

Представление результатов оценки учебных достижений: почему нельзя полностью доверять среднему баллу теста.

Сергей Анатольевич Боченков, эксперт Независимого агентства оценки качества образования «Лидер», г. Чебоксары
s_bochenkov@mail.ru
8 9051998891



Данная статья подготовлена по материалам выступления автора по проблеме представления результатов оценки учебных достижений в рамках вебинара Российского тренингового центра Института управления образованием РАО «Представление результатов оценки учебных достижений: почему нельзя полностью доверять среднему баллу теста», проведённого 14 марта 2013 года. С материалами вебинара можно ознакомиться здесь <http://www.rtc-edu.ru/trainings/webinar/224>. Представля-

ется, что материал окажется полезным не столько тем, кто на основе результатов оценочных процедур собственно производит расчёты, применяя необходимый ассортимент математических инструментов - они и так хорошо понимают, что такое средние величины и как их использовать, а иногда и манипулировать. Этот материал больше адресован тем, кто на основе уже рассчитанных показателей формулирует определённые оценочные суждения, выводы, готовит или принимает управленческие решения, управляет ресурсами, чтобы верно оценивать качество этой информации.

ЕГЭ согласно принятой на международном уровне классификации оценочных процедур относится к государственным экзаменам, причём совмещает в себе выпускной экзамен из средней общеобразовательной школы и вступительный в учреждения профессионального образования. Это экзамен с очень высокими ставками, так как очень значим для выпускников школы и их родителей. Прямое назначение ЕГЭ – сертификация учащихся на освоение образовательного стандарта среднего полного образования и селекция (отбор) для получения профессионального образования, по сути своей результаты ЕГЭ – это оценка уровня индивидуальных образовательных достижений выпускника средней школы. И не более того. Здесь на использовании результатов ЕГЭ нужно было бы поставить точку и закрыть список возможных областей их использования... но существующая практика не позволяет нам сделать этого. Дело в том, что с момента своего введения ЕГЭ позиционируется и как оценочная процедура, которая позволяет получить информацию о состоянии образования в стране, в регионе, в отдельных образовательных учреждениях, судить о тенденциях в изменениях, происходящих в образовании, со-

здавать основу для управления качеством образования на разных уровнях. Эту функцию вообще-то в мировой практике выполняют национальные мониторинги, результаты которых не значимы для самих учащихся, но по этим результатам принимаются решения по отношению к образовательным учреждениям и системам. Но в системе образования Российской Федерации национальные мониторинги ещё только зарождаются в проектах, а вот



Почему результаты ЕГЭ широко используют для оценки образовательных результатов:

- Потому что нет никакой другой информации об образовательных результатах, которая была бы сравнима по надёжности, защищённости, достоверности с результатами ЕГЭ.
- Потому что ресурсы на процедуру ЕГЭ уже потрачены. Почему бы не получить из базы результатов ЕГЭ некоторую информацию, пока нет ресурсов на новую процедуру.
- Потому что ЕГЭ - это итоговая аттестация выпускников, основная процедура оценки результата школы на выходе.

практика использования результатов ЕГЭ в данном контексте имеет уже многолетнюю историю и, к сожалению, уже накоплено достаточно кейсов, когда анализ и интерпретация результатов ЕГЭ делается некорректно, что приводит к ложным заключениям и выводам и к неадекватным управленческим решениям. Поэтому на данном этапе актуальным является обсуждение вопроса о корректной интерпретации результатов единого государственного экзамена как основной формы итоговой аттестации выпускников средней школы в оценке отдельных характеристик качества образования на уровне учителя (учебного предмета), на уровне образовательного учреждения, на уровне образовательных систем.¹

Средний балл ЕГЭ – наиболее распространённый и принятый в системе образования РФ показатель, отражающий образовательные результаты.



Главным показателем результатов ЕГЭ в России является средний балл по 100-балльной шкале. Следует отметить, что вообще средние величины очень удобны, потому что позволяют одной цифрой выразить всё многообразие единичных данных. Поэтому средний балл ЕГЭ – это самый распространённый и принятый в системе образования РФ показатель образовательных результатов.

Наиболее распространённый пример использования среднего балла на слайде, когда школа показывается в сравнении с муниципалитетом, с регионом и с Россией в целом. «Столбики» по каждому учебному предмету позволяют показать, что результаты по школе «лучше всех».

