



ЧТО ЯВЛЯЕТСЯ НАИБОЛЕЕ ВАЖНЫМ В СИСТЕМАХ ОЦЕНКИ ДОСТИЖЕНИЙ УЧАЩИХСЯ: ОСНОВНЫЕ ОРИЕНТИРЫ*

КЛАРК МАРГАРИТ

Старший специалист по образованию в Сети по развитию человеческого потенциала Всемирного банка, менеджер Трастового фонда российской программы содействия образованию в целях развития (READ TF), д-р пед. наук.

E-mail: mclarke2@worldbank.org

Вашингтон, округ Колумбия, США

АННОТАЦИЯ. Цель настоящей статьи – в обзорном порядке представить наиболее важные факторы, способствующие созданию более эффективной системы оценки достижений учащихся. Главное внимание уделяется системам оценки обучения и достижений учащихся на уровнях начального и среднего образования**. В качестве источников изложенных в статье принципов и методических рекомендаций автор использовал опыт различных стран, стандарты профессионального тестирования и сложившуюся на настоящее время научно-исследовательскую базу. Задача статьи — дать политическим руководителям стран, руководителям министерств образования, сотрудникам организаций, занимающихся проблемами развития, и другим заинтересованным лицам рамочную основу и набор ключевых показателей для диагностирования, обсуждения и выработки согласия в отношении путей создания надежной и устойчивой системы оценки достижений учащихся, которая будет способствовать повышению качества образования и выполнению задачи «обучение для всех».

КЛЮЧЕВЫЕ СЛОВА: системы оценки достижений учащихся, типы оценки, качество образования, обучение для всех.

* Материалы статьи впервые опубликованы в: Кларк М. Что является наиболее важным в системах оценки достижений учащихся: основные ориентиры // Серия «Системный подход к улучшению результатов образования (SABER) — оценка достижений учащихся». Вашингтон, Округ Колумбия: Всемирный Банк — 2012. — No 1.

** В статье не рассматриваются аспекты психологического и личностного тестирования. Она также не содержит прямого рассмотрения оценок успеваемости и достижений учащихся высших учебных заведений, хотя многие из поднятых здесь вопросов применимы и к этому уровню образования.

ВВЕДЕНИЕ

Оценка представляет собой процесс¹ сбора и анализа информации о том, какими знаниями, пониманиями и практическими навыками обладают учащиеся, которые позволили бы принять обоснованное решение о последующих шагах в их образовательном процессе. Используемые методы могут быть совсем простыми и строиться на модели «вопрос-ответ» (например, «Какой город является столицей Эфиопии?») или более сложными и строиться уже на моделях компьютерного адаптивного тестирования с использованием многогранных рейтинговых алгоритмов и прогрессий обучения.² Круг основанных на полученных результатах решений широк и включает в себя решения о методике разработки всесистемных программ по совершенствованию преподавания и обучения в школах, о выявлении следующих шагов преподавания в классе и даже о том, кого из подавших заявления следует принимать в университеты.

Система оценки — это комплекс методик, структур, практических действий и инструментов для генерирования и использования информации об обучении и достижениях учащихся. Эффективными считаются те системы оценки, которые обеспечивают получение информации, достаточной по качеству и количеству для удовлетворения потребностей заинтересованных лиц в информационной основе для принятия решений, направленных на повышение качества образования и улучшение результатов обучения учащихся (Равела и др., 2009 г.).³ Удовлетворение таких потребностей в информационной основе для принятия решений в мере, обеспечивающей поддержку ключевых политических и иных групп общества, будет вносить вклад в обеспечение более долгосрочной устойчивости и повышение эффективности этой системы оценки.

Правительства, международные организации и другие заинтересованные лица все более активно признают важность оценки для повышения результативности обучения учащихся и уровня их достижений, а следовательно, необходимость развития надежных систем оценки достижений учащихся (ГНО (IEG), 2006 г.; МакКинзи и Компания (McKinsey & Company), 2007 г.; ЮНЕСКО, 2007 г.). Это признание связано с накапливающимися доказательствами того, что многие преимущества образования (культурные, экономические и соци-

1 Понятие «оценка» может обозначать «конкретный инструмент», например тест.

2 Список компьютерных адаптивных программ размещен по следующему адресу: <http://www.psych.umn.edu/psylabs/catcentral/>.

3 Система оценки достижений учащихся обеспечивает целый ряд таких потребностей информационного характера, как создание информационной основы для обучения и преподавания, определение достигнутого прогресса, измерение достигнутых результатов, а также частичное повышение ответственности. Реализация этих информационных задач и принятие решений на основе такой информации в конечном итоге должны привести к повышению качества и уровней обучения в системе образования.

альные) становятся преумножающимся достоянием общества, только когда происходит обучение (ОЭСР, 2010 г.). Например, прирост одного нормативного отклонения по рейтингу международных оценок уровней достижений в чтении и математике связывается с 2-процентным ежегодным приростом ВВП на душу населения (Ханушек и Вёссманн, 2007 г., 2009 г.).

Некоторые утверждают, что оценки в целом и крупномасштабные их варианты, в частности, слишком дорогостоящи. На самом деле всё как раз наоборот: *доказано, что тестирование является одной из самых недорогих инноваций в реформировании образования* и обычно обходится значительно дешевле, чем повышение зарплат учителям или сокращение числа учащихся в классах. Хоксби (2002 г.) пришел к выводу, что затраты даже на самые дорогостоящие программы улучшения отчетности на основе тестирования, проводимые на уровне штатов в США, составляют менее 0,25 процента от расходов на школьника. Эту точку зрения подтверждает Волфф (2007 г.), заявляя, что ни в одной из рассмотренных им латиноамериканских стран на тестирование не было потрачено более 0,3 процента государственного бюджета на образование на проверяемом уровне (начальном или среднем). Хотя такие цифры по эффективности затрат и кажутся привлекательными, не стоит позволять им затмевать собою другие важные факторы: например, обеспечение равенства и достижение социальных целей, которые необходимо учитывать при принятии любого решения о проведении или непроведении конкретной программы оценивания.

За последние 20 лет многие страны начали реализовывать оценочные мероприятия или наращивать существующие системы оценки (ЮНЕСКО, 2007 г.). Кроме того, наблюдался огромный рост числа стран, участвующих в таких международных программах сравнительной оценки, как «Тенденции в изучении математики и естествознания в разных странах» (TIMSS) и «Программа оценки достижений учащихся в разных странах» (PISA).⁴ Неправительственные организации также все активнее начали обращаться к оценкам достижений учащихся с целью привлечения внимания общественности к низким уровням достижений и создания импульса к изменениям.

Несмотря на эту заинтересованность в оценках достижений учащихся очень немногие страны располагают методиками, структурами, наборами практических знаний и инструментов, которые все вместе составляли бы эффективные системы оценки. *Это особенно актуально для стран с низким уровнем доходов, которые могут только выиграть, прилагая систематические усилия для измерения учебных результатов.* Некоторые из этих стран экспериментировали с крупномасштабными или иными стандартизирован-

⁴ Например, количество стран, участвующих в программе PISA, увеличилось с 43 в 2000 г. до 66 в 2007 г. Сравнительно небольшое число развивающихся стран приняло участие в международных оценках достижений учащихся. Эти страны по результатам постоянно занимали последние места в таблице участников, ограничивая тем самым объем информации, который они могли извлечь из получаемых данных для формирования лучшего понимания и совершенствования собственных систем образования.

ными оценками обучения учащихся и уровней их достижений, но зачастую это были лишь одноразовые мероприятия, не ставшие частью образовательной стратегии и не выдержавшие испытания временем. Ключевым признаком, отличающим одноразовые оценки от устойчивой системы оценивания, является то, что первые дают лишь одномоментный срез (единичный кадр) уровней достижений учащихся, тогда как последняя предоставляет возможность отслеживать тенденции в уровнях обучения и достижений на протяжении времени (серия последовательных фотографий) и способствует лучшему пониманию относительного влияния различных вновь вводимых факторов и образовательных практик на изменения в этих тенденциях. Одноразовые оценки могут оказывать шоковое воздействие и инициировать дискуссии о качестве образования, и это может стать краткосрочной стратегией постановки проблемы обучения на повестку дня.⁵ Однако в конечном итоге правительства вынуждены сталкиваться с проблематичной, но необходимой задачей формирования и запуска в работу систем, при помощи которых можно будет осуществлять регулярный мониторинг успеваемости и достижений учащихся и оказывать поддержку в повышении их уровней. И это — единственный способ заставить программы оценки работать с полной отдачей.

