

Приложение 7. Методические рекомендации по разработке подходов к оцениванию итогового сочинения образовательными организациями высшего образования, доработанные по результатам общественно-профессионального обсуждения

1. Общие требования к разработке схем оценивания развернутых ответов и проблемы, связанные с их разработкой.

Что такое схема оценивания?

Как правило, описательные схемы оценивания разрабатываются для руководства анализом продукта или процессов, характеризующих учебные достижения. Обычно они применяются там, где необходимо качественное суждение в оценивании широкого круга субъектов или действий. Оценивание письменного текста можно привести в качестве общего примера применения схем оценивания. Суждения относительно качества определенного письменного текста могут варьироваться в зависимости от **критериев**, которых придерживается конкретный оценивающий. У одного оценивающего могут превалировать лингвистические характеристики текста, в то время как другой оценивающий будет больше уделять внимание степени убедительности аргументации. Высококачественное эссе, скорее всего, будет отличать комбинация этих и других факторов. Разработав предварительно схему для оценочного процесса, мы сможем понизить его субъективность.

Когда схемы оценивания являются приемлемым оценочным инструментом?

Схемы оценивания – одна из альтернатив, доступных для оценки учебных работ. Они основаны на дескриптивных шкалах и позволяют оценить степень, в которой выражено соответствие определенному критерию.

Присвоение числового веса вспомогательным умениям внутри процесса – еще один оценочный прием, который можно применять для определения степени соответствия критерию. Числовые величины, однако, в обратной связи не представляют никакой информации испытуемым, с помощью которой они могли бы продолжить работу по совершенствованию своих умений. Учащийся, получивший 70 из 100, может оставаться без малейшего представления о том, над чем ему необходимо работать дополнительно для совершенствования. Схемы оценивания позволяют решить эту проблему, представляя описания ожидаемого на каждом определенном уровне. Эти описания позволяют испытуемым понять, почему они получили тот или иной балл и, что им нужно делать, чтобы исправить проявившиеся недостатки.

То, является ли применение схем оценивания обоснованным, зависит от цели оценивания. Оценочные схемы имеют как минимум два достоинства в оценочном процессе. Во-первых, они позволяют исследовать степень соответствия определенному критерию: во-вторых, они предоставляют учащимся обратную связь, позволяющую им совершенствоваться на основе полученной информации. Если эти достоинства созвучны цели оценивания, тогда оценочные рубрики, вероятно, являются адекватным оценочным средством.

Применяются разные типы схем оценивания. То, какой тип приемлем в конкретной ситуации, также зависит от цели оценивания. Как правило, рассматриваются различия между, с одной стороны, холистическими и аналитическими схемами оценивания, с другой стороны, обобщенными и специализированными.

В начальной фазе разработки оценочной схемы разработчик должен определить, что станет оценочным **критерием**. Например, два фактора, которые можно рассматривать при оценке письменной работы, это грамматическая корректность и степень убедительности аргументации. Аналитическая оценочная рубрика, во многом напоминающая контрольный лист, позволяет отдельно оценивать каждый из этих факторов. Каждый критерий оценивается по отдельной дескриптивной шкале.

Холистическая схема может быть трансформирована и включать отдельный набор критериев для оценки, например, убедительности аргументации. Такая трансформация приведет к появлению аналитической схемы с двумя факторами: качество письменной речи и убедительность аргументации. Каждый фактор будет оценен отдельным баллом. Часто числовые значения приписываются в оценке каждого критерия. Как пояснялось выше, преимущество применения схемы оценивания в отличие от просто взвешенного балла, заключается в том, что схемы содержат описание того, что ожидается от испытуемого на каждом уровне. Испытуемые могут использовать эту информацию для совершенствования своих достижений в будущем.

В некоторых случаях бывает невозможно разбить оценку на отдельные факторы. Когда наблюдается наложение наборов критериев для оценки разных факторов, холистическая схема оценивания может быть предпочтительней аналитической. В холистической схеме критерии рассматриваются в сочетании на единой дескриптивной шкале. Холистические оценочные схемы способствуют более широким суждениям относительно качества процесса или продукта.

Выбор применения аналитической схемы оценивания не исключает возможность присутствия холистического фактора. Холистическое суждение может быть встроено в аналитическую схему как одна из оцениваемых категорий. Одной из проблем здесь является то, что наложение критериев, установленных для холистического суждения, на другие оцениваемые факторы становится неизбежным. Когда одной из целей оценивания является присвоение общего балла, такое наложение должно быть внимательно отслежено и проконтролировано. Оценивающий должен определить, не получают ли в результате наложения определенные критерии больший вес, чем предполагалось изначально. Другими словами, оценивающий должен проследить за тем, чтобы испытуемый не был непредумышленно излишне наказан за определенную ошибку.

Обобщенные против специализированных

Оценочные схемы могут предназначаться для оценивания отдельного задания или более широкой категории задач. Если задачей учебного курса было развитие письменных коммуникативных умений, обобщенная оценочная схема может быть разработана и использована для оценивания всех письменных текстов определенного учащегося.

Если каждый письменный текст посвящен разным историческим событиям или литературным произведениям и целью оценивания является определение владения испытуемым данным материалом, то использование обобщенной схемы для оценивания последовательности текстов может быть не вполне адекватным. Для того, чтобы оценить фактологические и концептуальные знания учащихся, возможно, придется разработать отдельные оценочные схемы для каждого текста. «Специализированная» схема оценивания предназначена для оценивания достижения учащегося в отдельном оценочном случае.

Оценочные схемы могут быть составлены таким образом, что будут содержать как обобщенный, так и специализированный компонент.

Как разрабатываются схемы оценивания?

Первый шаг в разработке схемы оценивания представляет собой четкое определение качеств, которые должны быть продемонстрированы испытуемым для подтверждения достижения требуемого уровня умения в соответствующей области. Эти качества, выраженные в **критериях**, будут представлять их высший уровень. Затем нужно будет принять решение относительно того, как будет лучше получить требуемую информацию, с помощью холистической или аналитической схемы. Если принимается решение в пользу

аналитической схемы оценивания, тогда каждый критерий рассматривается отдельно, и разрабатываются описания разных уровней на шкале. Эта процедура приводит к появлению отдельных описательных схем для каждого оцениваемого фактора. Для холистических оценочных рубрик рассматривается набор критериев для всех уровней единой шкалы, которая становится результатом процедуры разработки.

