

А.А. Малыгин¹, С.В. Щаницина²

СОВРЕМЕННАЯ ТЕОРИЯ ТЕСТОВ КАК ТЕОРЕТИЧЕСКАЯ ОСНОВА СОВРЕМЕННЫХ ПОДХОДОВ К ОЦЕНИВАНИЮ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ

¹Ивановский государственный химико-технологический университет

²Ивановский государственный университет

Email: a_malygin@mail.ru, swetlana145@mail.ru

В статье описаны возможности оценивания результатов обучения с позиций компетентностного подхода и требований теории педагогических измерений. Приводятся различные виды измерителей, обосновывается эффективность контроля методом адаптивного тестирования.

Ключевые слова: современная теория тестов, компетентностный подход, эвалюация, педагогические измерения, эффективность, адаптивное тестирование, результаты обучения.

В настоящее время существует две парадигмы оценивания качества образования – целевая и динамическая. При целевой парадигме результаты обучения выпускников учебных заведений трактуются как определенный уровень знаний и умений, умственного, физического и нравственного развития, которого достигли выпускники образовательного учреждения в соответствии с планируемыми целями обучения и воспитания. При этом под результатами обучения понимают компетенции, которые представляют собой комбинацию знания, понимания, умений и навыков, а также включают в себя установки, мотивацию, ценности. В рамках динамической парадигмы оценка качества учебных достижений строится на выявлении тех изменений в подготовке обучаемых, которые идентифицируются как улучшение знаний и умений, для чего необходимо накапливать данные о познавательной творческой активности и других учебных достижениях каждого учащегося или групп учащихся на протяжении всего периода обучения и фиксировать эти данные для анализа в виде дескриптивной статистики и различных документов на количественном и качественном уровнях оценивания. Динамические модели, достаточно широко распространенные за рубежом, пока в полной мере не реализуются в России, поэтому на данном этапе развития педагогической науки преобладает целевая парадигма, согласно которой

при оценивании определяется соответствие образовательных результатов нормативным требованиям (государственного стандарта, образовательной программы), ожиданиям учащихся, родителей и общества. Следовательно, результаты обучения – это формулировки того, что, как ожидается, будет знать, понимать и будет в состоянии продемонстрировать обучающийся после завершения процесса обучения.

Оценивание результатов обучения (эвалюация) включает целый «спектр теоретико-методологических и практических работ по систематическому исследованию ценности и позитивности качеств обучаемых, анализируемых на основе единой методологии, сочетания количественных и качественных методов для отслеживания характера и динамики изменений качества процесса и результатов образования» [1, 3]. Применение количественных методов в эвалюации требует использования преимущественно стандартизированных тестов, данные которых обрабатываются и анализируются с применением аппарата математики, статистики, мета-анализа, дисперсионного анализа, факторного анализа и т. д. Количественные показатели фиксируются в баллах или процентах. Качественные методы сбора информации основаны на привлечении экспертов, наблюдении, проведении анкетирования и интервью. Качественные оценки выражаются в суждениях типа «хорошо», «удовлетворитель-

но» и т. п. Часто каждому оценочному суждению приписывается определенный, заранее согласованный (установленный) балл, показатель (например, оценочному суждению «отлично» – балл 5). Очень важно при этом понимать, что оценка – это не число, получаемое в результате измерений и вычислений, а приписанное оценочному суждению значение, поэтому количественные манипуляции с оценочными суждениями (баллами) недопустимы. Чтобы избежать соблазна использовать оценки как числа, во многих странах мира они имеют буквенное обозначение, например А, В, С и т. д. Методологические основы современной теории эвалюации строятся на идее сочетания количественных и качественных данных, представленных в одной из трех шкал теории измерений (номинальная шкала, порядковая шкала, интервальная шкала).

Компетентностный подход к трактовке качества результатов обучения, отражающий стремление участников образовательного процесса к повышению его эффективности в условиях прагматизма без ущерба для фундаментальности образования, предусматривает оценивание ключевых, общепредметных, инструментальных и предметных компетенций [5]. Ключевые компетенции рассматриваются как способности индивидуума, характеризующие его взаимодействие с социумом, другими людьми и самим собой. Общепредметные компетенции включают способности по решению проблем на основе обобщенных междисциплинарных знаний и умений, основанные на переносе знаний из одной предметной области в другую и генерирующие на этой основе новую способность по применению знаний в практических ситуациях на основе обобщенной естественнонаучной и гуманитарной картин мира. Инструментальные компетенции можно трактовать как способности к применению навыков работы с информацией с целью ее интеграции и обработки, навыки организации и планирования труда, навыки по эффективной презентации результатов своего труда. Предметные компетенции рассматриваются как способности обучающихся привлекать для решения практических и жизненных проблем знания и умения, полученные в рамках конкретного предмета.