ТЕОРИЯ И ФАКТЫ ОБ ОЦЕНКЕ ДОСТИЖЕНИЙ УЧАЩИХСЯ

Основополагающий замысел проведения исследования оценки достижений учащихся заключается в том, что правильный выбор оценочных мероприятий и правильное использование полученных за счет этих мероприятий данных способствуют улучшению результатов, будь то более высокий уровень обучения или более качественные политические решения (например, Хойберт и Хаузер, 1999 г.).⁶ «Правильность» во многом предопределяется набором теоретических и технических указаний для

5 Одна из наиболее популярных инициатив такого рода известна как «EGRA» (оценка навыков чтения в начальных классах). Согласно информации, опубликованной на вебсайте USAID (Агентства США по международному развитию) (<https://www.eddataglobal.org/>), «оценка навыков чтения в начальных классах (EGRA) является устным оценочным мероприятием, предназначенным для измерения самых базовых основных навыков получения грамотности в начальных классах... и для предоставления министерствам и донорам необходимой информации о потребностях системы в совершенствовании преподавания».

6 Равела и др. (2008 г.) отмечают, что оценка достижений учащихся является необходимым, но недостаточным условием для совершенствования образования. Имеются определенные доказательства того, что уже само существование оценок и распространение информации по оценкам в определенной степени влияет на некоторых действующих лиц. Но оценка — это лишь один из ряда ключевых элементов образовательной политики, а ряд этот включает в себя подготовку учителей до начала профессиональной карьеры и повышение их профессионального уровня, условия работы учителей, управление школой и руководство, особенности учебного плана, учебники и образовательные материалы, инвестирование ресурсов пропорционально потребностям различных групп населения и объединенные действия ответственных за образование лиц, направленные на решение любых возникающих проблем.

разработчиков тестов и пользователей оценочной информации (AERA, APA и NCME, 1999 г.).

Также проведен значительный объем эмпирических исследований, демонстрирующих преимущества конкретных видов оценочных мероприятий в отношении успеваемости учащихся при условии их правильной реализации и использования. Например, исследования показывают наличие прочной взаимосвязи между высококачественными мероприятиями по формирующему оцениванию в классе и улучшением результатов обучения учащихся, измеряемых результативностью выполнения учащимися стандартизованных тестов-проверок образовательных достижений. Предпринятый Блэком и Уильямом (1998 г.) синтетический анализ более 250 эмпирических исследований воздействия высококачественного формирующего оценивания в классе, проведенных в разных странах мира, показал, что прирост достижений учащихся составил от половины до 100% среднего квадратичного отклонения по стандартизованным тестам достижений, при этом самый большой прирост демонстрировали учащиеся с низкой успеваемостью.⁷ Блэк и Уильям (1998 г.) приходят к выводу: «Приросты в достижениях представляются вполне значительными и ... попадают в число крупнейших приростов, о которых когда-либо сообщалось как о результатах образовательных вмешательств. Чтобы проиллюстрировать, насколько они велики, достаточно сказать, что при величине эффекта 0,7, если таковая могла бы быть достигнута на общегосударственном уровне, страна из таких «средних» по рейтингу математических достижений, как Англия, Новая Зеландия или Соединенные Штаты, попала бы в пятерку лучших в мире и заняла бы свое место сразу после Сингапура, Кореи, Японии и Гонконга» (стр. 61).

Беннетт (2011 г.), однако, отмечает, что предстоит проделать еще немалую работу по определению и выделению специфических параметров мероприятий по формирующему оцениванию в классе, которые приводят к росту учебных результатов учащихся.⁸

Корреляционные исследования выпускных экзаменов за курс средней школы или гимназический курс показывают, что наличие в стране такой методики связано с более высокими уровнями успеваемости учащихся, оцениваемыми в рамках таких международных систем оценки, как PISA или TIMSS (например, Бишоп, Мейн и Бишоп, 2001 г.). В других исследованиях указывается на наличие связи между конкретными параметрами тестов, используе-

⁷ Родригес (2004 г.) сообщает о воздействиях сравнимого масштаба на оцениваемую по системе TIMSS в США успеваемость по математике, обусловленных эффективным проведением оценивания в классе (этот вывод основан на анализе ответов учителей из стран-участниц TIMSS на вопросы по теме методики проведения оценивания в классе).

⁸ В результате одного метаанализа 21 контролируемого исследования (Фукс и Фукс, 1986 г.), в которых рассматривалась *частота* мероприятий по оцениванию в классе, был сделан вывод о том, что систематическое проведение мероприятий по формирующему оцениванию в классе (еженедельно или даже чаще) может оказывать мощное положительное воздействие на достижения учащихся (например, два оценочных мероприятия дают эффект размера 0,85, или процентильный прирост в 30 пунктов).

мых в этих экзаменационных программах, с учебными результатами учащихся, при этом экзамены на основе учебного плана или предметные экзамены (в отличие от более общих тестов на проверку умственных и иных способностей) рассматриваются как наиболее эффективные средства улучшения результатов обучения учащихся (Ауб, 2007 г.; Хилл, 2010 г.).

В то же время было показано, что такого типа экзамены с высокими ставками имеют отрицательное воздействие на учащихся из слабо успевающих групп по причине непропорционального ограничения возможностей для таких учащихся перейти на следующий уровень образовательной системы или воспользоваться определенными видами образовательных возможностей (Грини и Келлаган, 1995 г.; Мадаус и Кларк, 2001 г.). По причине наличия таких связанных с неравенством проблем применение экзаменов и их результаты должны подвергаться тщательному мониторингу на системном, групповом и личностном уровнях, кроме того, необходимо принимать меры по уменьшению или смягчению нежелательных негативных последствий.

Результаты проводящихся на системном уровне крупномасштабных оценок общих уровней достижений учащихся все чаще становятся фундаментом программ повышения уровня ответственности на основе тестов во многих странах. Исследования дают возможность заметить в целом слабую, но позитивную связь между способами использования данных, полученных в результате этих оценок, для укрепления подотчетности и повышения ответственности школ и учителей (например, путем присвоения категорий, выплаты денежных вознаграждений или принятия кадровых решений) и улучшением успеваемости учащихся (например, Карной и Лёб, 2002 г.). В то же время данные исследований демонстрируют, что простое представление отчетной информации о средних показателях школы по результатам таких оценок может также приводить к повышению успеваемости учащихся (Ханушек и Реймонд, 2003 г.), что наводит на мысль о том, что пока далеко не всё известно об оптимальной комбинации стимулов в моделях повышения ответственности на основе тестов, которая давала бы наилучшие результаты с минимальными негативными побочными эффектами. Проведенные до сих пор исследования свидетельствуют о том, что в группу ключевых факторов, определяющих преобладание позитивных или негативных результатов от мероприятий по повышению ответственности при помощи тестирования, входят техническое качество самих тестов, соответствие структуры теста методике использования его результатов и степень развитости механизмов помощи школам и учителям, демонстрирующим слабую результативность (Равела, 2005 г.).⁹

⁹ Равела (2005 г.) описывает использование результатов крупномасштабной национальной оценки в Уругвае для оказания помощи учителям в совершенствовании их преподавательской деятельности. Упор на формирующие пути использования оценки на уровне класса помог повысить готовность учителей к принятию результатов; он также повлиял на структуру оценки с точки зрения необходимости использовать подход к сбору данных на основе переписи и использования фоновых обстоятельств для регулирования воздействия на достижения со стороны внешкольных факторов.

Исследователи уделяют все более пристальное внимание характеристикам эффективных *систем* оценки, которые охватывают вышеупомянутые виды оценочной деятельности и пути их использования (т.е. оценивание в классе, экзамены и крупномасштабные оценки системного уровня). Данное исследование опирается на *принципы и передовые практические методики*, описанные в литературе, посвященной оцениванию, а также почерпнутые из аналитических работ по функционированию оценочных систем в странах, демонстрирующих высокие результаты. Дарлинг-Хаммонд и Вентворт (2010 г.) исследовали практическое функционирование высокоэффективных образовательных систем различных стран мира (например, Австралии, Финляндии, Сингапура, Швеции и Великобритании) и отметили, что мероприятия по оценке достижений учащихся в данных системах:

- показывают важность оценки *самого процесса обучения*, важность оценки *для* этого процесса и важность ее *в ходе* процесса обучения, а не оценки процесса обучения как отдельного изолированного элемента образовательной системы;
- дают учащимся, учителям и школам *отзыв* о том, что достигнуто в обучении, но в то же время и *вверх по цепи* передают информацию, которая в будущем может использоваться для формирования системы обучения, а также служить основанием для принятия решений по продолжению образования и построению карьеры;
- тесно *увязывают* ожидаемые по учебному плану результаты, предметные критерии, критерии эффективности и желательные результаты обучения;
- *вовлекают учителей* в разработку оценивания и подсчет рейтингов для совершенствования их профессиональных практических навыков и развития их способности помогать учащимся в достижении результатов;
- *вовлекают учащихся* в достоверное оценивание с целью укрепления их мотивации и повышения эффективности учебы;
- стремятся к продвижению учебной деятельности учащихся в направлении развития *навыков мышления более высокого порядка и решения задач* посредством использования более широкого круга стратегий преподавания и оценивания;
- ставят *качество* стандартизированного тестирования¹⁰ выше его *объемов*;
- являясь крупной и увеличивающейся частью их экзаменационных систем, используют *неограниченные задания на результативность и оценки*

¹⁰ Иными словами, некоторые страны демонстрируют хорошие результаты по итогам международных оценочных мероприятий, но в собственных системах образования используют не так много механизмов стандартизированного тестирования (например, Финляндия). Другие страны уделяют много внимания стандартизированному тестированию (например, Соединенные Штаты), но по тем же самым международным оценочным мероприятиям демонстрируют более слабые результаты.

