

## Раздел 3

# ПРОБЛЕМЫ СОВРЕМЕННОЙ ПРАКТИКИ РАЗВИВАЮЩЕГОСЯ ОБРАЗОВАНИЯ

### 3.1. ПСИХОЛОГИЧЕСКИЕ ПРОБЛЕМЫ ОЦЕНКИ ОБРАЗОВАТЕЛЬНЫХ РЕЗУЛЬТАТОВ

*П.Г. Нежнов*

Оценка учебных достижений школьников относится к числу ключевых элементов организации образовательной практики. Данные об успешности освоения учащимися учебных программ представляют собой едва ли не главную контрольную точку взаимодействия всех участников этой практики, начиная с детей и заканчивая управленцами высшего звена. Очевидно, что, функционируя на стыке стольких интересов, оценка должна отвечать весьма высоким требованиям, а именно быть объективной, количественно точной и содержательно осмысленной. Кроме того, она должна предполагать информативность, отвечающую задачам, которые решаются на разных уровнях образовательной системы.

Эти задачи можно разделить на две группы разной направленности. Первая связана с управлением сетью образовательных учреждений. Вторая – с организацией и сопровождением собственно образовательных процессов. В обоих случаях оценка образовательных результатов должна отвечать всем указанным выше требованиям, однако в первом из них особую значимость имеют надежность и точность количественной характеристики результатов, открывающие возможность их сравнения и ранжирования, а во втором – научная основательность их категоризации (качественного определения), дающая ориентиры для педагогических стратегий. Противоположность этих акцентов и очевидная трудность их совмещения, по-видимому, и задают один из важных импульсов развития современной практики оценивания учебных достижений школьников.

Указанное противоречие проявляется в своеобразном расщеплении традиционной школьной оценки и специализированном развертывании ее количественного и содержательного аспектов. При этом одну крайность представляют традиционные педагогические тесты как средства измерения обученности детей, а другую – представления о качественных уровнях этой самой обученности и соответствующие контрольно-диагностические методики. Однако бескомпромиссное разведение по полюсам диалектической пары «качество – количество» сразу обнаруживает свои слабости. Так, абсолютизация измерения неизбежно приводит к редукции качественной характеристики объекта, упрощению сложной содержательной картины результатов образовательного процесса. В итоге суммарный балл, полученный учащимся за решение некоторого массива задач, плохо поддается углубленной интерпретации. Если же акцент ставится на качественной характеристике, приводящей в итоге к описанию индивидуальных профилей, то проблемой становятся целостное сравнение и ранжирование достижений учащихся. Таким образом, в системе оценивания качественная и количественная характеристики образовательных результатов предполагают друг друга, оказываются нерасторжимым единством.

Очевидная необходимость удержания какого-то баланса в этом единстве стимулировала попытки построения тестов, сочетающих надежность и точность измерения с возможностями качественной характеристики тестируемой способности. Примером такой тенденции может служить методология разработки тестового инструментария международных мониторингов TIMSS и PISA [Ковалева, 2002, 2004].

В отличие от суммативных педагогических тестов, инструментарий указанных мониторинговых исследований предполагает не только покрытие определенного предметного содержания, но и учет качественных различий в его присвоении. Последнее означает, что одной из опор при разработке теста выступает та или иная таксономия, устанавливающая гипотетическую последовательность качественных ступеней (или уровней) в освоении учащимися предметного содержания. Например, в тестовом пакете TIMSS по математике введено различие четырех видов деятельности, которые фактически образуют иерархию:

- знание фактов и процедур,
- применение понятий,

- решение стандартных задач,
- рассуждения.

Все эти виды деятельности имеют детальные описания, на основе которых разрабатывается соответствующий массив заданий, покрывающий определенную часть школьной программы по математике. Что касается качественной оценки достижений, то она получается следующим образом.