Ключевые компетенции носят надпредметный характер и необходимы выпускнику учебного заведения в любой области жизнедеятельности, однако, в силу латентного характера и опосредованной связи с содержанием образования они трудно поддаются измерению. Для оценивания общепредметных, предметных и инструментальных компетенций можно использовать комплексы различных измерителей. К числу таких измерителей, в первую очередь,

можно отнести тесты для оценки практической деятельности (Performance assessment) и портфолио. Кейс-измерители, состоящие из набора заданий (обычно не более 10 в случае использования мини-кейсов), в основном развиваются в высшем профессиональном образовании, поскольку они включают специальные проблемные задачи, в которых студенту предлагают осмыслить реальную профессиональную (или квазипрофессиональную) ситуацию. Наиболее перспективными с позиций компетентностного подхода представляются компетентностные тесты, которые носят межпредметный (комплексный) характер и требуют при оценке подготовленности обучающихся использования методов многомерного шкалирования и специальных методов интеграции оценок, полученных по отдельным шкалам.

В настоящий момент на фоне традиционных оценочных средств и тестов на количественном уровне педагогических измерений возникает необходимость перехода к наиболее эффективным, а по сути эффективным, методам контроля. Традиционное тестирование с фиксированным числом заданий и временем на их выполнение не всегда приводит к точным, высоко объективным (надежным), обоснованным (валидным) и сопоставимым результатам. В этом случае имеет смысл применять в контрольно-оценочной деятельности адаптивное тестирование, неразрывно связанное с компьютерными и информационными технологиями и адаптивным обучением.

Адаптивное тестирование основывается на современной теории конструирования и применения тестов (Items Response Theory – IRT) и концепции Л.С. Выготского, согласно которой развитие в процессе обучения достигается путем постоянного преодоления грани между сферой актуального развития и зоной ближайшего развития обучаемых при предъявлении им достаточно трудных заданий, побуждающих каждого учащегося работать на грани своих возможностей. Предъявление заданий в адаптивном тесте представляет собой итерационный процесс, сходимость которого к оптимальным оценкам латентных параметров испытуемых обеспечивается путем минимизации ошибок измерения, меняющихся после выполнения каждого очередного задания, что позволяет меньшим числом заданий быстрее и точнее оценить уровень подготовленности учащихся. С развитием информационных технологий стало возможным говорить и о компьютерном адаптивном тестировании, хотя в современном понимании оба словосочетания предполагают использование компьютеров и соответствующего программного обеспечения.

Адаптивное тестирование следует рассматривать как эффективный метод за счет предъ-

явления каждому испытуемому таких заданий, которые ему посильны для выполнения, что, в свою очередь, приводит к повышению точности измерения, минимизации числа заданий и времени на контроль, а также позволяет достигать его индивидуализации и дифференциации. При этом адаптивность не следует ошибочно понимать как пошаговый отбор заданий, основанный на дихотомической оценке результатов их выполнения (верно – неверно). В действительности такое понимание никак не связано с научными подходами к определению адаптивного тестирования. Достичь эффективности контроля методом адаптивного тестирования возможно теми заданиями, которые не оцениваются дихотомически и имеют устойчивые характеристики (трудность и дифференцирующая способность).

