

Прошлое, настоящее и возможное будущее российской системы оценки качества образования

В. А. Болотов

Болотов Виктор Александрович доктор педагогических наук, профессор, президент Евразийской ассоциации оценки качества образования. Адрес: 119019, Москва, ул. Новый Арбат, 21. E-mail: vikbolotov@yandex.ru

Аннотация. Рассматриваются основные этапы становления российской системы оценки качества образования: проведение аттестаций образовательных учреждений, участие в международных сравнительных исследованиях, внедрение ЕГЭ и ОГЭ, формирование сообщества профессионалов, занимающихся оценкой качества образования. Насущные задачи развития отечественной системы оценки качества образования автор видит в переходе на компетентностно ориентированный ЕГЭ и ОГЭ с сохранением предметной составляющей; в создании национальных мониторин-

гов, которые помогали бы сравнивать регионы и муниципалитеты, а также отслеживать социализацию выпускников школ; в проработке различных моделей внутриклассного и внутришкольного оценивания и создании инструментария для измерения индивидуального прогресса учащихся. Главной же проблемой в оценке качества образования автор считает отсутствие на всех уровнях грамотной интерпретации результатов измерений.

Ключевые слова: школьное образование, российская система оценки качества образования, аттестация образовательного учреждения, международные сравнительные исследования, ЕГЭ, ОГЭ, мониторинги, рейтинги, рэнкинги, Всероссийские проверочные работы.

DOI: 10.17323/1814-9545-2018-3-287-297

Статья поступила в редакцию в марте 2018 г.

В последнее время все больше внимания уделяется оценке качества образования, происходит переход от надзора и контроля к управлению качеством образования. И естественно возникает все больше спекуляций на эту тему. В связи с этим мне кажется целесообразным коснуться истории вопроса. Я буду говорить только о школьном образовании — сюжеты с дошкольным, средним профессиональным и высшим образованием должны быть описаны отдельно.

**Основные
элементы
российской
системы оценки
качества
образования**

Задолго до того как в лексикон образовательного сообщества вошло словосочетание «российская система оценки качества образования», в отечественном образовании уже обсуждались отдельные элементы этой системы. Еще в начале 90-х годов прошлого века международные эксперты Всемирного банка, специалисты из Нидерландов и Великобритании предлагали изучить существующие системы национальных экзаменов и мониторингов и подумать о создании такой системы в России. Но в это время у нас даже зарплата учителям не выплачивалась — какие национальные экзамены и мониторинги?!

Однако в Законе об образовании 1992 г. уже появилось понятие «аттестация образовательного учреждения», которая предполагала необходимость установить, насколько качество подготовки выпускников соответствует требованиям государственных образовательных стандартов (это понятие, кстати, тоже впервые появилось в законе 1992 г.). Учреждение считалось аттестованным, если не менее 50% выпускников показывали освоение стандарта. Но стандарты не принимались (почему — тема отдельного разговора, замечу лишь, что первый стандарт школьного образования появился только в 2004 г.), а аттестация учреждений проводилась. Каждый регион создавал для выпускников школ свои оценочные материалы для проверки соответствия не стандартам, а базисному учебному плану и примерным программам. Качество этих материалов обсуждению не подлежало, да и специалистов, способных разрабатывать и оценивать и средства, и процедуры измерения, просто не было. Соответственно, и речи не могло быть о каких-то общих суждениях о качестве школьного образования.

С 1995 г. Россия стала регулярно участвовать в международных сравнительных исследованиях: сначала в TIMSS (математическая и естественнонаучная грамотность школьников в 4-м и 8-м классе), потом, с 2000 г., в PISA (проверка общих компетенций у девятиклассников) и с 2001 г. — в PIRLS (различные виды прежде всего читательской грамотности у четвероклассников), и у нас стали появляться специалисты, которые могли создавать и оценивать и измерительные материалы, и процедуры измерений. В TIMSS и PIRLS Россия занимала достойные места, в PISA показывала результат ниже среднего по странам ОЭСР. По итогам выполнения российскими школьниками заданий международных сравнительных исследований готовились рекомендации методистам, обновлялись программы повышения квалификации для учителей, вносились коррективы в учебники, в частности, в них стали появляться задания, позволяющие оценить различные виды грамотности и компетенций. Позднее подходы, принятые при составлении заданий в международных сравнительных исследованиях, использовались при проектировании стандартов школьного образования. Таким об-

разом, можно сказать, что результаты международных мониторингов стали использоваться в управлении качеством школьного образования в России.

