
П. Г. Нежнов, Е. Ю. Карданова, Б. Д. Эльконин

Статья поступила
в редакцию
в декабре 2010 г.

ОЦЕНКА РЕЗУЛЬТАТОВ ШКОЛЬНОГО ОБРАЗОВАНИЯ: СТРУКТУРНЫЙ ПОДХОД¹

Аннотация

Излагается новый подход к разработке тестового инструмента для измерения и квалификации достижений школьников в освоении предметных дисциплин. Кратко описано устройство теста SAM (School Achievements Monitoring), создаваемого группой российских специалистов в Центре международного сотрудничества по развитию образования (Center for International Cooperation in Education Development, CISED) Российской академии народного хозяйства и государственной службы при президенте Российской Федерации. Приведены примеры тестовых заданий и данные предварительной апробации теста.

Ключевые слова: общее образование, образовательные результаты, оценка качества, уровни присвоения содержания.

Оценка учебных достижений — открытая тема

Вопрос о том, на каких основаниях, по каким критериям и каким образом должна осуществляться оценка учебных достижений школьников, без преувеличения относится к числу ключевых для образовательной практики. Потенциально оценка — это эффективно мотивирующий и содержательно направляющий ориентир для ученика, но одновременно это и результирующая деятельности педагога, и важный показатель для школьной администрации. Наконец, данные об успеваемости — существенный элемент статистической отчетности школ, на который опираются не только органы управления образованием, но и специалисты, анализирующие тенденции развития образовательной системы в целом. Иными словами, оценка учебных достижений школьников — одна из важнейших опорных точек взаимодействия всех субъектов образовательной системы.

Очевидно, что для успешного выполнения столь ответственной функции оценка должна отвечать определенным требованиям:

¹ Выражаем глубокую признательность В. А. Болотову, А. Е. Волкову, Е. А. Ленской и И. Д. Фрумину за многократные содержательные обсуждения данной работы и советы по ее позиционированию. Мы также благодарны Б. Б. Соловьеву и возглавляемому им коллективу сотрудников ЦМСПО за обеспечение организационной стороны наших разработок. Благодарим за соавторство и сотрудничество наших коллег, участников данного проекта, вложивших в реализацию общего замысла свои силы и творческий потенциал.



быть объективной, прозрачной, достоверной и точной. Именно такая система обратной связи с ее инструментарием может в доходчивой материализованной форме обозначить поставленные перед школами образовательные цели и, фиксируя меру их достижения, дать основания для принятия согласованных педагогических и управленческих решений.

Традиционная для нашей школы пятибалльная система опиралась на крайне неопределенный набор критериев, оставлявший широкое поле для субъективизма в оценивании результатов обучения. Как следствие имела место проблема сравнимости (а точнее, несравнимости) данных об образовательных результатах в разных школах, да и в разных классах одной и той же школы. Все это, в свою очередь, вело к размыванию целевых ориентиров и ослаблению их консолидирующей роли в построении российского образовательного пространства.

В последнее время в отечественном школьном образовании предпринимаются настойчивые попытки модернизировать систему оценки достижений учащихся. Важным фактором интенсификации этих усилий послужили международные мониторинговые исследования, в которых Россия приняла активное участие [TIMSS, 2003; PIRLS, 2006; PISA, 2003 и др.]. Сравнительные данные, полученные при тестировании школьников по ряду предметов, пробудили интерес общественности к объективным методам оценки образовательных результатов и стимулировали, с одной стороны, широкое обсуждение итогов исследования, в частности в СМИ, а с другой — внимательное изучение и опробование доступных измерительных материалов в российских школах.

Все эти процессы имели важные культурно-педагогические последствия. Так, общественность вплотную познакомилась с зарубежной практикой мониторинга школьных достижений и получила возможность оценить уровень российского образования на фоне других стран [Каспржак, Митрофанов, Поливанова и др., 2005; Венгер, Калимуллина, Каспржак, 2006; Ковалева, 2006; Ковалева, 2004а, Ковалева, 2004б]. Кроме того, в сознании российских педагогов утвердилось представление о необходимости коррекции целевых ориентиров образования с учетом компетентностного подхода [Ковалева, Каспржак и др., 2004; Рыжаков, 1999; Фрумин, 2003; Шишов, 1999; Шишов, Агапов, 2001].

На данный момент благодаря усилиям органов управления образованием российская практика объективной оценки школьных достижений сделала весьма важные шаги в своем развитии: введены Единый государственный экзамен и Государственная итоговая аттестация². Нельзя сказать, что эта инновация происходит при всеобщем одобрении, но можно надеяться, что осознание архаичности традиционной школьной оценки на фоне современного

² ЕГЭ 2010. Официальный информационный портал единого государственного экзамена <http://www.ege.edu.ru/>

измерительного инструментария будет способствовать необратимым переменам в общественном мнении.

Определенные сдвиги происходят и в осмыслении компетентного подхода, которым заинтересовалось большое число специалистов. Появились публикации, примеривающие новую идеологию и терминологию к отечественному педагогическому опыту и традициям, началась работа по созданию пробного инструментария для измерения школьной успешности в компетентностном ключе [Ковалева, 2000; Ковалева, Каспржак и др., 2004; Фрумин, 2005].

