

Принцип разумной достаточности при оценке качества образования

Б.В. Илюхин

В научном и педагогическом сообществах ведутся дискуссии о том, каким должно быть образование, его стандарты, а механизмы достижения поставленных целей, измерения степени соответствия стандартам, необходимые ресурсы обычно остаются «за кадром».

• педагогические измерения • измерительные материалы • лучевые диаграммы • коррекция по результатам оценивания

Широко используемое понятие «качество образования» не может иметь одного конкретного значения (формулы), поскольку оно всегда контекстно (для конкретного объекта — страны, школы, человека) и должно сопровождаться системой отсчёта, а ещё лучше — измерения. Очевидно, что нельзя использовать обыденное толкование качества образования, часто используемое в школах: процент учеников, имеющих отметку не ниже «четвёрки», процент выпускников, поступивших в вузы, число победителей курсов разного уровня и т.п.

На VIII Международной научно-практической конференции «Тенденции развития образования: проблемы управления и оценки качества образования» остро обсуждался вопрос целей развития образовательной системы России. Большинство исследователей и практиков сходятся в том, что понятие «качество образования» различно для разных уровней системы образования (страны, региона, муниципалитета, школы), многомерно и не определяется только учебными результатами; должно измеряться, т.е. подлежать оценке, иначе невозможно и бессмысленно выдвигать тезисы о его (образования) ухудшении или улучшении, возрастании или падении.

Качество образования определяется как интегральная характеристика системы общего образования, отражающая степень соответствия образовательных программ, условий осуществления образовательного процесса, реально достигаемых образовательных результатов государственным нормативным требованиям и требованиям социального заказа, сформированного потребителями образовательных услуг; а *оценка качества образования* — установление степени (уровня) соответствия образовательных результатов учащихся, образовательных программ, условий образовательного процесса общепризнанным эталонам и требованиям, зафиксированным в нормативных документах.

Проблема использования этих определений в том, что до сих пор потребителями образовательных услуг так и не сформирован внятный социальный заказ, а государству с трудом удаётся принимать нормативные требования в виде образовательных стандартов, причём в этих стандартах чётко прописаны условия (ресурсы) и нечётко — результаты образования, т.е. те самые образовательные достижения.

Исходя из заказа на образование личности (почти всегда индивидуального), общества и государства (стандарты) каждая школа должна разработать свои образовательные программы, в которых описывается не только содержание, но и результаты. И только после этого можно говорить о степени соответствия результатов, программ и условий эталонам и требованиям.

Кроме того, сегодня очевидна подмена понятий. Государство требует от школы (заложено в новых образовательных стандартах) развития, социализации личности, воспитания членов гражданского общества и многого другого — того, что должны делать совместно школа, семья и общество. Рискну утверждать, что эта тенденция опасна, она может сформировать в обществе определённый инфантилизм, развитый ещё с советских времён, — «школа всему научит». Не говоря уже о том, что практически везде ставки школьных психологов ликвидированы, воспитательная работа часто сводится к «мероприятиям», а социализация происходит как раз вне школы — в Интернете, сообществах по интересам и т.п.

Констатируем факт: государственные нормативные требования к образовательным программам и условиям обеспечения образовательного процесса зафиксированы в ряде документов (стандарты, СанПиНы и пр.), требования к образовательным результатам, повторюсь, весьма расплывчаты, но зафиксированы в тех же стандартах. Возникает вопрос «измеряемости» этих самых образовательных результатов.

Педагогические измерения

Этим занимается весьма неразвитая в России область науки под названием «Педагогические измерения». Сразу оговорюсь: я далёк от мысли операционализировать и формализовать каждое повседневное действие каждого ребёнка и каждого учителя — это невозможно. Но если мы хотим что-либо (те самые образовательные достижения) с чем-либо сравнивать, то сначала необходимо получить хотя бы приближённые значения этих величин. Основной мой тезис звучит следующим образом: если мы хотим сделать что-либо лучше, то мы должны улучшать *конкретные* характеристики системы или объекта, причём для того, чтобы потом иметь возможность аргументированно доказывать, что этот объект стал лучше или хуже, необходимо уметь измерять эти самые характеристики.

И с качеством, и с количеством предоставляемой информации для удовлетворения информационных потребностей есть существенные проблемы.