С 1996 г. началось централизованное тестирование выпускников школ по единым измерительным материалам (это были задания с множественным выбором ответа) и единообразной процедуре. Его введение мотивировалось тем, что у выпускников должна быть возможность не сдавать дважды (сначала в школе, а потом в вузе) с разрывом в месяц экзамены на знание школьной программы. Поскольку централизованное тестирование было для выпускников не обязательным (более того, было платным для участников), а школы и вузы сами принимали решение о зачете результатов в качестве выпускных или вступительных экзаменов, то и особых дискуссий о качестве измерительных материалов, об объективности процедур не было. Вузы учитывали результаты централизованного тестирования в основном по непрофильным для данной специальности дисциплинам, в частности, многие инженерные вузы с удовольствием засчитывали баллы по русскому языку. Никаких суждений о качестве образования по результатам централизованного тестирования высказать было нельзя.

Всерьез о проблемах оценки качества образования заговорили только в 2001 г., когда начал разворачиваться эксперимент по введению Единого государственного экзамена.

Стоит напомнить причины появления ЕГЭ. К 2000 г. для общества стало совершенно очевидно, что оценки в аттестатах выпускников даже двух соседних школ не дают никакой возможности сопоставить уровень знаний этих учащихся. Резко выросло число золотых медалистов, поскольку тогда золотая медаль давала льготы при поступлении в вузы. В этой ситуации говорить о качестве школьного образования было просто невозможно.

Не меньшую тревогу в обществе вызывала ситуация с приемом в вузы. Каждый вуз сам производил задания для вступительных экзаменов, что вело к росту «серого» репетиторства: подготовительные курсы при конкретном вузе или отдельные репетиторы из числа преподавателей данного вуза изучали с будущими абитуриентами только те темы, по которым будут вопросы на экзаменах. В прессе много писали про злоупотребления при проведении самих вступительных экзаменов, когда «своим» абитуриентам гарантировалось поступление, а абитуриенты «с улицы» имели шанс поступить только при наличии свободных мест. У многих вузов были и свои «договорные» школы, в которых выпускные экзамены, проходившие в камерной обстановке, засчитывались как вступительные. Понятно, что попасть в эти школы «просто» ребенку было почти невозможно. В результате число иногородних студентов в ведущих вузах страны резко упало. Если в советские времена среди перво-

курсников было 75% иногородних, то к 2000 г. их стало всего 25%. Да и в региональных вузах численность студентов из сельской местности тоже сильно сократилась.

ЕГЭ был призван дать независимую от конкретной школы и вуза оценку индивидуальных достижений выпускников и, соответственно, абитуриентов в школьных дисциплинах. После изучения мировых практик и долгих обсуждений было принято решение, что два выпускных экзамена — русский язык и математику — школьники будут сдавать только в форме ЕГЭ. По остальным выпускным экзаменам ученик мог выбирать форму сдачи — ЕГЭ или традиционная. Понятно, что этот выбор предопределялся перечнем вступительных испытаний для конкретной специальности в конкретном вузе. После определения перечня обязательных экзаменов и экзаменов по выбору встал вопрос о формате контрольно-измерительных материалов. Опять-таки после изучения опыта различных международных сравнительных исследований, а также национальных экзаменов и мониторингов в разных странах, прежде всего в Австралии, Великобритании, Нидерландах и США, была выбрана следующая структура контрольно-измерительных материалов: часть А — задания с выбором ответа; часть В — задания с кратким ответом; часть С — задания с развернутым ответом (рассуждение, эссе, решение задачи и т.п.). При этом задания частей А и В проверяются компьютером, задания части С — экспертами.