Таким образом, можно констатировать, что современный зарубежный опыт в части нормирования и оценки учебных достижений школьников творчески ассимилируется отечественной системой образования. В то же время важным итогом проделанной работы стало понимание того, что общий подход к оценке школьных достижений, реализуемый в рамках международных мониторинговых исследований, не является универсальным и не в состоянии обеспечить решение всего спектра вопросов по управлению образовательной системой.

Проблема качества в контексте педагогических измерений

Существуют как минимум два класса организационно-управленческих задач, специфика которых должна учитываться при разработке средств оценивания. Так, с одной стороны, есть задачи, связанные с регулированием отношений между субъектами образовательной системы. И здесь первостепенную важность имеет надежное измерение учебных достижений, дающее возможность управленцам объективно сравнивать эффективность разных (региональных, муниципальных) образовательных систем, учреждений, программ, подходов, и на этой основе принимать мотивированные решения о поддержке тех или иных направлений, о распределении материальных ресурсов и т. д. С другой стороны, есть задачи, относящиеся к организации и управлению собственно образовательными процессами, в решении которых участвуют разработчики программ, методисты, администрация школ, учителя — т. е. лица, напрямую отвечающие за образовательные результаты, за повышение продуктивности работы школьных учреждений. И здесь особое значение приобретает качественная составляющая оценки, ее содержательный аспект, задающий основу для направления, коррекции и совершенствования образовательных процессов.

Конечно, оценка образовательных достижений всегда носит качественно-количественный характер, однако сбалансированное удержание этих двух аспектов, по-видимому, представляет определенные трудности. Во всяком случае, на данный момент основные разработки инструментария для оценки школьных достижений ведутся в русле методологии, которая, обеспечивая высокую надежность измерения образовательных результатов, одновременно сужает возможности получения информации об их качестве.



Иллюстрацией могут служить образцы инструментария для оценки результатов школьного образования, которые используются в международных мониторинговых исследованиях³.

Начнем с того, что тестовые материалы упомянутых исследований изначально создаются с опорой на ту или иную классификацию, а точнее, таксономию, определяющую виды деятельности, характерные для разных уровней присвоения учебного содержания. Иными словами, отправной точкой разработки теста выступает гипотетическое представление о качественных уровнях оцениваемой характеристики (обученности, предметной компетентности). Так, например, для проверки усвоения математики введено различие четырех видов деятельности, которые образуют некую иерархию: знание фактов и процедур; применение понятий; решение стандартных задач; рассуждения. Этим видам деятельности поставлена в соответствие типология заданий, на основе которой и разрабатывается соответствующий пакет тестовых материалов. Однако далее пакет проходит апробацию со статистической обработкой полученных данных и построением метрической шкалы, на которой исходная классификация видов деятельности уже не находит отражения. Вместо этого на шкале намечаются четыре уровня математической подготовленности (продвинутый, высокий, средний и низкий), отвечающие количественным критериям.

Правда, в заключение проводится статистический анализ решения задач группами детей разного уровня подготовленности и выявляются задачи, которые посильны для детей каждой из групп. Тем самым каждому уровню с некоторой вероятностью ставится в соответствие набор умений, т. е. придается вероятностная содержательная характеристика. А отнесенные к этой шкале результаты тестирования дают основание не только для ранжирования тестируемых индивидов, но и для деления их на четыре категории, которым приписано овладение определенными умениями [Ковалева, 2004а]. Однако уяснение качественной специфики этих категорий требует дополнительной концептуализации, возможность которой в данном случае выглядит проблематичной. Как следствие результаты, полученные с помощью указанного инструментария, предоставляют ограниченные возможности для их интерпретации и определения направлений совершенствования образовательной практики.

Описанный подход, при котором исходное представление о качественных уровнях освоения учебного содержания не удерживается в роли итоговой нормативной рамки для оценки результатов тестирования, сегодня безусловно доминирует. Можно сетовать на эту ситуацию, но она имеет свои причины. И главная состоит

³ PIRLS 2006 Assessment Framework and Specifications / I. V. S. Mullis, A. M. Kennedy, M. O. Martin, M. Sainsbury. IEA, Boston College ISC, 2006; PISA 2003 Assessment Framework — Mathematics, Reading, Science and Problem Solving Knowledge and Skills, OECD, 2003; TIMSS Assessment Frameworks and Specifications 2003. 2nd ed. IEA, Boston College ISC, 2003.

в критическом положении дел в области построения таксономий образовательных целей — иерархических классификаций, моделирующих этапы образовательного процесса и выступающих необходимой предпосылкой качественной оценки его результатов [Лернер, 1978. С. 7]. Так, по мнению известного эксперта А. Н. Майорова, сегодня отсутствуют отечественные разработки, которые обладали бы качествами, необходимыми для использования в практике создания тестового инструмента, т. е.

- «достаточно однозначно могли бы восприниматься педагогическим сообществом;
- давали взаимно однозначное соответствие конкретного тестового задания и уровня;
- перекрывали все возможное поле умений или навыков, или знаний, или способов деятельности» [Майоров, 2002. С. 67].