Следующий этап – разработка и написание дескрипторов для выбранного вида шкалы. Общие требования к дескрипторам можно представить следующим образом:

- Позитивность
- Определенность
- Ясность
- Краткость
- Независимость

Первое требование означает, что желательно чтобы дескриптор описывал то, что умеет делать участник, а не то, что он не сделать пока не в состоянии.

Второе требование означает, что дескриптор не должен содержать неопределенности, быть четким («Участник в той или иной форме...» - пример не вполне удачной формулировки).

Третье означает, что дескриптор должен быть понятен тому, кому он предназначен.

Четвертое требованию относится к тому, чтобы дескриптор не содержал нерелевантной информации. И, наконец, важное требование №5. Описание дескрипторов не единый текст, и каждый должен быть понятен без ссылок на другие. Еще одна, к сожалению достаточно распространенная ошибка – использование в описании одной характеристики, ссылок на параметры, оцениваемые в этой же схеме отдельно.

После определения критериев высшего уровня решения, внимание разработчика может быть переключено на определение минимально допустимого уровня. Какое решение будет представлять очень ограниченное понимание оцениваемого концепта? Различие между критериями для высшего уровня решения и минимально допустимого с большой долей вероятности обозначит критерии для среднего уровня. Такой подход приведет к трем уровням оценивания.

Если есть потребность в более широком разграничении, можно начать сравнение критериев для каждого из определенных оценочных уровней. Такая процедура сравнения может продолжаться до получения желаемого количества уровней оценки или до того момента,

когда проведение различий между критериями становится невозможным. Если значимых различий между оцениваемыми категориями уже нельзя сделать, создание дополнительных категорий оценки следует прекратить. Лучше иметь меньшее количество значимых уровневых категорий, чем большое количество таких, разницу между которыми очень трудно обозначить.

Каждую оценочную категорию следует определять, используя описание деятельности, а не суждения о деятельности. Например, «текст испытуемого не содержит орфографических ошибок», предпочтительнее, чем «хорошая орфография». Слово «хорошая» требует от оценивающего суждения, в то время, как «не содержит ошибок» позволяет сделать однозначное определение. Формулировки рубрик требуют проверки, с привлечением группы экспертов. Различия в оценке, произошедшие вследствие нечетких формулировок должны быть устранены.

Валидность схем оценивания.

Валидизация является процессом сбора свидетельств в поддержку приемлемости и адекватности выводов, делаемых на основе результатов применения оценочных инструментов. Валидность соотносится со степенью корректности интерпретации этих результатов. Как правило, три типа свидетельств в поддержку валидности оценочного инструмента подвергаются рассмотрению: содержательное, конструктивное и критериальное. Рассмотрим указанные свидетельства и то, как их надлежит учитывать в разработке схем оценивания.

Содержательное свидетельство

Содержательные свидетельства отражают степень, в которой ответы испытуемого в процессе применения определенного оценочного инструмента соотнесены с содержанием соответствующего предмета. Например, если в экзамене по истории вопросы построены с использованием сложных синтаксических конструкций, то невольно могут измеряться умения испытуемых, связанные с пониманием текста, а не со знанием истории. Интерпретируя результат экзамена, экзаменатор может сделать вывод о том, что у испытуемого отсутствуют определенные исторические знания, в то время, как он просто не понял вопрос. Экзаменатор интерпретировал свидетельство не верно – сделал интерпретацию невалидной.

Содержательное свидетельство также связано со степенью в которой оценочный инструмент адекватно представляет содержательный домен. Экзамен по арифметике, в котором доминируют задачи на сложение, будет не достаточно адекватен для выводов о способности испытуемого решать задачи на вычитание, умножение и деление. Правильное решение пятидесяти задач на сложение и двух на умножение не представляет убедительного свидетельства о том, что испытуемый умеет решать задачи на вычитание, умножение или деление.

Содержательное свидетельство следует рассматривать и при разработке схем оценивания. Например, при оценке определенного математического знания испытуемого итоговый оценочный критерий может выделить структуру предложение или орфографию перед демонстрируемым математическим знанием. Результат испытуемого, интерпретируемый как отражение его математического знания, будет на самом деле, отражать его грамматические умения. Благодаря схеме оценки, итоговый результат будет неадекватно измерять математические знания испытуемого. Не имеется в виду, что в данном задании не могут измеряться языковые умения испытуемого. Но если задание предназначено для измерения синтаксиса, орфографии и математики, оценочная рубрика должна отражать все эти области.

Конструктивное свидетельство

Конструкты можно представить как процессы, внутренне присущие индивидууму. Примером конструкта может служить процесс рассуждения индивидуума. Хотя рассуждение проходит внутри индивидуума, его можно частично представить через результаты и пояснения. Изолированный правильный ответ, однако, не предоставляет четкого убедительного свидетельства природы, лежавшего в его основе процесса рассуждения. Хотя ответ и является итогом процесса рассуждения испытуемого, правильный ответ может стать итогом некорректного рассуждения. Когда целью является оценка рассуждения, то и продукт (т.е. ответ) и процесс (т.е. рассуждение) необходимо потребовать к предъявлению и исследованию.

Оценочные критерии представленные внутри схемы могут быть не связанными с интересующим конструктом. Это подобно примеру, в котором орфографические ошибки рассматривались в математическом задании. Однако, то, что интересует нас в данном случае, это, являются ли оцениваемые элементы решения адекватными индикаторами оцениваемого конструкта. Если исследуемым конструктом является размышление, тогда орфографические ошибки в объяснении испытуемых нерелевантны целям оценивания и не должны включаться

в оценочные критерии. С другой стороны, если цель оценивания включает орфографию и рассуждение, тогда оба эти элемента должны быть отражены в оценочных критериях. Конструктивное свидетельство, это свидетельство, которое подтверждает то, что оценочный инструмент полно измеряет заявленный конструкт и только его.

В ходе текущего формативного оценивания могут измеряться различные конструкты. Решение проблем, креативность, письменные навыки, самоуважение, отношения входят в число возможных конструктов в измерении которых может быть заинтересован преподаватель. Вне зависимости от конструкта, необходимо будет определить фасеты конструкта, которые могут быть предъявлены и станут убедительным свидетельством процессов, происходящих внутри учащегося. Эти фасеты затем следует тщательно встроить в оценочный инструмент и отразить в критериях оценки.