Если при разработке структуры контрольно-измерительных материалов можно было использовать мировой опыт, то технология проведения ЕГЭ создавалась практически с нуля: прецедентов проведения национальных экзаменов практически одновременно во всей стране, чья территория охватывает десять часовых поясов, не существовало.

Эксперимент по введению ЕГЭ был начат в 2001 г., добровольное участие в нем приняли 4 субъекта Российской Федерации, контрольно-измерительные материалы были приготовлены только по 8 дисциплинам, результаты учитывали всего 16 вузов, и было проведено 45 тыс. человеко-экзаменов. Целью эксперимента была отработка и технологии создания измерительных материалов, и процедур проведения ЕГЭ — от единого регламента взаимодействия создаваемых для его организации федеральных и региональных структур, подключения общественных наблюдателей до порядка рассмотрения подаваемых апелляций. В 2008 г., когда эксперимент по введению ЕГЭ завершался, в Едином государственном экзамене добровольно приняли участие 84 субъекта Российской Федерации, контрольно-измерительные материалы были по 13 дисциплинам, результаты учитывали 1800 вузов и их филиалов, и было проведено более 2665 тыс. человеко-экзаменов.

Параллельно с ЕГЭ с 2003 г. проводился экзамен с целью независимой оценки индивидуальных достижений выпускников (аббревиатуры — сначала ГИА, потом ОГЭ) 9-х классов. Измерительные материалы разрабатывались Федерацией, но, в отличие от ЕГЭ, здесь процедуру проведения определял и обеспечивал каждый субъект Федерации отдельно, поскольку для выпускников основной школы межрегиональная мобильность минимальна и вопрос об эквивалентности результатов оценки в разных субъектах фактически не стоит.

Таким образом был построен основной элемент российской системы оценки качества образования — независимая оценка индивидуальных достижений выпускников основной и старшей школы, причем для ее осуществления была создана сеть региональных центров обработки информации, которые затем превратились в центры оценки качества образования. Помимо повышения доступности качественного высшего образования для детей из удаленных регионов, из сельской местности, ЕГЭ выполнял еще одну важную функцию: он предоставлял учителям и методистам ценную информацию об уровне освоения отдельных тем в составе учебных предметов, включенных в ЕГЭ. Каждый год публиковались материалы с подробным анализом выполнения выпускниками заданий ЕГЭ. Эти материалы активно использовались в системе повышения квалификации. Аналогичная работа проводилась на уровне субъектов РФ, муниципалитетов и школ, и эта деятельность, безусловно, влияла на качество российского образования.

Вместе с тем необходимо отметить, что с самого начала эксперимента по введению ЕГЭ и ОГЭ не прекращалось неправомерное использование результатов этих процедур. И это несмотря на многочисленные письма федеральных органов управления образованием, разъясняющие недопустимость такой практики. Результаты измерений с высокими ставками — так принято называть измерения, результаты которых могут сказаться на значимых для проходящих эти измерения событиях: в нашем случае на получении аттестата и/или поступлении в вуз, — стали использоваться для «лобовой» оценки эффективности деятельности учителя, школы, муниципалитетов, регионов и даже губернаторов, на их основании стали строить всевозможные рейтинги. Говорить об эффективности чьей-либо деятельности без учета контекстных обстоятельств (социально-экономического статуса семьи; соответствия языка, на котором проводится обучение, языку, на котором говорят дома и в ближайшем окружении; уровня развития доступной образовательной инфраструктуры и т. п.) просто некорректно — в этом случае мы неизбежно будем награждать не причастных и наказывать невиновных! На самом деле анализ показывает, что 100-балльники почти всегда использовали для подготовки ре-

сурсы за пределами своей школы (занимались на специальных курсах, в том числе в Интернете, работали с репетиторами и т. д.), а школьники, которые не набирают пороговый балл, очень часто оказываются выходцами из семей с низким социально-экономическим статусом, да еще и проживающих в депрессивных или удаленных местах. Нецелое использование результатов ЕГЭ и ОГЭ обусловило стремление школ, муниципалитетов и регионов любой ценой дать высокие баллы у выпускников, и с этой целью предпринимались многочисленные попытки искажения результатов.