Другой наиболее распространённый пример – когда показывают динамику среднего балла по учебным предметам по годам. При этом обсуждаются не только направления стрелок (вверх или вниз), но и сопоставляя величины столбиков делаются оценочные суждения о том, по каким предметам результаты оказываются лучше, а по каким – хуже.

Оставим пока эти примеры без комментариев, ограничившись лишь текстом одного из законов Мерфи: «Если что-то делается неправильно достаточно часто, оно становится правильным».

¹ Боченков С.А. [Электронный ресурс]: Анализ и интерпретация результатов ЕГЭ-2012//Управление образованием: теория и практика» /ФГНУ ИУО РАО. < <http://www.iuorao.ru/20120930/216-----2012>>. 2012. – № 3. – С. 9–34.

Вопросы, которые предлагается обсудить:

1. Что такое средняя величина и как она отражает совокупность?
2. Надёжность средних величин. Когда использование средних некорректно?
3. Средние величины бывают разные. Какую выбрать?
4. Средние величины и целевые показатели. Может ли средний балл быть целевым показателем?
5. Средние величины и шкалирование. Совпадают ли шкалы оценочных процедур?
6. Способы выравнивания шкал. Как сопоставить результаты разных оценочных процедур?
7. Условия, когда можно доверять среднему баллу теста.

Перечень вопросов, которые имеет смысл обсудить для того, чтобы корректно использовать результаты ЕГЭ в целом и средний балл в частности представлены на слайде.

Средней величиной называют показатель, который характеризует **обобщенное значение признака** или группы признаков в исследуемой совокупности.

Средняя величина **заменяет большое число индивидуальных значений** признака, обнаруживая **общие свойства**, присущие всем единицам совокупности.

Среднее - такое значение признака, которое имело бы каждая единица совокупности, если бы общий итог всех значений признака был распределен равномерно между всеми единицами совокупности.

Средняя величина — это **обобщающая характеристика** размера изучаемого признака. Она **позволяет одним числом количественно охарактеризовать** качественно **однородную совокупность**.

Что такое «средняя величина» и как она отражает совокупность?

На слайде представлены четыре наиболее часто используемых определения средней величины. При этом очень важно, чтобы средняя величина максимально адекватно отражала свойства всей совокупности отдельных измерений (в нашем случае – индивидуальных результатов ЕГЭ). Известны два условия, при которых можно доверять средней. Во-первых, *средняя должна быть типической*. Средняя, рассчитываемая для совокупностей, состоящих из качественно однородных единиц, называется типической средней. Получению типической средней должен предшествовать анализ того, насколько данная совокупность качественно однородна. Если совокупность состоит из отдельных частей, следует разбить ее на типические группы. Во-вторых, средняя должна вычисляться для совокупностей,

состоящих из достаточно большого числа единиц. Соблюдение этого условия необходимо для того, чтобы вступил в силу закон больших чисел, в

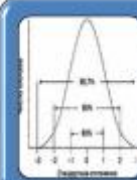
Что такое средняя величина и как она отражает совокупность?

Средней величиной называют показатель, который характеризует **обобщенное значение признака** или группы признаков в исследуемой совокупности.

Средняя величина **заменяет большое число индивидуальных значений** признака, обнаруживая **общие свойства**, присущие всем единицам совокупности.

Среднее - такое значение признака, которое имело бы каждая единица совокупности, если бы общий итог всех значений признака был распределен равномерно между всеми единицами совокупности.

Средняя величина — это **обобщающая характеристика** размера изучаемого признака. Она **позволяет одним числом количественно охарактеризовать** качественно **однородную совокупность**.



Средняя величина должна быть типична, а измеряемая ею совокупность - однородна

- Совокупность должна быть однородной, не распадающейся на несколько самостоятельных совокупностей
- Средняя вычисляется для признаков, присущих всем членам совокупности, для признаков, качественно однородных и различающихся только количественно



Объективность средней определяется достаточно большим числом единичных измерений, составляющих совокупность

- Рассчитываются на основе массовых данных – достаточного числа анализируемых единиц
- Достаточность анализируемых единиц обеспечивается корректным определением границ исследуемой совокупности.

результате действия которого случайные отклонения индивидуальных величин от общей тенденции взаимно погашаются. Таким образом, можно считать, что *увеличение числа наблюдений практически всегда ведет к росту надежности средней величины. Это один из способов борьбы с пагубным влиянием случайных факторов на анализируемые статистические показатели.*

Использование средней величины допускается только после оценки её типичности и надёжности. Как оценить типичность и надёжность средней?