Созданию системы, позволяющей предоставить объективную информацию о состоянии образования всем заинтересованным категориям пользователей, посвящена деятельность нашего Центра оценки качества образования в последние несколько лет. При этом наши подходы к построению этой системы, её элементам, их составу и количеству существенно изменились. Совершенно очевидна недостаточность для сколь-нибудь серьёзного анализа данных, предоставляемых одним измерением, даже если предположить, что оно качественное. В достаточно короткие сроки стало понятно, что и перевод в новую форму процедуры государственной итоговой аттестации выпускников девятых классов, так же как ЕГЭ, не позволяет (равно как и любая другая процедура с «высокими ставками») сформировать ту информацию, которая необ-

ходима для оценки деятельности как отдельного педагога, так и школы в целом. Отрадно сознавать, что в это же время (или раньше) к подобному выводу пришли и авторы идеи комплексного проекта модернизации образования (КПМО), декларировавшие необходимость ввести «портфолио» ребёнка, основанного на оценке динамики его учебных и не только достижений, «портфолио» учителя с выходом на стимулирующую часть его оплаты труда и проч. К сожалению, КПМО не оставил после себя сколько-нибудь значимых объективных механизмов решения описанных выше проблем. В подавляющем большинстве школ при распределении стимулирующих выплат основываются на результатах ЕГЭ и ГИА, а также используют безлико оцениваемые «уровень владения ИКТ», «число участников олимпиад и конкурсов различного уровня» и т.п. Например, несмотря на то, что мониторинговые исследования уровня обученности проводятся в Томской области с 1999 года, динамику их результатов по отдельным предметам реально используют не более 20% школ.

Нормативная база

Нормативно-правовая база системы оценки качества образования регионального уровня Томской области мало чем отличается от аналогичных документов в других субъектах Федерации. Её стержнем стало «Положение о региональной системе оценки качества образования в Томской области», утверждённое Приказом Департамента общего образования Томской области № 1447 от 26.11.2008 года.

Согласно этому документу, основной целью формирования региональной системы оценки качества в Томской области является «...удовлетворение потребностей субъектов образовательной деятельности и потребителей образовательных услуг в получении объективной информации о результатах образовательной деятельности, о состоянии и развитии системы общего образования Томской области».

Пунктом 2.3. этого документа «...Потребителями информации, представляемой функциональными структурами СОКО ТО, определяются:

- учащиеся и их родители (законные представители);
- педагогические работники образовательных учреждений и преподаватели вузов;
- методические службы и система повышения квалификации работников образования;
- органы управления образованием;
- работодатели;
- средства массовой информации и другие общественные институты».

Принципиально важно, что в число потребителей включены школьники и работники методических служб. В ситуации жесточайшей экономии многие методические кабинеты были сокращены или реорганизованы (закрыты). Видимо, поэтому крайне редко в последнее время появляются методические рекомендации для педагогов, разработанные в муниципалитетах, да и вообще на уровне области.

Свою главную задачу мы видим в создании и трансляции инструментов и методик, позволяющих получить *объективную* информацию об учебных достижениях учащихся, технологическом, методическом, аналитическом обеспечении использования инструментов и методик. Говоря о внеучебных достижениях, нельзя не сказать о том, что на сегодня нам неизвестен их разумный и достаточный перечень, и, самое главное, механизмы их оценки и возможного сопоставления. Как только научной, педагогической общественностью будет сформулирован этот перечень, можно будет пытаться разработать инструментарий оценки.

Граничные условия нашей деятельности

Примеры международного опыта и не только говорят о бессмысленности тотального контроля знаний как цели системы. Мы пришли к выводу о необходимости использовать педагогические измерения исключительно для совершенствования процессов преподавания учебного предмета: *измерение, оценка с целью коррекции*, а не рейтингования. Региональные мониторинговые исследования уровня обученности для нас не инструмент контроля, а инструмент совершенствования системы. Необходимо различать понятия оценки, контроля, мониторинга и диагностики применительно к качеству образования.

Итак, основные задачи для организации мониторинга на уровне региона мы видим в: разработке качественных измерительных материалов; разработке и внедрении современной технологии сбора, обработки и хранения данных, не отягощающей дополнительными существенными усилиями учителя и завуча; разработке и внедрении механизмов обеспечения достоверности (доверия к результатам), методики анализа полученной информации. И последнее, но главное — отработка рекомендаций и определение граничных условий по использованию результатов.