на базе школы, что требует от учащихся писать много и дает им возможность развивать навыки «двадцать первого века».¹¹

Хотя Дарлинг-Хаммонд и Вентворт в своем исследовании дают широкий взгляд на то, как должна выглядеть эффективная система оценивания, они не указывают, что необходимо сделать, чтобы ее создать. В других исследованиях проводится тщательный анализ проблем планирования, проработки и реализации на практике. Например, Ферре (2006 г.) советует создавать устойчивые и эффективные системы оценивания на основе его анализа таких систем, действующих в Латинской Америке. Брей и Стюард (1998 г.) проводят подобный анализ в отношении экзаменов средней школы. Другие (например, Локхид, 2009 г.) оценивают состояние донорской деятельности в области оценивания и обсуждают пути совершенствования эффективности такой помощи странам. Третьи погружаются в политические тонкости создания устойчивых и эффективных систем оценивания (МакДермот, 2011 г.).

В данной статье сливаются воедино все эти потоки научных обоснований, а ключевые проблемы и факторы организовано выстраиваются в единую основу понимания того, как должна выглядеть эффективная система оценки достижений учащихся и как страны могут начать построение таких систем.

ОСНОВА ДЛЯ СИСТЕМ ОЦЕНКИ ДОСТИЖЕНИЙ УЧАЩИХСЯ

Чтобы подойти к построению основы со стратегических позиций, нам необходимо выявить ряд ключевых параметров систем оценивания. Здесь обсуждаются два основных параметра: *виды/цели* оценочных мероприятий и *качество* этих мероприятий.

ПАРАМЕТР 1. ВИДЫ/ЦЕЛИ ОЦЕНИВАНИЯ

Системы оценивания преимущественно состоят из трех основных видов оценочных мероприятий, соответствующих трем главным информационным потребностям или целям. К ним относятся:

- *оценивание в классе* для получения информации в реальном времени в помощь совершенствованию процессов преподавания и учебы в отдельных классах;
- *экзамены* для принятия решений в отношении динамики прохождения через образовательную систему конкретного учащегося (например, аттеста-

¹¹ Результаты стандартизированных заданий на результативность вводятся как часть общей экзаменационной оценки в широком круге систем: GCSE в Великобритании, в экзаменационной системе Сингапура, сертификационной системе Виктории и Квинслэнда в Австралии и в системе «Международный бакалавриат», действующей в более чем 100 странах мира. Поскольку данные оценки являются неотъемлемой частью учебного плана, они влияют на повседневную деятельность по преподаванию и учебе, ориентируя ее на использование знаний для решения проблем.

ция или решения об отборе), включая предоставление «дефицитных» образовательных возможностей;

- *крупномасштабные исследования на уровне системы* для мониторинга и предоставления актуальной для руководителей и практиков информации об общих уровнях результативности в системе, изменениях в этих уровнях и связанных с ними или влияющих на них факторах.

Несомненно, эти виды оценок не являются ни совершенно независимыми друг от друга, ни всеобъемлющими (то есть существуют определенные мероприятия по оцениванию, которые не совсем вписываются в данные выше категории). В то же время они представляют собой основные виды оценочных мероприятий, проводимых в большинстве образовательных систем мира.

Оценивание в классе, также называемое постоянным или формирующим оцениванием, проводится учителями и учащимися в ходе ежедневной учебной деятельности (Айразян и Рассел, 2007 г.). Оно включает в себя целый ряд стандартизированных и нестандартизированных инструментов и процедур для сбора и интерпретации фактов об успеваемости и успехах учащихся в письменной, устной и иных формах. Примерами оценивания в классе можно считать устный опрос, домашние задания, самостоятельные доклады учащихся, диагностические тесты и опросы по итогам изучения раздела. Главной целью такого оценивания является получение информации в режиме реального времени в поддержку процессам преподавания и получения знаний.

Экзамены различного статуса, определяемого такими дополнительными определениями, как «государственные», «внешние» или «итоговые», дают информацию для принятия в отношении каждого учащегося важного решения, например, сможет ли учащийся обучаться в конкретном типе школы или по какой-либо конкретной программе, получить аттестат об окончании средней школы или быть принят в университет (Грини и Келлаган, 1995 г.; Хойберт и Хаузер, 1999 г.). Экзамены могут быть внешними или (что получает все большее распространение) проводиться на базе школы, но по сути своей они обычно стандартизированы с целью гарантировать всем учащимся равные возможности продемонстрировать знания и умения, полученные в рамках официального учебного плана или некоей другой установленной программы по формированию знаний и приобретению навыков (Мадаус и Кларк, 2001 г.). Хорошим примером тому являются экзамены на аттестат или выпускные экзамены, проводимые в конце периода обязательного образования в рамках многих образовательных систем. Как уже отмечалось ранее, огромная важность большинства экзаменов и сопряженность их с риском для испытуемого означает, что они могут оказывать неблагоприятное влияние на систему образования с точки зрения того, что преподается и усваивается, и так или иначе воздействовать на весь комплекс знаний и умений выпускников (Вест и Крайтон, 1999 г.). Эти обстоятельства необходимо учитывать при определении целесообразности¹² таких тестов и возможности или способа их ком-

12 Грини и Келлаган (1995 г.) отмечают, что в связи с высокой важностью результатов экзамена учи-

бинирования с другими источниками информации с целью обеспечения максимально справедливого использования их результатов по отношению к отдельным лицам, группам лиц и обществу в целом. Важно подчеркнуть, что существуют очень конкретные профессионально-технические нормы, касающиеся приемлемых и неприемлемых вариантов использования экзаменов (и тестов в целом) как основы для принятия важных и сопряженных с риском решений в отношении отдельных учащихся (AERA, APA и NCME, 1999 г.).

Крупномасштабные исследования на уровне системы предназначены для получения информации об уровнях результативности работы системы и о факторах, которые связаны с результативностью или влияют на нее (Грини и Келлаган, 2008 г.; Кифер, 2001 г.). Они обычно связаны с согласованным набором стандартов или учебных целей и направлены на предоставление информационной основы для принятия политических и практических решений в образовании. В качестве примеров подобных исследований можно назвать такие международные оценки уровней достижений учащихся, как TIMSS, PIRLS и PISA; такие региональные исследования, как PASEC во франкоговорящей Африке, SACMEQ в англоговорящих странах Африки и LLECE в Южной Америке, исследования национального уровня, например SIMCE в Чили, и такие субнациональные оценки, как оценки на уровне штатов в США и провинций в Канаде.¹³ Эти исследования отличаются по проверяемым классам или возрастным группам, охвату целевых групп населения (на основании выборки или переписи), внутренней или внешней направленности (например, контрольные показатели могут быть национальными или международными), охватываемым предметам или группам навыков, видам собираемой справочной информации и частоте проведения. Они также различаются по методике отчетности о результатах и их использования. Например, как отмечалось выше, если одни из них завершаются представлением отчетов о результатах политикам или населению, другие имеют продолжение в виде использования результатов для привлечения к ответственности определенных групп работников системы образования (Кларк, 2007 г.).¹⁴

теля часто преподают так, как будто они готовят учащихся к экзамену. В результате учащиеся, уходящие из школы на более раннем этапе, не получают адекватных знаний и навыков. Связанные с экзаменами практические явления, которые могут ставить некоторых учащихся в неравное положение с другими, включают в себя выставление оценок (создание рейтинга), требование платы от кандидатов, частное репетиторство, проведение экзамена на незнакомом для учащихся языке и различные злоупотребления. Использование систем квот (для сглаживания различий в уровнях успеваемости), связанных с географическим положением, этнической принадлежностью или языковым признаком, также ставит некоторых учащихся в относительно неравное положение с другими.