♦ Средняя величина должна быть типичной, т. е. должна отражать основную совокупность, из которой она получена.

♦ Типичность средней величины обратно-пропорциональна степени колеблемости (рассеянности) измеряемой величины.

♦ Чем более рассян ряд, тем менее типична средняя.

♦ Чем больше число измерений составляет совокупность, тем надёжнее средняя.

Среднее квадратическое (или стандартное) отклонение:

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$$

• Среднее квадратическое отклонение является мерой надёжности средней. Чем меньше эта величина, тем лучше средняя арифметическая отражает собой всю представляемую совокупность.

Коэффициент вариации:

$$V = \frac{\sigma}{\bar{x}} \cdot 100\%$$

• Коэффициент вариации – это отношение среднеквадратического отклонения к среднеарифметическому, рассчитывается в процентах.

• Критерийным значением коэффициента вариации служит 33,3%, то есть если V меньше или равен 33,3% – вариация считается слабой, а если больше – сильной. В случае сильной вариации изучаемая статистическая совокупность считается неоднородной, а средняя величина – нетипичной и её нельзя использовать как обобщающий показатель этой совокупности.

В левой части слайда приведены признаки, когда средняя типична и надёжна и может отражать свойства всей оцениваемой совокупности.

В правой части даны две формулы, при помощи которых необходимо оценить качество средней. Очень удобно использовать коэффициент вариации, который показывает на сколько процентов отклоняется от среднего каждое из измерений в среднем. Если это отклонение

больше трети шкалы, то совокупность неоднородна и её нужно разбить...



Однородность совокупности рассмотрим на следующем примере.

На слайде представлены две совокупности, имеющие одинаковые средние величины – по 50 баллов. Но, рассчитав стандартное отклонение и коэффициент вариации мы наглядно видим, что первая совокупность неоднородна и распадается на две самостоятельные группы...

Использование средней величины допускается только после оценки её типичности

Предмет	Доля сдававших	Средний балл	Стандартное отклонение	Коэффициент вариации
Регион А				
Русский язык	99,9%	61,32	13,09	21,34%
Математика	99,9%	43,94	13,15	29,93%
Физика	26,7%	43,53	11,67	26,82%
Химия	7,1%	54,65	14,82	27,12%
Биология	12,1%	51,18	12,98	25,35%
Иностранный язык	5,8%	57,33	18,14	31,65%
Обществознание	52,9%	54,43	11,84	21,76%
Информатика и ИКТ	8,8%	62,41	15,48	24,80%
Регион Б				
Русский язык	99,9%	46,01	14,54	31,59%
Математика	99,9%	37,69	14,74	39,11%
Обществознание	75,0%	42,88	14,29	33,33%

Кстати, ограничением для дробления неоднородных совокупностей является число измерений. Делая среднюю более типичной мы проигрываем в её надёжности. Необходимо помнить сразу о двух характеристиках – надёжность и типичность.

Итак, первая фиксация: *Прежде чем принимать значения средних величин за характеристику качества некой совокупности нужно убедиться, что средние проверены на типичность и за ними стоит достаточное число единичных измерений.*

Что делать, если средняя – нетипична? Разделять совокупность на несколько частей и пользоваться значениями средних каждой из них.

В практике анализа результатов ЕГЭ даже при использовании результатов по региону (несколько сотен или даже тысяч индивидуальных результатов!) часто оказывается, что средний балл по одному или нескольким

14 марта 2013 года

предметам не является типической средней величиной, т.е. совокупность индивидуальных результатов, лежащих в основе расчёта среднего балла неоднородна. Часто это связано с тем, что смешиваются результаты двух крайне различающихся групп выпускников: выпускников средних (полных) школ и выпускников вечерних (сменных) школ. Уровень подготовки этих двух групп сильно различен, поэтому результаты ЕГЭ этих групп имеет смысл считать отдельно.

Средние величины бывают разные. Простая и взвешенная форма средних.

Пример. Расчёт среднего балла по муниципалитету:

Наименование школ:	ЕГЭ-2012, средний балл по муниципалитету	Кол-во выпускников
Школа №1	67,3	33
Школа №2	54,5	58
Школа №3	48,6	15
Школа №4	45,4	9
Школа №5	56,9	11
Школа №6	56,0	3
Школа №7	33,5	4
Школа №8	60,4	13
Школа №9	54,2	38

Простая форма средней применяется, когда средняя вычисляется по первичным данным.

Взвешенная форма – при расчете средней по вторичным (сгруппированным) данным!

