**Концептуальная схема (рамка) построения эффективной системы оценивания[[1]](#footnote-1)**

Маргарит Кларк

**Оглавление**

Введение

Теория и практика оценивания

Рамка для систем оценивания

Параметр 1. Виды/цели оценивания

Параметр 2. Факторы качества/выполнения

Наполнение рамки

Ступени и стратегии разработки

Заключение

Ссылки

Приложение 1.Виды оценивания и их ключевые различия

Приложение 2. Техническое задание для проведения самодиагностики

Приложение 3. Вопросники для проведения системы оценивания

Приложение 4. Рубрики, представляющие этапы развития оценивания

*«Оценивание имеет первостепенное значение в образовании: оно влияет не только на показатели количества учащихся, охваченных образованием, и показатели тех, кто окончил школу, но и на конечную цель обучения учащихся» (Всемирный Банк, 2010).*

**Введение**

*Оценивание (assessment)* это процесс[[2]](#footnote-2) сбора и оценки информации о том, что учащийся знает, понимает и умеет, он направлен на принятие обоснованного решения, что делать дальше в образовательном процессе. Сбор данных и методы оценки могут быть такими простыми, как вопросы и ответы (например, «Какова столица Эфиопии?») или же сложными, как компьютерные программы адаптивного тестирования, основанные на многомерных алгоритмах подсчёта баллов и учебном прогрессе[[3]](#footnote-3). Принимаемые на основе результатов оценивания решения могут варьироваться от того, как создать системную программу по совершенствованию преподавания и обучения в классе, как определить последующие шаги в процессе преподавания в классе до определения тех абитуриентов, которые должны быть приняты в университет.

Система *оценивания* представляет собой сочетание политики, различных структур, практик и инструментов, направленных на получение и использование информации об успеваемости учащихся. Эффективные системы оценивания - это те, которые предоставляют информацию надлежащего качества и в необходимом количестве для того, чтобы удовлетворить информационные потребности всех заинтересованных групп и тех, кто принимает решения с целью повышения качества обучения учащихся (Ravela и др., 2009).[[4]](#footnote-4)

Правительства, международные организации и другие заинтересованные группы все больше признают важность оценивания для мониторинга и улучшения обучения учащихся, и как следствие необходимость создания сильных систем оценки учащихся (IEG, 2006; McKinsey & Company, 2007; UNESCO, 2007). Такое осознание связано с растущим подтверждением того, что выгода образования для общества возрастает только тогда, когда имеет место обучение (Hanushek & Woessmann, 2007, 2009; OECD, 2010). Например, увеличение баллов в международных тестах по чтению и математике на одно стандартное отклонение даёт с 2-х процентный прирост годового ВВП на душу населения.

Некоторые возражают, что оценивание, особенно крупномасштабное, слишком дорогое мероприятие. На деле все оказывается наоборот, тестирование зарекомендовало себя как наименее затратная инновация в реформе образования, требующая намного меньше средств, чем увеличение заработной платы учителей или сокращение наполняемости классов, наоборот, у тестирования самый высший показатель по соотношению преимущества-затраты. Hoxby (2002) показал, что даже самые дорогостоящие национальные программы отчетности на базе тестирования в Соединенных Штатах стоили меньше 0.25 центов на одного ученика. Аналогично, ни в одной Латиноамериканской стране, изученной Wolff (2007), на тестирование не было потрачено более 0.3 процента бюджета образования страны при проведении тестирования на уровне начального или среднего образования.[[5]](#footnote-5)

За последние 20 лет беспрецедентное количество стран начали проводить мероприятия по оцениванию или расширять существующие системы оценки (ЮНЕСКО, 2007). В дополнение к этому имеет место огромный рост числа стран, участвующих в международных сравнительных исследованиях по оцениванию, таких как Международные исследования по математике и естествознанию (TIMSS) и Программа международного оценивания учащихся (PISA).[[6]](#footnote-6) Неправительственные организации все больше участвуют в проведении оценивания учащихся с тем, чтобы привлечь общественное внимание к плохому уровню достижений учащихся и подтолкнуть систему к переменам (например, Pratham в Индии).

Несмотря на признание важности обучения и необходимости измерять его результаты, очень немногие страны проводят соответствующую политику, имеют необходимые структуры, практики и инструменты, которые и составляют эффективную систему оценивания. Это особенно актуально для стран с низкими доходами, которые получат наибольшую выгоду в результате систематических усилий по измерению результатов обучения. Некоторые из этих стран экспериментируют с крупномасштабными или другими стандартизированными оценками обучения учащихся, но зачастую это разрозненные усилия, которые не являются частью образовательной стратегии и не являются продолжающимися. Главное различие между системой разового оценивания и системой постоянного оценивания состоит в том, что первая дает, в лучшем случае, статическое представление (снимок) достижений ученика, в то время, как последняя, если проведена должным образом, дает возможность мониторинга тенденций (как серия снимков), дает возможность лучшего понимания вклада различных составляющих и образовательных практик с целью изменения имеющихся трендов. Разовые оценивания могут иметь шоковое значение и послужить толчком для обсуждения качества образования, а это может быть приемлемым в качестве стратегии на краткосрочную перспективу с целью включения обучения в повестку дня.[[7]](#footnote-7) В конечном счете, однако, правительствам придется иметь дело с задачей, которая требует напряжения всех сил, но является необходимой, с задачей создания систем, которые позволят проводить регулярный мониторинг и поддержку процесса обучения учащихся. Это единственный путь для использования всех возможностей оценивания.

Назначение данной статьи – помочь странам понять некоторые *ключевые принципы и характеристики эффективной системы оценивания.* Акцент делается на оценке достижений учащихся на уровне К-12.[[8]](#footnote-8) В работе приводятся принципы и основные положения из опыта разных стран, профессиональные стандарты и основа современных исследований, направленных на то, чтобы предоставить тем, кто вырабатывает политику в области образования, кто отвечает за развитие системы, *рамку и ключевые индикаторы для диагностики, обсуждения и выработки консенсуса относительно того, как строить надежную и рациональную систему оценивания, которая будет способствовать качеству образования и обучения учащихся*.

**Теория и практика оценивания**

Основная предпосылка исследований в области оценивания заключается в том, что правильная оценка и правильное использование полученных в ходе оценки данных, приводит к лучшим результатам, касается ли это улучшения качества обучения или более взвешенного принятия политических решений (например, Heubert & Hauser, 1999).[[9]](#footnote-9) Что понимать под «правильным», это набор теоретических и технических положений для тех, кто разрабатывает тесты и использует информацию тестирования (AERA, APA & NCME, 1999).

Существует значительный объем эмпирических исследований, которые показывают преимущества определенного вида тестов обучения учащихся, когда они проводятся и используются правильно. Например, исследования демонстрируют связь между высококачественным формативным внутриклассным оцениванием и более хорошими результатами обучения учащихся, которые измеряются по показателям стандартизированных тестов. Обобщения Блэка и Уильяма (1998) более 250 исследований, проведенных по всему миру о влиянии эффективного внутриклассного оценивания, показывают приращение от половины до целого стандартного отклонения при проведении стандартизированных тестов, причем наибольшее приращение наблюдается у плохих учеников.[[10]](#footnote-10) Это имеет важное значение для устранения различий в достижениях различных групп учащихся.

Изучение результатов выпускных экзаменов показывает связь между странами, которые проводят такую политику и имеют более высокие показатели в международных сравнительных исследованиях, таких, как PISA и TIMSS (Bishop, Mane & Bishop, 2001). В то же время экзамены такого рода были предложены ученикам из неблагоприятных групп и они имели непропорционально отрицательное воздействие, тем самым ограничивая их возможности продолжения обучения на следующей ступени системы образования или извлечения пользы из каких-то образовательных возможностей (Greaney & Kellaghan, 1995; Madaus & Clarke, 2001). Поэтому использование результатов экзаменов и сами результаты должны тщательно отслеживаться на уровне системы, группы и на индивидуальном уровне и необходимо предпринять усилия по сокращению или смягчению непредвиденных негативных последствий.

Изучение показывает слабую, но несомненную связь между применением данных крупномасштабного оценивания для того, чтобы школы и участники образовательного процесса были подотчетны, и более хорошими результатами обучения учащихся (Carnoy & Loeb, 2002). Ключевыми составляющими того, является ли влияние оценивания скорее положительным, чем отрицательным, оказались степень того, насколько соотносятся дизайн теста и использование полученных при тесте баллов, и степень возможности оказать поддержку школам или учителям, которые пытаются хорошо работать, но имеют низкие показатели (Ravela, 2005).

Исследования все больше фокусируются на характеристике *систем* эффективного оценивания, которые включают вышеперечисленные типы оценивания и использования его результатов (например, внутриклассное оценивание, экзамены, крупномасштабное оценивание). Данное исследование рассматривает принципы и лучшие практики оценивания, которые имеются в литературе по оцениванию, а также содержит анализ систем оценивания стран, показывающих высокие результаты. Darling-Hammond и Wentworth (2010) сделали обзор практических действий, используемых в системах с высокими показателями по всему миру (например, Австралия, Финляндия, Гонконг, Сингапур, Швеция, Великобритания) и сделали вывод, что системы оценивания в этих странах:

* иллюстрируют важность оценивания самого обучения, для обучения и оценивания как обучение, а не оценивания как отдельного оторванного элемента образовательного производства;
* обеспечивают *обратную связь* с учащимися, учителями и школами относительно того, что было изучено, и *«выдают»* информацию, которая способна формировать планы дальнейшего обучения, а также помогать в принятии решений относительно выбора высших учебных заведений и построения собственной карьеры;
* тесно *увязывают* ожидания от учебного плана, предмета и критерии показателей с желаемыми результатами обучения;[[11]](#footnote-11)
* *привлекают учителей* к разработке оценивания и его проведению чтобы повысить их профессионализм и способность помогать учащимся в обучении и достижении результатов;
* *привлекают учащихся* к аутентичному оцениванию чтобы повысить их мотивацию и результаты учебы;
* стремятся продвигать обучение учащихся навыкам, которые стимулируют мышление высокого порядка и умение решать проблемы с помощью широкого круга образовательных и оценочных стратегий;
* оказывают привилегию качеству, а не количеству стандартизированного тестирования;[[12]](#footnote-12)
* используют все больше и больше в системе экзаменов *открытые задачи* и *оценку школы*, а это требует, чтобы ученики тщательно писали свои работы и дает им возможность овладеть навыками 21-го века.[[13]](#footnote-13)

В то время как данное исследование знакомит нас с тем, как выглядит эффективная система оценивания, в нем не говорится, что заимствуется для того, чтобы проникнуть туда. Другие исследования углубляются в вопросы создания потенциала и выработки стратегии (например, Greaney & Kellaghan, 2008). К примеру Ferrer (2006) дает рекомендации по созданию надежной и рациональной системы оценивания на основе анализа существующих в Латинской Америке систем. Steward и Bray (1998) проводят аналогичный анализ экзаменов в средней школе. Другие (к примеру, Lockheed, 2009) оценивают статус донорской деятельности в данной области и обсуждают, как усовершенствовать эффективность такого рода поддержки странам.