13 IMSS — международное исследование тенденций математического и естественнонаучного образования; PIRLS — международное исследование достижений в чтении; PISA — международная программа оценки учебных достижений; PASEC — Programme d'Analyse des Systèmes Educatifs (программа анализа систем образования); SACMEQ — Консорциум стран Южной и Восточной Африки по мониторингу качества образования; LLECE — Латиноамериканская лаборатория оценки качества образования; Sistema de Medición de Calidad de la Educación (Система измерения качества образования).

14 В своей поддержке оценочной деятельности Всемирный банк за последние 20 лет (Ларч и Локхид, 1992 г.; Либерман и Кларк, 2012 г.) перенес основное внимание с реформы экзаменаци-

Один из способов понимания различий между этими тремя видами оценки сводится к тому, что оценивание в классе преимущественно представляет собой оценку *по мере* обучения и делается *для* обучения (и, следовательно, по сути своей является прежде всего формирующим), тогда как экзамены и опросы в основном представляют собой оценку самого обучения (получения знаний), а следовательно, по сути носят суммирующий характер. Эти различия не всегда четко срабатывают на практике, и все большее распространение получают гибридные подходы. Например, в Сингапуре существует система оценки, построенная вокруг государственных экзаменов, но при этом вокруг этой системы создана целая инфраструктура, работающая на принцип «для обучения» (Л. Бенвенисте, личная переписка, март 2010 г.). Другие гибридные оценки связаны с адаптацией инструментов, разработанных для одного вида оценивания (например, инструменты оценки в классе для укрепления информационной базы преподавания), для других целей (например, документирование успеваемости на системном уровне). Одна из наиболее известных инициатив подобного рода — «Оценка навыков чтения в начальных классах» (EGRA), механизм, разработанный при поддержке агентств-доноров и экспертов для использования в развивающихся странах (<https://www.eddataglobal.org/>). Основанная на инструменте, изначально разработанном для проведения оценивания в классе, EGRA уже некоторое время используется для сбора данных системного уровня об успехах учащихся в деле формирования навыков чтения на ранних этапах с целью предоставления информации министерствам и донорам о потребностях системы в улучшении преподавания (Гоув и Квелич, 2011 г.).

Системы образования могут иметь достаточно разные характеристики с точки зрения того, какое внимание в них уделяется различным типам мероприятий по оцениванию. Например, в рамках системы образования Финляндии основное внимание уделяется оценкам в классе как ключевому источнику информации по обучению и достижениям учащихся, а экзаменам и крупномасштабным системного уровня исследованиям придается намного меньшее значение. В Китае традиционно достаточно серьезно относились к экзаменам как средству сортировки и отбора из большой массы учащихся и уделяли относительно меньше внимания оцениванию в классе или крупномасштабным опросам (хотя подход в настоящее время меняется).¹⁵ Существует целый ряд факторов, влияющих на степень значимости систем оценки в разных системах образования. Этот ряд факторов включает в себя офици-

онных систем на реализацию крупномасштабных системного уровня оценочных мероприятий, нацеленных на мониторинг тенденций в учебных достижениях и формирование информационной базы для политических и практических решений.

15 К числу других оказывающих воздействие факторов относятся историческое наследие в конкретной системе образования, которое может предопределять склонность к конкретному виду оценивания (Мадаус, Кларк и О'Лири, 2003 г.), способность различных заинтересованных сторон в системе эффективно проводить различные виды оценивания (Грини и Келлаган, 2008 г.) и воспринимаемые или реальные затраты на проведение оценочных мероприятий (Вольф, 2007 г.).

альное восприятие системы образования и ее целей (и роль оценивания в формировании этого восприятия) и экономические структуры и возможности в стране, и соответствующие информационные потребности ключевых заинтересованных сторон. Неясно, существует ли *единственный* идеальный вариант системы оценивания, который был бы одинаково эффективен для любой ситуации.

ПАРАМЕТР 2. ОПРЕДЕЛЯЮЩИЕ ФАКТОРЫ КАЧЕСТВА

Ключом к решению задачи является не попытка возведения в идеал какого-либо одного набора характеристик для системы оценки достижений учащихся, а рассмотрение качества индивидуальных оценочных мероприятий и качества их комбинаций с точки зрения адекватности получаемой в результате информации для принятия решений (Мессик, 1989 г.; Шепард, 2000 г.).

Существует три главных фактора, определяющих качество информации в системе оценки (AERA, APA и NCME, 1999 г.; Дарлинг-Хаммонд и Венворт, 2010 г.):

- стимулирующая среда;
- согласованность системы;
- качество оценки.

Несмотря на наличие тесной связи между этими параметрами, в данной работе они представлены раздельно в целях их подробного рассмотрения.

Понятие «стимулирующая среда» относится к широкому пространству, в котором имеет место деятельность по оцениванию, и к степени содействия и поддержки оцениванию, получаемых от этого пространства. Понятие охватывает такие области, как законодательная или политическая база для оценочной деятельности, руководство, воздействующее на процесс оценивания (в том числе вопросы наличия политической воли на проведение оценки, несмотря на осведомленность в том, что результаты могут вскрыть серьезные проблемы или факты неравенства в обучении), вовлеченность общественности в оценочную деятельность, институциональные механизмы для разработки и проведения мероприятий по оценке и дальнейшего использования их результатов,¹⁶ наличие достаточных и стабильных источников финансирования, а также наличие компетентных сотрудников в оценочных группах и учителей.

16 Идут многочисленные споры в отношении того, должны ли группы специалистов, проводящих экзамен или крупномасштабную оценку, формироваться из штатных сотрудников министерств образования. На самом деле, институциональная принадлежность здесь не так важна, как важна культура преобладания и приложение.

Такая культура утверждается, когда у оценки имеются четкие задачи и прочная структура, а для этого необходимо, чтобы данная система оценки опиралась на некий нормативно-правовой документ.

Стимулирующая среда важна для достижения правильного результата, так как она является ключевым фактором, определяющим качество и эффективность оценочной системы на долгосрочную перспективу, и без нее, как планета без почвы, воды и воздуха, любая система оценки нежизнеспособна (Всемирный банк, 2010 г.). В большинстве случаев на правительстве лежит ответственность, как минимум, за очерчивание общих перспектив, исполнение лидирующей роли и создание политической основы для формирования такой стимулирующей среды (в то же время следует иметь в виду, что относительная свобода от политического влияния является одним из основных признаков более зрелой системы оценки), которая позволит впоследствии обеспечить реализацию мероприятий по оцениванию через частно-государственные партнерства (например, передача исполнения программы оценки по контракту внешней фирме). Некоторые системы образования, особенно в федеральных контекстах, создают стимулирующую среду совместными усилиями, объединяя ресурсы или институциональные механизмы для разработки и проведения тестов, а также для анализа результатов и составления отчетности по ним (например, когда штаты или системы объединяют усилия для разработки совместного банка тестовых заданий, которым впоследствии может пользоваться каждый/ая из них для собственных целей, таким образом сокращая затраты для каждого отдельного штата или каждой отдельной системы). Такие региональные мероприятия по оценке, как SACMEQ, PASEC и LLECE, представляют собой еще одну форму сотрудничества в деле создания стимулирующей среды. Коэффициент полезного действия, достигнутый за счет увеличения масштаба в результате этого сотрудничества, позволяет с большей экономической эффективностью разрабатывать более качественные тесты и внедрять в процесс тестирования технологические новинки.

Согласованность системы оценивается с точки зрения степени согласованности и скоординированности оценивания с другими компонентами системы образования. Данное понятие включает в себя связь оценивания с учебными целями системы, стандартами, учебным планом и возможностями получения педагогического образования до начала преподавательской деятельности и профессионального совершенствования работающих учителей (Фурман и Элмор, 1994 г.; Смит и О'Дэй, 1991 г.). Согласованность системы оценивания с остальными элементами системы образования важна и необходима для того, чтобы информация, получаемая в результате оценивания, была полезна для повышения качества образования в рамках системы и способствовала созданию синергизма.