простое среднее арифметическое: 52,98

взвешенное среднее арифметическое: 55,94

Способы расчёта средних величин.

Для расчёта средних применяют разные способы. Рассмотрим пример со средней арифметической. Простая форма средней арифметической применяется, когда средняя вычисляется по первичным данным, а взвешенная форма – при расчете средней по вторичным (сгруппированным) данным! Если не различать эти способы расчёта, то есть опасность некорректных сравнений. Допустим, необходимо рассчитать средний балл по муниципалитету.

Средние величины бывают разные. Средние степенные и структурные.

Средние величины. Пример манипуляций.

11 «А» класс		11 «Б» класс	
Ученик 1	3	Ученик 1	2
Ученик 2	3	Ученик 2	2
Ученик 3	3	Ученик 3	2
Ученик 4	3	Ученик 4	2
Ученик 5	3	Ученик 5	2
Ученик 6	4	Ученик 6	5
Ученик 7	4	Ученик 7	5
Ученик 8	4	Ученик 8	5
Ученик 9	4	Ученик 9	5
Ученик 10	4	Ученик 10	5

средняя арифметическая = 3,5

средняя гармоническая = 3,4

структурные средние:

- Мода – это наиболее часто встречающееся значение
- Медиана – это значение, которое делит совокупность на две равные части

Мо Мо 1

Очень часто вместо того, чтобы использовать индивидуальные результаты всех выпускников школ муниципалитета ошибочно берётся средняя из среднего балла всех школ. Ввиду того, что в школах разное число выпускников получается достаточно значимая ошибка.

Кроме средней арифметической, которую используют чаще всего есть и другие средние. Так на слайде показано, что применение средней гармонической позволяет оценить выше более однородные совокупности.

родные совокупности.

Средние величины как целевые показатели. Очень часто увлекаясь математикой, мы забываем для чего все наши расчёты. Ведь одна задача – показать лучше или хуже оказался результат и совсем другая – а каким он

Формулировка требований к результату

Для каждого объекта оценки в зависимости от задач формулируется рамка требований к результату.

Преобразование требований в целевые показатели

Система требований к результату прописывается в виде показателей с закреплением источника данных, формул для расчётов, периодичности и т.д.

Распределение зон ответственности в зависимости от условий и ресурсов

Показатели должны отражать состояние того участка работы и в таком ракурсе, в котором работает лицо, работу которого оценивают, т.к. ориентируясь на эти показатели он не только выстраивает свою деятельность, но и использует для их «улучшения» все имеющиеся в его распоряжении ресурсы.

14 марта 2013 года

должен быть? Поставить чёткую цель в виде конкретной цифры. Это будет целевой показатель – это необходимый уровень достижений или ожидаемая норма, установленные для конкретного показателя. И тут возникает вопрос – можно ли пресловутый средний балл ЕГЭ использовать в качестве целевого показателя?

Из логической схемы видно, что *хороший (тот, с которым можно работать) целевой показатель тот, из которого сразу становится понятно, куда надо приложить усилие, чтобы изменить ситуацию в желаемом направлении. А вот средний балл трудно использовать в качестве целевого показателя, т.к. неясна точка приложения необходимых ресурсов для достижения.*

Что значит – «хорошая школа» по результатам ЕГЭ?
Признаки хорошей по результатам ЕГЭ школы:

- это школа, выпускники которой активно выбирают экзамены в форме ЕГЭ из числа предметов профиля школы (которые необходимы им для продолжения образования и которые они изучали в 10-11 классах на профильном уровне);
- это школа, все выпускники которой успешно освоили образовательный стандарт не ниже минимального уровня по двум обязательным экзаменам в форме ЕГЭ (необходимо для получения аттестата о среднем (полном) образовании) и экзаменам по выбору независимо от профиля, т.е. получили полноценное свидетельство о результатах ЕГЭ для конкурса при поступлении на выбранную специальность;
- это школа, в которой значительная часть выпускников преодолевают порог профильного уровня по всем предметам, которые они изучали на профильном уровне;
- это школа, где результаты текущего оценивания коррелируют с результатами внешней оценки;
- это школа, показывающая положительную динамику или стабильность в высоких результатах по вышеназванным показателям среди школ города (района) и среди аналогичных (кластер) школ региона.

Представление результатов оценочной процедуры должно максимально полно раскрывать картину в этой системе координат

Сравнение результатов двух оценочных процедур:

Средний балл по математике по РФ = 44,6

Средний балл по русскому языку = 61,1

Какие выводы можно сделать об уровне подготовки, сравнивая результаты по двум предметам?