Данная работа сводит воедино все эти факты чтобы создать унифицированную рамку для понимания, что же такое эффективная система оценивания и как страны могут приступить к созданию таких систем.

**Рамка для систем оценивания**

Чтобы подойти к рамке с точки зрения стратегии нам необходимо определить некоторые ключевые параметры систем эффективного оценивания. Два основных параметра обсуждаются здесь: (i) виды/цели оценочных мероприятий и (ii) качество данных мероприятий.

**Параметр I. Виды/цели оценивания**

Существует такая тенденция, что системы оценивания созданы из трех главных видов оценочной деятельности, которые соответствуют трем основным информационным потребностям или целям (см. приложение 1). Эти оценочные мероприятия и потребности в сопутствующей информации/цели следующие:

* *внутриклассная оценка* для предоставления информации в реальном времени с целью оказания поддержки преподаванию и обучению в конкретном классе;
* *экзамены* для принятия решений относительно индивидуального прогресса учащегося в рамках системы образования (например, сертификация или отбор), включая предоставление «редких» образовательных возможностей; и
* *крупномасштабное оценивание* для мониторинга и предоставления релевантной для политиков и практиков информации о деятельности системы на всех уровнях, об изменениях на этих уровнях и о факторах, которые с ними связаны или которые способствовали этим изменения.

Безусловно, данные виды оценивания/цели не являются совершенно независимыми друг от друга; они также не всеобъемлющие (например, есть такие виды оценивания, которые не целиком подходят под данные определения). В то же время они представляют три основные вида оценочных мероприятий, которые проводятся в большинстве систем образования по всему миру.

Внутриклассное оценивание, которое также рассматривается как непрерывное, проводится учителями и учащимися с процессе ежедневных занятий (Airasian & Russell, 2007). Они включают в себя разнообразие стандартизированных и не стандартизированных инструментов и процедур по сбору и интерпретации письменных, устных материалов и других форм данных о достижениях учащихся. К ним относятся вопросы и ответы, домашние задания, презентации учеников, диагностические тесты и контрольные опросы . Главная задача этих оцениваний ─ дать информацию «в реальном времени» для того, чтобы помочь преподаванию и обучению. К ним относится оценивание *для* обучения (например, определяя последующие шаги в процессе обучения, основанном на том, что ученик уже знает и умеет) и оценивание *как* обучение (помогая ученикам реагировать на реалии обучения так, чтобы они лучше знали, что они изучают, как они изучают и что им помогает учиться).

Экзамены под всевозможными названиями «публичные», «внешние» или «по завершении курса» дают информацию для принятия важных решений о судьбе ученика, например, перейдет ли он в следующий класс, надо ли ему посещать особую школу или учебную программу, окончит ли он старшую школу или поступит в университет ( Greany & Kellaghan, 1995; Heubert & Hauser, 1999). Руководство экзаменами может быть внешним или (как все чаще происходит) внутришкольным, их стандартизированная природа подразумевает, что всем учащимся даются равные возможности показать то, что они знают и умеют в рамках учебного плана или другого определенного свода знаний и умений (Madaus & Clarke, 2001). Хорошим примером являются выпускные экзамены на аттестат зрелости по окончании обязательного образования во многих системах образования. Экзамены в основе своей относятся к программам оценки с высокими ставками и, как следствие, оказывают негативное влияние на систему, ориентируя на то, чему учить (приводя к тому, что развивается практика «учить выполнять тесты» или даже «учить сами тесты») и что изучать, влияя в худшую или лучшую сторону на знания и умения выпускников (West & Crighton, 1999). Такие тесты имеют потенциально негативные последствия для конкретного ученика, особенно из неблагополучных групп, поскольку его могут исключить из школы, которую он выбрал (или вообще исключить из системы образования) на основе полученных отметок (Greaney & Kellaghan, 1995). При решении вопроса об использовании подобных тестов необходимо рассматривать такие последствия.[[14]](#footnote-14) Важно обратить внимание, что существуют очень специфические профессиональные и технические стандарты должного и неправильного использования экзаменов (и тестов вообще) для принятия определяющих решений о судьбе учащегося (AERA, APA, & NCME, 1999).

Крупномасштабные оценивания создаются для предоставления информации о работе системы на всех уровнях и связанных с этим факторах (Greaney & Kellaghan, 2008; Kifer, 2001), обычно относительно согласованного набора стандартов или целей обучения чтобы информировать образовательную политику и практику. Примеры включают в себя международные оценивания достижений учащихся, такие, как TIMSS, PIRLS и PISA; региональные оценивания, такие, как PASEC во франкоговорящей Африке, SACMEQ в англоговорящей Африке и LLECE в Южной Америке; национальные оценивания как SIMCE в Чили; территориальные оценивания, такие, как оценивания на уровне штатов в США и Канаде. Такие оценивания отличаются по возрасту учеников и классам, которые тестируются, по охвату тестируемого населения (выборка или перепись), по предметам или навыкам и умениям, по типам собираемых исходных данных и по частоте их проверки. Они также разнятся по тому, как представляются и используются результаты. Например, когда одни перестают отчитываться о результатах перед теми, кто вырабатывает политику или перед общественностью, другие используют результаты чтобы делать подотчётными специальные группы в системе образования (Clarke, 2007). Ravela (2005) описывает использование результатов крупномасштабного национального оценивания в Уругвае для помощи учителям в улучшении процесса преподавания. Акцент на созидательное применение результатов в классе способствовал восприятию результатов учителями; это также повлияло на разработку дизайна оценивания с точки зрения необходимости применить «принцип переписи» для сбора данных и использовать фоновые факторы для контроля за внешкольными факторами, влияющими на достижения учащихся. Помощь со стороны Всемирного Банка в проведении оценивания за последние 20 лет (Larch & Lockheed, 1992; Liberman, in progress) изменилась, и ее акцент сместился с реформы экзаменов на проведение крупномасштабных оцениваний с целью мониторинга тенденций в достижениях учащихся и информирования политиков и практиков.

Один из возможных путей дифференцировать три перечисленных выше типа оценивания состоит в том, что при внутриклассном оценивании речь идет в основном об оценивании *как* процессе обучения или оценивании *для* обучения (следовательно, по существу формативное оценивание), тогда как экзамены и обследования относятся к оцениванию самого изучения (следовательно, по сути суммативное оценивание). На практике такие различия не всегда четко обозначены и все чаще встречаются гибридные подходы. Например, у Сингапура система оценивания структурирована вокруг публичных экзаменов, но создана целая инфраструктура поддержки обучения вокруг нее (L.Benveniste, personal communication, March 2010). Другие разнородные действия включают адаптацию инструментария, разработанного для одного вида оценивания (например, внутриклассные инструменты для передачи данных о преподавании), для другой цели (например, документального подтверждения успеваемости на системном уровне). Одна из наиболее известных инициатив такого рода - Оценка чтения на ранней ступени (EGRA), инструментарий разработан при поддержке донорских агентств и экспертов для использования в развивающихся странах (<http://www/eddataglobal.org/>). EGRA изначально основывалась на инструментарии, специально разработанном для внутриклассного использования, а теперь EGRA используется для сбора данных системного уровня об успеваемости учащихся по освоению навыков раннего чтения для того, чтобы информировать министров и доноров о потребностях системы с целью совершенствования преподавания. В последнее время EGRA функционирует как отдельно стоящее оценивание, которое ценится в первую очередь тем, что способно ставить на повестку дня вопросы качества образования и обучения учащихся. Однако в долгосрочной перспективе это оценивание скорее всего будет работать лучше, если будет использоваться как часть вертикально выстроенных серий оценивания системного уровня, проводимых на ключевых этапах обучения, что позволит анализировать тренды во времени или при переориентировании будет инструментом для внутриклассного использования.

Системы образования могут иметь очень разные конфигурации в этих трех типах оценивания, они могут отличаться настолько, насколько разнятся цели оценивания и связанное с ними применение результатов. Нет ни одной идеальной конфигурации. Например, система образования Финляндии делает акцент на внутриклассное оценивание, как ключевой источник информации об успеваемости ученика, и обращает меньше внимания на экзамены и крупномасштабные оценивания. Индия и Китай традиционно уделяют больше внимания на экзамены, как средство сортировки и отбора большого числа учащихся, и относительно мало внимания уделяют внутриклассному или крупномасштабному оцениванию (хотя ситуация меняется).[[15]](#footnote-15)

**Параметр 2. Факторы/ параметры качества**

Вместо того, чтобы рассматривать один идеальный «профиль» для системы эффективного оценивания, возьмем за главное для рассмотрения индивидуальное и объединенное качество оценочных действий с точки зрения адекватности получаемой информации для помощи в принятии решений (Messick, 1998; Shepard, 2000).

Есть три основных фактора, которые обуславливают значимость/качество информации в системе оценивания (AERA, APA, & NCME, 1999; Darling-Hammond & Wentworth, 2010):

* *благоприятная внешняя среда,*
* *координация и согласованность с элементами образовательной системы,*
* *технология*.

Несмотря на то, что эти показатели тесно связаны между собой, они представлены отдельно для разъяснения и обсуждения.

Благоприятная внешняя среда относится к более широкому контексту, в котором происходит оценивание и к тому, насколько такой контекст способствует и помогает проведению оценивания. К нему относятся такие области, как более широкие законодательные и политические рамки оценочной работы; институциональные и организационные структуры для создания, выполнения оценивания и использования его результатов[[16]](#footnote-16); наличие стабильных и достаточных источников финансирования; и обученный для проведения оценивания персонал и школьные учителя. Важно понять правильно, что такое благоприятная внешняя среда, поскольку это один из ключевых факторов долгосрочного качества и эффективности системы оценивания и – так же, как земля, вода и воздух, растение должно расти – ни одна система оценивания не будет устойчивой при ее отсутствии (Всемирный Банк, 2010). В большинстве случаев долг правительства обеспечить видение, руководство и политическую основу для установления благоприятной внешней среды, которая потом будет претворена в жизнь через общественно-частные партнерства. Некоторые образовательные системы, особенно в федеральном контексте, объединяют усилия для создания благоприятной внешней среды путем объединения ресурсов или институциональных соглашений для разработки тестов, их проведения, анализа или составления отчетов об их результатах. Мероприятия по оцениванию на региональном уровне, такие как SACMEQ, PASEC и LLECE представляют другую форму сотрудничества ради создания благоприятной среды. Продуктивность, достигнутая таким сотрудничеством, способствует большей рентабельности при разработке тестов высокого качества и включению в процесс тестирования технологических новинок.[[17]](#footnote-17)

Координация и согласованность с образовательной системой показывает, насколько система оценивания скоординирована и согласована с остальными компонентами системы образования. Она включает связь оценочных мероприятий и системных целей обучения, стандартами и учебным планом, учебниками и методическими материалами, профессиональным развитием учителя (Fuhrman & Elmore, 1994; Smith & O’Day, 1991). Для оценочных мероприятий важно быть скоординированными с другими компонентами системы образования чтобы информация, которую они предоставляют, использовалась для повышения качества всей системы образования.