Сказанное в определенной мере справедливо и по отношению к зарубежным разработкам, включая классическую таксономию Б. Блума, которая, по признанию самих авторов, не получила исчерпывающего эмпирического подтверждения и, насколько нам известно, не используется в качестве опорного конструкта в образовательных тестах [Bloom, 1956].

Таким образом, для построения инструментария, обеспечивающего качественную оценку образовательных результатов, необходимо адекватное и надежное определение уровней освоения программного содержания, что возвращает нас к вопросу о педагогических таксономиях.

Первое, что характеризует разработки в этой области, — это большое количество вариантов, отчасти перекликающихся, но тем не менее конкурирующих друг с другом. Вслед за когнитивной классификацией Б. Блума было создано множество уровнейых схем, в той или иной степени отличающихся от своего прототипа. В частности, в российском образовании появились версии И. Я. Лернера, В. П. Симонова, М. Н. Скаткина и других авторов [Лернер, 1980; Лернер, Журавлев, 1994; Симонов, 1999; Скаткин, Краевский, 1978]. А в международном образовательном поле известность приобрели классификации, развиваемые в контексте мониторингов школьной успешности TIMSS, PIRLS, PISA.

Вторая характерная черта указанной области исследований — это отсутствие той теоретической основы, которая дала бы возможность рационального сопоставления и взаимной критики разных таксономий с возможностью выхода на принципиальные обобщения и выработку приемлемой базовой схемы. Традиционная ориентация педагогики на атомистические модели психического (см.: [Давыдов, 1972]), уводящие от целостной интерпретации образовательных процессов, делает существующие таксономии равноценными в плане теоретико-психологического обоснования. В этих условиях предмет дискуссии перемещается в поле педагогических интуиций, что объясняет устойчивость сложившегося многообразия.



На наш взгляд, заложить основы надежной оценки качества образовательного результата возможно лишь на пути развертывания и критического сопоставления объемлющих психолого-педагогических подходов, позволяющих, во-первых, свести в систему основные понятия, описывающие образовательный процесс (такие как «обучение», «развитие», «образовательный результат», «компетенция» и др.), а во-вторых, построить на этом основании уровневую модель процесса, допускающую эмпирическую проверку и пригодную для построения и совершенствования тестового инструмента. Мы полагаем, что к числу таких подходов правомерно отнести культурно-историческую теорию Л. С. Выготского.

Основные положения теории Л. С. Выготского, существенные в контексте заданной проблемы, сводятся к следующему.

- Взросление ребенка есть особый процесс, который психологически следует интерпретировать как «культурное развитие». Образовательный процесс в широком значении этого понятия есть организованная форма культурного развития. Обучение — необходимое условие такого развития: оно стимулирует этот процесс и задает его содержательное направление.
- Обучаясь, ребенок присваивает знаковые структуры (системы понятий, схем, правил и т. д.), в которых кристаллизованы культурно-нормативные образцы человеческого действия. Эти знаковые структуры выполняют функцию психологических средств (орудий), обеспечивающих ориентировку действий.
- Акт передачи ребенку знаковой структуры есть лишь начало образовательного процесса. Семена знаний закладывают «зону ближайшего развития» (ЗБР) и запускают процесс освоения этой зоны. Освоение выступает как процесс реконструкции и ассимиляции содержания присвоенного знания, а его итогом выступает обобщенный способ действия — способность разумно решать определенный класс задач.
- Реконструкция и ассимиляция культурных образцов действия есть тот внутренний спонтанный процесс (по Выготскому — процесс функционального развития), результаты которого (промежуточные и конечные) представляют первостепенный интерес для педагогов, и должны быть объектом мониторинга [Выготский, 1982. С. 188–202, 244–267, 305; Выготский, 1983. С. 78–85, 128, 147–151, 225, 292–316].
- Функциональное развитие — не монотонное продвижение от несовершенного умения к совершенному, а процесс, в котором можно зафиксировать три ступени (или уровня) становления культурного способа действия, связанные с возможными способами удержания знака как носителя этого способа:
 - 1) овладение «внешней структурой знака»; «внешнее», «ассоциативное», «магическое» его использование [Выготский, 1982. С. 115; Выготский, 1983. С. 157–161];

Образовательный процесс как культурное развитие

- 2) овладение содержанием знака, установление соответствующей «связи» [Выготский, 1983. С. 159];
- 3) «вращивание» знака в структуры мышления; «заключительная фаза интериоризации» [Выготский, 1982. С. 108–110, 115–117; Выготский, 1983. С. 157–163].

Благодаря работам учеников и последователей Л. С. Выготского [Гальперин, 1998; Давыдов, 1996, С. 228–237; Львовский, 1988; Магкаев, 1974, 1995; Нежнов, 2007; Нежнов, Медведев, 1988; Савельева, 1989; Эльконин, 1994; Эльконин, 1989], сегодня указанным уровням можно дать более конкретную интерпретацию, связав их с тремя основными аспектами ориентировочной основы способа действия, зафиксированного в знаке:

- первый уровень — репродуктивный, или формальный, — ориентация (опора) на внешние характеристики способа действия (алгоритм, правило, внешне заданный образец);
- второй уровень — рефлексивный, или содержательный, — ориентация на существенное отношение, лежащее в основе способа действия;
- третий уровень — функциональный — ориентация на поле возможностей способа действия.