Отмена решения об оценке губернаторов, а затем и мэров, по итогам ЕГЭ (ранее средний балл региона по ЕГЭ учитывался в оценке эффективности работы губернаторов и мэров), а также беспрецедентные меры по усилению безопасности при проведении ЕГЭ, принятые в последние два года, повысили объективность ЕГЭ и вернули доверие к его результатам.

Приходится признать, однако, что неправомерное использование результатов имеет под собой определенные основания: у лиц, принимающих решения, просто нет другой сколько-нибудь надежной информации, которая позволяла бы оценить эффективность деятельности учителя, школы, муниципалитета и т. д. Именно поэтому в 2006 г. была разработана концепция Общероссийской системы оценки качества образования, но в силу кадровых изменений в отрасли она так и не была утверждена. В концепции предусматривалось, помимо совершенствования измерителей и процедур ЕГЭ и ОГЭ (в том числе перехода на базовый и профильный уровни обязательных экзаменов), сформировать национальные мониторинги не только учебных достижений, но и социализации учеников и выпускников, отработать различные модели внутришкольного оценивания, создать программы обучения учителей и управленцев использованию и интерпретации результатов различных видов измерений с целью повышения качества образования. Ряд позиций из концепции вошел в Закон об образовании 2012 г. и в другие нормативные документы, и они, безусловно, способствовали развитию российской системы оценки качества образования.

Важную роль в создании отечественной системы оценки качества образования сыграла Российская программа содействия образованию в целях развития (Russia Education Aid for Development, READ), запущенная в 2008 г. в рамках сотрудничества между правительством Российской Федерации и Всемирным банком с целью помощи развивающимся странам. Одним из ведущих направлений этой программы было создание пространства для формирования профессионального сообщества, которое занимается проблемами оценки качества образования, проводит измерения и исследования в России и странах СНГ.

Для этого регулярно проводились вебинары по актуальным вопросам оценки качества образования, организовывались конференции по этой проблематике, издавался журнал. В мероприятиях, проводимых в рамках этой программы, участвовали сотни специалистов из российских регионов и стран СНГ. При поддержке READ был разработан ряд измерительных материалов, которые могут использоваться для оценки как индивидуального прогресса учащихся, так и эффективности деятельности образовательных организаций, а также для проведения мониторингов. К сожалению, федеральные органы управления образованием, в отличие от региональных, интереса к этой программе фактически не проявляли.

Под мониторингом большинство исследователей понимают регулярное наблюдение и описание состояния объекта (объектов) по небольшому числу определенных параметров. Результаты мониторинга часто оформляются в форме различных рейтингов (рейтинг — одномерное сопоставление по выбранному критерию).

Рейтинги в образовании, как правило, используются для выделения лучших практик с целью их последующего распространения и для выявления зон риска — школ, муниципалитетов или субъектов РФ, устойчиво показывающих низкие результаты, так что для них надо проектировать программы действий, направленных на изменение ситуации.

Для нашей страны образовательными мониторингами являлись прежде всего различные международные сравнительные исследования, в которых Россия принимала регулярное участие. По итогам этих исследований создавались рекомендации по улучшению качества российского образования. После введения ЕГЭ немедленно стали предприниматься попытки использовать его результаты как основу для страновых, региональных и муниципальных мониторингов качества школьного образования. И посыпались всевозможные рейтинги. При этом допускались две принципиальные ошибки. Во-первых, игнорировалось то обстоятельство, что экзамены с высокими ставками не могут служить основой для оценки эффективности деятельности школы, муниципалитета или региона без учета контекстных показателей. Во-вторых, не учитывался тот факт, что в связи с изменением принципов составления экзаменационных заданий (появление базового и профильного уровней экзаменов по математике, исключение по непонятным основаниям заданий с выбором ответов) и все большим ужесточением (абсолютно правомерным) в последние годы процедур проведения ЕГЭ сравнивать баллы разных лет просто невозможно. Другими словами, из сравнения баллов ЕГЭ в разные годы не может следо-

**Мониторинги
в рамках
системы оценки
качества
образования**

вать суждение об улучшении или ухудшении качества образования. При этом отказ по «политическим» мотивам от заданий с выбором ответов, которые активно использовались и будут использоваться во всех международных исследованиях, привел к постоянному понижению порога удовлетворительного выполнения тестов ЕГЭ.