Рассмотрение согласованности и координации систем оценивания включает в себя:

* Сфера охвата ─ насколько широко мероприятия по оцениванию предоставляют информацию о достижениях учащихся и их обучении в части учебного плана вообще и ключевых знаний, умений и компетенций, в частности;
* Охват ─ насколько полно деятельность по оцениванию предоставляет информацию обо всех учащихся во всех классах;
* Практичность ─ насколько оценочные действия соотносятся или полезны/применимы к целям обучения и приоритетам заинтересованных групп.

Несомненно, что согласованность с элементами образовательной системы включает в себя гораздо больше, чем соответствие того, что тестируют, с тем, что в учебном плане. Поэтому в то время как соответствие между содержанием образования в стране и тем, что тестируется в ходе международных сравнительных исследований, таких, как PISA, и TIMSS, может быть низким, оценивание все равно может быть согласовано и скоординировано (и полезно для информирования) с общими целями системы образования. Несомненно, использование данных TIMSS, PIRLS и PISA для определения того, что способствует успеваемости, и для мониторинга влияния реформ на успеваемость во временном отрезке, является ключевым для совершенствования уровня достижений стран, таких разных как Бразилия, Иордания и Польша.

Технология относится к психометрическому качеству инструментария, процессов и процедур, применяемых при оценочной работе (AERA, APA, & NCME,1999). Важно отметить, что технология относится не только к крупномасштабному оцениванию, а к любому оцениванию. Технология охватывает такие вопросы, как *дизайн и проведение* оценивания, экзаменационные вопросы, анкеты; анализ и интерпретация ответов учащихся на оценочные действия, вопросы или задания; и правильность того, как результаты оценивания, экзамена или опроса *представляются в отчетах и используются* (Heubert & Hauser, 1999; Shepard, 2000). В зависимости от оценочной деятельности используются строго соответствующие критерии, и они каждый раз разные. Технология важна, т.к. если оценивание ненадежное в части дизайна, проведения, анализа, интерпретации или использования, то его результаты повлекут некорректные решения в отношении обучения учащихся и качества системы (Messick, 1989; Wolff, 2007). Для любого оценивания важны два тесно перекликающихся технологических вопроса: надежность и валидность.

*Надежность* относится к тому, дает ли оценивание безошибочные данные, это особенно критично для судьбоносных экзаменов и мониторинга трендов во времени.

*Валидность* – если баллы, полученные в ходе тестирования, отражают то, что они должны отражать и если они могут быть использованы так, как было намечено. Одной общей угрозой валидности баллов тестирования является различие в языке преподавания и языке тестирования, когда ребенку может быть трудно показать что он знает и умеет. Применение – очень важное направление, когда дело касается достоверности, требуется тщательное рассмотрение последствий использования баллов тестирования, включая социальные, экономические и другие последствия и их влияние на различные группы населения.

Комбинируя факторы, обуславливающие качество, с различными видами/функциями оценивания мы получим концептуальную схему (рамку), показанную в табл.1.

**Таблица 1. Рамка создания эффективной системы оценивания**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Факторы, обуславливающие качество |  | Виды/функции оценивания | | |
| Внутриклассное оценивание | Экзамены | Крупномасштабное оценивание |
| Благоприятная внешняя среда |  |  |  |
| Координация и согласованность с элементами образовательной системы |  |  |  |
| Технология |  |  |  |

Источник: автор

Далее в статье будет представлено содержание данной рамка и будет обсуждаться ее использование в качестве «дорожной карты» для построения более эффективной системы оценивания. Дорожная карта может применяться к системе оценивания любой страны как путь, определяющий в чем система сильна и над чем нужно больше работать.

**Наполнение рамки**

Рамка, представленная в таблице 1 служит отправным пунктом для определения индикаторов, которые могут быть использованы для пересмотра систем оценивания и планирования их улучшения. Индикаторы можно определить на основе набора критериев, например,

* профессиональные стандарты для оценивания; и
* эмпирические исследования характеристик эффективных систем оценивания, включая анализ характеристик, которые отличаются у стран с низкими и высокими показателями.

Там, где нет профессиональных стандартов или ограничены эмпирические исследования, мы можем выбрать индикаторы на основе дополнительных трех критериев: (i) теория – есть общее мнение экспертов, что конкретно способствует эффективному оцениванию; (ii) ресурсы – большинство правительств делают ощутимые вложения в данную область; и (iii) применимость на практике – шаги могут быть предприняты в надежде, что в результате улучшится качество оценивания.

Фактическая база сильнее в одних областях по сравнению с другими. Например, существует много профессиональных стандартов для технологии проведения внутриклассной оценки, экзаменов и крупномасштабных оцениваний (APA, AERA, & NCME, 1999). Есть также значительная исследовательская база по соответствию и координации с другими элементами системы (например, Fuhrman & Elmore, 1994; Hamilton, Stecher, & Klein, 2002). Намного меньше данных об основе профессионального или эмпирического исследования благоприятной внешней среды, поэтому нам приходится полагаться на анализ словесного описания существующего оценивания в данной области.

Вышеперечисленные критерии были использованы для определения индикаторов, которые обуславливают благоприятную внешнюю среду, согласованность с другими элементам системы и технологию, как показано в табл.2. Эти индикаторы наиболее релевантны для проведения экзаменов и крупномасштабных оцениваний, при некоторой модификации могут применяться и для внутриклассной оценки.   
**Таблица 2. Рамка создания эффективной системы оценивания с индикаторами**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Факторы, обуславливающие качество |  | Виды/функции оценивания | | |
| Внутриклассное оценивание | Экзамены | Крупномасштабное оценивание |
| Благоприятная внешняя среда | Политика  Финансовые ресурсы  Организационные структуры  Человеческие ресурсы | | |
| Координация и согласованность с элементами образовательной системы | Учебный план  Учебники и методические материалы  Подготовка и повышение квалификации учителя | | |
| Технология | Дизайн  Администрирование  Анализ  Применение | | |

Источник: автор

Данные по некоторым из этих индикаторных зон можно найти в официальных документах, опубликованных в отчетах (например, Ferrer, 2006), научных статьях (например, Braun & Kanjee, 2005) и в базах данных в режиме online. [[18]](#footnote-18) Однако,[[19]](#footnote-19) большей частью данные не собраны в каком-то систематизированном полном виде. Те, кто хочет изучить такого рода информацию об определенной системе оценивания, скорее всего должен будет самостоятельно собирать данные. Необходимую информацию можно собрать несколькими путями, например,

* Самодиагностика
* Стандартизированные обследования
* Интервью экспертов и их оценка.

Некоторые из этих подходов (например, стандартизированные обследования) больше подходят для сбора данных, целью которого является отображение и сравнение всей структуры системы оценивания, тогда как другие (например, самодиагностика) больше подходят для обзора, обсуждения и планирования на национальном уровне.

В приложении 2 приведено Техническое задание на проведения самодиагностики, которое было разработано для использования на уровне страны.[[20]](#footnote-20) Команды стран могут использовать данный инструмент для лучшего понимания, каков статус и потребности оценивания в их стране и для планирования, куда идти дальше. В Приложении 3 даны анкеты для сбора данных по проведению внутриклассного оценивания, экзаменов и мероприятий по крупномасштабному оцениванию в конкретной стране. Как самодиагностика, так и опросники (анкеты) основаны на рамке и индикаторных зонах, показанных в Табл.2

**Ступени и стратегии разработки**

Приложение 4 содержит набор рубрик, которые тоже основаны на рамке и индикаторах, показанных в Табл. 2. Есть отдельная рубрика для каждого из трех оцениваний: внутриклассного оценивания, экзаменов и крупномасштабного оценивания. Эти рубрики могут использоваться вместе с данными, полученными в ходе анкетных опросов (Приложение 3) для анализа качества проведения оценивания в любой заданной стране Анализ дает понятие уровня качества оценивания в стране относительно лучшей или рекомендуемой практики проведения данного оценивания. Базовая структура таких рубрик представлена в Табл.3.

**Таблица 3. Рамка для рубрик качества оценивания**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Уровень развития/качества проведения оценивания | | | | | |
| Область/Размер | Зрелая (лучшая практика) | Созданная (минимум желаемых стандартов) | Создаваемая (на пути к тому, чтобы отвечать стандартам) | Латентная (отсутствие определяющих признаков) | Ключевые вопросы | Логическое обоснование/ обоснованность |
| **БС – Благоприятная внешняя среда** | | | | | | |
| БС1 - Политика |  |  |  |  |  |  |
| БС2 – Финансовые ресурсы |  |  |  |  |  |  |
| БС3 – Организационные структуры |  |  |  |  |  |  |
| БС4 – Человеческие ресурсы |  |  |  |  |  |  |
| **КС – Координация и согласованность с элементами образовательной системы** | | | | | | |
| КС1 – Учебный план |  |  |  |  |  |  |
| КС2 – Учебники и методические материалы |  |  |  |  |  |  |
| КС3 –  Подготовка и повышение квалификации учителей |  |  |  |  |  |  |
| **T - Технология** | | | | | | |
| T1 – Дизайин |  |  |  |  |  |  |
| T2 -Администрирование |  |  |  |  |  |  |
| T3 - Анализ |  |  |  |  |  |  |
| T4 - Использование |  |  |  |  |  |  |

Источник: автор и M.Ramirez

Для каждого индикатора рубрика показывает четыре уровня системы оценивания – *латентная, на стадии создания, уже созданная и зрелая*.[[21]](#footnote-21). Каждый уровень сопровождается описанием характерных индикаторов. *Латентный у*ровень – самый низкий, он показывает отсутствие определяющих признаков. Следующий уровень – *создаваемый*, здесь частично присутствуют определяющие признаки. *Созданный* - уровень с минимальным набором желаемых показателей и *зрелый* уровень, представляющий идеальные или лучшие практики. Не все вопросы из опросников Приложения 3 представлены в рубриках Приложения 4; это связано с тем, что не все вопросы на практике демонстрируют связь уровня и индикаторов с изменением качества оценочных мероприятий.

В дополнение к изучению того, как работают конкретные индикаторы, может быть полезно провести разметку опорных точек (benchmrking) и сравнить систему оценивания в целом по индикаторами, представленными в Приложении 4, со стилизованным наброском или профилем систем оценивания, как они выглядят на разных стадиях развития. В Табл. 4 приведены некоторые общие профили моделей систем оценивания на уровне создания, существования и зрелости системы (латентный уровень опускается, т.к. при нем отсутствуют какие-либо оцен6очные действия). При стилизации они достаточны для того, чтобы выделить ключевые различия между менее и более продвинутыми системами оценивания.