- Если тесты по двум предметам сопоставимы по сложности и шкалы выровнены, то уровень подготовки по русскому всех выпускников страны в среднем выше, чем по математике.
- Если тесты по двум предметам не выровнены по сложности, то возможно несколько вариантов:
 - Уровень подготовки по русскому выше, чем по математике.
 - Уровень подготовки по математике выше, чем по русскому языку: просто тест сложнее.
 - Уровень подготовки по двум предметам примерно равный.

Для того, чтобы оценить качество того или иного объекта, необходимо представить рамку требований к качеству. На слайде представлен набор характеристик хорошей по результатам ЕГЭ школы. Напомним, что ЕГЭ – это лишь одна из многих оценочных процедур а результаты ЕГЭ по школе – это небольшая часть образовательных и других результатов по которым оценивается качество школы.

Средние величины и шкалирование. Совпадают ли шкалы оценочных процедур?

При анализе результатов ЕГЭ невозможно обойти вопрос межпредметного сравнения результатов. По результатам ЕГЭ-2012 средний балл ЕГЭ по математике по Российской Федерации равен 44,6 баллов, а по русскому языку – 61,1 балла. Какие выводы можно сделать об уровне подготовки, сравнивая результаты по

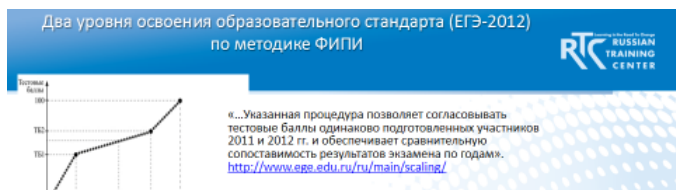
двум предметам? Если тесты по двум предметам сопоставимы по сложности и шкалы выровнены, то уровень подготовки по русскому всех выпускников страны в среднем выше, чем по математике. А вот если тесты по двум предметам не выровнены по сложности, то возможно несколько вариантов: либо уровень подготовки по русскому выше, чем по математике, либо уровень подготовки по математике выше, чем по русскому языку, просто тест сложнее, либо уровень подготовки по двум предметам примерно равный...

Когда мы пытаемся сравнивать результаты двух оценочных процедур, мы должны понимать, что они измеряют сопоставимые характеристики и результаты их представлены в сопоставимых величинах... Ведь если длина помещения измерена в сантиметрах, а высота – в дюймах, то мы же не будем сравнивать значения этих измерений без учёта различия в единицах измерения?

Для корректной интерпретации результатов ЕГЭ необходимо понимание некоторых особенностей этой

оценочной процедуры.

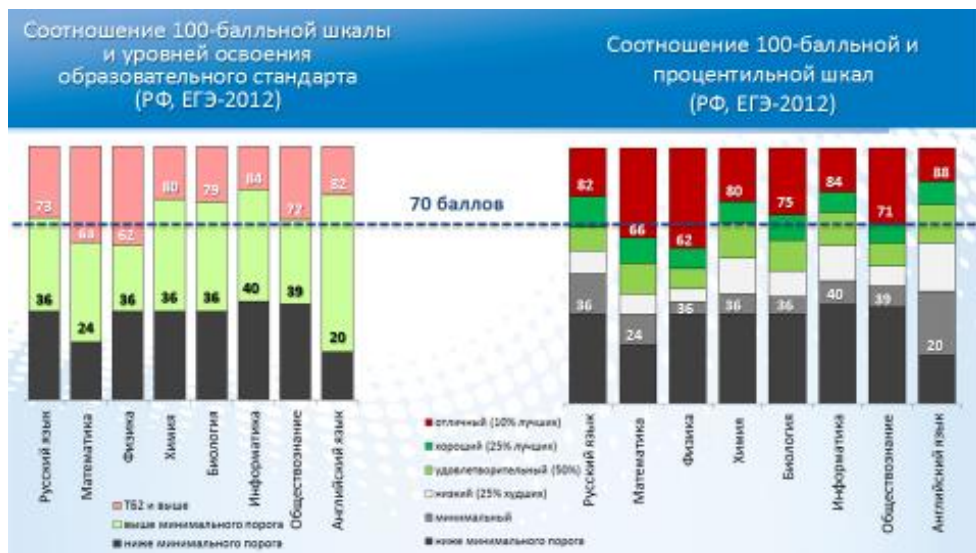
Необходимо понимать, что выполняя задания контрольно-измерительных материалов ЕГЭ (далее – КИМы) выпускник набирает определённое количество первичных баллов. Например, успешное выполнение задания базового уровня из части А (выбор одного правильного ответа из четырёх) как правило оценивается одним первичным баллом, задания части В (повышенный уровень) и С (высокий уровень сложности) как правило позволяют получить несколько первичных баллов. Набранная выпускником сумма первичных баллов после процедуры шкалирования² переводится в балл по 100-балльной шкале и выставляется в свидетельство о результатах ЕГЭ. Этот результат чаще всего и берётся за основу в существующих сегодня моделях анализа и интерпретации результатов ЕГЭ, а *между тем несёт в себе заметную ошибку*.