Согласованность — это больше, чем просто соответствие предмета тестирования требованиям официальных стандартов или учебного плана (в то же время важно обеспечить ситуацию, при которой в результате большинства оценочных мероприятий собиралась бы хотя бы частично информация по официальным стандартам и учебному плану). Следовательно, несмотря на возможно не очень выраженное соответствие между учебным планом страны и тем, что проверяется в ходе таких международных оценок, как PISA

и TIMSS, оценка тем не менее может быть увязана (и полезна для создания информационной базы) с общими целями и ожиданиями системы образования и соответствующих реформ. При условии такого сценария развития событий оценивание фактически может играть ведущую роль в повышении качества системы образования, а не просто служить инструментом пассивного мониторинга изменения качества (несмотря на тот факт, что использование данных оценок TIMSS, PIRLS и PISA для мониторинга воздействия национальных реформ на успеваемость с течением времени сыграло ключевую роль в повышении уровня достижений в столь разных странах, как Бразилия, Иордания и Польша).

Качество оценки — понятие, охватывающее психометрическое качество инструментов, процессов и процедур, используемых в ходе проведения оценочных мероприятий (AERA, APA и NCME, 1999 г.). Важно заметить, что качество оценки актуально для любого мероприятия по оценке, будь то оценка в классе, экзамены или широкомасштабная оценка на уровне системы. Понятие охватывает такие проблемы, как *структура и реализация* мероприятий по оценке, экзаменационные вопросы или темы опроса, *анализ и интерпретация* ответов на вопросы, задаваемые в ходе оценивания, вопросы, темы, приемлемость методики *отчетности* по результатам оценки, экзамена или опроса и *использования* этих результатов (Хойберт и Хаузер, 1999 г.; Шепард, 2000 г.). Точные критерии, используемые для таких заключений, различны и зависят от типа оценочного мероприятия. Качество оценки имеет большое значение, поскольку если оценка не доброкачественна с точки зрения структуры, реализации, анализа, интерпретации, отчетности или использования результатов, это может негативно отразиться на принятии решений в отношении обучения учащихся и качества системы (Мессик, 1989 г.; Вольф, 2007 г.). По сути низкое качество оценки может привести к провалу всего оценочного мероприятия, если оно вызывает недоверие с точки зрения используемых подходов.

Надежность и достоверность — вот два технических параметра, которые необходимо учитывать при любом анализе качества оценки. *Надежность* предполагает, что информация, получаемая в результате оценки, должна быть точной, что является особо актуальным фактором для важных и связанных с высоким риском экзаменов, а также для мониторинга изменения тенденций с течением времени. *Достоверность* предполагает, что оценки по тесту представляют то, что они должны представлять, и могут использоваться по назначению. Существует одна общая угроза достоверности результатов теста, которая возникает, если язык тестирования отличается от языка преподавания. В этом случае ребенку может быть трудно продемонстрировать свои знания и умения. «Использование» является очень важным понятием, связанным с достоверностью. Необходимо тщательно рассматривать последствия использования результатов тестирования с точки зрения социального, экономического и иных видов их воздействия на различные группы населения.

Рассматривая эти определяющие факторы качества в совокупности с различными видами оценок, преследующих различные цели, мы приходим к следующей матрице, показанной в таблице 1.

Таблица 1. Матрица для построения более эффективной системы оценки достижений учащихся

	ВИД ОЦЕНКИ / ЦЕЛИ		
	Оценивание в классе	Экзамены	Широкомасштабная оценка системного уровня
Стимулирующая среда			
Согласованность системы			
Качество оценки			

Источник: Всемирный банк

Оставшаяся часть этой статьи посвящена заполнению данной матрицы и ее использованию для построения более эффективной системы оценивания. Матрицу можно применять для системы оценивания в любой стране как отправную точку дискуссии для выяснения сильных сторон существующей системы и направлений, по которым, возможно, предстоит провести дополнительную работу.

ЗАПОЛНЕНИЕ МАТРИЦЫ

Матрица, приведенная в таблице 1, является отправной точкой для определения показателей, которые могут использоваться для пересмотра систем оценивания и планирования действий по их совершенствованию. Показатели могут быть выявлены на основе комбинации критериев, включающей в себя:

- профессиональные стандарты оценки;
- эмпирическое изучение характеристик эффективных систем оценивания, включая анализ характеристик, по которым различаются системы оценивания в более и менее эффективно функционирующих государствах;

- теорию, которая по общему согласию экспертов вносит вклад в повышение эффективности оценивания.

Доказательная база в некоторых из этих областей прочнее, чем в других. Например, существует множество профессиональных стандартов в области качества оценки (APA, AERA и NCME, 1999 г.),¹⁷ но значительно меньше таких стандартов, касающихся стимулирующей среды. Кроме того, некоторые эмпирические исследования ограничены своей корреляционной природой, следовательно, мы должны остерегаться приписывания несуществующих характеристик и преувеличения взаимосвязи между существующими характеристиками. Несмотря на эти ограничения, факты из целого ряда различных источников достаточно убедительно показывают, чем является лучшая по качеству оценка (и чем не является).

Упомянутые выше критерии и соображения были использованы для расширения трех определяющих факторов качества в широких областях показателей, представленных в таблице 2. Эти области показателей особенно актуальны для широкомасштабных оценок системного уровня, но могут изменяться также и в условиях оценки в классе.

Таблица 2. Матрица для построения более эффективной системы оценки достижений учащихся с широкими областями показателей

	ВИД ОЦЕНКИ / ЦЕЛИ		
	Оценивание в классе	Экзамены	Широкомасштабная оценка системного уровня
Стимулирующая среда	Политика Руководство и вовлечение общественности Финансирование Институциональные механизмы		
Согласованность системы	Учебные цели/целевые показатели качества Учебный план Возможности подготовки учителей и повышения их квалификации		
Качество оценки	Обеспечение качества (структура, проведение, анализ) Обеспечение эффективного использования		

Источник: Всемирный банк

¹⁷ Кроме того, существует достаточно объемная научная база по согласованию системы (например, Фурман и Элмор, 1994 г.; Хамилтон, Стетчер и Кляйн, 2002 г.).

Данные по некоторым из этих областей показателей можно найти в официальных документах, опубликованных отчетах (например, Феррер, 2006 г.), научных статьях (например, Браун и Канджи, 2006 г.) и в базах данных, размещенных в Интернете.¹⁸ Однако по большей части комплексный и системный сбор актуальных данных до сих пор не проводился.¹⁹ Тем, у кого возникнет желание проанализировать этот вид информации для конкретной системы оценивания, вероятнее всего, придется собирать данные самостоятельно. Пытаясь удовлетворить эту потребность, Всемирный банк разработал набор стандартизированных анкет и рубрик для сбора и оценки данных по трем видам оценивания (оценивание в классе, экзамены и оценка системного уровня) и соответствующим определяющим факторам качества (стимулирующая среда, согласованность системы, качество оценки). Инструментарий, постоянно обновляемый и корректируемый на основании новых научных данных и опыта различных стран, доступен в Интернете по адресу <http://www.worldbank.org/education/saber>. Страны могут использовать данные инструменты, созданные на базе матрицы и расширенных областей показателей, представленных в таблице 2, для систематического изучения и повышения уровня понимания сильных и слабых сторон своих систем оценки достижений учащихся и планирования последующих направлений деятельности. Важно отметить, что инструментарий преимущественно направлен на установление контрольных показателей политики, практики и механизмов страны для проведения оценивания в классе, экзаменов и широкомасштабных оценочных мероприятий системного уровня на системном уровне. Для определения фактических направлений практических действий на местах со стороны учителей и учащихся потребуются дополнительные инструменты.

В 2011 году Группа организаций Всемирного банка начала реализацию «Стратегии в сфере образования на период до 2020 года: обучение для всех», в которой содержится план достижения цели «Обучение для всех» в развивающемся мире в течение предстоящего десятилетия. Для поддержания стратегии Всемирный банк начал многогодичную программу, в рамках которой странам оказывается содействие в систематическом анализе и повышении результативности работы их систем образования. В рамках данной научно обоснованной инициативы, названной сокращенно SABER (от англ. Systems Approach for

18 Две полезные базы данных можно найти по адресам: <http://www.inca.org.uk/> и <http://epdc.org/>.

19 Бринкли, Гатри и Вайэт (1991 г.) изучили практику проведения широкомасштабных оценок системного уровня и экзаменов в странах ОЭСР. Ларах и Локхид (1992 г.) провели подобное исследование оценок, осуществленных при поддержке Всемирного банка. Макинтош (1994 г.) провел исследование в 10 странах (Австралии, Бахрейне, Англии и Уэльсе, Гватемале, Израиле, Малайзии, Намибии, Польше, Шотландии и Словении).