**Таблица 5. Уровни развития систем оценивания**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Создаваемая** | **Созданная** | **Зрелая** |
| **Благоприятная внешняя среда** | ● Нет или ограниченная политическая рамка | ● Наличие политической рамки |  |
|  | ● Мало подготовленного персонала; высокая текучесть кадров | ● Учебные программы/обученный персонал и низкая текучесть кадров | Те же, что и для Созданной  + сильный акцент на: |
|  | ● Ненадежное финансирование | ● Стабильное финансирование |  |
|  | ● Неясные институциональные системы/мероприятия | ● Четкие институциональные структуры/мероприятия | * Оценивание для обучения * Внутриклассное оценивание |
|  |  |  |  |
| **Координация и согласованность с элементами системы образования** | ● Оценивание не полностью соответствует целям обучения | ● Оценивание соответствует целям обучения | * Роль учителей * Инновационные и основанные на исследованиях практики |
|  | ● Оценивания не сихронизированы с реформами в других областях | ● Оценивание сихронизировано с реформами в других областях |  |
|  |  |  |  |
| **Технология** | ● Ограниченное использование результатов | ● Систематическое использование результатов для информирования политиков и практиков |  |
|  | ● Ограниченное знание и применение технологических или профессиональных стандартов | ● Знание и применение технологических или профессиональных стандартов |  |

Источник: автор

Системы оценивания, которые находятся на стадии *Создания* как правило имеют благоприятную внешнюю среду, так же, как соответствие другим элементам системы образования и технологию, которые находятся в процессе возникновения или становления. Для таких систем характерны нестабильность и неопределенность выбора, частоты проведения оценивания и применения его результатов; неясное представление оценивания на уровне системы и неопределенное и недостаточное финансирование оценочных мероприятий. В таком контексте оценивание функционирует как некое «добавление» к системе, без систематических усилий, направленных на то, чтобы система соответствовала учебному плану, учебным материалам или профессиональной подготовке учителей. Создание потенциала для проведения оценивания непланомерное и имеет ограниченную эффективность, поскольку люди уходят в другие организации или в частных сектор после того, как прошли подготовку. Технологическое качество мероприятий по оцениванию имеет тенденцию к низкому уровню, т.к. отсутствует знание профессиональных стандартов и внимание к ним.

Системы оценивания на *Созданной* стадии имеют тенденцию к наличию благоприятной внешней среды, а также соответствие и технологию, которые стабильны, гарантированы или «созданы». Для таких систем характерны непрерывность и определенность выбора, частоты проведения и использования мероприятий по оцениванию; а также стабильные и достаточные финансовые ресурсы; что является признаком видения и «выгодной сделки» на системном уровне. В такой внешней среде оценивание работает скорее как нераздельная часть системы, предпринимаются систематические усилия для того, чтобы привести его в соответствие с учебным планом, учебными материалами и подготовкой учителей. Создание потенциала имеет тенденцию быть сфокусированным, устойчивым и эффективным, текучесть кадров низкая. Мероприятия по оцениванию имеют хорошую технологию благодаря знанию профессиональных стандартов и вниманию к ним. Эта стадия рассматривается как минимальный стандарт для действительно эффективной системы оценивания.

Системы оценивания на *Зрелой* стадии имеют благоприятную внешнюю среду, а также соответствие и координацию с другими элементами системы и технологию, которая хорошо разработаны, продвинутая и «зрелая» по своей природе. В дополнение к параметрам, которые характерны для *Созданной* системы, *Зрелые* системы имеют высокий уровень инноваций и практик, которые основаны на исследовании. В такой среде оценивание работает как неотъемлемая часть всей системы. Создание потенциала во многом сконцентрировано на учителях, а не только на «техническом персонале», что свидетельствует об акценте на внутриклассное оценивание.

Стоит отметить, что система может быть на разных стадиях развития в зависимости от типов оценочных мероприятий, т.е. система может быть *Созданной* в области проведения экзаменов, но *Создаваемой* в области проведения опросов и наоборот. Хотя обычно лучше чтобы система оценивания была продвинутой как можно в больших областях, нет ясности относительно жизненной важности функционирования на *Зрелой* стадии в каждом аспекте. Таким образом, можно рассматривать *Созданный* уровень, как минимально желаемый для достижения во всех областях оценивания, а в тех областях оценивания, которые вносят наибольший вклад в национальное видение или приоритеты образования, стремиться к бόльшему.

Полезно иметь представление о том, что представляют собой системы оценивания на разных стадиях, но также важно понимать, каким образом можно достичь прогресса в переходе от одной стадии к другой. Поэтому нам необходимо понимать ключевые реформы или усилия, которые страны предпринимают для создания более эффективных систем оценивания.

Главным фактором того, что система переходит от *Создаваемой* ступени к *Созданной*, является осознанный фокус на реформы, усилия и практики, которые укрепляют благоприятную внешнюю среду оценивания (Ferrer, 2006).[[22]](#footnote-22) При анализе проектов Всемирного Банка в области оценивания, которые проходили в разных странах, Larach и Lockheed (1992) обнаружили, что проекты, которые фокусировались на улучшении институционального качества прежде чем переходить к вопросам качества технологии или диссеминации, были более успешны чем проекты, где сначала пытались улучшить качество технологии или диссеминации. Аналогично, при изучении реформ оценивания в странах Восточной Европы и Центральной Азии West и Crighton (1999) заметили, что реформы имеют лучшие шансы там, где есть общественный консенсус относительно необходимости преобразований и ясная и постоянная политическая поддержка преобразований, достаточное выделение ресурсов и усилия по привлечению педагогов-профессионалов.

Главный фактор, характеризующий переход системы от *Созданной* к *Зрелой* стадии, является акцент на реформы, усилия и практики, которые во главу угла ставят класс, учителей и учащихся, как ключевых действующих лиц процесса оценивания (Darling-Hammond & Wentworth, 2010; Shepard, 2000).

**Заключение**

В данной работе выделены принципы и установки, которые используются в опыте разных стран, а также современная исследовательская база чтобы обозначить рамку создания более эффективной системы оценивания. Данная рамка дает политикам и другим заинтересованным группам структуру для дискуссии или консенсуса при обсуждении приоритетов и ключевых усилий, необходимых для их систем оценивания.

Хотя не следует занижать значимость данного набора принципов и установок, но важно также обращать внимание на контекст конкретной страны, на ожидания и потребности, прежде чем будут приняты решения, с чего начинать, какой подход использовать и как долго его осуществлять. Страны, следовательно, должны использовать данную рамку как средство, которое дает им высокую степень гибкости в вопросах куда, когда и каким образом двигаться. Измерение успеха в конце дня и есть система оценивания, которая вносит вклад в более высокий уровень качества образования и обучения учащихся.

Приложение 1

**Виды оценивания и их ключевые различия**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Внутриклассное оценивание** | **Крупномасштабные оценки** | | **Государственные экзамены** | |
|  |  | **национальные** | **международные** | **выпускные** | **вступительные** |
| **Цель** | Моментальная обратная связь для информирования классного руководителя | Представление данных об общем здоровье системы на уровне конкретной возрастной группы/класса.  Мониторинг тенденций в обучении | Представление сравнительных данных системы образования на уровне конкретного класса/возраста | Сертификация учеников по мере продвижения с одного уровня система образования на следующую (или на рабочее место) | Отбор учащихся для получения дальнейшего образования |
| **Частота** | Ежедневно | По конкретному предмету рекомендуется на регулярной основе (каждые 3 -5 лет) | По конкретному предмету рекомендуется на регулярной основе (каждые 3 -5 лет) | Ежегодно и чаще, где система позволяет повторять | Ежегодно и чаще, где система позволяет повторять |
| **Кто тестируется** | Все учащиеся | Выборочно или все учащиеся определенного класса/возраста | Выборка учащихся определенного класса/возраста | Все выпускники | Все поступающие, которые хотят сдавать данный экзамен |
| **Формат** | Варьируется от наблюдения до опросов и письменных работ учащихся | Обычно это множественный выбор и короткий ответ | Обычно это множественный выбор и короткий ответ | Обычно сочинение или множественный выбор | Обычно сочинение или множественный выбор |
| **Охват учебного плана** | Все предметы | Обычно ограничивается несколькими предметами | Обычно ограничивается одним или двумя предметами | Охватывает основные предметные области | Охватывает основные предметные области |
| **Дополнит. информация, собранная от учащихся** | Да, как часть учебного процесса | Часто | Да | Редко | Редко |
| **Подсчёт баллов** | Обычно неформальное и простое | Варьируется от простых до более сложных статистических методов | Обычно задействованы сложные статистические методы | Варьируется от простых до более сложных статистических методов | Варьируется от простых до более сложных статистических методов |

Источник: Всемирный банк

Приложение 2

**Техническое задание для проведения самодиагностики национальной системы оценки качества образования**

Определение потенциала страны и потребностей в области оценивания с целью совершенствования качества образования и обучения учащихся

Состояние проблемы

Все чаще считается необходимым измерение результатов обучения не только для мониторинга качества образования, но и для его улучшения. Данные о достижениях учащихся могут использоваться для информирования как тех, кто вырабатывает политику в образовании, так и практиков, начиная от разработки и осуществления образовательных программ, направленных на улучшение преподавания и обучения в классе, и, заканчивая обеспечением того, чтобы слабые школы и учителя получили адекватную поддержку и подготовку и тем самым гарантировали всем учащимся равные возможности в обучении.

В октябре 2008 года стартовала Российская программа помощи развитию в области образования (READ), которая реализуется совместно правительством Российской Федерации и Всемирным Банком и нацелена, главным образом, на повышение качества образования в развивающихся странах.

READ Трастовый Фонд (далееs READ) является частью этой программы. Его цель – оказывать поддержку странам в улучшении результатов обучения школьников посредством разработки, осуществления и использования жизнеспособных систем оценивания достижений учащихся. Подобные системы включают в себя:

• Тесты, разработанные для определения уровня подготовки учащихся, например, анкетирование по образцам для выявления уровня достижений учащихся;

• Тесты, разработанные для принятия решений о конкретном учащемся, например, выпускные и вступительные экзамены в университет;

• Ежедневное оценивание в классе;

• Предоставление информации о результатах обучения всем заинтересованным сторонам;

• Анализ результатов оценивания, направленный на улучшение учебных программ, политики и практики преподавания и обучения;

• Материалы для профессионального роста педагогов по применению результатов оценивания.

Главная цель READ – оказывать поддержку странам в усилении потенциала тех организаций и институтов, которые отвечают за измерение учебных результатов или за использование информации, полученной в ходе оценивания, для улучшения преподавания и обучения. READ будет поддерживать аналитическую работу и оказывать техническое содействие странам в (a) становлении систем или институтов — или усилении уже существующих — которые формулируют образовательные цели и проводят оценивание учебных результатов; (b) совершенствовании существующих или развитии новых инструментов измерения результатов обучения учащихся; и (c) укреплении существующих или выработке новых механизмов (политики) использования результатов измерения с целью улучшения преподавания и обучения.

READ предлагает систематический подход к решению этих задач в каждой конкретной стране, определяя порядок шагов, обозначенных ниже, которые вместе взятые создают комплексную рамку действий, направленных на выработку логической системы оценивания как базы для дальнейших усилий по повышению качества образования и обучения учащихся:

• Проведение *самодиагностики* с целью определения потенциала страны и ее потребностей в области оценивания учащихся и ее применение для улучшения преподавания и обучения;

• Разработка *Плана действия,* основанного на заключении самодиагностики, для укрепления систем оценивания учащихся и их использования;

• Осуществление *Плана действий*, включая использование информации, полученной в ходе оценивания, для совершенствования преподавания и обучения в классе; и

• Подведение итогов *Полученных уроков* для информирования тех, кто принимает политические решения и предпринимает конкретные действия.