Предмет	ПБ1 (в скобках ТБ1 – балл по 100-балльной шкале)	ПБ2 (в скобках ТБ2 – балл по 100-балльной шкале)
Русский язык	17 из 64 (36 из 100)	54 из 64 (73 из 100)
Математика	5 из 32 (24 из 100)	15 из 32 (63 из 100)
Обществознание	15 из 59 (39 из 100)	48 из 59 (72 из 100)
История	13 из 58 (32 из 100)	46 из 58 (72 из 100)
Физика	12 (11 – min) из 51 (39 из 100)	33 из 51 (62 из 100)
Химия	14 из 65 (36 из 100)	58 из 65 (80 из 100)
Биология	17 из 69 (36 из 100)	60 из 69 (79 из 100)
География	14 из 54 (37 из 100)	44 из 54 (69 из 100)
Информатика	8 из 40 (40 из 100)	35 из 40 (84 из 100)
Иностранные языки	16 из 80 (20 из 100)	65 из 80 (82 из 100)
Литература	8 из 42 (32 из 100)	36 из 42 (73 из 100)

В таблице приведены пороговые значения первичных баллов и баллов по 100-балльной шкале из методики шкалирования. Величины ПБ1 и ПБ2 устанавливаются разработчиками, а минимальный порог – распоряжение Рособрнадзора (он как правило совпадает с ПБ1).

Отметим, что величина ПБ1 (наименьший первичный балл, получение которого свидетельствует об усвоении участником экзамена основных понятий и методов по соответствующему общеобразовательному предмету) в первичных баллах между предметами не совпадает (и это нормально, т.к. в каждом учебном предмете можно набрать разное максимальное число первичных баллов). Но не совпадает и значения ТБ1 по 100-балльной шкале! А это значит, что чтобы преодолеть минимальный порог



по математике достаточно 24 баллов, а по информатике – 40 баллов из 100 возможных. Аналогичная ситуация наблюдается и при сравнении величин ТБ2. Различия в величинах ТБ1 и ТБ2 по предметам позволяет сделать вывод, что «линейки» в каждом предмете свои.

Использование подхода, основанного на сравнении процентильных групп участников по результатам ЕГЭ-2012 по Российской Федерации и соответствующие им тестовые баллы по 100-балльной шкале по учебным предметам также позволяют выявить несовпадение шкал. Например, чтобы попасть в группу 10% лучших результатов

² Подробнее о методике шкалирования можно посмотреть здесь http://www.ege.edu.ru/ru/main/legal-documents/index.php?id_4=17904

14 марта 2013 года

ЕГЭ в 2012 году по РФ по математике нужно было набрать 66 баллов, по русскому – 82, по английскому языку – 88. А это значит, что 70 баллов по математике не равны 70 баллов по русскому языку и, вероятно, 70 баллов по математике ЕГЭ-2011 не равны 70 баллам по математике ЕГЭ-2012. Таким образом, **при интерпретации данных важно понимать, что 100-балльные шкалы по разным учебным предметам по сути своей не совпадают и баллы (максимальные, минимальные, средние и т.д.) нельзя сравнивать как между предметами, так и в пределах одного предмета в динамике по годам.**



Итак, когда мы видим перед собой значения средних баллов по нескольким предметам по школе, району, региону или стране, то прежде чем делать выводы о высоких или низких результатах данные необходимо привести к сопоставимому виду. В противном случае мы будем сравнивать длину, измеренную в сантиметрах с длиной в дюймах и удивляться, почему в дюймах величина всегда меньше, чем в сантиметрах.

