

Системы оценки в Англии и Уэльсе — четыре вопроса и пять уроков



Дэвид Хокер,
профессор, педагогический колледж (College of Teachers),
Великобритания

Четыре вопроса

1. Не слишком ли много тестирования?

На протяжении многих лет Англия считалась одной из самых тестируемых стран (возможно, за исключением США). Верно то, что в системе образования для 16–19-летних учащихся доминируют экзамены, несмотря на многочисленные неуспешные попытки по созданию более эффективной модульной структуры квалификаций, не требующей трёх масштабных экзаменационных систем для школьников в возрасте 16, 17 и 18 лет. Однако за последние годы наблюдается значительное уменьшение объёма формального национального тестирования для учеников младшего возраста. Проблема состоит в том, что результаты тестов 11-летних школьников бросают тень на начальные школы, так как результаты этих тестов публикуются в таблицах рейтингов школ.

При подготовке к тестам большое количество времени урока тратится на изучение техники сдачи теста, а многие аспекты учебной программы при этом игнорируются.

И такая ситуация наблюдается во всём мире. Хотя результаты оценивания учителем теоретически эквивалентны тестам, ответственность не относится к ним так серьёзно, и **Англия не нашла способ решения проблемы доминирования тестов над обучением**. В Уэльсе, где тесты были отменены несколько лет назад, существует **кризис доверия к системе оценки**, так как регион показывает худшие результаты по Великобритании в международном исследовании PISA. Школы обвиняются в недостатке потенциала и отсутствии нацеленности на результат в обучении читательской грамотности, умению работать с числами и понимания естественно-научных явлений.

Итак, можно утверждать, что в английских школах слишком много тестирования, а в валийских школах недостаёт подотчетности. Оба региона в настоящий момент пересматривают свои системы оценки для того, чтобы они работали более эффективно. Однако это осуществляется в различных направлениях.

Проблема тестирования не может быть отделена от проблемы оценивания в классе. А это тесно связано с **качеством подготовки учителей и ясностью целей обучения**. Другими словами, требуется комплексный подход ко всей системе образования. На протяжении последних лет была проведена большая ра-

бота по улучшению педагогических навыков учителей, особенно в области оценивания и отслеживания прогресса учащихся, вовлечения учеников в постановку и мониторинг своих собственных целей обучения. Этот процесс получил название **оценивание для обучения**. Многие международные исследования показали, что эти техники положительно влияют на учебные достижения учащихся, а некоторые учёные утверждают, что со временем они могут заменить формальные национальные тесты. Однако маловероятно, что правительство Великобритании это сделает из-за продолжающейся практики публикации таблиц рейтингов школ.

2. Не пытаемся ли мы достичь слишком многих целей при помощи тестирования?

Национальная система оценки ставит перед собой **четыре основные цели**:

- измерить и отследить прогресс отдельного ученика; это может быть сделано при помощи комбинации стандартизированных тестов, оценивания в классе и самостоятельного оценивания себя учениками;
- общественно признанная аккредитация учебных достижений отдельного ученика для его дальнейшего продвижения на следующий уровень обучения или для поступления на работу;
- предоставить данные для сравнительной оценки деятельности школ, учителей и региональных образовательных систем относительно заданного стандарта;
- отследить результаты работы системы образования, используя комбинацию национальных измерений и международных сравнительных исследований.

При введении национальных тестов в начале 1990-х достигалась первая цель. Однако, как только результаты экзаменов стали собираться централизованно, они стали служить третьей цели и частично четвёртой.

Публикация результатов тестов привела к тому, что учителя стали улучшать результаты тестов за счёт натаскивания учеников на их сдачу, а не за счёт глубокого изучения материала.

В погоне за лучшими результатами тестов в жертву приносились глубина и объём изучаемого материала. Результаты тестов должным образом улучшались, а результаты страны в международных сравнительных исследованиях либо не изменялись, либо ухудшались. Некоторые специалисты утверждали, что жажда улучшения результатов тестов в целом неблагоприятно сказалась на качестве образования.

Представлен **классический случай системы, пытающейся использовать один инструмент для достижения слишком многих целей**. Национальное тестирование сейчас служит средством оценки эффективности работы школ. И это более действенно, чем изначальная задача диагностирования прогресса школьников. Последующие правительства уменьшили количество тестирования: остались только тесты для измерения необходимых математических навыков и грамотности.

Это поле деятельности оставляется школам: они должны вводить способы оценки инди-

видуального прогресса учащихся. Но случится это или нет, зависит от того, насколько эти процессы будут учитываться в общей оценке деятельности школ. Многое будет зависеть и от того, как будет работать система инспекции в будущем.