По результатам тестирования выборки, репрезентативной по отношению к генеральной совокупности, формальным образом, т. е. абсолютно условно, намечаются несколько уровней математической компетентности. На 1000-балльной шкале TIMSS этих уровней четыре, и они имеют следующие пороговые значения:

продвинутый уровень	– 625 баллов и более,
высокий уровень	– 550–624 балла,
средний уровень	– 475–549 баллов,
низкий уровень	– 400–474 балла.

В результате успешность обучения помимо метрической характеристики получает и уровневую, позволяющую разбить выборку учащихся на группы разной успешности. Кроме того, статистически выявляются задачи, которые посильны для детей каждой из групп, и каждому уровню освоения ставится в соответствие комплекс умений, т. е. дается вероятностная содержательная характеристика. Таким образом, идея качественного различия учебных результатов получает новое воплощение, причем подкрепленное данными эмпирического исследования.

Однако это решение проблемы напоминает хождение по кругу. Так, первый шаг состоял в том, чтобы множество конкретных математических умений представить в более крупной категориальной сетке, отражающей качественные ступени становления математической компетенции. А последний шаг этого пути возвращает нас к перечню предметных умений. Например, высокий уровень математической компетентности учащихся 4-х классов характеризуется следующим образом: «Учащиеся могут применить свои знания к решению поставленных проблем. Они могут решать многошаговые текстовые задачи на сложение, вычитание, умножение и деление; применять свое понимание поместного значения цифр в записи многозначного числа, а также несложных обыкновенных дробей для решения поставленных проблем; выделить числовые данные, которые характеризуют представленную в задаче ситуацию. Они

показывают понимание пространственных фигур, разбиения фигур на части и составления из них новых фигур, а также простейших движений на плоскости; демонстрируют умение производить различные измерения, могут интерпретировать и использовать данные, представленные в таблицах и на графиках, для решения поставленных проблем» [Ковалева, 2004, с. 23–24].

Сходным образом, через совокупность умений, не обобщенных в некоторой целостной характеристике, представлены и остальные уровни компетентности. Правда, теперь все математические умения разнесены по четырем группам как умения разного порядка. Однако категории исходной таксономии за этой иерархией уже не просматриваются, а какая-то новая категоризация не предлагается, да и вряд ли возможна. Последнее позволяет с большой вероятностью предполагать, что уровни эти «фантомны» [Бодалев, Столин, Аванесов, 2000].

Описанный подход реализован TIMSS и в исследовании естественнонаучной компетентности, для качественной характеристики которой предложены уже не четыре, а три вида деятельности. Как и в первом случае, результаты структурируются в соответствии с четырехуровневой шкалой, построенной на тех же условных основаниях. Аналогичный подход к оценке реализован и в мониторинге PISA [Ковалева, 2002].

Таким образом, анализ методологии современных мониторинговых исследований обнаруживает, что исходные таксономии, послужившие опорой при подборе тестовых задач, были в итоге успешно «преодолены» и замещены классификацией формально-конвенционального характера. То есть можно уверенно констатировать, что обсуждаемая модель оценивания отдает приоритет задаче измерения и не обеспечивает сбалансированной качественно-количественной характеристики результатов обучения.

Ограниченность этой методологии сегодня осознается специалистами в области оценки качества. Например, американский эксперт У. Попхем отмечает, что приоритетная ориентация на «измеряемость» учебных достижений провоцирует существенное сужение образовательных целей в пользу наиболее простых (легко измеряемых) знаний, умений и навыков. Как следствие, происходит дробление образовательной цели на множество мелких, что делает процесс управления образовательным процессом крайне нетехнологичным и даже невозможным [Pophem, 1987]. По сути, Попхем говорит про упомянутую нами редукцию содержательной сто-

роны объекта, в результате которой системное содержание трактуется как сумма его элементов, а процесс умственной реконструкции учащимся этого содержания, имеющий ступенчатый характер, рассматривается как постепенное накопление умений.