Критериальное свидетельство

И, наконец, еще один тип свидетельства – критериальное. Этот тип свидетельства отражает то, насколько результат оценивания коррелирует с текущим или будущим событием. Иначе, критериальное свидетельство можно представить как степень в которой результат выполнения определенной задачи может быть обобщен для вывода о способности испытуемого выполнить другие, более релевантные действия.

Другими словами, высокий балл, полученный за контрольную работу, должен предполагать успешное функционирование вне текущего учебного заведения, в учебном заведении более высокого уровня или на будущем рабочем месте.

Вопросы валидности в разработке схем оценивания

Вопросы валидности интерпретации результатов оценки должны рассматриваться до выбора или разработки заданий или инструмента оценки. Хорошо разработанная схема не компенсирует недостатков инструмента оценки. Так как установление валидности связано с целью оценивания, заинтересованные стороны должны четко заявить о том, что они надеются узнать об испытуемых (т.е. цель), и как испытуемые проявят соответствующие умения (т.е. задачи). Заявленные цель и задачи будут определять разработку оценочной схемы.

Для того чтобы оценочный инструмент выявлял свидетельства, согласующиеся с целью, рекомендуется пронумеровать предполагаемые задачи определенного оценивания, а затем

обозначать соответствующим номером вопрос или задание, которые соответствуют данной задаче. Таким образом станет четко видно те задачи, которые не нашли отражение в оценивании. Этот метод оценки адекватности инструмента может быть применен по отношению к схеме оценивания. Во-первых, четко изложить цель и задачи оценивания. Затем, определите критерии, соотносимые с каждой задачей. Если какая-либо из задач не нашла отражения ни в одной из оцениваемых категорий схемы, она вряд ли предоставит необходимые свидетельства, для решения соответствующей задачи. Если какие либо оценочные критерии не соотносятся ни с одной из задач оценивания, точно также можно будет поставить под сомнение адекватность представленной схемы.

Шаги в оценке адекватности оценочных категорий заявленной цели можно представить следующим образом:

Оценка адекватности оценочных категорий заявленной цели



определить цель и задачи оценивания



определить критерий для каждой задачи



проверить:

- | |
|--|
| <ol style="list-style-type: none">1) Все ли задачи измеряются представленными критериями?2) Есть ли критерии, которые не связаны ни с одной из задач? |
|--|

Размышление о цели и задачах оценивания должно привести к выбору того, какие формы свидетельств – содержательные, конструктивные и/или критериальные – следует рассмотреть. Например, если в задачи оценивания входит получение свидетельства о знаниях испытуемого в рамках определенной содержательной области, например, знание исторических фактов, следует рассмотреть приемлемость содержательного свидетельства. Если инструмент предназначен для измерения способности к рассуждению или другим внутренним процессам и требующим косвенного исследования, следует рассмотреть приемлемость конструктивных свидетельств. Если цель оценивания определить то, как

испытуемый будет действовать вне учебных условий или в иной ситуации, рассмотрению подлежит приемлемость критериальных свидетельств.

Осознание в ходе всего процесса разработки схемы оценивания существования разных типов свидетельств, поддерживающих валидность, вполне вероятно повысит адекватность интерпретаций, когда она будет использована. Свидетельства валидности могут быть исследованы после того, как будет представлен предварительный вариант схемы. Ниже в таблице представлен список вопросов, которые могут оказаться полезными в оценке адекватности представленной схемы оценивания в отношении заявленной цели. Эта таблица разделена в соответствии с типом рассматриваемого свидетельства.

Вопросы для исследования типов свидетельства валидности схемы оценивания

Содержательные	Конструктивные	Критериальные
<ol style="list-style-type: none">1. Относятся ли критерии к какому либо нерелевантному элементу содержания?2. Отражены ли в оценочной рубрике все предполагаемые аспекты содержания?3. Есть ли какой либо элемент содержания, отраженный в задании, но не отраженный в оценочной рубрике, хотя следует быть отражен в ней?	<ol style="list-style-type: none">1. Все ли важные фасы конструкта оцениваются предложенными критериями?2. Является ли какой либо из предложенных критериев оценивания нерелевантным по отношению к заявленному конструкту?	<ol style="list-style-type: none">1. Как критерии оценивания отражают компетенции, наличие которых связано с успешностью в соответствующей деятельности в будущем или в сходных условиях?2. Какие важные компоненты будущей деятельности могут быть оценены с помощью данного инструмента?3. Как критерии оценивания измеряют важные

		<p>компоненты будущей деятельности?</p> <p>4. Есть ли фасеты будущей деятельности, которые не нашли отражение в критериях оценивания?</p>
--	--	---

Надежность оценивания

Надежность соотносится с последовательностью в результатах оценивания. Например, при использовании надежного теста испытуемый вправе полагать, что его результат не будет зависеть от времени выполнения теста, времени его проверки, личности проверяющего... В случае ненадежного экзамена результат испытуемого может варьироваться в зависимости от факторов никак не связанных с его целью.

Две формы надежности, которые наиболее связаны с оценочными схемами, основаны на надежности оценивающего (или репитера). Надежность репитеров связана с тем насколько последовательно несколько репитеров оценивают одну работу, или как один репитер в разное время оценивает одну работу. Первый тип надежности называют «интер-репитер надежность», в то время как второй тип – «интра-репитер надежность».

Интер-репитер надежность

Интер-репитер надежность связана с той проблемой, что результат испытуемого может варьироваться от репитера к репитеру. Учащиеся часто критикуют экзамены в которых их результат кажется зависимым от субъективных суждений оценивающего. Например, одним из способов анализа экзамена в форме эссе является прочтение работы и вынесение суждения о качестве умений письма испытуемого и качестве представленного текста. Два независимых репитера, не опираясь на набор критериев, направляющих процедуру оценивания, часто по-разному оценивают представленную работу. Каждый репитер опирается на собственную систему критериев. Оценочные схемы решают эту проблему, формализуя критерии на каждом оцениваемом уровне. Хотя они полностью не способны устранить

вариации между рейте́рами, хорошо разработанная схема способна значительно сократить расхождения.