3. Хороши ли тесты с точки зрения их технического качества, дизайна и администрирования?

Дизайн тестов — сложная тема, включающая проблемы валидности и надёжности. Изначально национальные тесты в Англии и Уэльсе ставили перед собой задачу достижения хорошей **валидности**, и они значительно влияли на процесс обучения. В их основе лежала необходимость как можно более справедливой оценки того, насколько хорошо усвоены предметы школьной программы. Эти же соображения легли в основу систем национальных экзаменов для 16–19-летних учащихся.

Тесты не являются надёжным инструментом измерения степени усвоения школьной программы.

Проблема **надёжности** также важна, если результаты тестов предназначены для публичного использования. Так же как и в случае с инструментами PISA, тест представлял собой комбинацию вопросов с вариантом выбора ответа и открытых вопросов. Таким образом, учащиеся могли продемонстрировать как знание предметного материала, так и возможность использования своих знаний.

Всестороннее испытание тестов гарантировало удаление некорректных вопросов, а интенсивная подготовка специалистов, оценивающих результаты тестирования, улучшала надёжность системы оценок. Для обеспечения стандартизированного проведения тестов создавались детальные инструкции и проводилось обучение школьных администраторов. Опыт Великобритании по этим техническим вопросам — один из лучших в мире. Британские тесты и экзамены, учитывая баланс валидности и надёжности, возможно, лучшие в мире.

Тем не менее очень сложно корректно оценивать некоторые вопросы учебной программы.

Доказано, что невозможно создать надёжный тест по творческому письму или пониманию литературы.

Сложно тестировать предметы естественно-научного цикла из-за объёма материала, поэтому часто вопросы по этим предметам оказываются либо слишком банальными, либо непонятными.

Эти проблемы присущи практике тестирования во многих странах, включая Великобританию, и они подчёркивают необходимость проведения оценивания учителем как необходимой части национальной системы оценки. Ключевым вопросом является проблема того, **как объединить оценку учителя и баллы, полученные за тест**, в общий результат. За последние несколько лет для решения этой проблемы применялись различные подходы. Иногда эти две оценки объединя-

ют различными способами (как в модульном подходе к большинству квалификаций для 16–19-летних); или их выставляют отдельно (как в случае оценивания 11-летних школьников, когда баллы за тест и оценку учителя указывают отдельно, так как считается, что они оценивают различные вещи).

Основной задачей для руководителей является достижение оптимального соотношения валидности и надёжности в свете того, что пытаются достигнуть система оценки.

4. Улучшают ли тесты стандарты обучения?

Тесты, от результатов которых зависит дальнейшая судьба ученика, несомненно, **улучшают способность учащихся их сдавать** и способность учителей готовить к ним своих учеников. Является ли это повышением стандартов, зависит от того, насколько тесты близки к тем стандартам, которые собираются повысить.

Так как учителя учат тому, что будет проверяться в тестах, то ответом на вопрос, повышает ли тестирование стандарты, является ответ на вопрос, насколько содержание тестов отражает учебную программу.

С одной стороны, это больше вопрос образовательной политики, чем собственно системы оценки. В случае Англии и Уэльса национальные тесты и экзамены сосредоточены на оценке базовых навыков, которые

настолько фундаментальны, что овладение ими является необходимым условием достижения высоких стандартов в других областях знания. С другой стороны, более широкий опыт обучения игнорируется, и мотивация учеников к обучению снижается, хотя они обладают всеми необходимыми для этого навыками. Важно не только обладать техническими навыками, настолько же важно обладать желанием учиться. Другими словами, навыки не повысят стандарты без мотивации к обучению.

Эта область, без сомнения, лежит вне зоны любой системы оценки, но именно она наиболее важна для будущего успеха. На протяжении многих лет в Великобритании выражали обеспокоенность относительно того, что **избыток тестирования демотивирует наименее способных детей**. Эти дети воспринимают себя неудачниками в системе образования, что затем губительно сказывается на их последующей жизни. **В любой системе оценки необходимо создать рамку успешности для тех, кто не успевает**, чтобы они могли продемонстрировать свой потенциал. Тогда такая система образования поможет вырастить следующее поколение здоровых, продуктивных граждан.

Пять уроков

1. Ясность цели

Первый урок заключается в том, что нужно ясно понимать, что предоставляет система оценки. Когда в Англии и Уэльсе были введены экзамены на сертификат об общем среднем образовании (GCSE), предполагалось, что это будет выпускным школьным экзаменом, сдав который школьники могли

бы продолжать обучение на более высоком уровне или пойти работать, имея признанную квалификацию.