Рейтинг, составленный по результатам одного года, тоже малоинформативен, поскольку из общего списка образовательных учреждений не выделялись группы однородных прежде всего по социально-экономическому и культурному параметрам школ, муниципалитетов и регионов. В итоге в рейтингах на первых местах обычно стояли так называемые губернаторские школы, а на последних — школы, где большинство учащихся составляют дети из семей с низким социально-экономическим статусом. Вряд ли имеет смысл по результатам такого рейтинга говорить о распространении лучших практик «губернаторских» школ на малокомплектные сельские школы в депрессивных районах, и, соответственно, проектировать улучшение качества образования, опираясь на эти рейтинги, было бы очень трудно.

Лишь в 2014 г. был запущен национальный мониторинг — Национальное исследование качества образования (НИКО), предусматривающее проведение регулярных исследований качества общего образования по отдельным учебным предметам на конкретных уровнях общего образования, в определенных классах. При этом сфера использования полученных результатов пока остается крайне ограниченной. Национальная локализация НИКО не позволяет сравнивать его результаты на международном уровне. Локальная выборка школ не репрезентативна для субъектов Российской Федерации и не дает возможность сопоставлять их результаты и делать обобщенные выводы о качестве образования на уровне регионов, муниципалитетов и отдельных образовательных организаций. Также периодичность проведения НИКО по конкретным предметам не установлена, что не позволяет фиксировать тенденции в обучении по предметным областям. Как следствие — не очень понятно, кто и как может использовать результаты этого мониторинга для повышения качества образования.

Ряд регионов проводит собственные мониторинги, в которых, как правило, используется инструментарий, созданный в рамках программы READ либо применяемый Центром оценки качества образования Института стратегии развития образования.

**Актуальные
проблемы
в развитии
российской
системы оценки
качества
образования**

Начну с новшества, которое, пожалуй, более всего волнует образовательное сообщество, — с введения Всероссийских проверочных работ, которые представляют собой контрольные для

школьников по завершении обучения в каждом классе. Всероссийские проверочные работы — самая массовая оценочная процедура в российской системе образования. В апреле — мае 2017 г. в них участвовали около 3 млн школьников и почти 40 тыс. школ. Весной 2017 г. эти работы писали учащиеся 4-х, 5-х, 10-х классов. Организация этой оценочной процедуры вызывает немало вопросов.

Во-первых, утверждается, что во Всероссийских проверочных работах используются задания, разработанные на федеральном уровне в соответствии с Федеральным государственным образовательным стандартом, а также обеспечиваются единые критерии оценивания, при том, что эти контрольные школы проводят самостоятельно. Однако стандарты описывают результаты не по классам, а по ступеням, в них нет предметных результатов, и каждая школа по закону имеет право создавать собственные образовательные программы. А значит, Всероссийские проверочные работы опять начинают избыточно регламентировать деятельность школы, и авторские образовательные организации опять попадают в категорию «плохих». Что касается оценки по одним критериям, по опыту школ международного бакалавриата и опыту введения ЕГЭ можно с уверенностью сказать, что обеспечение такого оценивания требует очень серьезной работы по выработке этих критериев и обучению ими пользоваться.

Во-вторых, непонятно, кто и как будет применять результаты Всероссийских проверочных работ. Утверждается, что их будут использовать учителя для оценки индивидуальных учебных достижений конкретного ученика с целью планирования индивидуальной работы с ним. Другими словами, эти контрольные дают материалы для диагностики, т. е. представляют собой сервис для учителя и школы. Но зачем тогда такая регламентация работы с этим сервисом? Она может привести и уже приводит к тому, что учителей и школы сравнивают по итогам выполнения Всероссийских проверочных работ. Последствия предсказуемы: наказание невиновных и фальсификации — ведь контрольные проводят сами школы! И очень может быть, что по их итогам будут выстраиваться рейтинги.