В первом случае мера обобщенности способа действия минимальна и охватывает узкий спектр типовых ситуаций. Вторым вариантом дает принципиальную возможность решать весь класс задач, отвечающих данному способу. Наконец, при третьем варианте ориентировки, психологическая природа которого раскрыта лишь отчасти [Выготский, 1983. С. 463–464; Пиаже, 1969. С. 61–100; Запорожец, 2000. С. 518–659; Гальперин, 1998. С. 364, 388; Нежнов, 2007; Эльконин, 1989. С. 494–495], способ действия характеризуется функциональностью, т. е. возможностью свободного применения в разных контекстах. Отметим, что этот вариант владения способом действия близко подходит под известное определение такой категории ключевых компетенций, как *интерактивное использование социокультурных средств (орудий)*. К этой категории относятся прежде всего когнитивные орудия — язык, символы, текст, — обеспечивающие такие базисные компетенции, как владение устной и письменной речью, средствами математического действия⁴.

Названные три уровня освоения знака как носителя обобщенного способа действия образуют таксономию образовательных целей, в которой за уровнями усматриваются культурно-психологические структуры, фиксирующие опорные моменты в освоении полной ориентировочной основы этого способа.

Предложенная гипотетическая модель в настоящее время конкретизируется и проходит эмпирическую проверку в рамках многоступенчатого проекта по разработке тестового инструментария для оценки учебно-предметных компетенций учащихся начальной школы [Нежнов, Фрумин, Хасан, Эльконин, 2009].

⁴ Rychen D.S., Salganik L. (eds). Defining and selecting key competencies. Bern, 2001; The definition and selection of competence. Executive summary, 2005, www.deseco.admin.ch.



Принципы построения теста

Исходным моментом создания теста на основе предложенной модели является разработка матрицы предметного содержания, дающего компактное представление о совокупностях когнитивных средств действия (понятий, принципов, схем, алгоритмов и др.), составляющих содержание основных разделов некоторой учебной дисциплины. Мера усвоения этих совокупностей средств, обеспечивающих решение соответствующих классов задач, и подлежит тестированию.

Главная особенность инструмента состоит в том, что для каждого раздела учебного предмета (т.е. определенной совокупности когнитивных средств) разрабатываются задачи, соответствующие критериям названных выше трех уровней. Три задачи разного уровня, построенные на одном и том же предметном материале, образуют естественную иерархию по трудности и выступают как диагностический блок. Каждый такой блок выполняет функцию детектора, определяющего качественный уровень присвоения соответствующего раздела учебной программы (определяется по наиболее трудной задаче блока, которую решил учащийся). Разработка блоков опирается на систему индикаторов (типологию задач), соответствующую заявленным выше обобщенным критериям уровней освоения способа действия.

Так, индикатором освоения способа на первом уровне является выполнение задач, в которых связь условий с искомой схемой действия может быть установлена непосредственно. Этому требованию удовлетворяют задачи, в которых: а) описание условий с очевидностью указывает на их принадлежность к определенному классу с отработанной процедурой решения (типовые задачи); б) описание условий непосредственно подводит к правильной схеме действия.

Индикатором освоения способа действия на втором уровне является решение задач, в которых нельзя непосредственно применить типовые схемы действия, а необходимо выявить существенное отношение, определяющее принцип, на основе которого строится адекватная ситуации схема решения. Этому требованию удовлетворяют задачи: а) с «косвенной» формой условия, когда искомая последовательность операций отличается от той, которая «подсказывается» описанием условий задачи; б) абстрактного характера, исключающие манипулирование конкретными данными и прямое нащупывание решения; в) предполагающие прикидку решения на основе обобщенной ориентировки в условиях; г) с условиями, представленными в разных формах (например, одна часть условий дана текстом, а другая с помощью чертежа); д) предполагающие преобразование объекта или средства действия.

Индикатором освоения способа действия на третьем уровне является решение задач, в которых требуется оперирование полем возможностей в рамках некоторых контекстных условий. В частности, этому требованию соответствуют задачи, предполагающие обыгрывание существенного отношения с выявлением

веера возможных вариантов действия для выбора того, который удовлетворяет некоторому внешнему требованию. Например, математическая задача на поиск максимально возможного значения некоторого выражения.

Ниже представлены примеры блоков задач по математике и русскому языку.

Блок задач по математике

Сколько квадратных сантиметров составляет площадь прямоугольника $ALKM$?

Ответ: _____

Измерь площадь фигуры заданной единицей.

фигура единица измерения

Ответ: _____

Дети измеряли площадь одной и той же фигуры своими единицами площади.

Вася Коля Саша Таня

Результаты трех измерений приведены в правой колонке таблицы. Запиши имена этих троих детей в левую колонку таблицы.