Но сейчас большинство молодых людей продолжают обучение в школе или колледже до 18 или 19 лет, то есть GCSE больше не является выпускным школьным экзаменом. Он служит мостиком для получения следующих квалификаций (дополнительного сертификата продвинутого уровня и сертификата продвинутого уровня, экзамены на получение которых сдаются в 17 и 18 лет). Поэтому экзамен GCSE может быть значительно сокращён и он может стать модулем пакета квалификаций для 16–19-летних учащихся. Это будет более продуктивно и не так дорого. Но вместо этого результаты экзамена GCSE служат основным критерием оценки работы средних школ, он выступает **скорее как контролирующий фактор для школ, чем квалификация для молодых людей**. Эта ситуация требует основательного пересмотра структуры системы оценки для 16–19-летних учащихся, но, учитывая другие принципиальные соображения, это вряд ли произойдёт.

Аналогично национальные тесты, изначально создававшиеся для отслеживания прогресса детей, стали служить для мониторинга деятельности школ и системы образования. Здесь больше свободы для реформирования системы для достижения изменяющихся целей, и это постепенно происходит. К сожалению, в стране не было авторитетных заявлений о цели национальной системы оценки, и пока остаётся неясным, что правительство пытается достигнуть. Недавний пересмотр тестов для 11-летних учащихся был сделан по поручению правительства. Это даёт основание надеяться, что заявление о целях обучения будет сделано.

Урок для национальных руководителей состоит в том, что они должны иметь ясное представление о различных целях и задачах, решаемых при помощи национальной системы оценки. Вначале необходимо определить цель, которую собираются достигнуть с помощью системы оценки, и только после этого можно начинать работу по созданию такой системы. Во многих странах, включая Россию, национальные экзамены создавались для того, чтобы обеспечить более справедливый отбор студентов в высшие учебные заведения. Такая ясная цель, несомненно, облегчает задачу технического дизайна и администрирования национальных экзаменов. Но если эта цель изменится или появятся еще другие задачи, которые надо решить с помощью системы оценки, тогда необходимо пересматривать дизайн тестов. Аналогично страны, пытающиеся создать свой национальный инструмент для мониторинга успеваемости школьников наподобие PISA, должны задуматься о влиянии такого инструмента на учебный план, особенно при публикации результатов экзаменов.

2. Техническая надёжность

Практически не обсуждается тот факт, что никакая система оценки, в особенности формальные тесты, не может считаться заслуживающей доверия без технической надёжности. Однако всегда есть альтернатива. Например, при создании национального тестирования в Англии и Уэльсе специально не вводилось объективное тестирование (этот вид тестирования распространён в США). Тесты содержали открытые вопросы, так как это в большей степени соответствовало учебной программе и было более благоприятно для обучения. Другими словами, **мы сознательно пожертвовали надёжностью в пользу**

валидности. Однако при публикации результатов школ эта система продемонстрировала свою слабость: результаты одних школ могли быть несправедливо оценены учителями других школ. Для того чтобы уменьшить этот риск, была проделана огромная работа по убеждению общественности в том, что формальные результаты экзаменов — не единственный критерий оценки работы школ. Учитывая то, что возникал риск недоверия к тестам вообще, выполнение этой задачи было непростым делом.

Техническую надёжность можно достигнуть разными способами, но её невозможно достигнуть, используя только один критерий измерения.

Образовательные системы, так же как и человек, являются сложными организмами, которые нужно рассматривать с различных точек зрения. Поэтому системы оценки, заявляющие о своей эффективности, должны включать различные критерии измерений, если они хотят представить справедливую и полную картину.

3. Эффективное управление

Так как оценивание обычно происходит крупномасштабно и затрагивает большое количество учебных заведений и практически всех учащихся, то управление такой системой является комплексным и многосторонним. Для решения этой задачи необходимо эффективное управление. Национальные центры тестирования должны управляться опытными менеджерами, которые понимают всю сложность системы и готовы создавать такие

структуры управления, которые охватывают все аспекты проведения тестирования.

Обычно делается **три вида ошибок** по отношению к национальным центрам тестирования.

1. Эти центры могут иметь **недостаточное количество ресурсов**. У них может не хватать квалифицированных сотрудников. Их менеджеры могут быть неопытными или они могут быть перегружены и не обращать внимания на различные элементы системы оценки. У этих менеджеров могут быть другие обязанности помимо проведения тестирования, которые для них являются приоритетными. Это приводит к тому, что система никогда не будет работать эффективно.