Описанный выше подход к построению тестового инструментария оценки учебных достижений, при котором измерение осуществляется за счет упрощения качественной характеристики объекта, сегодня является доминирующим. Можно сетовать на эту ситуацию, но она имеет свои причины. Одна из них состоит в том, что на данный момент органы управления образованием, располагающие основными финансовыми и организационными возможностями, пока что не ощущают острой необходимости учета качественного аспекта результатов обучения. Другая и, пожалуй, более глубокая причина состоит в размытости тех представлений о качественной стороне образовательных результатов, которые показаны сегодня в педагогическом поле. Ближайшим подтверждением последнего тезиса является положение дел в области построения таксономии педагогических целей – уровневой схемы, моделирующей принципиальные этапы научения и соответствующие им качественно различные варианты овладения учебным содержанием.

Главное, что характеризует сегодняшнюю ситуацию в разработке таксономии педагогических целей, – это большое количество предлагаемых версий. Напомним, что вслед за классической таксономией Б. Блума [Bloom, 1956] в российской педагогике появились версии В.П. Беспалько, В.В. Гузеева, И.Я. Лернера В.П. Симонова, М.Н. Скаткина и др. [Беспалько, 2002; Гузеев, 2001; Лернер, 1980; Симонов, 1999; Скаткин, Краевский, 1978]. Еще более выразительная ситуация сложилась в области международных мониторинговых исследований. Так, в рамках только одного проекта PISA фигурируют несколько таксономий, отличающихся не только характеристиками уровней присвоения некоторого содержания, но и количеством этих уровней: математическая компетенция представлена тремя категориями действий, а естественнонаучная компетенция и грамотность чтения – пятью [Ковалева, 2002].

Такое избыточное разнообразие вариантов решения проблемы является следствием относительной независимости этих разработок от развивающейся психологической теории, обращение к которой носит в основном фрагментарный и сугубо вспомогательный характер. В отсутствие же общезначимой психологической платформы, задающей рамку для критики и экспериментальной провер-

ки разных позиций, все уровневые модели оказываются формально равноправными и равноценными. В свою очередь, это объясняет отсутствие тенденции к сокращению их количества и ставит под сомнение возможность использования данных моделей в качестве опорных конструкторов в системе оценивания.

Сказанное подводит к выводу, что обеспечение сбалансированной качественно-количественной характеристики результатов обучения возможно лишь на базе теории, раскрывающей психологические закономерности процесса присвоения культурного содержания. Попыткой реализации такого подхода стала разработка оценочной модели, получившей название SAM (Student Achievement Monitoring) [Нежнов, Хасан, Эльконин, 2007; Нежнов, Карданова, Эльконин, 2011].

Основанием этой модели послужила гипотеза о внутренней структуре культурного способа действия и этапах его присвоения ребенком, предложенная Л.С. Выготским [Выготский, 1982] и развернутая в работах его коллег [Гальперин, 1998; Леонтьев, Запорожец, 1945; Эльконин, 1989] и многочисленных последователей [Давыдов, 1996; Магкаев, 1995; Максимов, 1987; Микулина, Савельева, 1987, 1997; Нежнов, 2005а, 2005б; Нежнов, Карданова, Рябинина, 2013; Нежнов, Медведев, 1988; Савельева, 1989; Фруммин, 2003; Эльконин, 1994]. Согласно этой гипотезе:

а) культурный способ действия держится на системе обобщений, закрепленных в знаковых структурах;

б) полная ориентировочная основа способа действия включает три категории обобщений – эмпирические, теоретические и смысловые;

в) при освоении способа роль реальной опоры действия сначала играют эмпирические обобщения, затем теоретические и, наконец, смысловые;

г) указанные три варианта ориентировки маркируют три качественно различных уровня освоения культурного способа действия – формальный, рефлексивный и функциональный.

Приведенная таксономия трех теоретически заданных уровней освоения способа действия и была положена в основу модели уровневого педагогического теста, которую можно представить следующим образом.