Better Education Results — системный подход к улучшению результатов образования), создается набор диагностических инструментов для исследования образовательных систем и направлений политики в сфере образования на основе глобальных стандартов и передовых практических методов, а также проведения сравнительного анализа их с методиками и практическими подходами стран в различных частях земного шара. За счет привлечения этого глобального ресурса научных знаний программа SABER ликвидирует дефицит данных о наиболее важных факторах, влияющих на повышение качества образования и достижение лучших результатов, и научного обоснования этих факторов. Дополнительную информацию можно найти по адресу: www.worldbank.org/education/saber.

SABER — оценка достижений учащихся — одна из составляющих программы Всемирного банка, содержит разработанные инструменты для анализа методик и систем оценки учащихся во всем мире и установки соответствующих контрольных показателей с целью укрепления оценочных систем, способствующих повышению качества образования и обеспечивающих достижение цели «Обучение для всех». Для повышения качества изучения научных достижений в данной сфере группа, занимающаяся оценкой достижений учащихся в рамках SABER, привлекла ведущих ученых, специалистов и практиков из развивающихся и индустриально развитых стран для обсуждения проблемных вопросов оценки, важных для повышения качества образования и учебных результатов.

Исследования и подготовка всех документов по оценке достижений учащихся по системе SABER стали возможны благодаря поддержке со стороны трастового фонда Российской программы содействия образованию в целях развития (READ TF). READ TF является совместным специализированным фондом Российской Федерации и Всемирного банка, предназначенным для оказания содействия в улучшении результатов обучения в странах с низкими уровнями доходов за счет создания надежных систем оценки достижений учащихся. Дополнительную информацию о деятельности трастового фонда можно найти по адресу: www.worldbank.org/readtf.

УРОВНИ РАЗВИТИЯ

Основная структура рубрик для оценки данных, собранных при помощи стандартизированных анкет, кратко представлена в таблице 3. Цель рубрик состоит в том, чтобы в стране сложилось некое восприятие уровня развития, на котором она находится в сфере оценивания. Для этого проводится сравнение существующей практики оценивания с лучшими или рекомендуемыми практическими методами в данной области.

Таблица 3. Базовая структура рубрик для оценки данных, собранных в рамках системы оценивания достижений учащихся

<i>Параметр</i>	УРОВЕНЬ РАЗВИТИЯ				
	НЕРАЗВИТЫЙ (отсутствие показателя или отклонение от него)	НА СТАДИИ ФОРМИРОВАНИЯ (движение к соответствию минимальным стандартам)	СЛОЖИВШИЙСЯ (приемлемое соответствие минимальным стандартам)	ПЕРЕДОВОЙ (использование передовой практики)	Обоснование
ЕС — Стимулирующая среда					
ЕС11—государственная политика					
ЕС2—лидерство, вовлечение общества					
ЕС3—финансирование					
ЕС4—институциональные механизмы					
ЕС5—кадровые ресурсы					
SA — Согласованность системы					
SA1—цели в области обучения и качества					
SA2—учебный план					
SA3—получение педагогического образования до начала преподавательской деятельности и профессиональное совершенствование работающих учителей					
AQ — Качество оценки					
AQ1—обеспечение качества (структура, исполнение, анализ)					
AQ2—обеспечение эффективного применения					

Источник: Всемирный банк

По каждому показателю в рубрике указано четыре уровня развития: неразвитый, на стадии формирования, сложившийся, передовой.²⁰ Данные уровни являются искусственно построенными категориями, которые представляют собой ключевые фазы непрерывного цикла развития по каждому показателю. По каждому из уровней имеется пояснение, в котором указывается, как выглядит показатель результативности на данном уровне. «Неразвитый» является самым низким уровнем по результативности. Неразвитость означает отсутствие показателя либо отклонение от него. Уровень «на стадии формирования» является следующим уровнем и свидетельствует о частичном наличии признаков показателя. Уровень «сложившийся» относится к состоянию, при котором выполняется минимальный стандарт по показателю, а уровень «передовой» представляет собой идеальное положение дел в применении передовых практических методов, сложившихся на настоящий момент. Не все вопросы из анкет представлены в рубриках. Сделано это по той причине, что не все вопросы имеют под собой доказательную базу, которая показала бы наличие взаимосвязи между повышением уровня результативности по определенному показателю и качеством или эффективностью оценочных мероприятий.

Очень важно признать, что многие из проблем, которые мы пытаемся решить при помощи показателей и связанных с ними уровней развития, иногда достаточно трудно измерить. В некоторых случаях существуют четкие стандарты, которые могут помочь в деле измерения (например, международные стандарты для определения, годится ли страна для включения в международный список по TIMSS). В других случаях все происходит на уровне суждений (например, измеряется уровень поддержки определенному виду оценки со стороны населения). Для того чтобы повысить уровень общей надежности и обеспечить возможность проведения сравнения между показателями и уровнями развития на системном уровне, анкеты и рубрики на максимально возможном уровне сосредоточиваются на объективных мерах.

В дополнение к оценке результативности по индивидуальным показателям может быть вполне целесообразно сравнивать общие характеристики систем оценивания по качественным показателям с характеристиками систем оценивания, находящихся на разных уровнях развития. В таблице 4 представлены типовые оценки с опорой на информацию, представленную в таблице 2 и приложении 2, по системам оценивания на уровнях «на стадии формирования», «сложившийся» и «передовой» (уровень «неразвитый» опущен, поскольку он подразумевает отсутствие какой-либо деятельности по оцениванию).

Системы оценивания, попадающие в категорию «на стадии формирования», характеризуются как системы, работающие в стимулирующей среде,

20 Понятие «неразвитый» применимо к странам, где нет официальной системы оценивания или система образования не работает по причине наличия войны или иного конфликта.

при этом начинает формироваться определенный уровень согласованности системы и строиться качество оценивания. Для данных систем характерна нестабильность и неопределенность в отношении выбора, частоты и проведения оценочных мероприятий, что указывает на отсутствие четкого видения оценивания на уровне системы, а также на неопределенность в отношении финансирования или на дефицит финансовых средств, необходимых для проведения мероприятий по оценке. В такой ситуации оценка, вероятнее всего, будет некой «добавкой» к системе, при этом не будут прилагаться усилия на системном уровне по согласованию оценки со стандартами, учебным планом или подготовкой учителей.

Таблица 4. Стилизованные характеристики систем оценки достижений учащихся на различных уровнях развития

	На стадии формирования	Сложившийся	Передовой
Стимулирующая среда	<ul style="list-style-type: none"> • Отсутствие или минимальное наличие политических основ • Слабый уровень лидерства/вовлечения населения • Мало подготовленных кадров; высокая текучесть кадров • ненадежное/нерегулярное финансирование • Неясные или нестабильные институциональные механизмы 	<ul style="list-style-type: none"> • Наличие четких политических основ и руководящих принципов • Сильное лидерство/вовлечение населения • Учебные программы/подготовленные кадры, минимальная текучесть • Стабильное/регулярное финансирование • Ясные и стабильные институциональные механизмы 	<p>Так же, как на уровне «сложившийся»</p> <p>+внимание сосредоточено на:</p> <ul style="list-style-type: none"> • Оценивании для обучения • Оценке в классе и школе • Роли учителей
Согласованность системы	<ul style="list-style-type: none"> • Оценки не полностью согласованы с целями обучения/качеством, стандартами, учебным планом • Оценки не согласованы с педагогическим образованием и повышением квалификации учителей 	<ul style="list-style-type: none"> • Оценки согласованы с целями обучения/качеством, стандартами, учебным планом • Оценки согласованы с педагогическим образованием и повышением квалификации учителей 	<ul style="list-style-type: none"> • Практических методов, основанных на научных исследованиях и инновациях
Качество оценки	<ul style="list-style-type: none"> • Ограниченная осведомленность в отношении применения технических или профессиональных стандартов для обеспечения качества оценки и эффективного использования результатов 	<ul style="list-style-type: none"> • Осведомленность и использование технических или профессиональных стандартов для обеспечения высокого качества оценки и использования ее результатов 	

Источник: Всемирный банк

Примечание: «неразвитый» уровень опущен, поскольку на данном уровне имеет место отсутствие какой-либо деятельности в данном направлении.

Наращивание потенциала обычно происходит медленно, поскольку подготовленные профессионалы уходят в другую организацию или в частный сектор экономики. Следовательно, оценочные мероприятия проводятся на низком качественном уровне в связи с незнанием профессиональных стандартов.