Таким образом, Всероссийские проверочные работы — это какой-то странный гибрид материалов для внутришкольного оценивания (но зачем тогда такая внешняя регламентация?) с мониторингом (но зачем тогда проверять всех детей? и как можно обходиться без внешнего контроля?) и с экзаменами с высокими ставками (ведь будут отличники и двоечники).

На этого оценочного монстра уходят очень большие финансовые, временные и человеческие ресурсы. Он отвлекает внимание от действительно насущных проблем российской системы оценки качества образования, к которым, по моему мнению, относятся:

- переход на компетентностно ориентированный ЕГЭ и ОГЭ в соответствии со стандартами старшей и основной школы и с сохранением предметной составляющей;
- создание национальных мониторингов, которые помогали бы сравнивать регионы и муниципалитеты, причем не только по уровню знания предметов у учащихся, но и по уровню развития у них «мягких» навыков, а также отслеживать социализацию выпускников школ и подростков из различных групп риска;
- проработка различных моделей внутриклассного и внутришкольного оценивания и создание инструментария для измерения индивидуального прогресса учащихся, причем не только предметно ориентированного, но и прогресса в развитии различных компетенций.

Главной же проблемой сегодня является отсутствие на всех уровнях грамотной интерпретации результатов измерений. Главным направлением в использовании этих результатов становится переход от рейтингов к рэнкингам — многопараметрическим сравнительным оценкам с возможностью сортировать результаты оценок по интересующим конкретного пользователя параметрам и тем самым строить множество рейтингов, а также подготовка на их основании грамотных управленческих решений по улучшению качества образования.

По всем этим направлениям есть неплохие наработки, но они сделаны энтузиастами и не находятся в зоне приоритетного внимания органов управления образованием. Перспективы развития российской системы оценки качества образования обусловлены решением поставленных выше проблем.

Что касается самого главного на сегодня элемента этой системы — Единого государственного экзамена, на мой взгляд, в нынешнем виде он уйдет. На базе созданной для него инфраструктуры в перспективе будут развернуты центры по независимой оценке уровня предметных знаний, различных видов грамотности, уровня развития «мягких навыков» (в этом направлении начали работать и в OECD, и в движении WorldSkills). Эти центры будут выдавать сертификаты, которые человек будет собирать в портфолио и использовать и при поступлении на учебу, и в конкурсе при приеме на работу. Прототипы таких центров есть, один из примеров — центры, где проводится TOEFL.

В целом значимость проблематики оценки качества образования как в мировой, так и в российской повестке развития образовательных систем будет нарастать.

The Past, Present, and Possible Future of the Russian Education Assessment System

Viktor Bolotov

Author

Doctor of Sciences in Pedagogy, Tenured Professor, President of the Euroasian Association for Educational Assessment. Address: 21 Novy Arbat Str., 119019 Moscow, Russian Federation. E-mail: vikbolotov@yandex.ru

The article describes the key stages in the development of the education quality assessment system in Russia: certification of educational institutions, participation in international comparative studies, implementation of the Unified State Examination (USE) and Basic State Examination (BSE), and the emergence of a community of education assessment experts. The most urgent goals in the development of the Russian education assessment system are seen to be switching to competency-oriented USE and BSE (with the subject-specific component preserved), developing national monitoring studies to compare education quality among regions and municipalities, tracing the socialization patterns of school graduates, elaborating various models of in-class and in-school assessment, and providing tools to measure the individual progress of students. Meanwhile, the lack of competent interpretation of measurement results appears to be the main challenge in education quality assessment.

Abstract

schooling, Russian education quality assessment system, certification of educational institutions, international comparative studies, USE, BSE, monitoring, rating, ranking, Russian nationwide tests.

Keywords