Данное ТЗ направлено на то, чтобы помочь странам READ провести *самодиагностику*.

Назначение Самодиагностики

По завершении самодиагностики в стране должна быть бóльшая ясность относительно адекватности проводимой политики, существующих структур и имеющихся для проведения оценивания ресурсов; степени совпадения проводимых по оцениванию мероприятий с общими целями и деятельностью системы образования в целом; качества инструментария оценивания; и определенные идеи по поводу следующих шагов, которые можно было бы предложить для финансирования за счет средств программы READ.

Основные результаты Самодиагностики

Главным результатом будет:

I. Технический Отчет

• о потенциале страны и о потребностях в области оценивания учащихся, включая

– ключевые проблемы/наработки

– возможные варианты последующих шагов на основе приоритетов страны и лучшей мировой практики с фокусом на стратегии институционального усиления

• который может быть использован для диалога с правительством/заинтересованными сторонами и для разработки Плана Действий в рамках READ.

II. Семинар для Заинтересованных Сторон

• для обсуждения проекта вышеуказанного отчета

• для проведение мозгового штурма/принятия решения по поводу последующих шагов

Ключевые Задачи Самодиагностики

Проведение самодиагностики включает в себя следующие восемь задач:

***Задача 1. Создание Команды в Стране***

Каждая страна должна создать команду специалистов для проведения самодиагностики. В идеале в нее должны входить от 4 до 6 специалистов в области технических знаний/выработки политики и иметь опыт в таких областях как:

• крупномасштабные оценочные обследования;

• тесты/экзамены высокой значимости (с высокими ставками, например, выпускные или вступительные экзамены);

• учебные планы и преподавание;

• инспектирование школ;

• подготовка учителей/профессиональное развитие учителей;

• мониторинг системы и оценивание.

*Продукт*: Команда для проведения самодиагностики в стране.

***Задача 2. Доработка ТЗ для Проведения Самодиагностики***

Созданная Команда специалистов рассматривает данное ТЗ на проведение самодиагностики и вносит необходимые изменения с тем, чтобы ТЗ лучше соответствовало контексту страны. Внесенные изменения рассматриваются Руководителем проекта со стороны Всемирного Банка в стране (TTL) и координационным центром READ в данной стране.

*Продукт*: Доработанное ТЗ с объяснениями всех внесенных изменений

***Задача 3. Сбор Основной Информации о Системе Образования Страны***

Команда собирает следующую информацию о системе образования страны (от дошкольного до высшего образования):

• Продолжительность и фазы школьного образования;

• Организация учебного года, дня в школе;

• Количество школ, учебных классов, учащихся и учителей (квалифицированных/неквалифицированных);

• Доля учащихся, охваченных образованием по типу, уровню и расположению школы;

• Распределение функций между организациями, которые разрабатывают и претворяют в жизнь образовательную политику;

• Проводимые/планируемые реформы в области образовательной политики и другие важные инициативы;

• Вложения в образование – государственные/другие средства;

• Политика, проводимая в отношении учителей/директоров школ (подготовка, сертификация, распределение кадров, заработная плата).

Команда специалистов должна всецело использовать имеющиеся источники информации (например, План развития образовательного сектора, базы данных Всемирного Банка или отчеты и базы данных Министерства образования), выверять/обновлять их с помощью партнеров в других странах или при поддержке Всемирного Банка. Данные должны быть самыми последними, которые имеются в наличии, они должны браться по возможности из одних и тех же источников, таких как План развития образовательного сектора. Если есть данные о наличии тенденций, то динамика должна быть представлена за последние 10 лет.

*Материалы для использования*: План развития образовательного сектора, базы данных и аналогичные документы Всемирного Банка/министерства образования, интервью с ключевыми фигурами.

*Материалы, которые необходимо собрать*: План развития образовательного сектора и другие основные документы (или информация в других источниках)

*Продукт*: База данных с релевантной информацией и краткий отчет.

***Задача 4. Сбор и анализ информации о политическом климате, структуре и ресурсах для проведения оценивания учащихся/экзаменов***

Команда специалистов собирает и анализирует информацию о политической ситуации в стране в целом, о структурах и имеющихся ресурсах для проведения оценивания учащихся/экзаменов:

• Правовые/политические рамки

Ключевые правовые/политические шаги в отношении проведения оценивания/экзаменов

• Структуры

– Агентства, подразделения и организации, отвечающие за оценивание/экзамены на разных уровнях системы, и следующая информация о них:

 Ответственность/полномочия и структуры управления

 Взаимоотношения друг с другом и с другими государственными/негосударственными структурами

 История и стабильность

 Требования к отчетности; прозрачность

• Человеческие ресурсы

– Для каждой организации, проводящей оценивание/экзамены собрать следующую информацию:

 Количество персонала, его квалификация, образования и профессиональная подготовка

 Резюме персонала и их потенциал для выполнения поставленных задач

 Типовая заработная плата (в сравнении с заработной платой других министерств/аналогичных организаций) и ротация кадров

 Подходы к подготовке персонала и наращивания его потенциала

– Опора на внешних экспертов в проведении оценивания/экзаменов

• Финансовые ресурсы

– Источники и размер финансирования

 Годовое финансирование/источники финансирования (на каждую организацию, проводящую оценивание/экзамены и на каждое мероприятие

 Стабильность финансовых источников и выделяемых сумм

 Затраты на проведение оценивания/экзаменов (в % от затрат на образование/государственного бюджета , а также разбивка по затратам на разработку теста, его проведение, обсчет/анализ результатов, распространение результатов)

*Материалы для использования*: Базы данных министерствам и других организаций, официальные и неофициальные документы и отчеты, интервью с ключевыми фигурами

*Материалы, которые необходимо собрать*: Главные документы/отчеты и информация из источников

*Продукт*: База данных, содержащая релевантную информацию и краткий отчет

***Задача 5. Сбор и анализ информации о целях системного обучения и деятельности по проведению оценивания/экзаменов***

Команда специалистов собирает и анализирует информацию о целях системного обучения и деятельности по проведению оценивания/экзаменов:

• Понятие “качественного образования”

– Ожидания государства относительно того, каким должно быть “качественное образование”

 Составляющие качества

 Необходимые условия и процессы

 Результаты “качественного образования” на уровне индивидуума, на социальном и экономическом уровне

• Цели обучения

– Что должны знать и уметь учащиеся на уровне определенного класса и ступени обучения в школе

 Наличие официальных стандартов или учебных программ по разным предметам (плюс дата последнего или планируемого их пересмотра)

 Степень соответствия между планируемыми учебными программами, теми, которые даются на практике, и теми, которые усваивают учащимися

 Степень соответствия между официальными стандартами или учебными программами и запросом общества/рынка труда

• Оценивание учащихся/экзамены

– Оценивание учащихся/экзамены, которые призваны определить, усвоил ли учащийся желаемые знания и навыки/компетенции

 Перечень всех формальных мероприятий — например, региональные/национальные обследования, государственные экзамены, и т.д.

 По каждому оцениванию/экзамену представить такие детали, как:

• годы проведения/график, цель, формат заданий, оцениваемые умения и знания, классы/

учащиеся, шкала применения/учет численности или система выборки, группы учащихся, которые не тестировались, организация, проводившая оценивание, механизмы обеспечения качества

• типы собранной информации по характеристике учеников или исходных условий (например, время обучения, учебные материалы, технические средства)

• ключевые выводы тех, кто руководил последним тестированием, в целом, по группам учащихся, по предметам/навыкам; включая данные по имеющим место тенденциям

• виды отчетов и получатели информации; временные рамки и формат предоставления результатов; диссеминация/уровни системы, которые получают результаты; публичные публикации/доступность тестовых заданий или данных

 Рассмотреть главные отличия проводимых оценок/экзаменов с точки зрения учебного материала, отчетности разных заинтересованных лиц, результатов. Использовать все возможности чтобы избежать дублирования и применить все имеющиеся технологические приемы

• Использование результатов оценивания/экзаменов

– По каждому оцениванию/экзамену собрать данные, относительно:

 кем и как в основном используются результаты

 сказывается ли использование результатов на улучшении качества обучения, есть ли сведения, что данные отрицательно влияют на качество обучения

– Источники/поддержка для использования данных оценивания/экзаменов:

 Учителя/Преподавание

• Виды информации, предоставляемой учителям, и как они ее используют

• Источники/обучение для понимания или использования данных оценивания/экзаменов для разработки своих собственных оценок

 Учащиеся/Обучение

• Виды информации, предоставляемой учащимся, и как они ее используют

• Источники/поддержка для понимания или использования данных оценивания/экзаменов для улучшения своей успеваемости

 Школы/Планирование

• Виды информации, предоставляемой школам, и как они ее используют

• Источники/тренинг/поддержка, которая может быть оказана школам в использовании данных оценивания/экзаменов для улучшения результатов учащихся

 Выработка политики/Разработка программ

• Виды информации, предоставляемой политикам/тем, кто занимается планированием

• Как эта информация используется для информирования: например, при

o Разработке/постановке целей новых реформ, программ, инициатив

o Принятии решений по выделению ресурсов

 (Информирование других лиц: родителей, граждан, тех, кто отвечает за обучение/оценивание/

аккредитацию учителей, исследователей, доноров)

*Материалы для использования*: Базы данных министерствам и других организаций, официальные и неофициальные документы и отчеты, интервью с ключевыми фигурами , материалы тестов

*Материалы, которые необходимо собрать*: Ключевые документы/отчеты (например, копии публикаций результатов оценивания/экзаменов) или информация из других источников; все имеющиеся материалы тестирования (рамки, инструменты, базы данных, отчеты)

*Продукт*: База данных, содержащая релевантную информацию и краткий отчет.