Стобалльная шкала ЕГЭ не позволяет сравнивать результаты по предметам и оценить динамику по годам, т.к.:

НЕ СОВПАДАЕТ:

- величина наименьшего тестового балла, получение которого свидетельствует об усвоении участником экзамена основных понятий и методов по предмету (ТБ1)
- величина наименьшего тестового балла, получение которого свидетельствует о высоком уровне подготовки участника экзамена (а именно, о наличии системных знаний, овладении комплексными умениями, способности выполнять творческие задания по соответствующему общеобразовательному предмету) ТБ2

МЕНЯЕТСЯ ПО ГОДАМ:

- время, отведённое на выполнение работы без изменения её содержания (русский язык в 2012 и в 2013)
- количество, качество и состав заданий, уровень сложности

ЕДИНЬЙ государственный экзамен – это ряд самостоятельных предметных экзаменов, которые не согласованы между собой ни по фреймам тестов, ни по форме представления результата

В этой ситуации при интерпретации результатов ЕГЭ имеет смысл использовать либо относительные показатели 100-балльной шкалы (или пересчитанные, приведённые её значения), либо показатели доли выпускников успешно справившихся (преодолевших минимальный порог) с минимальными требованиями стандарта по предмету и доли успешно преодолевших порог ТБ2 – по-

казавших высокий уровень подготовки и наличие системных знаний, овладение комплексными умениями, способности выполнять творческие задания по соответствующему общеобразовательному предмету.

Таким образом, обобщая вышесказанное, фиксируем, что **стобалльная шкала ЕГЭ не позволяет сравнивать результаты по предметам и оценивать динамику по годам, а единый государственный экзамен представляет собой ряд самостоятельных предметных экзаменов, которые не согласованы между собой ни фреймом контрольно-измерительных материалов, ни по формату представления результатов.**

Рассмотрим выявленные различия в шкалировании на примере. Допустим, поставлена задача сравнения результатов двух школ по качеству подготовки их выпускников по профильным предметам. Одна школа даёт инженерный профиль с углубленной подготовкой по физике и математике с достаточно хорошим русским язы-

ком, результаты которого также востребованы для поступления. Вторая школа ведёт подготовку по медицинскому профилю с углубленной подготовкой по биологии и химии и чтобы русский тоже был на хорошем уровне. В каждой школе одинаковое число выпускников – по 30 человек.

Профиль школы	Кол-во выпускников	Средний балл ЕГЭ, при условии, что все выпускники все предметы сдали с одинаковым результатом, равным ТБ2	Вывод о результатах профильной подготовки	Вывод об эффективности вузов, по качеству подготовки поступивших в них выпускников этих школ
Школа с инженерным профилем (математика, физика + русский язык)	30	(математика 63 + физика 62 + русский 73) / 3 = 66,0	При одинаковом качестве подготовки (равном ТБ2 -балл, получение которого свидетельствует о высоком уровне подготовки участника экзамена, а именно, о наличии системных знаний, овладении комплексными умениями, способности выполнять творческие задания) средний балл инженерного класса оказывается НИЖЕ показателя медицинского класса на 11,3 балла или на 17%	Все выпускники поступили в соответствующие профильные вузы на инженерное и медицинское направление подготовки. Таким образом, медицинский вуз оказывается на 17% лучше инженерного по показателю среднего балла, при этом качество подготовки абитуриентов не отличается, а разницу определяет невыровненная столбчатая шкала
Школа с медицинским профильным (биология, химия + русский язык)	30	(русский 73 + биология 79 + химия 80) / 3 = 77,3		

Представим ситуацию, что подготовка в этих школах ведётся настолько хорошо, что каждый выпускник по каждому предмету сдаёт ЕГЭ с результатом, равным ТБ2 (наименьший тестовый балл, получение которого свидетельствует о высоком уровне подготовки участника экзамена, а именно, о наличии системных знаний, овладении комплексными умениями, способности выполнять творческие задания по соответствующему общеобразовательному предмету - этот порог выведен специалистами ФИПИ). А эти баллы по названным предметам следующие: русский язык – 73, математика – 63, биология – 79, химия – 80, физика – 62. Для наших расчётов упростим ситуацию, указав, что все выпускники сдали профильные предметы с одинаковым результатом, равным по каждому предмету значению ТБ2.

Посчитаем средний балл профильных предметов школы с инженерной подготовкой: в нашем случае, все 30 выпускников сдали ЕГЭ одинаково по всем предметам, поэтому средний балл оказался равен (математика 63+ физика 62+русский 73) / 3=66 баллов. Аналогично средний балл медицинского класса окажется равным (русский (73) + биология (79) + химия (80)) / 3=77,3

Таким образом, сравнивая средние баллы «в лоб» мы получаем, что при одинаковом качестве подготовки (равном ТБ2) средний балл инженерного класса оказывается ниже показателя медицинского класса на 11,3 балла или на 17%!