Измерительные материалы для региональных исследований, внутришкольного контроля

За последние несколько лет разработка измерительных материалов для оценки уровня обученности в России превратилась в достаточно понятную науку. Федеральный институт педагогических измерений (ФИПИ), внедрив передовой международный опыт, использовав наработки отечественных учёных, создал технологию, позволяющую создавать вполне валидные измерители для ЕГЭ и ГИА. Не случайно в последнее время серьёзных нареканий содержание контрольно-измерительных материалов ЕГЭ не вызывает, ушли в прошлое вопросы типа «цвет собачки Герасима?» и тому подобные. К сожалению, практика разработки качественных измерительных материалов существует пока почти исключительно лишь на федеральном уровне: для массового производства качественных измерителей (например, для внутришкольного мониторинга) необходимы специалисты, компетентные не только в своих предметах, но и обладающие знаниями в области разработки тестов (педагогических измерений), а таких специалистов у нас мало.

Надо отдать дань уважения тем муниципалитетам, где проводятся собственные такого рода исследования, но, к сожалению, в силу недостаточности подготовки, опыта и знаний в этой области часто появляются «шедевры». Например, в рамках муниципального мониторинга по биологии методисту было предложено за месяц самостоятельно разработать по два варианта «измерительных материалов» для шести учебников (УМК) по биологии восьмого класса — суммарно более 300 заданий! Как сравнивать результаты таких «измерений», каково их качество, насколько реально можно оценить уровень обученности в данной ситуации, когда даже в рамках одного учебника есть разница в темпах прохождения учебной программы в разных классах?

Рисков при разработке тестов и тестовых заданий очень много — это и отсутствие системы стандартизации измерительных материалов, и не принятые до сих

Б.В. Илюхин

Принцип разумной достаточности при оценке качества образования

пор образовательные стандарты, но главное — крайне низкая культура профессионального сообщества в области педагогических измерений.

Основная идея нашего мониторинга состоит в выявлении пробелов в знаниях не в результате итоговой аттестации, а в ходе входного или рубежного контроля, что даёт возможность внести коррективы в учебный процесс школы не по итоговому результату, а непосредственно по результатам промежуточной оценки. Понимая, что бесполезно ожидать создания органа по масштабной разработке большого количества качественных материалов, за последние три года в Центре мониторинга и оценки качества образования Томского областного института повышения квалификации работников образования создали рабочие группы для анализа содержания разрабатываемых измерителей. Внедрили технологию разработки и комплексной экспертизы измерительных материалов: она предусматривает троекратную содержательную, двукратную тестологическую экспертизы; планируется ввести межрегиональную экспертизу. Ежегодно создаются измерители (более 50 комплектов по два варианта), не зависящие от учебников, для входного и рубежного контроля в 4–10-х классах. Разрабатываются измерительные материалы компетентностного характера (в некоторой степени аналог PIRLS).

Разработка измерителей — крайне серьёзное дело. Себестоимость разработки одного варианта составляет сегодня более 10 тысяч рублей (по крайне низким расценкам) и занимает более полугода.

Любая процедура оценки качества образования затратна, а невнимательное отношение хотя бы к одному элементу оценки приводит к её недостоверности.

Технологические отличия

Давно на всех уровнях говорится, что ЕГЭ и ГИА-9 не единственные инструменты оценки качества образования: они создавались для решения конкретных задач ранжирования выпускников по уровню знаний и определения факта освоения образовательного минимума конкретным выпускником; другие варианты использования этих процедур «притянуты за уши». Результаты ЕГЭ и ГИА очень опосредованно могут быть использованы для коррекции образовательного процесса. Как только речь заходит о технологиях организации массовых процедур оценивания, возникает множество технологических и технических нюансов. Под массовыми процедурами здесь понимается проводимая по единым измерителям и правилам их применения процедура оценивания с общим числом участников от 15 человек. Под это определение попадают региональные мониторинговые исследования, внутришкольный мониторинг, оценивание в рамках процедур контроля качества, часть процедуры аккредитации школы, внутриклассный рубежный контроль, испытания педагогических работников на соответствие занимаемой должности и т.д.

Существуют серьёзные технологические отличия организации различного рода процедур оценивания. Например, технологии бланчного тестирования, используемые для итоговой аттестации школьников, мало приемлемы для внутришкольного контроля из-за сложности технологических процессов обработки бланков ответов. Также неэффективно использовать стандартные типы бланчного тестирования для изучения уровня обученности младших школьников из-за их неготовности многократно переключать внимание с бланка задания на бланк ответов и т.д. При контроле уровня обученности в рамках аккредитации образовательных учреждений измерительные материалы, как правило, калиброванные, используются многократно, и их распространение или разглашение крайне нежелательно.