***Задача 6. Оценка технического качества ключевых мероприятий по оцениванию/проведению экзаменов***

Команда собирает и анализирует информацию относительно технического качества ключевых мероприятий по оцениванию/проведению экзаменов. Команда должна отобрать пару мероприятий по оцениванию/экзаменов, которые являются ключевыми для улучшения качества образования и результатов обучения учащихся, или которые вызывают обеспокоенность и проблемы. Для оценивания по типу опроса надо собрать следующие данные:

• Разработка и валидация инструментария, включая адекватность схемы тестирования и формат заданий для целевой аудитории, а также применение инструментария

• Механизмы, используемые для мониторинга соответствия стандартизированным условиям, которые должны превалировать во время теста (например, качество подготовки тех, кто проводит тестирование, контроль качества проведения теста, мотивация учащихся к прохождению тестов, распределение технических средств, обеспечение возврата материалов тестирования)

• Документация по техническим процедурам, которые применялись при разработке инструментария, оценке точности измерений, разработке образцов и составлении отчетов о результатах, при прохождении тестирования и оценке качества его организации, при определение уровня результатов и минимально допустимого уровня показателей, что делает результаты сопоставимыми с предыдущими оценками

• Точность анализа, включая отдельные показатели, проводимые сравнения, наблюдения и выводы

• Отчетность и распространение результатов, с данными относительно их правомерности (справедливости) и охватываемом периоде для определенной аудитории

• Мониторинг влияния результатов и их использования, включая информацию, насколько широко используются полученные данные там, где их предполагалось

Для значимых государственных экзаменов (с высокими ставками) надо собрать информацию по следующим позициям:

• Разработка и валидация инструментария, включая адекватность схемы тестирования/ охвата и формат заданий для целевой аудитории, а также применение инструментария

• Возможность учащихся освоить проверяемый материал, включая наличие учебников, учебных материалов, необходимого оборудования , и подготовку учителей

• Возможность для учащихся пересдать экзамен (в случае неудовлетворительного результата) с использованием альтернативных форм

• Результаты экзамена – это только один компонент, который используется для подтверждения обучения учащегося (выдачи ему сертификата) или для принятия других важных решений

• Механизмы, используемые для мониторинга соответствия стандартизированным условиям, которые должны превалировать во время теста (например, качество подготовки тех, кто проводит тестирование, контроль качества проведения теста, распределение технических средств, обеспечение безопасности, хранения и возврата материалов тестирования)

• Документация по техническим процедурам, которые применялись при разработке инструментария, оценке точности измерений, разработке образцов и составлении отчетов о результатах, при прохождении тестирования и оценке качества его организации, при определение уровня результатов и минимально допустимого уровня показателей, что делает результаты сопоставимыми с предыдущими оценками

• Мониторинг влияния результатов и их использования, включая информацию, насколько широко используются полученные данные там, где их предполагалось

*Материалы для использования*: Копии общей схемы оценивания/проведения экзамена, инструментарий, отчеты и другая, относящаяся к вопросы техническая информация (например, базы данных) , интервью с ключевыми фигурами

*Материалы, которые необходимо собрать*: Ключевые документы/отчеты или информация из других источников; все имеющиеся материалы тестирования

*Продукт*: База данных, содержащая релевантную информацию и краткий отчет по каждому рассмотренному оцениванию/экзамену.

***Задача 7. Написание отчета на основе полученных данных***

Команда специалистов должна использовать материалы /продукты Задач 3-6 для написания Технического Отчета. Отчет будет использоваться как основа для диалога с правительством/другими заинтересованными сторонами, а также для разработки План действий в рамках READ. В отчет должен войти анализ потенциала страны и её потребностей в области оценивания учащихся /проведении экзаменов, а также возможные последующие шаги, основанные на приоритетах страны и лучшей мировой (с фокусом на стратегии институционального усиления). С отчетом необходимо ознакомить ключевых работников Министерства образования , персонал, работающий в сфере оценивания успеваемости/экзаменов, и членов местного образовательного сообщества, включая основных доноров и других партнеров.

*Материалы для использования*: Материалы Задач 3-6

*Материалы, которые необходимо собрать*: нет в наличии

*Продукт*: Технический Отчет

***Задача 8. Проведение семинара для обсуждения отчета и его завершения***

Команда специалистов проводит семинар на основе полученных выводов и наработок, обозначенных в Техническом Отчете. Команда должна обеспечить обсуждение отчета участниками семинара и использовать обратную связь для подготовки финальной версии отчета, уделяя особо внимание договоренностям относительно последующих шагов.

*Материалы для использования*: Материалы Задач 3-7

*Материалы, которые необходимо собрать*: Отзывы участников

*Продукт*: Финальная версия Технического Отчета

Приложение 3

**Вопросники для проведения диагностики системы оценивания**

*Комментарий*. Вопросники будут представлены после проведения апробации и доработки.

Пример. Вопросник оценки благоприятной среды для внутриклассного оценивания.

**Оценивание в классе**: Процесс сбора, обобщения и толкования информации во время классных занятий, прежде всего, силами учителей и учащихся в целях принятия решений в классе, поддержки учащихся в процессе обучения (текущее оценивание) и оценки успеваемости (итоговое оценивание). Этот процесс включает формальную оценку работы учащихся и неформальные наблюдения.

**БC. Благоприятная среда для оценивания в классе**

**1. Какие документы, если таковые имеются, обеспечивают концептуальную основу на национальном уровне для мероприятий по оцениванию в классе?**

a. ( )Национальный юридический документ, требования которого подлежат выполнению в обязательном порядке, например, закон, приказ или нормативный акт

b. ( ) Oфициальный документ, например, принципы/стандарты/методические рекомендации по оцениванию учащихся, не имеющий обязательной юридической силы

c. ( ) Неофициальный неопубликованный документ

d. ( ) Документы, образующие концептуальную основу для мероприятий по оцениванию в классе на национальном уровне отсутствуют

**2. В чем заключаются основные цели/задачи оценивания в классе в данной стране?**

a. ( ) Поддержка учащихся в процессе обучения

b. ( ) Диагностика проблем возникающих в процессе обучения

c. ( ) Предоставление информации учащимся о том, насколько успешно они усваивают учебный материал

d. ( ) Информирование родителей о том, как учится их ребенок

e. ( ) Определение дальнейших методических шагов

f. ( ) Оценивание учащихся

g. ( ) Принятие решения о переводе учащихся в следующих класс (на следующий уровень)

h. ( ) Ведение официального учета учащихся

i. ( ) Мониторинг уровня качества образования

j. ( ) Выполнение требований по оцениванию на уровне всей школы

k. ( ) Обеспечение контроля за работой школы

l. ( ) Обеспечение контроля за работой учителей

**3. Имеется ли орган/ведомство, который/которые инструктирует учителей и школы в отношении того, какие данные о достижениях учащихся предоставляются родителям, общественности, и органам образования, а также как и в какие сроки такие данные должны собираться и предоставляться?**

**4. Какие факторы способствовали тому, что оценивание в классе стало одной из приоритетных задач образовательной программы в данной стране? Выберете все применимые факторы.**

a. ( ) Оценивание в классе не является частью программы

b. ( ) Позиция политического руководства

c. ( ) Позиция руководства системы образования

d. ( ) Были выделены финансовые ресурсы

e. ( ) Hаличие требований со стороны специалистов системы образования

f. ( ) Рекомендации внешних организаций или результаты экспертных оценок системы образования

g. ( ) Результаты исследований, включая установление зависимости с успеваемостью в рамках международных исследований оценивания

Приложение 4

**Рубрики, представляющие этапы развития оценивания**

**Благоприятная среда**

**Общие принципы политики, организационные структуры и бюджетное и кадровое обеспечение как условия для работы по оценке в стране и их значение для или непосредственное влияние на работу по оценке**

|  |  |  |  |
| --- | --- | --- | --- |
| **Описания уровня сформированности базы для проведения *внешних экзаменов*** | | | |
| **ЛАТЕНТНЫЙ**  **Отсутствие или отклонения от необходимых условий** | **НА СТАДИИ СТАНОВЛЕНИЯ**  **На пути к установлению минимальных требований** | **СФОРМИРОВАННЫЙ**  **Приемлемые минимальные требования** | **РАЗВИТЫЙ**  **Передовые подходы** |
|  |  |  |  |
| **БС1 -- Четкое определение политики** | | | |
| Официальных программных документов, регулирующих принятие решений по проведению экзаменов, нет. Заинтересованные стороны не участвуют в принятии решений по экзаменам. В большинстве случаев процессы проведения экзаменов и принятия решений недостаточно прозрачны. | Программные документы недостаточно детальны, как правило, не публикуются и широко не распространяются. В них частично рассматриваются цели/задачи проведения экзаменов, или даются соответствующие указания, а иногда и процессы принятия решений по другим ключевым направлениям. Иногда обеспечивается частичное участие заинтересованных сторон, но, как правило, оно весьма ограничено. Информацию о процессах можно получить по запросу, но только по очень конкретным вопросам. | Официальные программные документы содержат адекватные методические указания по принятию ключевых решений в отношении экзаменов. В них описываются функции и обязанности по общему управлению, финансированию, видам целям (использования результатов), разработке, организации, анализу и отчетности по экзаменам. Обычно в них нет краткого описания процедур расследования и рассмотрения утечек информации или нарушений установленных правил контроля проверки знаний учащихся, и порядка тестирования групп учащихся с особыми потребностями/ограниченными возможностями. Некоторые заинтересованные стороны участвуют в создании и изменении этих документов. Информация о процессах предоставляется по запросам. | Официальные программные документы содержат адекватные методические указания по принятию ключевых решений в отношении экзаменов. В них содержится четкое описание функции и обязанности по общему управлению, финансированию, видам целям (использования результатов), разработке, организации, анализу и отчетности по экзаменам, а также краткое описание процедур расследования и рассмотрения утечек информации, нарушений установленных правил контроля проверки знаний учащихся, и других форм недостойного поведения и краткое описание порядка тестирования групп учащихся с особыми потребностями/ограниченными возможностями. Заинтересованные стороны участвуют в создании и изменении этих документов, а связанные с ними процессы прозрачны и предусматривают общедоступность информации. |
| **БС2 -- Сильная позиция руководства** | | | |
| У руководства нет четкого представления об оценке и позиции в этом отношении. Нельзя назвать какого-либо руководителя, который играл бы ведущую роль в разработке экзаменов. Не хватает полномочий и влияния для того, чтобы что-то начало происходить. Проведению экзаменов препятствует сопротивление ключевых заинтересованных сторон (например, руководителей, отвечающих за формирование политики, профсоюза учителей, учащихся, родителей). | Определение видения и ведение работы, в основном, зависят от одного человека. Если он уйдет со своего поста, система экзаменов будет ослаблена. Способность трансформирования видения в комплекс конкретных задач и стратегий весьма ограничена. Не хватает полномочий и влияния для того, чтобы что-то начало происходить. Ключевые заинтересованные стороны противодействуют проведению экзаменов (в таких формах, как бойкоты, протесты, отказ принимать участие в работе), что ставит под угрозу саму возможность проведения такой работы и стабильность программы. Тем не менее, программу экзаменов можно осуществить. | Сильная позиция руководства, но она обеспечивается в основном группой лиц, в которой представлены не все ключевые учреждения и категории заинтересованных сторон. Притом, что есть четкое представление о проведении экзаменов, его "перевод" в комплекс конкретных задач и стратегий остается неполным. Чтобы что-то начало происходить, можно, хотя это трудно, использовать имеющиеся полномочия и влияние. Большинство ключевых заинтересованных сторон согласны на проведение экзаменов. Сопротивление некоторых из ключевых заинтересованных сторон не ставит под угрозу саму возможность и стабильность проведения экзаменов. | Сильная позиция руководства в отношении экзаменов обеспечивается в основном представителями ключевых учреждений и категорий заинтересованных сторон. Есть четкое общее представление об экзаменах, которое "переведена" в комплекс конкретных задач и стратегий с продуманными механизмами их реализации. Есть потенциал для организации совместной работы по реализации концепции. Для того, чтобы что-то начало происходить можно использовать имеющиеся полномочия и влияние. Ключевые заинтересованные стороны поддерживают или, как минимум, не препятствуют проведению экзаменов. |
| **БС3 -- Обеспечение бюджетного финансирования в достаточных объемах** | | | |
| Нет источников финансирования для экзаменационной деятельности. | Есть финансирование, но его объемы и источники недостаточны и нестабильны. Финансирование можно получать из других источников на ограниченные периоды времени, но без гарантий продолжения оказания поддержки. | Объемы и источники финансирования стабильны. Финансирование позволяет покрывать затраты на основные мероприятия, связанные с экзаменами (разработка, управление, анализ, отчетность). | Объемы и источники финансирования стабильны. Финансирование позволяет покрывать затраты на основные мероприятия, связанные с экзаменами (разработка, управление, анализ, отчетность) в рамках долгосрочной программы работы, а также затраты на научно-техническую деятельность. |
| **БС4 -- Наличие соответствующих организационных структур** | | | |
| Экзаменационный орган если он есть, не является достаточно самостоятельной структурой и недостаточно хорошо контролируется. Процессы принятия решений непрозрачны. Нет необходимой материально-технической базы для проведения экзаменов. | Авторитет экзаменационного органа только формируется. Это - новая организация, которая еще не достигла надлежащего баланса между самостоятельностью и контролем со стороны внешнего органа. Принятие решений основано преимущественно на политических, a не на технических критериях. Экзаменационный орган не имеет адекватной материально-технической базы для проведения экзаменов, однако при этом способен осуществлять управление ключевыми процесами проведения экзаменов. | Экзаменационный орган является авторитетной организацией в стране, она достаточно самостоятельна для того, чтобы принимать решения преимущественно на основе технических критериев. Это возможно даже, если нет контроля со стороны внешнего органа. Имеет достаточно большие площади/помещения; установлен порядок выполнения минимальных требований по безопасности. Адекватные технические возможности для сбора, обработки, оценки и учета данных. | Экзаменационный орган является стабильной организацией, признанной на международном уровне. Он обеспечивает надлежащий баланс между самостоятельностью и контролем: может самостоятельно принимать решения на основе технических критериев, но контролируется внешним органом/ведомством. Есть необходимые помещения, охрана и ресурсы для сотрудников. Для сбора, обработки, оценки и учета данных используются современные технологии, что позволяет минимизировать вероятность возникновения ошибок и максимизировать эффективность. |
| **БС5 -- Кадровое обеспечение** | | | |
| Экзаменационный орган если он есть, не имеет персонала, удовлетворяющего минимальным квалификационным требованиям для проведения экзаменов. У персонала нет опыта практической работы, не говоря уже о формальном обучении методам оценивания. Штат сильно недоукомплектован или вообще неукомплектован, поэтому основные функции по проведению экзаменов он выполнять не может. | Несколько сотрудников экзаменационного органа прошли неформальное обучение в ходе практической работы. Скорее всего никто из них не проходил формальной подготовки по методам оценивания, учебным планам, анализу, психометрии или статистике. Экзаменационный орган недоукомплектован, поэтому сотрудники перегружены работой, и наблюдается сильная текучесть кадров. Нет постоянного пула квалифицированных специалистов для выполнения краткосрочных заданий; для их выполнения обычно нанимают новых людей, которых приходится обучать каждый год. | У экзаменационного офиса адекватный технический персонал. Все они прошли неформальную подготовку в ходе практической работы. Некоторые из них также могли пройти формальную подготовку по некоторым аспектам оценивания, учебных планов, анализа, психометрии или статистики. Есть относительно достаточное число специалистов для выполнения основных функций. Текучесть кадров умеренная. Есть пул квалифицированных специалистов для оперативного выполнения краткосрочных заданий; но его не хватает ввиду больших объемов работы. | У экзаменационного офиса есть высококвалифицированный технический персонал для выполнения основных функций: разработки, реализации, и проведения отчетности. Персонал имеет формальную подготовку по оцениванию, составлению учебных планов, анализу, психометрии или статистике, а также неформальное обучение в ходе практической работы. Достаточно персонала для выполнения основных функций по проведению экзаменов (штатных сотрудников и подрядчиков). Незначительная текучесть кадров. Также есть пул квалифицированных специалистов для выполнения краткосрочных заданий. |

