивание латентных параметров испытуемых и заданий тестов на основе математико-статистических моделей измерения, является частью более общей теории *латентно-структурного анализа* (LSA), хотя каждое из этих направлений и имеет свои характерные особенности и свою сферу применения. В частности, в LSA оцениваемые значения параметров рассматриваются как некие дискретные точки на оси латентной переменной, в то время как в IRT распределения переменных предполагаются непрерывными. Область использования LSA — социально-психологические исследования, тогда как IRT применяется в основном при педагогических измерениях.

### Современная технология для оценочной деятельности

Необходимость использования IRT для сбора информации при проведении эвалюации достаточно очевидна. Эта теория обладает рядом неоспоримых преимуществ по сравнению с классической теорией тестов.

К наиболее значимым преимуществам IRT обычно относят:

- ♦ устойчивость и высокую объективность оценок параметра, характеризующего уровень учебных достижений или компетентность испытуемых<sup>3</sup>;
- ♦ устойчивость и высокую объективность оценок параметра трудности заданий измерителя, не зависящих, в свою очередь, от подготовленности репрезентативной выборки испытуемых, на которой выполнялось измерение;
- ♦ возможность измерения значений параметров испытуемых и трудности заданий теста в одной и той же шкале, имеющей свойства интервальной, что предполагает наличие одной единицы измерения для оценок обоих параметров и значительно повышает возможности интерпретации результатов измерений при эвалюации;
- ♦ наличие дифференцированных ошибок измерения для оценок параметров, позволяющих повысить надежность и валидность информации о результатах воздействия проекта на качество результатов обучения при проведении эвалюации.

*Первоначально в IRT вводится основное предположение о существовании некоей функциональной взаимосвязи между наблюдаемыми результатами из-*

<sup>2</sup> Выделение этих переменных и их категоризация являются задачей эвалюаторов — специалистов по эвалюации, существующих в профессиональном статусе в странах с высокоразвитыми системами образования.

<sup>3</sup> Эта устойчивость — важное преимущество IRT. Ее источником является относительная инвариантность оценок параметра испытуемых от трудности заданий измерителя.