Интра-рейтер надежность

Факторы, являющиеся внешними по отношению к цели оценивания, могут влиять на то, как конкретный рейте́р оценивает работы испытуемых. Например, рейте́р может устать от процедуры оценивания и меньше уделять времени анализу каждой последующей работы. Так некоторые работы могут быть оценены иначе, если бы их оценивали в более раннее время. Настроение рейте́ра в конкретный день или знание автора проверяемой работы могут также повлиять на процедуру оценивания. Хорошая работа ученика считающегося слабым может быть оценена строже, чем идентичная работа «сильного» учащегося. Интра-рейтер надежность относится к каждой ситуации, в которой манера оценивания конкретного рейте́ра меняется во времени в разных направлениях. Непоследовательность в оценках определяется причинами присущими рейте́ру, а не истинным различиям в качестве представленных работ. Качественные схемы оценивания помогают в решении проблемы интра-рейтер надежности заранее предоставляя описания оценочных критериев. На протяжении всей процедуры оценки рейте́ру следует возвращаться к ним для того, чтобы поддерживать последовательность решений.

Вопросы надежности в разработке схем оценивания

Ясность и четкость схемы оценивания способна улучшить как интер-рейтер, так и интра-рейтер надежность. Схема с четко определенными категориями оценки должна способствовать поддержке последовательности в оценивании независимо от того, кто проводит оценку или когда она проводится. Следующие вопросы могут быть использованы для оценки того, насколько ясной является определенная схема.

- 1) Четко ли определены критерии оценки?
- 2) Достаточно ли ясны различия между критериями оценки и степенями выражения описываемых ими качеств?
- 3) Могут ли два независимых рейте́ра прийти к одинаковому результату оценивания одной работы на основе предложенной схемы?

Ответ «нет» на любой из вопросов должен привести к пересмотру соответствующих категорий схемы.

Другой способ уточнения схемы оценивания связан с использованием «якорных» работ. Якорные работы, это предварительно оцененные работы, которые иллюстрируют нюансы оценочных схем. Отдельный рецензент имеет возможность обратиться к якорной работе в любой момент процесса оценивания для прояснения различия между уровнями оценки.

После того, как все возможное сделано для уточнения оценочных категорий, оценочные схемы и якорные работы могут быть представлены для апробации отдельной группой преподавателей. Любые расхождения в их оценках будут указывать на то, какие компоненты оценочной схемы нуждаются в дальнейшей доработке и разъяснении. Все расхождения в интерпретации дескрипторов необходимо обсуждать и вносить необходимые уточнения. Хотя этот процесс достаточно времязатратен, он может значительно повысить надежность оценивания.

Другая проблема надежности связана с приемлемостью отдельной схемы для определенной группы испытуемых. Схема оценивания, с помощью которой последовательно измеряются результаты одной выборки испытуемых, может приводить к непоследовательности в оценке другой группы испытуемых. Например, если задача вписана в определенный контекст, одна группа испытуемых может быть в большей степени знакома с этим контекстом, чем другая. Последняя группа получит более низкие результаты лишь за счет недостаточного знакомства с контекстом задания. Если бы эти же самые испытуемые выполняли другое задание, основанное на том же учебном материале, но встроенное в знакомый контекст, их результат был бы выше. Когда причина вариативности в результатах не связана с целью оценивания, результаты считаются ненадежными.

Иногда в процессе оценивания рецензенты осознают, что начинают пользоваться имплицитными критериями, отсутствующими в схеме оценивания. Всегда, когда возможно, необходимо знакомить испытуемых с со схемами оценки, чтобы они могли строить ответ таким образом, чтобы предоставить убедительное свидетельство того, что их работа соответствует критериям. Если испытуемых заранее ознакомили со схемой оценки, они не должны нести ответственность за не продекларированный в ней критерий. Выявление имплицитных критериев поможет в уточнении рубрик.

Установление надежности является условием установления валидности. Хотя валидное оценивание по определению надежно, последнее не всегда так. Надежная оценка не обязательно является валидной. Схема оценивания может привести к невалидным

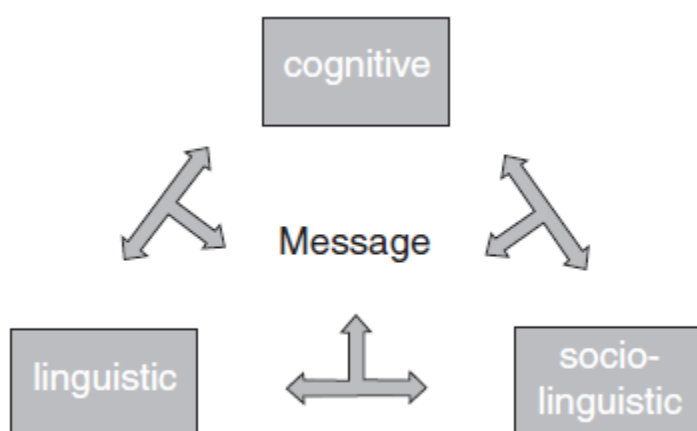
интерпретациям тогда, например, когда критерии оценки сфокусированы на элементе ответа, который не связан с целью оценивания. В то же время этот критерий может быть так четко сформулирован, что оценка по нему будет в высокой степени надежна и не зависеть от рейтера.

Схемы описывают общие, синтезированные критерии, проявляющиеся в большом числе индивидуальных ответов и, таким образом, не в состоянии отражать уникальные характеристики каждого ответа. Тот, кто зависит только от оценочного критерия во время оценивания с меньшей степенью вероятности обратит внимание на расхождение возникающее между конкретным ответом и конечным результатом.

(Barbara M. Moskal & Jon A. Leydens. Scoring Rubric Development: Validity and Reliability)

Рассмотрим, в качестве примера, краткое описание изменений в оценивании письменных работ в рамках батареи тестов, разработанных в Cambridge Assessment.

Работа началась с пересмотра основной модели измеряемого конструкта, который был в общем плане принят в следующей форме:



На схеме представлено взаимодействие элементов конструкта: в центре содержание (что хотел сказать испытуемый) и связанные с содержанием аспекты – когнитивный (когнитивные операции, используемые для формулировки ответа), социолингвистический (понимание адресата текста, умение сделать текст для него максимально читабельным) и лингвистический (арсенал используемых языков средств, его адекватность).

В связи с данным конструктом была принята соответствующая структура заданий: условно «Напишите письмо своему другу про А и Б и спросите о В», где каждая часть соответствует определенному элементу конструкта:

Напишите	лингвистический
Напишите про А	содержание
Напишите письмо своему другу про А и Б	социолингвистический
Напишите письмо своему другу про А и Б и спросите о В	когнитивный

Соответственно выбраны четыре параметра для аналитической схемы оценивания:

Содержание

Коммуникативная цель (социолингвистический)

Организация (когнитивный аспект нельзя проконтролировать напрямую, поэтому он оценивается косвенным путем через организацию текста)

Язык (лингвистический)

Каждый параметр оценивается по шкале 1-5 баллов.