Учитывая разнообразие запросов различных групп пользователей на автоматизацию любых оценочных мероприятий — от единичных до массовых, в нашем Центре создаются программные продукты и аппаратные средства автоматизации внутришкольного контроля; автоматизации подготовки к обработке результатов массовых оценочных мероприятий; автоматизации квалификационных испытаний педагогических кадров; сравнительного анализа данных различных оценочных мероприятий; автоматизации ввода результатов различных оценочных мероприятий без использования методов бланчного и компьютерного тестирования и многие другие.

Принципиальное отличие нашего подхода к автоматизации этих процессов — не просто ввод и проверка ответов испытуемых, а получение *детальной картины их уровня обученности*. Для этого каждое тестовое задание имеет привязку к элементам кодификаторов содержания и требований каждой отдельной дисциплины (предмета). В результате каждого оценочного мероприятия выявляется процент выполнения заданий по каждому элементу кодификатора (содержания, требований). Дополнительно, в качестве характеристики испытуемого, мы получаем логит его уровня подготовленности. В совокупности это позволяет выявить проблемные зоны (темы, умения), над которыми необходимо детально работать.

Такого рода диагностика должна проводиться регулярно. Её результаты либо не нужно вообще, либо крайне осторожно можно использовать с точки зрения выставления отметки. Её роль должна сводиться к инструменту диагностики проблем для корректирующих воздействий. При этом сами воздействия не обязательно должны быть массовыми (повторение отдельных тем), а скорее индивидуальными, например такими, как индивидуальная образовательная траектория ученика, обучение педагога на тематических курсах повышения квалификации, консультации в рамках учебно-методического объединения по предмету и т.п.

Понимая, что организация такого рода оценочных мероприятий, обработка их результатов в конечном итоге (если не считать дорогостоящие процедуры «с высокими ставками») ляжет на школу, т.е. на завуча, учителя, главной своей задачей в этом направлении мы видим разработку технологий проведения, методов сбора и хранения информации для внутришкольного мониторинга.

К сожалению, до сих пор основным предметом нашей деятельности остаётся оценка уровня обученности. Причина — отсутствие надёжного научно обоснованного инструментария, позволяющего оценить уровень сформированности ключевых компетенций. Те механизмы, которые апробируются и внедряются в настоящий момент, пока, к сожалению, недостаточно технологичны, хотя проводимые рядом исследовательских коллективов работы вызывают наш неподдельный интерес и, безусловно, часть из них будет внедрена нами уже в ближайшее время.

Оценивание для коррекции

Одна из главных проблем российского образования — оценивание для выставления отметок, а не для коррекции обучения. Сегодня в школе получается так,

Б.В. Илюхин

Принцип разумной достаточности при оценке качества образования

что измерений нет, а оценка есть. Образование в России сегодня часто не предполагает объективных измерений, а строится в основном на экспертных оценках. Внедрение измерений в образование — крайне сложный процесс. С другой стороны, есть понимание того, что основным звеном в любой системе оценки качества образования должна быть школа и учитель: именно в школе необходима основная коррекция образовательного процесса и его совершенствование. Рисков здесь очень много: это и невысокая степень заинтересованности и поддержки государством развития систем внутришкольного мониторинга, и боязнь членов коллектива любой школы независимой внешней оценки, и проблемы обеспечения корректности процедуры оценочных мероприятий: любой педагогический коллектив в той или иной форме сопротивляется процедуре оценки, подозревая возможные негативные последствия. Вопрос использования результатов мониторинга для распределения средств — крайне опасный вопрос. Очевидно, что как только мониторинг превращается в процедуру с «высокими ставками», его результаты сразу становятся такими «как надо», причём это «как надо» заложено в психологии нашего учителя ещё с советских времён.

Если на ЕГЭ в ряде субъектов Федерации (в том числе и в Томской области) удаётся обеспечить достоверность, о других процедурах этого, к сожалению, сказать нельзя. В Томской области технологически процедуры проведения ЕГЭ и ГИА-9 в новой форме весьма похожи: работы с открытыми ответами участников этих процедур проверяются анонимно и независимыми экспертами. Обе процедуры проводятся с использованием машиночитаемых бланков ответов, задания для обеих процедур разрабатываются ФИПИ. Разница в процедуре лишь в том, что ЕГЭ проводится с участием множества привлечённых сотрудников вузов в качестве руководителей пунктов проведения экзаменов и организаторов, а ГИА-9 в новой форме проводится территориальными (муниципальными) комиссиями.