Каждому уровню ставятся в соответствие определенные индикаторы, т. е. типы тестовых задач. Затем для каждого раздела содержания учебного предмета разрабатываются задачные блоки,

включающие по три задачи 1, 2 и 3-го уровней. Поскольку в рамках блока задачи образуют естественную иерархию по трудности, сам блок играет роль детектора, определяющего уровень присвоения соответствующего раздела учебного материала. Из подобных блоков составляются варианты теста, что обуславливает его двойную направленность: с одной стороны, он работает как измерительный инструмент (набор из 45 задач, покрывающих основные разделы предмета, позволяет строить метрическую шкалу учебных результатов); с другой – как батарея диагностических методик (15 задачных блоков).

Совокупность блоков, покрывающая основные разделы программы, позволила получать трехмерные профили, дающие выразительную структурную картину освоения этой программы учащимися, но неудобную для проведения сравнения между индивидами, классами и т. д. (рис. 1).

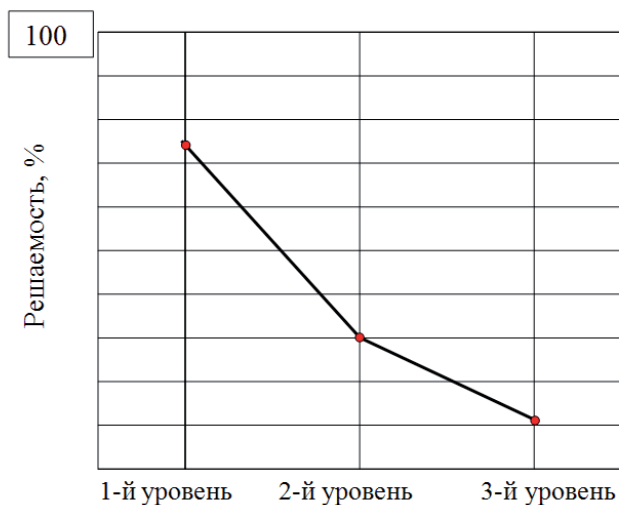


Рис. 1. Средний профиль по математике для выборки учащихся

Поэтому на следующем шаге для каждого предметного теста SAM был разработан ступенчатый вариант шкалы достижений, где каждой ступени приписана качественная характеристика, основанная на теоретически намеченных уровнях освоения способов действия. Всего выделено 3 ступени достижений, которые отвечают следующим критериям:

*первая ступень* – учащийся выполняет не менее 50% заданий 1-го уровня;

*вторая ступень* – учащийся выполняет не менее 50% заданий 2-го уровня;

*третья ступень* – учащийся выполняет не менее 50% заданий 3-го уровня.

В результате исследований для каждого предметного теста были определены пороговые значения, которые позволили наложить ступенчатую схему на метрическую шкалу (рис. 2).

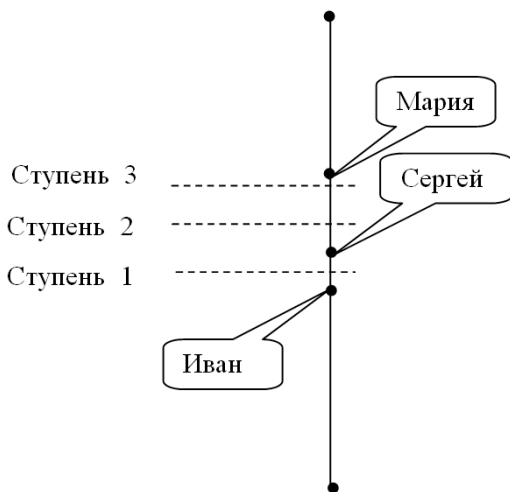


Рис. 2. Результаты тестирования на ступенчатой метрической шкале

Таким образом, задача разработки модели, обеспечивающей баланс между измерением и категоризацией результатов обучения, получила одно из возможных решений.

На данный момент созданы и апробированы два теста для оценки математической и языковой компетенций выпускников начальной школы [Нежнов, Карданова, Эльконин, 2011; Нежнов, Карданова, Рябинина, 2013]. Положительный результат исследований, посвященных проверке валидности тестов и соответствия их психометрических характеристик нормативным требованиям, свидетельствует о том, что намеченный нами подход к разработке системы оценки образовательных результатов имеет перспективу дальнейшего развития.