5 - полное соответствие, 3- среднее, 1 – минимальное. 2 и 4 балла выставляются если в ответе присутствуют элементы, представленные в дескрипторах 1 и 3 (2 балла), или 3 и 5 (4 балла).

Отдельно отметим, что эта схема не предусматривает автоматического наказания на слишком короткий или слишком длинный ответ, поскольку это противоречит оцениваемому конструкту. При этом баллы за такие тексты могут быть снижены при использовании существующих шкал. Так слишком короткий текст, скорее всего будет не соответствовать таким параметрам как эффективное использование связующих элементов (организация) или использование сложных грамматических форм (язык). А слишком длинный текст таким параметрам как релевантность содержания поставленной задаче (содержание) или удерживается ли внимание читателя (коммуникативная цель).

(Sad S Lim (2012) Developing and validating a mark scheme for writing. Cambridge Research Notes #49)

2. Подходы, заложенные в Рекомендуемые критерии оценивания итогового сочинения организациями, реализующими образовательные программы высшего образования

Для оценивания предлагается использовать аналитическую шкалу с выделением следующих критериев:

Соответствие теме

Аргументация. Привлечение литературного материала

Композиция

Качество речи

Оригинальность сочинения

Речевые нормы

Орфографические нормы

Пунктуационные нормы

Грамматические нормы

Фактическая точность в фоновом (не литературном) материале

Холистическая интерпретация схемы говорит о том, что от испытуемых ожидается сочинение, отвечающее следующим требованиям:

«Участник раскрывает тему сочинения или в той или иной форме рассуждает на предложенную тему, выбрав убедительный путь её раскрытия (например, отвечает на вопрос, поставленный в теме, или размышляет над предложенной проблемой и т.п.), коммуникативный замысел сочинения выражен ясно. Участник при раскрытии темы сочинения строит рассуждение на основе произведения (произведений) отечественной или мировой литературы по собственному выбору, показывая разный уровень осмысления литературного материала: от элементов смыслового анализа (например, тематика, проблематика, сюжет, характеры и т.п.) до комплексного анализа художественного текста в единстве формы и содержания; выбор литературного произведения и аспекты его анализа оправданны с точки зрения темы;

допущено не более 1 фактической ошибки, связанной со знанием литературного материала (ошибка в написании автора и названия произведения, имен персонажей и топонимов произведения, в изложении сюжетной линии, литературных и исторических фактов и т.п.). Сочинение отличается композиционной цельностью, логичностью изложения мыслей и соразмерностью частей, внутри смысловых частей нет нарушений последовательности и необоснованных повторов мысли. Сочинение характеризуется творческим, нестандартным подходом к раскрытию темы (присутствуют интересные мысли, или неожиданные и вместе с тем убедительные аргументы, или оригинальные наблюдения и проч.) или яркостью стиля. Речевых ошибок нет, или допущено 1–2 речевые ошибки. Орфографических ошибок нет, или допущена 1 ошибка. Пунктуационных ошибок нет, или допущена 1 ошибка. Грамматических ошибок нет, или допущена 1 грамматическая ошибка. Фактические ошибки отсутствуют.

Разработчики исходили из того, что критерии должны согласовываться с принципом объективности оценивания сочинения. Итоговое сочинение представляет собой литературоцентричную работу, которая, при этом носит надпредметный характер....

Критерики должны отражать специфические запросы вузов к абитуриентам, при этом. При этом нельзя требовать через критерии вузов того, что не требовалось при написании сочинения. Характерной ошибкой при составлении критериев является выход за пределы инструкции, предложенной участникам сочинения

3. Типичные ошибки при разработке критериев

В рамках заданной структуры вузы применяют (исходя из своих условий), варианты схем оценки сочинения, которые порою не вписываются в предложенные рекомендации. При корректировке рекомендуемых критериев или создании собственных критериев вузы вправе отражать свои специфические требования к абитуриентам через систему оценивания итоговых сочинений.

Так, некоторые вузы практикуют использование рекомендуемых критериев с изъятием части критериев (оценивание содержания без учета грамотности или изъятие содержательных критериев с выдвиганием на первый план оценки владения речью). Иногда вводятся специфические критерии («соблюдение этических норм»; «филологическая компетентность», «Знание текстов художественного произведения и /или фактов истории, Культуры»).

Многие решения не вызывают возражений. Вместе с тем анализ вузовских критериев позволил выявить ряд типичных ошибок, которые существенно искажают компаративность представляемых результатов.

1. Повышение планки требований (для 3 баллов требуется опора на 2 и более литературных текста; рекомендуется 400 слов)

3. Культурологическая и/или филологическая компетентность.

Обучающийся применяет термины и понятия литературоведения, культурологии, искусствоведения в качестве инструмента интерпретации и анализа – 1 балл.

2. Неоправданное объединение рекомендованных критериев («Речевое оформление и грамотность»; «Самостоятельность написания и соответствие теме» «Содержание и речь»)

Пример: Оценка качества речи, грамотности сочинения: стилистическое, речевое оформление текста, степень свободного владения речью, соблюдение основных орфографических, пунктуационных и грамматических норм (в совокупности не более 15 ошибок) от 0 до 1 балла

3. Отсутствие «отсекающих» требований (при 0 баллов за соответствие теме сочинение оценивается по другим критериям)
4. Отсутствие необходимой дифференциации абитуриентов.