Системы оценивания, которые характеризуются как сложившиеся, имеют стимулирующую среду, высокий уровень согласованности на уровне системы, а также высокий уровень качества оценивания, при этом данные уровни стабильны, обеспечены и консолидированы по своей сути. Данные системы характеризуются постоянством, определенностью выбора, частотой проведения оценок, использованием их результатов, а также стабильностью и достаточностью финансирования, что свидетельствует о наличии политического видения и «веры» в оценивание на системном уровне. В такой обстановке оценивание составляет неотъемлемую часть системы и соответствует стандартам, учебному плану и отвечает потребностям в профессиональном совершенствовании учителей. Наращивание потенциала обычно сосредоточено, стабильно, эффективно, текучесть кадров невелика. Оценочные мероприятия обычно высокого качества в связи с высоким уровнем осведомленности и внимания к профессиональным стандартам. Данный уровень может рассматриваться как минимальный приемлемый стандарт, делающий систему эффективной.

Системы оценки, попадающие в категорию «передовые», могут быть охарактеризованы как системы, имеющие стимулирующую среду, а также высокий уровень согласованности системы и качества оценивания, которые по сути своей сильно развиты. В дополнение к наилучшим качествам уровня «сложившийся» «передовой» уровень систем характеризуется высокими уровнями инновационной деятельности и практическими методиками, основанными на научных исследованиях. В такой среде оценивание является неотъемлемой частью системы. Наращивание потенциала сосредоточено, прежде всего, на учителях в дополнение к «технической» приверженности к оценкам на уровне школы и класса (и напоминает ключевые черты результативно работающих систем, описанных Дарлинг-Хаммонд и Вентвортом в их работе).

В реальной жизни системы оценки разнятся по уровню развития в разных областях. Например, система может попадать в категорию «сложившаяся» в области экзаменов, но при этом быть «на стадии формирования» в области широкомасштабного оценивания на системном уровне и наоборот. Хотя интуиция подсказывает, что лучше, возможно, продвигаться вперед во многих областях, тем не менее нет четких данных, подтверждающих необходимость во всех областях быть на передовом уровне. Следовательно, можно рассматривать «сложившийся» уровень как желательный минимальный уровень во всех областях (что наблюдается в системах оценивания в таких странах, как Финляндия и Австралия), но только в областях, которые вносят наибольший вклад в национальное видение приоритетов образования.

Хотя целесообразно иметь представление о том, как выглядят различные системы оценки и разные виды оценивания на различных уровнях развития, не менее (если не более) целесообразно понимать, как происходит продвижение от одного уровня к другому. Следовательно, нам необходимо понимать, как происходят некоторые из ключевых реформ или иных процессов, используемых странами для создания более эффективных систем оценивания. К сожалению, научные доказательства в этой области недостаточны, и требуется дополнительная исследовательская работа, чтобы наполнить содержанием конкретные стратегии.

Основываясь на том небольшом объеме доступных научных данных, можно сделать вывод, что главным фактором, способствующим переходу системы из категории «на стадии формирования» в категорию «сложившаяся» (в целом или в конкретной области оценки), является концентрация внимания на реформах и практических методах, которые создают благоприятную среду для оценивания (Феррер, 2006 г.).²¹ Например, Ларах и Локхид (1992 г.), рассматривая деятельность Всемирного банка по поддержке оценочных проектов в государствах-клиентах, пришли к выводу, что вероятность успеха наиболее велика у проектов, которые в первую очередь сосредоточены на совершенствовании институциональных механизмов (с точки зрения создания устойчивой программы оценивания в стране), чем у проектов, которые ориентированы на повышение технического качества существующих мероприятий оценки. В соответствии с данным выводом, анализируя попытки реформирования системы оценивания в странах Центральной и Восточной Европы, Вест и Крайтон (1999 г.) отметили, что реформы носят более устойчивый характер, когда имеет место согласие населения с необходимостью изменений, четкая и стабильная политическая поддержка изменений и выделение ресурсов в достаточном объеме.

Главным фактором, который характеризует системы, переходящие из категории «сложившиеся» в категорию «передовые», является сосредоточение на реформах, инновациях и практических методах, в центре внимания которых стоят класс, учителя и ученики как ключевые участники процесса оценивания (Дарлинг-Хаммонд и Вентворт, 2010 г.; Шепард, 2000 г.). Это связано с тем фактом, что самая мощная форма оценки, если оценка осуществляется правильно, это оценка, которая проводится учителями и учащимися в ходе их повседневной классной работы (то есть оценка на уровне класса). Однако чтобы осуществлять такую оценку правильно, требуется сильное наращивание потенциала и серьезное внимание к качеству подготовки учителей.

21 Хотя для системы может быть полезно в краткосрочной перспективе сосредоточить ресурсы на одном конкретном определяющем качестве факторе (например, на стимулирующей среде), это не является долгосрочной стратегией, так как каждый определяющий качество фактор является необходимым условием формирования эффективной системы оценивания.

ВЫВОДЫ

Оценка играет ключевую роль в формировании понимания того, дает ли система образования желаемые результаты для учащихся, экономики и общества в целом. Без эффективного оценивания невозможно узнать, получают ли знания учащиеся и правильным ли путем идут реформы.

В данной статье кратко изложены руководящие принципы, сложившиеся на опыте различных стран, и представлена научная база, позволяющая сформировать основу для построения более эффективной системы оценки достижений учащихся. Она дает политикам и иным участникам процесса научно обоснованную структуру для дискуссии и достижения согласия в отношении приоритетов и ключевых инноваций, вносимых в систему оценивания.

Важной характеристикой основных ориентиров, изложенных в статье, является оказание странам поддержки в выявлении ключевых определяющих факторов качества, которые необходимо поднять на высокий уровень для повышения уровня качества и полезности информации, получаемой в результате проведения различных оценочных мероприятий. Это имеет критически важное значение, поскольку главная цель любой системы оценивания заключается в предоставлении правдивой и своевременной информации целевым группам – учащимся, учителям, обществу и политикам, чтобы они принимали более действенные решения в поддержку более высокого качества и результативности обучения. Выбор системы оценивания должен соответствовать информационным потребностям этих целевых групп, на основании которых они принимают решения.

У данных основных ориентиров также имеется динамический параметр, отражающий траекторию перемещения с одного уровня развития на другой в каждой оценочной области. Важно помнить, что для продвижения с одного уровня на другой требуется время. Изучение конкретных примеров стран, которые совершенствовали свои системы оценки достижений учащихся, показывает, что требуется как минимум десятилетие для того, чтобы реформы и инновации устоялись и привели к ощутимым результатам. Следовательно, группы специалистов, работающие в конкретной стране, должны настраиваться на долгую работу, инвестиции, разработку политики, инновации и действия, которые потребуются для преобразования системы оценивания. Результатом станет система оценивания, которая будет служить хорошей основой для принятия решений и способствовать повышению качества образования и обучения для всех.

Приложение 1. Типы оценок и их основные отличия

Масштабные оценки системного уровня				
	На уровне класса	Национальные	Международные	Экзамены
Задача	Получить в режиме реального времени информацию о состоянии обучения для учёта в преподавании	Сообщить об общем состоянии системы на определённом уровне/для конкретных возрастных групп и отследить тенденции в обучении	Сообщить о сравнительной эффективности работы системы образования для определённого уровня/возраста	Отобрать/аттестовать учащихся по мере их продвижения к следующему уровню системы образования (или выхода на рынок труда)
Периодичность	Ежедневно	Регулярно по отдельным предметам (например, раз в 3-5 лет)	Регулярно по отдельным предметам (например, раз в 3-5 лет)	Ежегодно и чаще, если система позволяет повторную сдачу
Кого проверяют	Всех учащихся	Выборку из учащихся определённого года обучения/возраста	Выборку из учащихся определённого года обучения/возраста	Всех подлежащих проверке учащихся
Формат	Может быть разным: от наблюдения до опроса, от письменных тестов до выступлений учащихся	Обычно выбор из нескольких вариантов и краткие ответы на вопросы	Обычно выбор из нескольких вариантов и краткие ответы на вопросы	Обычно сочинение и выбор из нескольких вариантов
Охват учебного плана	Все учебные темы	Обычно ограничение несколькими предметами	Обычно ограничение одним-двумя предметами	Охватывает основные учебные темы
Запрашивается ли дополнительная информация у учащихся	Да, это составная часть процесса преподавания	Часто	Да	Редко
Выставление баллов	Как правило, неформально и просто	От простых до более сложных статистических методов	Обычно привлекаются сложные статистические методы	От простых до более сложных статистических методов