|  |  |  |  |
| --- | --- | --- | --- |
| **Описания уровня сформированности базы для *национальных крупномасштабных исследований* оценивающих учебные достижения учащихся на основании опросов** | | | |
| **ЛАТЕНТНЫЙ**  **Отсутствие или отклонения от необходимых условий** | **НА СТАДИИ СТАНОВЛЕНИЯ**  **На пути к установлению минимальных требований** | **СФОРМИРОВАННЫЙ**  **Приемлемые минимальные требования** | **РАЗВИТЫЙ**  **Передовые подходы** |
|  |  |  |  |
| **БС1 -- Четкое определение политики** | | | |
| Нет официальных программных документов, регулирующих принятие решений по проведению крупномасштабных исследований оценивающих учебные достижения учащихся. | Нет официальных программных документов, содержащих описание программы крупномасштабных исследований оценивающих учебные достижения учащихся. Однако в настоящее время, идет разработка планов по созданию такого документа, и уже имеется возможность познакомиться с проектом такого документа. | Есть официальные программные документы, в которых четко изложены цели и задачи осуществления программы крупномасштабных исследований оценивающих учебные достижения учащихся, как минимум, на ближайшую перспективу. | Есть официальные программные документы, основанные на репрезентативных, консультативных процессах в которых четко изложены цели, применения и основные характеристики программы крупномасштабных исследований оценивающих учебные достижения учащихся на ближайшую и долгосрочную перспективу. |
| **БС2 -- Сильная позиция руководства** | | | |
| У руководства отсуствует надлежащее видение и позиция в отношение проведения оценки. Нельзя назвать какого-либо руководителя, который играл бы ведущую роль в разработке программы крупномасштабных исследований оценивающих учебные достижения учащихся. Не хватает полномочий и влияния для того, чтобы что-то начало происходить. Проведению крупномасштабных исследований оценивающих учебные достижения учащихся препятствует сопротивление ключевых заинтересованных сторон (например, руководителей, отвечающих за формирование политики, профсоюза учителей, учащихся, родителей). | Определение видения и ведение работы, в основном, зависят от одного человека. Если он уйдет со своего поста, разработка программы крупномасштабных исследований оценивающих учебные достижения учащихся прекратится или задержится, способность трансформирования имеющегося видения в комплекс конкретных задач и стратегий весьма ограничен. Не хватает полномочий и влияния для того, чтобы что-то начало происходить. Ключевые заинтересованные стороны противодействуют проведению крупномасштабных исследований оценивающих учебные достижения учащихся (в таких формах, как бойкоты, протесты, отказ принимать участие в работе), что ставит под угрозу саму возможность проведения такой работы и стабильность программы. Тем не менее, программу оценки можно осуществить. | Сильная позиция руководства, но она обеспечивается в основном группой лиц, в которой представлены не все ключевые учреждения и категории заинтересованных сторон. Притом, что есть четкое представление о программе крупномасштабных исследований оценивающих учебные достижения учащихся, ее "перевод" в комплекс конкретных задач и стратегий остается неполным. Чтобы что-то начало происходить, можно, хотя это трудно, использовать имеющиеся полномочия и влияние. Большинство ключевых заинтересованных сторон солганcы на проведение оценки. Сопротивление некоторых их ключевых заинтересованных сторон не ставит под угрозу саму возможность проведения такой работы и стабильность программы. | Сильная позиция руководства в отношении программы крупномасштабных исследований оценивающих учебные достижения учащихся обеспечивается в основном представителями ключевых учреждений и категорий заинтересованных сторон. Есть четкое общее представление о том, какой должна быть программа крупномасштабных исследований оценивающих учебные достижения учащихся, которая "переведена" в комплекс конкретных задач и стратегий с продуманными механизмами их реализации. Есть потенциал для организации совместной работы по реализации концепции. Для того, чтобы что-то начало происходить можно использовать имеющиеся полномочия и влияние. Ключевые заинтересованные стороны поддерживают или, как минимум, не препятствуют осуществлении крупномасштабных исследований оценивающих учебные достижения учащихся. |
| **БС3 -- Обеспечение бюджетного финансирования в достаточных объемах** | | | |
| Нет источников финансирования для разработки крупномасштабных исследований оценивающих учебные достижения учащихся. | Есть финансирование, но его объемы и источники недостаточны и нестабильны. Финансирование может быть предоставлено внешним донором или программой на ограниченный период времени, но без гарантий продолжения оказания поддержки. | Объемы и источники финансирования стабильны, и, как правило, законодательно обеспечены. Финансирование, позволяет покрывать затраты на основные мероприятия (разработка, управление, анализ, отчетность) только в рамках программы крупномасштабных исследований оценивающих учебные достижения учащихся. | Объемы и источники финансирования, как правило, законодательно обеспечены. Финансирование, позволяет покрывать затраты на основные мероприятия (разработка, управление, анализ, отчетность) в рамках программы крупномасштабных исследований оценивающих учебные достижения учащихся, а также затраты на научно-техническую деятельность. |
| **БС4 -- Наличие соответствующих организационных структур** | | | |
| Специального подразделения или группы специалистов по оценке нет. | Специальное подразделение или группа специалистов по оценке не является достаточно самостоятельным и независимым органом, чтобы к нему относились как к политически и технически авторитетной структуре. Эта структура не контролируются никакими внешними органами. Ее стабильность ограничена: подразделения или группы специалистов, которым поручено проведение крупномасштабных исследований оценивающих учебные достижения учащихся, могут изменятся от раунда к раунду. | Специальное подразделение или группа специалистов по оценке является стабильной, самостоятельной и независимой организацией, но не контролируется никакими внешними органами. Нет статуса и технической авторитетности, необходимых для утверждения результатов, которые, по сути, являются спорными. Есть только необходимая материально-техническая база для работы в "достаточно хороших" условиях. | Специальное подразделение или группа специалистов по оценке является стабильной, политически и технически авторитетной организацией с высоким статусом. Является самостоятельной и независимой организаций, но при этом контролируется внешним органом. Имеет необходимую материально-техническую базу для работы по оцениванию в соответствии с самыми высокими стандартами/требованиями. |
| **БС5 -- Кадровое обеспечение** | | | |
| Специалистов, которые занимаются или могли заняться программой по оценке нет. | Подготовка технического персонала, занимающегося работой по программе, ограничивается только практическим опытом работы по оценке, анализу, психометрии или статистике. | Часть технического персонала, занимающегося работой по программе, прошла формальное обучение или ознакомление с методами оценки; но большинство прошло лишь неформальную подготовку в виде практической работы по оценке, анализу, психометрии или статистике. Есть относительно устойчивое и достаточное число специалистов для проведения ключевых мероприятий по оцениванию (разработка, реализация, анализ, отчетность). | Потенциал, квалификация и стабильность технического персонала структуры, которая занимается оцениванием, достаточно высоки. Большинство специалистов прошли необходимое формальное обучение или ознакомление с методами оценивания, анализа, психометрии или статистики, а также дополнительное обучение в ходе практической работы. Есть достаточно большой пул квалифицированных работников для оперативного выполнения краткосрочных заданий. |
| **БС6 -- Создание возможностей для формального обучения методам проведения крупномасштабных оценок** | | | |
| В стране нет возможностей для ознакомления c методами проведения крупномасштабных исследований оценивающих учебные достижения учащихся, или получения подготовки в применении таких методов. | Возможности для ознакомления c методами проведения крупномасштабных исследований оценивающих учебные достижения учащихся, а также возможности получения подготовки в применении таких методов ограничены. Несколько учебных курсов и ознакомительных семинаров для будущих специалистов по оценке на основные темы (такие, как статистика, анализ данных, политика оценивания, психометрия), но они организуются не регулярно. | Есть достаточно хорошие возможности для ознакомления с методами проведения крупномасштабных исследований оценивающих учебные достижения учащихся и соответствующего формального обучения. В некоторых вузах есть соответствующие программы, а также организуются учебные курсы и ознакомительные семинары для будущих специалистов по оценке на основные темы (такие, как статистика, анализ данных, политика оценивания, психометрия). | Существуют обширные возможности для ознакомления с методами проведения крупномасштабных исследований оценивающих учебные достижения учащихся и прохождения соответствующего формального обучения. В некоторых вузах есть программы на уровне получения мастерских и докторских степеней, и постоянно организуются учебные курсы и ознакомительные семинары для будущих специалистов по оценке на основные темы (такие, как статистика, анализ данных, политика оценивания, психометрия). Можно получать стипендии и гранты на обучение. |