Имена детей	Результаты измерений
	12
	24
	48

В приведенном блоке все три задачи относятся к измерению площадей. Первая предполагает действие по правилу — прямое наложение стандартной мерки на измеряемый объект (решаемость 63%). Вторая задача требует осознания неприменимости прямого правила действия, анализа условий, в которых представлено



отношение формы мерки и формы объекта, и выбора пути решения, например через мысленное преобразование фигуры, превращающее ситуацию измерения в типовую (32%). Третья задача предполагает относительно свободное владение отношением между числом, величиной и единицей, открывающее возможность отвлечься от факта незаданности измеряемой величины, проанализировать отношения между результатами измерения и на этой основе выявить соответствующий набор единиц измерения (15%).

Блок задач по русскому языку

Отметь предложение, в котором поставлены не все запятые.

1. Осенний листопад. Листья летят, скачут, плывут.
2. Мальчишки просидели в засаде до вечера, но ушли почти ни с чем.
3. Мама сидела за компьютером и писала какой-то доклад.
4. На массивном столе с зелёным сукном лежал ноутбук папки с бумагами, калькулятор.

Поставь знаки препинания в соответствии с описанными ситуациями.

1. Один ученик неверно расставил знаки препинания, и у него получилось, что море выбросило на берег предметы и людей.

Прибоем выбросило на берег корабль __ испанцев __ лодку __ рыбака __ катер.

2. Другой ученик верно расставил знаки препинания, и у него получилось, что море выбросило на берег только предметы.

Прибоем выбросило на берег корабль __ испанцев __ лодку __ рыбака __ катер.

Переформулируй предложение так, чтобы смысл его сохранился, но оно стало простым с однородными членами.

Вода непрестанно трудится, и время от времени края горных уступов обрушиваются.

Все три задачи этого блока связаны с расстановкой знаков препинания в простом предложении с однородными членами. В первой дан набор простых предложений, предполагающих прямое применение соответствующего правила (решаемость 60%). Вторая задача требует выделения однородных членов в условиях смысловой неоднозначности предложений, т. е. предполагает анализ и учет смысловых отношений (35%). Наконец, третья задача предполагает активное владение понятием простого предложения с однородными членами, позволяющее преобразовывать сложное высказывание в простое, сохраняя его содержание, т. е. рассчитана на прикидку вариантов и выбор адекватного (12%).

Тесты разрабатываются на школьном программном содержании, очерченном в действующем стандарте начального образования. При этом содержание представлено в виде относительно крупных единиц — совокупностей знаковых средств, — обеспечивающих ориентировку в некотором предметном разделе.

В качестве примера рассмотрим тест по математике, созданный группой специалистов под руководством С. Ф. Горбова. Предметное содержание этого теста разделено на пять разделов, или областей: числа и вычисления, измерение величин, закономерности, зависимости между величинами, элементы геометрии. В каждой предметной области выделены математические средства, овладение которыми обеспечивает ориентировку и построение соответствующих математических действий. Для каждой области специфицированы индикаторы уровней освоения содержания.

Применительно к разделу «Числа и вычисления» это выглядит следующим образом. Для определения освоенности содержания на первом уровне конструируются задачи, которые связаны с выполнением арифметического действия и некоторых стандартных приемов, используемых при вычислениях, таких как оценка результата, округление. Задачи второго уровня строятся не столько на самих вычислениях, сколько на выяснении и учете «строения» многозначного числа или выражения. К этому же уровню относятся и задачи, где надо определить программу вычислений. Наконец, для третьего уровня конструируются задачи, в которых абстрактное выражение подлежит конкретизации, удовлетворяющей некоторому условию. В качестве примера можно привести следующий блок.

Какое получится число, если 10472 разделить на 34?

Ответ: _____

Рассеянный Петя переписал из учебника пример на умножение двух чисел. Первый множитель он записал правильно: 7, а во втором множителе по ошибке переставил местами цифры. Из-за этого в ответе у него получилось число 147.

Какой ответ должен был получить Петя, если бы он правильно переписал пример?

Ответ: _____

Какой самый большой результат может получиться, если в выражении

$AB5 + BC2$ буквы заменить цифрами (разные буквы заменяются разными цифрами)?

Ответ: _____



Первая задача блока решается прямым применением правила (алгоритма) вычисления (решаемость 70%). Вторая требует анализа ошибочного арифметического действия (с учетом позиционного принципа) и построения программы по его коррекции (34%). Наконец, третья задача предполагает «обыгрывание» позиционного принципа для определения конкретных значений выражения, удовлетворяющих внешнему, контекстному условию — получению максимального значения выражения (15%).

Из подобных блоков и составляются варианты теста. Часть задач имеет открытую форму (с кратким ответом в виде числа, слова), а часть — закрытую (с выбором одного правильного ответа из четырех-пяти предложенных). Используются и другие формы тестовых заданий: на установление соответствия, требующие построений и т. д.

Таким образом, тест по любому предмету представляет собой набор блоков задач, покрывающих выделенные разделы предметного содержания. Он может быть рассмотрен как состоящий из трех субтестов: каждый субтест представляет собой набор задач одного уровня из различных разделов содержания. На основе обработки результатов тестирования отдельно по каждому уровню может быть построен профиль освоения программы по некоторому предмету для класса или большей выборки (рис. 1).

Результаты тестирования можно отнести и к интегральной шкале, что позволит получить количественную оценку уровня компетентности каждого участника в данной предметной области, а также ранжировать успешность выполнения теста отдельными учащимися, классами и т. д.