А теперь эту же ситуацию перевернём для оценки вузов – представим, что эти выпускники поступили в соответствующие профильные вузы на инженерное и медицинское направление подготовки. Таким образом, медицинский вуз оказывается на 17% лучше инженерного по показателю среднего балла, при этом качество подготовки абитуриентов не отличается, а разницу определяет «кривая» столбчатая шкала....

Условия, когда можно доверять среднему баллу теста:

Оценка результата в пределах одного теста (например, ЕГЭ-2012 по математике):	Сопоставление результатов нескольких тестов или динамика результатов:
★ Если совокупность данных, лежащая в основе расчётов средней однородна и средняя величина типична. ★ Если измерений в совокупности данных достаточно много, чтобы работал закон больших чисел. ★ Если ясна природа воздействия на средний балл контролируемых и случайных факторов. ★ Если способ расчёта средней выбран так, чтобы полученная оценка совокупности максимально отражала реальную ситуацию.	★ Если быть уверенным, что тест ЕГЭ по русскому и тест ЕГЭ по математике (и т.д. – по другим предметам) измеряют сопоставимые величины и результаты представлены по выровненным шкалам. ★ Если быть уверенным, что результаты ЕГЭ по математике (и т.д. – по другим предметам) по годам измеряют сопоставимые величины и результаты представлены по выровненным шкалам.

Подводя итоги вышесказанному зафиксируем общие условия, при которых имеет смысл доверять среднему баллу теста вообще и баллу ЕГЭ в частности. Оказывается, не так много случаев, когда среднему баллу ЕГЭ можно доверять и использовать в оценке.

И это абсолютно противоречит имеющейся практике, когда средние баллы ЕГЭ используют везде, где только можно себе вообразить, когда на их сравнении формируются оценочные суждения и принимаются управленческие решения. Иногда они касаются смены учителя, администрации школы, реорганизации сети школ и т.п., а в основе своей имеют некорректное сравнение несопоставимых величин.

Вместо среднего балла ЕГЭ при анализе результатов этой оценочной процедуры имеет смысл использовать иные индикаторы. Кратко они представлены на слайде, подробнее о системе показателей ЕГЭ можно узнать из публикаций автора здесь <http://www.rtc-edu.ru/sites/default/files/files/public/Bochenkov.docx> и здесь: Боченков С.А. [Электронный ресурс]: Анализ и интерпретация результатов ЕГЭ-2012//Управление образованием: теория и практика» /ФГНУ ИУО РАО. < <http://www.iuorao.ru/20120930/216-----2012>>. 2012. – № 3. – С. 9–34.

Если не средний балл, то что? Какие показатели имеет смысл использовать при анализе результатов ЕГЭ?

Для оценки результатов учителя:	Для оценки результатов школы:
Уровень освоения образовательного стандарта: ★ Доля выпускников, успешно сдавших экзамен в форме ЕГЭ по предмету из числа выпускников, допущенных к итоговой аттестации. Качество профильной подготовки: ★ Доля выпускников, выбравших экзамен по предмету в форме ЕГЭ от общего числа изучавших данный предмет на профильном уровне. ★ Доля выпускников сдавших экзамен по предмету с результатом равным Т52 (высокий уровень) и выше в общем числе изучавших данный предмет на профильном уровне. ★ Корреляция между результатами ЕГЭ и текущей отметкой учащихся.	Уровень освоения образовательного стандарта: ★ Доля выпускников школы успешно сдавших два обязательных экзамена в форме ЕГЭ из числа выпускников, допущенных к итоговой аттестации. ★ Доля выпускников, успешно сдавших все экзамены (обязательные и по выбору) в форме ЕГЭ. Качество профильной подготовки: ★ Доля выпускников, выбравших все экзамены (учитываются только экзамены по выбору) для сдачи в форме ЕГЭ из числа предметов изучавшихся на профильном уровне. ★ Доля выпускников, сдавших все предметы из изучавшихся на профильном уровне не ниже порога профильной подготовки (Т52).

Важно, что задача определения роли и места ЕГЭ в оценке работы учителей и оценке эффективности деятельности образовательных учреждений и образовательных систем остаётся на сегодня одной из самых актуальных задач в области оценки качества образования.