Пример: зачетная система (по 5 критериям ставятся не баллы, а зачет-незачет)

5. Высокий уровень обобщения аспектов оценивания

1-0 баллов - за содержание; 1-0 – качество речи (вариант 1-0 баллов – за грамотность)

6. Упрощение критериев

«Глубина баллов» по всем критериям: 1-0

7. Дробность оценивания

(1 –0,5-0 б)

8. Невнятность:

Пример1 1-2 ошибки – 5 баллов; 2-3 ош. – 4 б.; 4-5 ош. – 3 б.; 6-7 ош. – 2 б.; 8-9 ош. –1 б. (0 баллов нет)

Пример 2: В сочинении сделано допустимо/недопустимое количество ошибок

9. Громоздкость формулировок критериев

Пример: **К2. Аргументация. Привлечение литературного материала**

3 балла может быть выставлено, если есть прямой и счерпывающий ответ по теме, обнаружено **отличное** знание и **глубокое** понимание текста художественного произведения, а также умение пользоваться литературным материалом для раскрытия темы, давать оценку излагаемым фактам, композиционно стройно, логически последовательно и аргументировано излагать свои мысли, писать правильным и выразительным литературным языком. Орфографические, пунктуационные, речевые, грамматические и фактические **ошибки** либо **отсутствуют** вообще, либо обнаружены единичные нарушения норм письменной речи. <...>

2 балла ставится за сочинение, в котором дан **схематичный** ответ на тему сочинения, допущены **отклонения** от темы, **неточности** в изложении фактического материала, **нарушения** последовательности изложения мыслей. Обнаружено **не более 4** (суммарно) орфографических и пунктуационных ошибок. Присутствуют и другие виды ошибок.

1 балл ставится за сочинение, в котором преобладают **общие фразы**, не подтвержденные литературным материалом; изложение носит **трафаретный** характер или сводится к **простому пересказу** произведения (или учебника), имеются

серьезные нарушения последовательности в выражении мыслей. Обнаружено не более 8 (суммарно) орфографических и пунктуационных ошибок.

0 баллов ставится за сочинение, в котором не понята и не раскрыта тема, обнаруживается полное незнание литературного материала. Допущено более 8 (суммарно) орфографических и пунктуационных ошибок. Присутствуют и другие виды ошибок.

Разберем несколько конкретных критериев.

Один из Вузов представляет следующие критерии оценки сочинения:

Речевые нормы/ Орфографические нормы / Пунктуационные нормы/ Грамматические нормы
Таким образом оценка сочинения сводится только к двум из представленных в рекомендациях ФИПИ критериям (Качество письменной речи и Грамотность), причем исключая те, которые признаны основными.

В другом примере Вуз официально выделяет только два критерия (Содержание и речь/Грамотность) с попыткой в первом критерии объединить два из рекомендуемых (Соответствие теме и Композиция), при этом опять-таки выпадает один из основных критериев (№2). Идеальное выполнение работы представляется следующим образом:

1. Содержание работы полностью соответствует теме.
2. Фактические ошибки отсутствуют.
3. Содержание излагается последовательно.
4. Работа отличается богатством словаря, разнообразием используемых синтаксических конструкций, точностью словоупотребления
5. Достигнуто стилевое единство и выразительность текста.

При этом пункты дескрипторов фактически не являются таковыми, а скорее представляют собой дополнительный список критериев, остающихся без дескрипторов: (Полнота соответствия теме/ Наличие фактических (?) ошибок/ Последовательность изложения (?)/Лексико-грамматический репертуар/ Силевое единство и выразительность).

Остается впечатление, что разработчиков данной схемы больше привлекает холистический подход к оценке. Но в результате имеется схема с расплывчатым описанием подхода к оцениванию.

Сходный подход можно наблюдать и в следующем примере: «Сочинение оценивается по двум критериям и с учетом объема». Указанные критерии — это «оценка содержания и качества речи, грамотности. При том, что в случае недостаточного объема (менее 250 слов) сочинение считается невыполненным, максимальная оценка за него, всего два балла. Если граница между 0 и 1 по второму критерию изложена в количественных показателях, то оценка по «основному критерию» остается несколько загадочной, несмотря на перечисление различных параметров. Нелегко определить разницу перехода от 0 к 1 если нужно обратить внимание на «соответствие теме (проверка умения и навыков формулировать тезисы, анализировать текст), аргументацию, знание историко-литературного контекста (проверка умения логически мыслить, рассуждать, а также подтверждать свои доводы историко-литературными фактами, структуру (проверка умения последовательно выстраивать рассуждение на предложенную тему).»

В другом примере предлагаются критерии, соответствующие рекомендуем, с добавлением критерия «Оригинальность», который, как и следовало ожидать, описывается предельно субъективно: «**яркий стиль** написания работы, **интересные мысли**, **неожиданные** и **убедительные аргументы**».

Впрочем дескрипторы других критериев также вызывают вопросы. Например в критерии «Композиция», только в одном пункте возникает фраза о «точности выражения мысли», оставляя повод для размышления о том, как это соотносится с заявленным критерием.

Рассмотрим следующий пример:

«Сочинение оценивается по пяти критериям:

1. Глубина раскрытия темы.

Данный критерий (содержательный) является главным. Если при проверке сочинения по первому критерию поставлено 0 баллов, то задание считается невыполненным и дальше не проверяется, по всем остальным критериям выставляется 0 баллов.

При оценке следует учитывать объем написанного сочинения. Если в сочинении менее 200 слов, то такая работа считается невыполненной и оценивается на 0 баллов.

Максимальное количество слов в сочинении не установлено. Рекомендуемое максимальное количество слов в сочинении 400 слов.

Обучающийся раскрывает тему сочинения и/или отвечает на поставленный вопрос, аргументирует свои тезисы в соответствии с формулировкой темы, тезисы аргументированы обращением к тексту литературного произведения и/или фактам истории, культуры,

обращение к тексту того или иного литературного произведения и/или к тому или иному факту истории, культуры оправдано логикой и содержанием сочинения – 2 балла.

Обучающийся раскрывает тему сочинения поверхностно и/или отвечает на поставленный вопрос, и/или аргументирует свои тезисы в соответствии с формулировкой темы, но не обращается к тексту литературного произведения или фактам истории, культуры, и/или обращение к тексту литературного произведения и/или фактам истории, культуры не оправдано логикой и содержанием сочинения - 1 балл.

Обучающийся не раскрывает тему сочинения, не даёт ответа на вопрос, и/или обращение к тексту литературного произведения и/или фактам истории, культуры отсутствует - 0 баллов;»

Отметим здесь попытку объединить первые два из рекомендуемых критериев, но при этом совершенно абстрактно представлено и соотношении в плане «глубины раскрытия темы»

«2. Знание текста художественного произведения и/или фактов истории, культуры.

Обучающийся показывает знание текста, апеллирует к тексту в своих суждениях (интерпретирует, цитирует, комментирует, пересказывает, анализирует) текст художественного произведения, и/или обнаруживает знания фактов истории и культуры (интерпретирует, анализирует, комментирует их). Фактические ошибки и неточности отсутствуют - 2 балла.