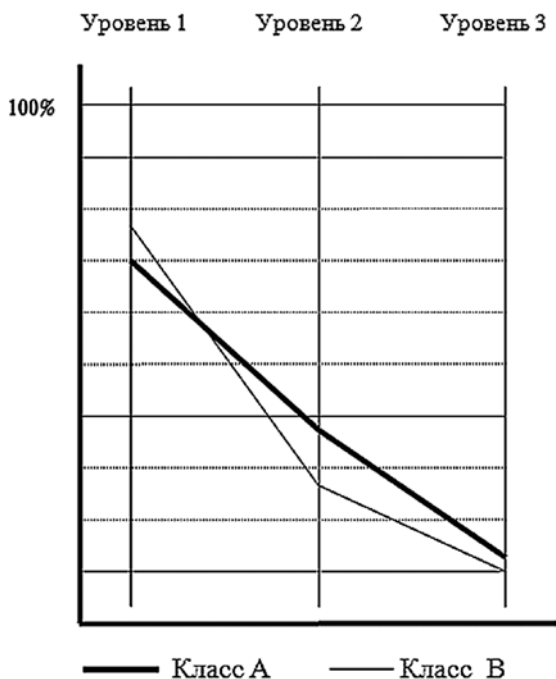


Рис. 1. Профили освоения математики двумя группами учащихся

Подобное структурное представление результатов тестирования расширяет возможности содержательной интерпретации тестовых данных и качественной характеристики оцениваемой компетенции. Так, приведенный на рис. 1 график, на котором представлены профили двух классов, получивших практически одинаковые интегральные баллы, позволяет констатировать, что класс А демонстрирует заметно лучшее понимание усвоенного материала, чем класс В. По первичным материалам можно установить и разделы содержания, которые каждой из групп не удалось освоить на втором уровне. Что касается шкалы третьего уровня, то она демонстрирует линию продвижения, которая, согласно возрастным представлениям, в начальной школе должна только намечаться.

Результаты апробации

Разумеется, для того чтобы тестовые баллы можно было правильно интерпретировать и на их основе делать обоснованные выводы, должна быть подтверждена валидность разрабатываемого инструментария и исследована его надежность. Решение этих задач требует апробации тестовых комплексов на достаточно большой и репрезентативной (по отношению к генеральной совокупности выпускников начальной школы) выборке испытуемых.

Цель апробационного тестирования — проверка функционирования заданий (анализ тестовых заданий), исследование системообразующих свойств теста, оценивание его надежности и обоснование валидности. Решение указанных задач предполагает специальный статистический анализ данных апробации, который, в соответствии с требованиями науки о педагогических измерениях, проводился как в рамках классической теории тестирования, так и на основе математического моделирования в рамках современной теории тестирования [Карданова, 2008; Крокер, Алгина, 2010]. Ниже приведены результаты апробации одного из вариантов теста по математике.

Тест включал 15 трехуровневых блоков, всего 45 заданий. Время выполнения теста — 90 минут. Все задания оценивались дихотомически: за правильное выполнение задания ученик получал 1 балл, за неправильное — 0 баллов.

Тест апробировался в Красноярском крае РФ. При составлении выборки учитывался тип учебного заведения (школа/гимназия) и его местоположение (город/село). Апробация проводилась в апреле 2010 г. на 418 учащихся выпускных классов начальной школы.

При классическом анализе основными характеристиками задания являются его трудность и дифференцирующая сила (дискриминативность). Трудность задания определяется как доля испытуемых выборки, выполнивших задание правильно. Дискриминативность задания характеризует его способность различать испытуемых с разным уровнем подготовки. Если на какое-либо задание теста отвечают почти все испытуемые, независимо от уровня их подготовки, то такое задание не дифференцирует



испытуемых, не отличает сильных от слабых. Аналогичная ситуация с заданием, на которое почти нет правильных ответов, и в этом случае ситуация усугубляется тем, что правильный ответ может быть получен слабым учеником случайно. Еще хуже, когда сильные испытуемые не отвечают на задание правильно, а слабые отвечают: про такие задания говорят, что они обладают противоположной дифференцирующей способностью, имеют отрицательную дискриминативность. В классической теории разработаны несколько индексов, характеризующих дискриминативность задания, из них в настоящей работе использовались показатель различительной способности задания (показатель дискриминативности) и коэффициент точечной бисериальной корреляции.

Приведем кратко основные результаты анализа заданий рассматриваемого варианта теста по математике. В табл. 1 представлены средние значения трудности и индексов дискриминативности заданий отдельно по уровням и по всему тесту. В целом задания первого уровня легче заданий второго уровня, которые, в свою очередь, легче заданий третьего уровня. Задания третьего уровня оказались намного более трудными по сравнению с заданиями других уровней для данной группы участников тестирования.

Индексы дискриминативности предпочтительны не меньше 0,3. Задания третьего уровня обладают меньшей различительной способностью по сравнению с заданиями первого и второго уровней, что частично может быть объяснено повышенной трудностью заданий третьего уровня для данной выборки испытуемых.