2. **Управление системой оценки может стать слишком жёстким**, будет всё контролироваться. Менеджеры центров тестирования будут слишком бояться сделать ошибку, они не будут обсуждать свои планы с директорами школ, университетами или местной администрацией. Они будут отдаляться от тех, чья поддержка им нужна. В результате система оценки будет искусственно навязываться и поддерживаться за государственный счёт.

3. Может быть обратный подход, при котором **участникам системы оказывается слишком большое доверие**. Это может привести к размытым стандартам управления, ненадёжным результатам и дискредитации системы в глазах общественности.

Золотая середина — это система, при которой центральное агентство тестирования хорошо оснащено, профессионально исполняет свои функции, контролирует систему, но в то же время работает в партнёрстве с другими

участниками образовательной системы. Происходит совместная (до некоторой степени) разработка тестового инструментария, делается упор на работу с местной администрацией, а не на жёсткий набор национальных спецификаций, существует хорошо структурированная программа тренингов и пиар-кампания, чётко объясняющая, что можно и чего нельзя ожидать от системы оценки.

4. Связи с системой образования

Нельзя забывать, что **любая система оценки является частью системы образования**. Успех системы оценки (в отличие от того, что происходит сейчас) зависит от того, в какой степени она отражает и использует такие факторы системы образования, как учебная программа, подготовка и переподготовка учителей, развитие лидерства, организация школы, политика в области высшего образования, требования экономики и гражданского общества. Важно вовлечь как можно большее количество заинтересованных сторон в обсуждение и согласование основных принципов системы оценки, потому что относительно того, что преподаётся и оценивается в школах, требуется своего рода социальный и политический консенсус.

Не существует системы оценки, свободной от оценочных суждений. Выбор того, что включить в тест, как его провести и каким образом сообщить результаты, представляет собой набор оценочных суждений. В идеале эти суждения должны определяться демократическим путем, хотя часто их определяют правительства.

Как менеджер национальной системы оценки в 1990-х, я всегда осознавал, насколько глубоко влияют мои решения на жизни целого поколения молодых людей. И, конечно, я

всегда искал такие решения, которые основывались на фактах и были благоприятны. Разумеется, я работал в рамках той образовательной политики, которая проводилась правительством. Но эта образовательная политика оживлённо обсуждалась как внутри правительства, так и между правительством, заинтересованными сторонами и широкой общественностью. Я не могу сказать, что все мои решения были правильными — это было бы слишком самонадеянно.

Верно одно — мы сделали больше правильных вещей, прислушиваясь к участникам системы образования. Если бы мы не делали этого, то ошибок было бы больше.

5. Опасность ложных стимулов и непредвиденных последствий

Мой последний урок связан с постоянным риском **вмешательства в социальную политику**. Как только вы обнаружили проблему и ввели меры по её решению, тут же возникает другая проблема как прямое следствие того, что вы сделали. В нашем случае **мы создали ложный стимул для школ по натаскиванию детей на выполнение тестов**. Произошло сокращение учебной программы, а это было обратное тому, что мы хотели. Смягчение этих непредвиденных последствий требует других мер, например мотивации школ по расширению учебного плана через участие в специальных программах по усовершенствованию школьной программы.

При вмешательстве в социальную среду надо иметь в виду **теорию эластичности организаций**. Организации похожи на живой организм, они по-своему развиваются и решают свои проблемы. Вмешательство, предписывающее определённые процедуры, может помочь, но может и ослабить организацию, подорвав её веру в способности решить проблему своими силами. Таким образом, вмешательства с добрыми намерениями могут ослабить организацию.

Национальные системы тестирования с этой точки зрения безвредны. Но если они становятся основой для вмешательства во внутренние дела школы, например указывают, как надо учить детей в классе, готовить их к экзаменам, то возникает опасность ослабления школы. В настоящее время в Англии проводится большая работа по укреплению потенциала школ, достижению понимания ими своего места в системе образования, то есть школьная система становится самосовершенствующейся.

Для трансформации школ в постоянно совершенствующиеся организации необходимо уделять внимание развитию лидерских качеств, самоуправлению, мотивации, пониманию педагогики. В этом контексте национальная система оценки будет играть важную, но **вторичную** роль в системе образования. Она будет выступать как один из механизмов сравнения результатов и будет использоваться разумными менеджерами для признания успехов и улучшения результатов обучения. В конце концов, система оценки необходима для обслуживания системы образования, но не наоборот.