Показывает недостаточное знание текста и/или неумение его анализировать и интерпретировать, подменяя анализ и интерпретацию пересказом и/или допускает не более 3 фактических ошибок в знании текста или фонового материала - 1 балл.

Показывает незнание текста, неумение его анализировать и интерпретировать и/или допускает более 3 фактических ошибок - 0 баллов.»

Очевидно, что данный критерий применим лишь к сочинению написанному по конкретному литературному произведению, указанному в задании, что не всегда работает (или в данном конкретном Вузе формулируются только такие темы, что уже не вопрос письменной речи, а именно знания конкретного литературного произведения).

«3. Культурологическая и/или филологическая компетентность.

Обучающийся применяет термины и понятия литературоведения, культурологии, искусствоведения в качестве инструмента интерпретации и анализа - 1 балл.

Не применяет термины и понятия литературоведения, культурологии, искусствоведения - 0 баллов.»

Данное требование должно быть специально оговорено и отражено в задании и в принципе может быть предъявляемо только к определенной группе абитуриентов.

«4. Композиционная цельность и логичность изложения.

Сочинение характеризуется композиционной цельностью, его части логично связаны, внутри смысловых частей нет нарушений последовательности и необоснованных повторов - 2 балла.

Сочинение характеризуется композиционной цельностью, его части логически связаны между собой, но внутри смысловых частей есть нарушения последовательности и необоснованные повторы и/или в сочинении прослеживается композиционный замысел, но есть нарушения композиционной связи между смысловыми частями, и/или мысль повторяется и не развивается - 1 балл.

В сочинении не прослеживается композиционного замысла; допущены грубые нарушения последовательности частей высказывания, существенно затрудняющие понимание смысла сочинения - 0 баллов.»

«Необоснованный повтор, прослеживается замысел, мысль не развивается, грубые нарушения, существенно затрудняющие» подобного рода абстрактные описания следует избегать в дескрипторах критериев.

«5. Следование нормам речи.

Речевых ошибок нет, или допущено не более 1 речевой ошибки - 3 балла. Допущены 2 - 3 речевые ошибки - 1 балл. Допущено 4 и более речевых ошибок - 0 баллов.»

В данном случае возникает подозрение о том, что под речевыми ошибками полагаются языковые и вопрос, на каком основании данный критерий стал (по максимальному баллу) основным.

В указанных примерах отражены некоторые из основных недостатков, связанных с разработкой и применением критериев и схем оценки развернутых ответов.

Приложение 1. Критерии оценивания итогового сочинения организациями, реализующими образовательные программы среднего общего образования

К проверке по критериям оценивания допускаются итоговые сочинения, соответствующие установленным требованиям:

Требование № 1. «Объем итогового сочинения»

Рекомендуемое количество слов – от 350.

Максимальное количество слов в сочинении не устанавливается. Если в сочинении менее 250 слов (в подсчет включаются все слова, в том числе и служебные), то выставляется «незачет» за невыполнение требования № 1 и «незачет» за работу в целом (такое сочинение не проверяется по критериям оценивания).

Требование № 2. «Самостоятельность написания итогового сочинения»

Итоговое сочинение выполняется самостоятельно. Не допускается списывание сочинения (фрагментов сочинения) из какого-либо источника или воспроизведение по памяти чужого текста (работа другого участника, текст, опубликованный в бумажном и (или) электронном виде, и др.).

Допускается прямое или косвенное цитирование с обязательной ссылкой на источник (ссылка дается в свободной форме). Объем цитирования не должен превышать объем собственного текста участника.

Если сочинение признано несамостоятельным, то выставляется «незачет» за невыполнение требования № 2 и «незачет» за работу в целом (такое сочинение не проверяется по критериям оценивания).

Итоговое сочинение, соответствующее установленным требованиям, оценивается по критериям:

1. «Соответствие теме»;
2. «Аргументация. Привлечение литературного материала»;
3. «Композиция и логика рассуждения»;
4. «Качество письменной речи»;
5. «Грамотность».

Критерии № 1 и № 2 являются основными.

Для получения «зачета» за итоговое сочинение необходимо получить «зачет» по критериям № 1 и № 2 (выставление «незачета» по одному из этих критериев автоматически ведет к «незачету» за работу в целом), а также дополнительно «зачет» по одному из других критериев.

Критерий № 1 «Соответствие теме»

Данный критерий нацеливает на проверку содержания сочинения.

Участник должен рассуждать на предложенную тему, выбрав путь ее раскрытия (например, отвечает на вопрос, поставленный в теме, или размышляет над предложенной проблемой и т.п.).

«Незачет» ставится только в случае, если сочинение не соответствует теме или в нем не прослеживается конкретной цели высказывания, то есть коммуникативного замысла. Во всех остальных случаях выставляется «зачет».

Критерий № 2 «Аргументация. Привлечение литературного материала»

Данный критерий нацеливает на проверку умения использовать литературный материал (художественные произведения, дневники, мемуары, публицистику, произведения устного народного творчества (за исключением малых жанров), другие литературные источники) для аргументации своей позиции.

Участник должен строить рассуждение, привлекая для аргументации не менее одного произведения отечественной или мировой литературы, избирая свой путь использования литературного материала; при этом он может показать разный уровень осмысления художественного текста: от элементов смыслового анализа (например, тематика, проблематика, сюжет, характеры и т.п.) до комплексного анализа произведения в единстве формы и содержания и его интерпретации в аспекте выбранной темы.

«Незачет» ставится при условии, если сочинение написано без привлечения литературного материала или в нем существенно искажено содержание произведения, или литературные произведения лишь упоминаются в работе, не становясь опорой для аргументации. Во всех остальных случаях выставляется «зачет».

Критерий № 3 «Композиция и логика рассуждения»

Данный критерий нацеливает на проверку умения логично выстраивать рассуждение на предложенную тему. Участник должен выдерживать соотношение между тезисом и доказательствами.

«Незачет» ставится при условии, если грубые логические нарушения мешают пониманию смысла сказанного или отсутствует тезисно-доказательная часть. Во всех остальных случаях выставляется «зачет».

Критерий № 4 «Качество письменной речи»

Данный критерий нацеливает на проверку речевого оформления текста сочинения.

Участник должен точно выражать мысли, используя разнообразную лексику и различные грамматические конструкции, при необходимости уместно употреблять термины.

«Незачет» ставится при условии, если низкое качество речи (в том числе речевые ошибки) существенно затрудняет понимание смысла сочинения. Во всех остальных случаях

выставляется «зачет».