Доверие к результатам оценивания

Весной в одном из муниципалитетов Комитетом по надзору в сфере образования Томской области, при активной поддержке Центра оценки качества образования был проведён внешний контроль качества, или «чистый контрольный срез». Привлекли студентов одного из вузов. Контрольные материалы были разработаны по спецификации регионального мониторинга, поэтому полученные результаты можно соотносить с мероприятиями региональных мониторинговых исследований. Отметим, что более шести лет назад такие оценочные мероприятия проводились исключительно для выявления общей картины обученности по области, а также для получения среднеобластного показателя, который потом использовался для аккредитации образовательных учреждений, причём на уровне знака (выше или ниже среднеобластного).

В последние два года основная цель исследования — получение школой достоверной картины уровня обученности для своевременной коррекции образовательного процесса. И получается, что дети, показывающие отличные результаты на мероприятиях регионального мониторинга, которые школа проводит «сама для себя», затем при внешнем контроле скатываются не на несколько процентов, а в несколько раз. Причём для этих оценочных мероприятий измерительные материалы составлены по весьма похожей спецификации. Возможная причина проста и банальна: мысль — «А вдруг за плохие результаты нас накажут?» заложена в нашем сознании ещё с советских времён. Таким образом, особенно остро встаёт вопрос о доверии к результатам различного рода оценочных мероприятий, особенно тех, которые школа проводит самостоятельно.

Анализ полученной информации

Такой анализ может быть весьма затруднён из-за высокой степени «зашумленности» результатов. Для «очистки» полученной информации рекомендуем несколько приёмов:

- использовать результаты, полученные различными способами, — педагогические измерения, статистическая информация, и т.п.;
- использовать математические методы выявления статистически недостоверных результатов;
- снизить уровень психологического давления на исполнителей;
- формализовать процедуры проведения оценочного мероприятия;
- вести просветительскую работу среди организаторов — педагогов, завучей, руководителей школ.

При анализе нас, прежде всего, интересует степень обученности школьника не вообще, а по отдельным элементам содержания дисциплины или по отдельным элементам кодификатора умений (требований). Механизмы анализа, применяемые в настоящий момент, предполагают тестологический анализ по результатам одного оценочного мероприятия с выдачей соответствующих пояснений и рекомендаций, в том числе ранжирование учащихся внутри класса и среди участников данного оценочного мероприятия по различным критериям. Также существует возможность отследить по результатам нескольких оценочных мероприятий показатели усвоения отдельных пунктов кодификатора в динамике.

Для повышения качества и наглядности анализа есть возможность строить лучевые диаграммы индивидуально для ученика или класса в целом, на лучах которых могут откладываться различные показатели (суммарный логит подготовленности, средний процент успешности выполнения заданий оценочных мероприятий и т.п.), по которым можно определить, например, склонность ученика к определённым учебным предметам. Механизмы и инструменты анализа позволяют выдать рекомендации для учителей, завучей, членов учебно-методических объединений.

Управленческие решения

На данном этапе развития необходимо предостеречь «горячие головы» от скоропалительных управленческих решений по результатам процедур оценки качества обученности. Педагогические измерения предполагают для получения достоверной оценки использование очень сложного и весьма дорогостоящего механизма. Исключение из этого механизма даже самого мелкого элемента неизбежно приводит к фатальным для всего измерения последствиям. Надёжных механизмов педагогических измерений с высокой степенью доверия к результатам в стране крайне мало. Управленческие решения типа «высечь», «выговор директору» могут привести только к негативным результатам. Управленческие решения по результатам оценочных мероприятий должны предполагать позитивный характер изменений: например, учителю, дети которого показали в рамках исследования низкий уровень обученности по отдельным разделам, необходимо повышать квалификацию по данным разделам, а не лишать его стимулирующей надбавки. Наш Центр оценки качества образования строит свою работу по использованию результатов оценки на следующих направле-

Б.В. Илюхин

Принцип разумной достаточности при оценке качества образования

ниях: формирующее оценивание; внутриклассное оценивание; корректирующие методики; оценка качества образовательных программ.

Для создания системы оценки качества образования на уровне школы (муниципалитета) необходимо в течение первых двух лет НЕ использовать механизмы рейтингования, финансирования или премирования по результатам оценочных мероприятий (средним баллам ЕГЭ и проч.). Не дай бог ситуации, когда от результатов напрямую зависит финансирование. Поощрение за результат сильных школ приводит к ещё большему расслоению.

Борис Валентинович Илюхин,

руководитель Томского областного центра оценки качества образования