**Пример карты состояния системы оценивания на уровне страны/региона**

**по конкретному виду программы оценки**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **СТРАНА / РЕГИОН** | **Уровни сформированности базы для *национальных крупномасштабных исследований,* оценивающих учебные достижения учащихся** | | | |
| **ПАРАМЕТРЫ**  **КАЧЕСТВА** | ЛАТЕНТНЫЙ  Отсутствие или отклонения от необходимых условий | НА СТАДИИ СТАНОВЛЕНИЯ  На пути к установлению минимальных требований | СФОРМИРОВАННЫЙ  Приемлемые минимальные требования | РАЗВИТЫЙ  Передовые подходы |
|  |  |  |  |  |
| **Благоприятная среда**  Общие принципы политики, организационные структуры и бюджетное и кадровое обеспечение как условия для работы по оценке в стране и их значение для или непосредственное влияние на работу по оценке | | | | |
| БС1 - Четкое определение политики |  |  |  |  |
| БС2 - Сильная позиция руководства |  |  |  |  |
| БС3 - Обеспечение бюджетного финансирования в достаточных объемах |  |  |  |  |
| БС4 - Наличие соответствующих организационных структур |  |  |  |  |
| БС5 - Кадровое обеспечение |  |  |  |  |
| БС6 - Создание возможностей для формального обучения методам проведения крупномасштабных оценок |  |  |  |  |
| **Координация и согласованность с элементами образовательной системы**  Степень, в которой система оценивания согласована с другими элементами образовательной системы | | | | |
| КС1 – Учебный план (соответствие целям обучения) |  |  |  |  |
| КС2 – Учебники и методические материалы |  |  |  |  |
| КС3 – Подготовка и повышение квалификации учителей |  |  |  |  |
| **Технология**  Степень, в которой система оценивания соответствует стандартам разработки теста, администрирования, анализа и использования результатов | | | | |
| TQ1 – Дизайн |  |  |  |  |
| TQ2 –Администрирование |  |  |  |  |
| TQ3 – Анализ |  |  |  |  |
| TQ4 – Использование результатов (честным и эффективным образом) |  |  |  |  |

1. Сlarke M. Framework for building an effective assessment system. The World Bank, 2011. [↑](#footnote-ref-1)
2. Когда используется, как существительное, *оценивание* может рассматриваться как определенный инструментарий, например, такой, как тест. [↑](#footnote-ref-2)
3. Перечень компьютерных программ адаптивного тестирования можно найти на сайте <http://www/psych.umn.edu/psylabs/catcentral/>. [↑](#footnote-ref-3)
4. Система оценивания учащихся служит целому ряду информационных целей или потребностей, таких как информирование тех, кто учится и тех, кто учит, определение прогресса учащихся, измерение достижений учащихся и предоставление неполной отчетной информации. Все эти цели и принимаемые на этой основе решения должны в конечном итоге привести к улучшению качества образования и уровня обучения в системе. [↑](#footnote-ref-4)
5. Другие возражают, что вложения в оценивание должны рассматриваться с точки зрения использования результатов, а не основываться на таких данных как стоимость на одного ученика. [↑](#footnote-ref-5)
6. Например, число стран, принимавших участие в PISA, возросло с 43 в 2000 до 66 в 2007. Сравнительно малое число развивающихся стран приняло участие в международной оценке достижений учащихся. Эти страны постоянно показывали низкие результаты, ограничивали количество информации, которую можно получить на основе данных, для лучшего понимания ситуации и усовершенствования своих образовательных систем. [↑](#footnote-ref-6)
7. Одной из наиболее популярных инициатив такого рода является Оценка чтения в начальных классах (EGRA). В соответствии с сайтом Американского агентства международного развития (<https://www/eddataglobal.org/>): “Оценка чтения в начальных классах – устное оценивание, разработанное для измерения самых базовых навыков и овладения грамотностью в начальных классах… Инструментарий данного оценивания обычно адаптируется для использования в конкретной стране и на определенном языке. Это индивидуально проводимое устное оценивание, на которое затрачивается порядка 15 минут на одного ребенка. Поэтому его лучше применять на выборке учащихся, а не использовать подход как при переписи населения. Цель такого оценивания – задокументировать успеваемость ученика по навыкам чтения в начальных классах для того, чтобы проинформировать министров и доноров о потребностях системы по улучшению обучения». [↑](#footnote-ref-7)
8. В данной работе не рассматривается тестирование природных способностей, психологическое тестирование или тестирование на рабочем месте; также как не рассматривается отдельно оценивание на уровне высшего образования (хотя многие из обсуждаемых вопросов могут применяться на этом уровне). [↑](#footnote-ref-8)
9. Ravela и др. (2008) отмечают, что оценка является необходимым, но не достаточным условием для улучшения образования. Есть доказательства того, что одно лишь наличие и распространение информации имеет определенный эффект на участников процесса. Но оценка – один из ключевых элементов образовательной политики; другие элементы включают подготовку и повышение квалификации преподавателей, условия труда учителей, школьное управление и руководство, разработку учебных планов, учебников и методических материалов, вложение ресурсов в соответствии с потребностями различных групп населения, и продуманные действия тех, кто отвечает за образование с тем, чтобы решать возникающие проблемы. [↑](#footnote-ref-9)
10. Rodriguez (2004) сообщает об изменениях в таких же размерах данных TIMSS по математике в США в результате эффективной организации внутриклассного оценивания. [↑](#footnote-ref-10)
11. Стандарты – это «что» - они определяют, что учащиеся должны знать и уметь. Учебный план – это «как» - он задает каркас учебного процесса, направленного на достижение стандартов. Обучение – это «каким образом» - оно обеспечивает конкретные учебные мероприятия и выделяет те, которые способны поддержать обучение учащихся. Оценивание – это «насколько хорошо» - оно измеряет результаты обучения чтобы иметь информацию для образовательной практики и составления учебного плана. [↑](#footnote-ref-11)
12. Некоторые страны имеют хорошие результаты обучения, но не проводят много тестов (Финляндия). Другие страны много тестируют (США), но не показывают такие уж хорошие результаты при международном сравнительном оценивании. [↑](#footnote-ref-12)
13. [↑](#footnote-ref-13)
14. Greaney и Kellaghan (1995) отмечают, что из-за того, что результатам экзамена придается большое значение, учителя готовят к ним, в результате неадекватные возможности в получении релевантных знаний и умений предоставляются ученикам, которые уходят из школы на ранней стадии. К практике, которая ассоциируется с экзаменами и может создавать несправедливые ситуации для каких-то учащихся, относятся выставление баллов, требование заплатить взнос, частное репетиторство, проведение экзамена на языке, незнакомом для ученика и разного рода злоупотребления. Неравенство для учеников также создает использование системы квот при рассмотрении разных показателей по успеваемости, обусловленных местоположением, этнической или языковой принадлежностью. [↑](#footnote-ref-14)
15. Существует несколько факторов, обуславливающих разницу параметров использования в отношении этих оцениваний разными странами. Один важный фактор – официальное видение и цели системы образования и осознанная роль оценивания в достижении такого восприятия (например, Lockheed, 2010). Другой фактор – исторически унаследованное оценивание в конкретной системе образования, что может обуславливать привлекательность определенного типа оценивания (Madaus, Clarke, & O’Leary, 2003). Еще один – способность разных заинтересованных сторон системы эффективно проводить различные типа оценивания (Greaney & Kellaghan, 2008). Наконец еще один – стоимость, осознаваемая или реальная проведения оценивания (Wolff, 2007). [↑](#footnote-ref-15)
16. Много споров относительно того, должны ли проводящие оценивание структуры располагаться в министерствах образования или вне их. На самом деле их расположение не так важно, как культура прозрачности и непрерывности проведения оценивания (Ravela и др., 2008). Такая культура достигается, когда оценивание имеет ясный мандат и прочную структуру, при этом необходимо, чтобы система оценивания подкреплялась определенным юридическим статусом. [↑](#footnote-ref-16)
17. С точки зрения сокращения расходов государственного сектора выгодным вариантом является приватизация функции оценивания. Это предполагало бы, что конкурс тех, кто предоставляет такую услугу, уменьшит цену и повысит качество оценивания. Однако, следует создать определенный задел прежде, чем ставить долгосрочную цель использования по крайней мере смешанных частно/государственных структур, которые будут предоставлять эти услуги. [↑](#footnote-ref-17)
18. Ссылки на две из наиболее полезных баз данных в режиме online: <http://www.inca.org.uk/> и <http://epdc.org/>. [↑](#footnote-ref-18)
19. Brinkley, Guthrie и Wyatt (1991) обследовали проведение крупномасштабных оцениваний и экзаменов в странах ОЭСР. Larach и Lockheed (1992) провели аналогичное исследование оценочных мероприятий, проведенных при поддержке Банка. Macintosh (1994) провел изучение в 10 странах (Австралия, Бахрейн, Англия и Уэльс, Гватемала, Израиль, Малайзия, Намибия, Польша, Шотландия и Словения) природы изменения практики и теории в оценивании достижений учащихся, а также социо-культурных и технических факторов, связанных с разрывом между теорией и практикой. [↑](#footnote-ref-19)
20. Данный инструмент, а также представленные в Приложении 3 и 4 инструменты были разработаны при поддержке Трастового Фонда в рамках программы Российской Поддержки Развитию в Области Образования (READ) и Всемирного Банка. Главное назначение программы READ – помощь развивающимся странам в создании устойчивой системы оценки достижений учащихся. [↑](#footnote-ref-20)
21. Термин «латентная» может быть применен к системам тех странам, где нет формальных мероприятий по оцениванию или где образование временно приостановлено из-за войны или других конфликтов. [↑](#footnote-ref-21)
22. Несмотря на то, что в краткосрочной перспективе концентрация ресурсов на прогресс в качестве одного параметра (например, благоприятная внешняя среда) может быть выгодна системе, но в долгосрочной стратегии качество каждого параметра является важной составляющей эффективной системы оценивания. [↑](#footnote-ref-22)