Критерий № 5 «Грамотность»¹

Данный критерий позволяет оценить грамотность выпускника.

«Незачет» ставится при условии, если на 100 слов приходится в сумме более пяти ошибок: грамматических, орфографических, пунктуационных.

Приложение. 2. Рекомендуемые критерии оценивания итогового сочинения организациями, реализующими образовательные программы высшего образования

Данные критерии носят рекомендательный характер. Образовательная организация высшего образования вправе разработать собственные критерии оценивания итогового сочинения, в том числе на основе предложенных.

Сочинение оценивается по десяти критериям и с учетом его объема и самостоятельности.

Критерии № 1 и № 2 являются основными.

Если при проверке сочинения по критерию № 1 или № 2 поставлено 0 баллов, то сочинение дальше не проверяется: по всем остальным критериям выставляется 0 баллов.

Критерии оценивания	Баллы
К1. Соответствие теме	
Участник раскрывает тему сочинения или в той или иной форме рассуждает на предложенную тему, выбрав убедительный путь её раскрытия (например, отвечает на вопрос, поставленный в теме, или размышляет над предложенной проблемой и т.п.), коммуникативный замысел сочинения выражен ясно.	2
Участник поверхностно рассуждает на предложенную тему или рассуждает на тему, близкую к предложенной, коммуникативный замысел сочинения прослеживается.	1
Сочинение не соответствует теме, и/или коммуникативный замысел сочинения не прослеживается.	0
К2. Аргументация. Привлечение литературного материала	

¹ Итоговое сочинение для лиц с ограниченными возможностями здоровья, детей-инвалидов и инвалидов может по их желанию проводиться в устной форме. Оценивание итогового сочинения указанной категории участников итогового сочинения проводится по двум установленным требованиям «Объем итогового сочинения» и «Самостоятельность написания итогового сочинения». Итоговое сочинение, соответствующее установленным требованиям, оценивается по критериям. Для получения «зачета» за итоговое сочинение необходимо получить «зачет» по критериям № 1 и № 2, а также дополнительно «зачет» по одному из критериев № 3- № 4.

<p>Участник при раскрытии темы сочинения строит рассуждение на основе произведения (произведений) отечественной или мировой литературы по собственному выбору, показывая разный уровень осмысления литературного материала: от элементов смыслового анализа (например, тематика, проблематика, сюжет, характеры и т.п.) до комплексного анализа художественного текста в единстве формы и содержания; выбор литературного произведения и аспекты его анализа оправданны с точки зрения темы;</p> <p>допущено не более 1 фактической ошибки, связанной со знанием литературного материала (ошибка в написании автора и названия произведения, имен персонажей и топонимов произведения, в изложении сюжетной линии, литературных и исторических фактов и т.п.)</p>	2
<p>Участник строит рассуждение с опорой на литературный материал, но ограничивается общими высказываниями по поводу художественного произведения,</p> <p>и/или привлекает литературный материал не вполне уместно с точки зрения выбранной темы,</p> <p>и/или ограничивается простым пересказом художественного произведения,</p> <p>и/или допущено 2–4 фактические ошибки, связанные со знанием литературного материала.</p>	1
<p>сочинение написано без привлечения литературного материала,</p> <p>или литературные произведения лишь упоминаются в работе, не становясь опорой для рассуждения,</p> <p>и/или сочинение содержит 5 и более фактических ошибок, связанных со знанием литературного материала.</p>	0
К3. Композиция	
<p>Сочинение отличается композиционной цельностью, логичностью изложения мыслей и соразмерностью частей, внутри смысловых частей нет нарушений последовательности и необоснованных повторов мысли.</p>	2
<p>Сочинение отличается композиционной цельностью, его части логически связаны между собой, но внутри смысловых частей есть <u>не более 2</u> нарушений последовательности и необоснованные повторы мысли,</p> <p>и/или в сочинении прослеживается композиционный замысел, но есть <u>не более 2</u> нарушений композиционной связи между смысловыми частями,</p> <p>и/или мысль не развивается.</p>	1
<p>Логические нарушения мешают пониманию смысла написанного, или</p>	0

отсутствует тезисно-доказательная часть.	
К4. Качество речи	
Сочинение характеризуется точностью выражения мысли, использованием разнообразной лексики и различных грамматических конструкций, уместным употреблением терминов.	2
Сочинение характеризуется точностью выражения мысли, но прослеживается однообразие грамматического строя речи.	1
Низкое качество речи существенно затрудняет понимание смысла, <u>и/ или сочинение отличается бедностью словаря и однообразием грамматического строя речи.</u>	0
К5. Оригинальность сочинения	
Сочинение характеризуется творческим, нестандартным подходом к раскрытию темы (присутствуют интересные мысли, или неожиданные и вместе с тем убедительные аргументы, или оригинальные наблюдения и проч.) или яркостью стиля.	1
В сочинении не продемонстрирован творческий, нестандартный подход, оригинальность стиля.	0
К6. Речевые нормы	
Речевых ошибок нет, или допущено 1–2 речевые ошибки.	2
Допущено 3–4 речевые ошибки.	1
Допущено 5 и более речевых ошибок.	0
К7. Орфографические нормы	
Орфографических ошибок нет, или допущена 1 ошибка.	3
Допущено 2–3 орфографические ошибки.	2
Допущено 4–5 орфографических ошибок.	1
Допущено более 5 орфографических ошибок.	0
К8. Пунктуационные нормы	
Пунктуационных ошибок нет, или допущена 1 ошибка.	3
Допущено 2–3 пунктуационные ошибки.	2
Допущено 4–5 пунктуационных ошибок.	1
Допущено более 5 пунктуационных ошибок.	0
К9. Грамматические нормы	
Грамматических ошибок нет, или допущена 1 грамматическая ошибка.	2

Допущено 2–3 грамматические ошибки.	1
Допущено 4 и более грамматических ошибок.	0
К10. Фактическая точность в фоновом (не литературном) материале	
Фактические ошибки отсутствуют.	1
Допущены фактические ошибки (1 и более) в фоновом материале.	0
МАКСИМАЛЬНЫЙ БАЛЛ	20

Рекомендации по переводу баллов в 10-балльную шкалу

Отметка по десятибалльной системе	0	1	2	3	4	5	6	7	8	9	10
Первичный балл	0-4	5-6	7-8	9-10	11-12	13-14	15-16	17	18	19	20