Источник: Всемирный банк

СПИСОК ЛИТЕРАТУРЫ

- 1) Airasian P., Russell M. Classroom Assessment: Concepts and Applications (6th ed.). New York: McGrath Hill, 2007.
- 2) Au W. High-Stakes Testing and Curricular Control: A Qualitative Metasynthesis // Educational Researcher. 2007. Vol. 36 (5). P. 258–267.
- 3) Bennett R.E. Formative Assessment: A Critical Review // Assessment in Education: Principles, Policy and Practice. 2011. Vol. 18 (1). P. 5–25.
- 4) Bishop J., Mane F. Bishop M. Secondary Education in the United States: What Can Others Learn from Our Mistakes? CAHRS Working Paper Series. Cornell Center for Advanced Human Resource Studies (CAHRS), 2001.
- 5) Black P., Wiliam D. Assessment and Classroom Learning // Assessment in Education: Principles, Policy and Practice. 1998. Vol. 5 (1). P. 7–73.
- 6) Braun H., Kanjee A. Using Assessment to Improve Education in Developing Nations // Cohen J., Bloom D., Malin M. (Eds.). Educating All Children: A Global Agenda. Cambridge, MA: American Academy of Arts and Sciences, 2006.
- 7) Bray M., Steward L. (Eds.). Examination Systems in Small States: Comparative Perspectives on Policies, Models and Operations. London: The Commonwealth Secretariat, 1998.
- 8) Brinkley M., Guthrie J., Wyatt T. A Survey of National Assessment and Examination Practices in OECD Countries. Lugano, Switzerland: OECD, 1991.
- 9) Carnoy M., Loeb S. Does External Accountability Affect Student Outcomes? A Cross-State Analysis // Educational Evaluation and Policy Analysis. 2002. Vol. 24 (4). P. 305–331.
- 10) Clarke M. State Responses to the No Child Left Behind Act: The Uncertain Link between Implementation and 'Proficiency for All' // Kaestle C., Lodewick A. (Eds.). To Educate a Nation: Federal and National Strategies of School Reform. Lawrence: University of Kansas Press, 2007. P. 144–174.
- 11) Darling-Hammond L., Wentworth L. Benchmarking Learning Systems: Student Performance Assessment in International Context. Stanford, CA: Stanford University, Stanford Center for Opportunity Policy in Education, 2010.
- 12) Education for All Global Monitoring Report 2008: Education for All by 2015. Will We Make It? Paris: UNESCO/Oxford University Press, 2007.
- 13) Ferrer G. Educational Assessment Systems in Latin America: Current Practice and Future Challenges. Washington, DC: Partnership for Educational Revitalization in the Americas, 2006.
- 14) Fuchs L.S., Fuchs D. Effects of Systematic Formative Evaluation on Student Achievement: A Meta-Analysis // Exceptional Children. 1986. Vol. 53. P. 199–208.
- 15) Fuhrman S., Elmore D. (Eds.). Governing Curriculum. Alexandria, VA: ASCD, 1994.

16) Gove A., Cvelich P. Early Reading: Igniting Education for All. A Report by the Early Grade Learning Community of Practice. Revised Edition. Research Triangle Park, NC: Research Triangle Institute, 2011.

17) Greaney V., Kellaghan T. Assessing National Achievement Levels in Education. Washington, DC: World Bank, 2008.

18) Greaney V., Kellaghan T. Equity Issues in Public Examinations in Developing Countries. Washington, DC: World Bank, 1995.

19) Hamilton L., Stecher B., Klein S. (Eds.). Making Sense of Test-Based Accountability in Education. Santa Monica, CA: RAND Corporation, 2002.

20) Hanushek E., Woessmann L. Education Quality and Economic Growth. Washington, DC: World Bank, 2007.

21) Hanushek E., Woessmann L. Schooling, Cognitive Skills and the Latin American Growth Puzzle (Working Paper 15066). Cambridge, MA: National Bureau of Economic Research, 2009.

22) Heubert J., Hauser R. High Stakes: Testing for Tracking, Promotion and Graduation. Washington, DC: National Academy Press, 1999.

23) Hill P. Examination Systems. Asia-Pacific Secondary Education System Review Series. Bangkok: UNESCO, 2010.

24) How the World's Best Performing School Systems Come Out On Top. London: McKinsey & Company, 2007.

25) Hoxby C. The Cost of Accountability (NBER Working Paper Series No. w8855) [Электронный ресурс] // Cambridge, MA: National Bureau of Economic Research. 2002. URL: <http://ssrn.com/abstract=305599> (дата обращения 12.05.2013).

26) Independent Evaluation Group (IEG). From Schooling Access to Learning Outcomes: An Unfinished Agenda. Washington, DC: World Bank, 2006.

27) Kifer E. Large-Scale Assessment: Dimensions, Dilemmas and Policy. Thousand Oaks, CA: Corwin Press, Inc, 2001.

28) Larach L., Lockheed M. World Bank Lending for Educational Testing (PHREE Background Paper, 92/62R). Population and Human Resources Department. Washington, DC: World Bank, 1992.

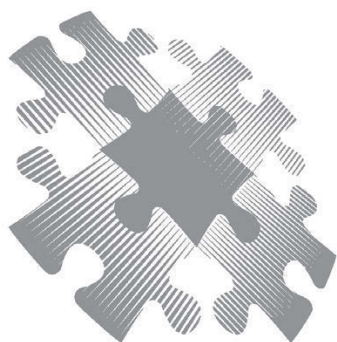
29) Liberman J., Clarke M. Review of World Bank Support for Assessment Activities in Client Countries. Unpublished manuscript. Washington, DC: World Bank, 2012.

30) Lockheed M. Review of Donor Support for Assessment Capacity Building in Developing Countries. Unpublished manuscript. Washington, DC: World Bank, 2009.

31) Macintosh H. A Comparative Study of Current Theories and Practices in Assessing Students' Achievements at Primary and Secondary Level (IBE Document Series, Number 4). Geneva, Switzerland: International Bureau of Education, 1994.

- 32)** Madaus G., Clarke M., O’Leary M. A Century of Standardized Mathematics Testing // Stanic G.M.A., Kilpatrick J., (Eds.). A History of School Mathematics. Reston, VA: NCTM, 2003. P. 1311–1434.
- 33)** Madaus G., Clarke M. The Impact of High-Stakes Testing on Minority Students // Kornhaber M., Orfield G. (Eds.). Raising Standards or Raising Barriers: Inequality and High Stakes Testing in Public Education New York: Century Foundation, 2001. P. 85–106.
- 34)** McDermott K.A. High-Stakes Reform: The Politics of Educational Accountability. Washington, DC: Georgetown University Press, 2011.
- 35)** Messick S. 1989. Validity // Linn R., (Ed.). Educational Measurement (3rd ed.). New York: American Council on Education/Macmillan, 1989. P. 13–103.
- 36)** P. Peterson, West M. (Eds.). No Child Left Behind? The Politics and Practice of Accountability. Washington, DC: Brookings Institution Press, 2003.
- 37)** Ravela P. A Formative Approach to National Assessments: The Case of Uruguay // Prospects. 2005. Vol. 35 (1). P. 21–43.
- 38)** Ravela P. et al. The Educational Assessments that Latin America Needs (Working Paper Series No. 40). Washington, DC: Partnership for Educational Revitalization in the Americas (PREAL), 2008.
- 39)** Ravela P. et al. The Educational Assessments that Latin America Needs. Washington, DC: PREAL, 2009.
- 40)** Rodriguez M.C. The Role of Classroom Assessment in Student Performance on TIMSS // Applied Measurement in Education. 2004. Vol. 17 (1). P. 1–24.
- 41)** Russia Education Aid for Development (READ) Trust Fund Annual Report 2009. Washington, DC: World Bank, 2010.
- 42)** Shepard L. The Role of Assessment in a Learning Culture // Educational Researcher. 2000. Vol. 29 (7). P. 4–14.
- 43)** Smith M.S., O’Day J. Systemic School Reform // Fuhrman S.H., Malen B. (Eds.). The Politics of Curriculum and Testing, 1990 Yearbook of the Politics of Education Association. London and Washington, DC: Falmer Press, 1991. P. 233–267.
- 44)** Standards for Educational and Psychological Testing. Washington, DC: AERA, 1999.
- 45)** The High Cost of Low Educational Performance. The Long-Run Economic Impact of Improving PISA Outcomes. Paris: OECD, 2010.
- 46)** West R., Crighton J. Examination Reform in Central and Eastern Europe: Issues and Trends // Assessment in Education. 1999. Vol. 6 (2). P. 271–280.
- 47)** Wolff L. The Costs of Student Assessment in Latin America. Washington, DC: PREAL, 2007.

ЕВРАЗИЙСКАЯ АССОЦИАЦИЯ
ОЦЕНКИ КАЧЕСТВА ОБРАЗОВАНИЯ



Е А О К О