Реализация адаптивного тестирования на практике предполагает использование вероятностных моделей IRT [3; 6]. Отметим, что IRT вместе с классической теорией тестов (Classical Test Theory) и теорией генерализации данных педагогических измерений (Generalizability Theory) рассматриваются как общая методология педагогических измерений, которая носит междисциплинарный характер и является ядром научной организации контрольно-оценочных процессов в образовании [6]. Помимо выбора модели IRT, обоснованной нами ранее [8], необходимо иметь обширный банк заданий с различными характеристиками, обеспечивающих высокую точность и валидность измерений. В этой связи крайне важны как качество начального тестового материала, которое обеспечивается правильным отбором проверяемого содержания и умением разработчика корректно отобразить его в заданиях теста, так и этап обработки эмпирических результатов тестирования, получаемых на стадии апробации. Для осуществления адаптивного тестирования необходимы программно-инструментальные средства, позволяющие реализовать алгоритмы и стратегии адаптивного тестирования. Наиболее оптимальной стратегией отбора и предъявления заданий следует считать многошаговую стратегию, при которой на каждом шаге происходит переоценка уровня подготовленности испытуемого и подбор каждого последующего задания в соответствии с вычисленным уровнем. Тем самым создается «ситуация успеха» и достигается максимальный индивидуализирующий эффект, при котором каждому контролируемому учащемуся не предъявляются слишком трудные и непосильные для выполнения задания, равно как и весьма легкие. Подробная информация о реализации стратегий адаптивного тестирования представлена в исследовании А.А. Малыгина [7].

В целом адаптивное тестирование можно определить как научно обоснованный метод педагогического контроля, реализуемый с помощью автоматизированных процессов генерации, предъявления и оценки результатов выполнения адаптивных тестов. Он обеспечивает более высокую (по отношению к традиционному тестированию с фиксированными временем и длиной теста) эффективность педагогических измерений за счет оптимального подбора характеристик заданий (трудности и дифференцирующей способности), их количества, последовательности и скорости предъявления применительно к уровню подготовленности каждого испытуемого.

Очевидно, контрольно-оценочная деятельность не сводится только лишь к тестам. Однако, какая бы форма оценивания результатов обучения ни была использована, как для количественных оценок, так и для качественных суждений должен использоваться инструментарий, разработанный в соответствии с современными подходами в теории педагогических измерений. Поскольку компетенции характеризуют способности, которые проявляются после окончания обучения, то при их оценивании необходимо получать высоко прогностические показатели, которые с большой долей вероятности будут предсказывать успехи в дальнейшей деятельности. В идеале это должны быть такие учебные задания, процесс получения ответа на которые и само содержание ответа к заданию будет связано с будущей профессиональной деятельностью выпускников вузов, которую не все преподаватели зачастую хорошо себе представляют. На практике это выражается в составлении большого числа заданий повышенной трудности, требующих развернутых ответов, что ведёт к снижению у обучающегося мотивации к выполнению заданий и увеличению утомляемости, а в итоге – приводит к росту ошибки измерения. Вот почему так важно выбирать и индивидуальный темп, и дифференцированный подход в педагогическом контроле.

Информация, полученная при оценивании результатов обучения на основе бипарадигмальных измерений (количественных и качественных), позволяет принимать различные управленческие решения – от построения индивидуальных траекторий обучения и корректировки учебных программ до оценки эффективности инноваций и политики в образовании в целом.

Список использованной литературы

1. Гуськова М.В. Эвалюация в образовании: монография. М.: Инфра-М, 2012. 152 с.
2. Загвоздкин В.Г. Теория и практика применения стандартов в образовании. М.: Народное

- образование, НИИ школьных технологий, 2011. 344 с.
3. Звонников В.И. Измерения и качество образования: монография. М.: Логос, 2006. 312 с.
 4. Звонников В.И. Оценка качества результатов обучения при аттестации (компетентностный подход): учеб. пособие / В.И. Звонников, М.Б. Челышкова. М.: Логос, 2012. 280 с.
 5. Ефремова Н.Ф. Формирование и оценивание компетенций в образовании: монография. Ростов-на-Дону: Аркол, 2010. 386 с.
 6. Крокер Л. Введение в классическую и современную теорию тестов: учебник / Л. Крокер, Дж. Алгина. М.: Логос, 2010. 668 с.
 7. Малыгин А.А. Адаптивное тестирование в дистанционном обучении: монография. Иваново: Иван. гос. хим.-технол. ун-т, 2012. 138 с.
 8. Малыгин А.А. Теория педагогических измерений как научная основа для повышения эффективности контрольно-оценочных процессов в образовании // Известия высших учебных заведений. Серия «Гуманитарные науки». 2010. Том 1. Вып. 2. С. 154-158.
 9. Щаницина С. В. Анализ дискриминационной валидности заданий теста при оценке качества результатов обучения // Вестник государственного университета управления. 2009. № 7. С. 106–109.

Статья поступила 25.09.2012

Принята в печать 09.10.2012