Таблица 1 Средние значения трудности и индексов дискриминативности заданий

	Трудность и диапазон, (%)	Показатель дискриминативности	Точечная бисериальная корреляция
Задания первого уровня	59 (43–75)	0,51	0,38
Задания второго уровня	20 (4–39)	0,33	0,33
Задания третьего уровня	6 (1–15)	0,10	0,18
Весь тест	28 (1–75)	0,31	0,30

Анализ каждого задания в отдельности с учетом статистической информации позволил наметить пути улучшения заданий. В целом все задания имеют удовлетворительные показатели качества. Более того, надежность анализируемого теста (в качестве коэффициента надежности рассматривался коэффициент Кронбаха альфа) достаточно высока — 0,85.

Из теоретической модели, на основе которой разрабатывается инструмент, логически следует, что задания трех уровней,

построенные на одном содержании, должны образовывать иерархию по степени решаемости. Так, если содержание освоено на втором уровне, то это предполагает, что оно освоено и на первом, что означает, что ученик, выполнивший задание второго уровня, должен был выполнить и задание первого уровня в этом блоке. В анализируемом тесте все блоки заданий удовлетворяют этому требованию: успешность решения заданий падает от первого уровня к третьему в каждом блоке, что служит аргументом в пользу валидности инструмента.

Дополнительно был проведен анализ дистракторов — неправильных вариантов ответов, присутствующих в заданиях с выбором ответа, — в закрытых заданиях. Все дистракторы признаны работающими, приблизительно одинаково привлекательными для незнающих учеников и не вводящими в заблуждение сильных учеников.

На втором этапе анализа данных апробации применялась современная теория тестирования, которая дает дополнительные возможности для анализа отдельных заданий и всего теста в целом, а также позволяет разработать модель оценивания участников тестирования. Анализ проводился как в рамках одномерного подхода (моделирование осуществлялось на основе однопараметрической модели Раша), так и с применением многомерного подхода (на основе многомерной модели — расширения модели Раша). Результаты анализа позволили сделать следующие выводы.

Тест может быть рассмотрен как существенно одномерный, т. е. нацеленный на измерение одной латентной переменной, в качестве которой выступает уровень математической компетентности учащихся. Поэтому по итогам тестирования может быть получена интегральная оценка математической компетентности участников тестирования. Надежность такой оценки равна 0,85, что является достаточно высоким показателем.

Многомерный (трехуровневый) подход даже при малом количестве заданий каждого уровня позволяет с достаточной надежностью оценить участников тестирования по первому и второму уровням (надежность измерений равна 0,83 и 0,8 соответственно). По третьему уровню надежность несколько ниже (0,65), что объясняется тем, что большинство заданий этого уровня оказалось очень трудным для данного контингента испытуемых. Эти затруднения обусловлены несформированностью функционального уровня освоения способа действия у большинства выпускников начальной школы. Для получения устойчивых и надежных характеристик заданий третьего уровня предполагается их апробировать на другой выборке испытуемых (на ступени основной школы).

Многомерная модель статистически лучше подходит для анализируемых данных, чем одномерная. Этот факт свидетельствует в пользу теоретических предположений авторов, что тест способен дифференцировать учащихся по трем уровням предметной компетентности.



В целом полученные статистические данные свидетельствуют о правомерности использования одномерной модели для получения интегрального показателя математической компетенции участников тестирования и многомерной трехуровневой модели — для оценки их математической компетентности по отдельным уровням.

Тестовые баллы участников тестирования, полученные с применением современной теории тестирования, находятся на метрической шкале, что позволяет сравнивать между собой результаты выполнения теста учащимися разных классов или любых других групп, а также использовать широкий спектр методов математической статистики для проведения исследований и проверки различных гипотез. Более того, метрический характер шкалы позволяет (при соблюдении некоторых дополнительных условий) выравнивать результаты тестирования по различным вариантам теста, и даже полученные в разное время.

Исследование показало, что предложенный теоретический подход в принципе поддается реализации и позволяет создать тестовый инструмент, отвечающий психометрическим требованиям. В ближайшей перспективе предполагается опробование создаваемого инструментария в контексте психолого-педагогических исследований и школьной практики для более детального определения способов интерпретации и использования результатов тестирования.

1. Выготский Л. С. Мышление и речь // Собр. соч. в 6 т. Т. 2. М.: Педагогика, 1982.
2. Выготский Л. С. История развития высших психических функций // Собр. соч. в 6 т. Т. 3. М.: Педагогика, 1983.
3. Гальперин П. Я. Психология как объективная наука. М.; Воронеж: Институт практической психологии, 1998.
4. Давыдов В. В. Виды обобщения в обучении. М.: Педагогика, 1972.
5. Давыдов В. В. Теория развивающего обучения. М.: ИНТОР, 1996.
6. ЕГЭ 2010. Официальный информационный портал единого государственного экзамена <http://www.ege.edu.ru/>
7. Запорожец А. В. Психология действия. М.; Воронеж: НПО «Модэк», 2000.
8. Карданова Е. Ю. Моделирование и параметризация тестов: основы теории и приложения. М.: ФГУ «Федеральный центр тестирования», 2008.
9. Каспржак А. Г., Митрофанов К. Г., Поливанова К. Н. и др. Новые требования к содержанию и методике обучения в российской школе в контексте результатов международного исследования PISA-2000. М.: Университетская книга, 2005.
10. Венгер А. Л., Калимуллина Г. Р., Каспржак А. Г. Российская школа: от PISA-2000 к PISA-2003. М.: Логос, 2006.
11. Ковалева Г. С. Основные тенденции изменения подходов к оценке результатов обучения в странах мира // На пути к 12-летней школе. М., 2000.

Литература

12. Ковалева Г. С. (ред.) Качество общего образования в российской школе. По результатам международных исследований. М.: Логос, 2006.
13. Ковалева Г. С., Каспржак А. Г. и др. Новый взгляд на грамотность. По результатам международного исследования PISA-2000. М.: Логос, 2004.
14. Ковалева Г. С. (рук. авт. колл.) Основные результаты международного исследования качества математического и естественнонаучного образования TIMSS-2003. М.: Центр оценки качества образования ИСМО РАО, НФПК, 2004а.
15. Ковалева Г. С. (рук. авт. колл.) Основные результаты международного исследования образовательных достижений учащихся PISA-2003. М.: Центр оценки качества образования ИСМО РАО, НФПК, 2004б.
16. Крокер Л., Алгина Д. Введение в классическую и современную теорию тестов. М., 2010.
17. Лернер И. Я. Качества знаний учащихся. Какими они должны быть. М., 1978.
18. Лернер И. Я. Процесс обучения и его закономерности. М., 1980.
19. Лернер И. Я., Журавлев И. К. (ред.). Современная дидактика: теория — практике. М., 1994.
20. Львовский В. А. Психологические требования к контролю и оценке знаний учащихся на основе решения системы задач: автореф. дис. ... канд. психол. наук. М., 1988.
21. Магкаев В. Х. Экспериментальное изучение планирующей функции мышления в младшем школьном возрасте // Вопросы психологии. 1974. № 5. С. 98–106.
22. Магкаев В. Х. Теоретические предпосылки построения метода исследования и объективно-нормативной диагностики развития основ рефлексивного мышления // Развитие основ рефлексивного мышления школьников в процессе учебной деятельности. / под ред. В. В. Давыдова, В. В. Рубцова. Новосибирск: Психологический институт РАО, 1995. С. 8–29.
23. Майоров А. Н. Теория и практика создания тестов для системы образования. М.: Интеллект-центр, 2002.
24. Нежнов П. Г. Опосредствование и спонтанность в модели «культурного развития» // Вестник Моск. ун-та. Сер. 14. Психология. 2007. № 1. С. 133–146.
25. Нежнов П. Г., Медведев А. М. Метод исследования содержательного анализа у школьников // Вестник Моск. ун-та. Сер. 14. Психология. 1988. № 2. С. 14–25.
26. Нежнов П. Г., Фрумин И. Д., Хасан Б. И., Эльконин Б. Д. (ред.). Диагностика учебной успешности в начальной школе. М.: ОИРО, 2009.
27. Пиаже Ж. Избранные психологические труды. М.: Просвещение, 1969.



28. Рыжаков М. В. Ключевые компетенции в стандарте: возможности реализации // Стандарты и мониторинг в образовании. 1999. № 4. С. 20–23.
29. Савельева О. В. Психологические критерии качества знаний младших школьников: автореф. дис. ... канд. психол. наук. М.: ИП РАО, 1989.
30. Симонов В. П. Диагностика степени обученности учащихся: учебно-справочное пособие. М.: МПА, 1999.
31. Скаткин М. Н., Краевский В. В. (ред.). Качество знаний учащихся и пути его совершенствования. М.: Педагогика, 1978.
32. Фрумин И. Д. Компетентностный подход как естественный этап обновления содержания образования // Педагогика развития: ключевые компетентности и их становление. Красноярск: ИЦ Инта естеств. и гуманит. наук, 2003.
33. Фрумин И. Д. и др. Оценка качества образования: обзор международных подходов и тенденций. М., 2005.
34. Шишов С. Е. Понятие компетенции в контексте качества образования // Стандарты и мониторинг в образовании. 1999. № 2. С. 30–35.
35. Шишов С. Е., Агапов И. Г. Компетентностный подход к образованию как необходимость // Мир образования — образование в мире. 2001. № 4. С. 8–19.
36. Эльконин Б. Д. Введение в психологию развития. М.: Тривола, 1994.
37. Эльконин Д. Б. Избранные психологические труды. М.: Педагогика, 1989.
38. Bloom B. (ed.) Taxonomy of educational objectives. Handbook 1: Cognitive domain. N. Y.: David McKay, 1956.
39. The definition and selection of competencie. Executive summary, 2005 www.deseco.admin.ch.
40. PIRLS 2006. Assessment framework and specifications / I. V. S. Mullis, A. M. Kennedy, M. O. Martin, M. Sainsbury. IEA, Boston College ISC, 2006.
41. PISA 2003. Assessment framework — mathematics, reading, science and problem solving knowledge and skills. OECD, 2003.
42. Rychen D. S., Salganik, L. (eds). Defining and selecting key competencies. Bern, 2001.
43. TIMSS Assessment frameworks and specifications 2003. 2nd ed. IEA, Boston College ISC, 2003.