

Можно ли сопоставить результаты ЕГЭ и ГИА

Сравнение показателей, рассчитанных на основе разных тестовых испытаний

М. Л. Агранович

Статья поступила
в редакцию
в июле 2013 г.

Агранович Марк Львович

кандидат экономических наук, руководитель Центра мониторинга и статистики образования ФИРО, исполнительный директор Межрегиональной ассоциации мониторинга и статистики образования. Адрес: Москва, 129319, ул. Черняховского, д. 9. E-mail: magran@firo.ru

результатов тестирования учащихся, когда показатели теста имеют разные параметры распределения или даже разные шкалы. Его использование показано на примерах сравнения результатов ЕГЭ по разным предметам и за разные годы, а также результатов ЕГЭ и ГИА.

Ключевые слова: школы, тестовые испытания, ЕГЭ, ГИА, распределение результатов, коэффициент детерминации, рейтинги.

Аннотация. Предлагается способ приведения к сопоставимому виду

Результаты Единого государственного экзамена, являясь показателем уровня освоения школьной программы, остаются сегодня единственным источником данных для анализа и оценки многих характеристик образовательной системы, связанных с качеством образования. По результатам ЕГЭ оценивают всё — от учителей до губернаторов, несмотря на неоднократные призывы и экспертов, и официальных лиц этого не делать. Результаты ЕГЭ абитуриентов стали даже одним из показателей эффективности вузов.

На практике часто сопоставляют результаты Единого государственного экзамена (ЕГЭ) за разные годы, рассчитывают средний балл ЕГЭ по разным предметам, ранжируют вузы разного профиля по уровню ЕГЭ абитуриентов. Делаются многочисленные попытки сравнить результаты ЕГЭ и Государственной итоговой аттестации (ГИА). То есть при решении многих задач, требующих оценки и анализа данных об образовательных достижениях учащихся, возникает необходимость сравнить показатели, полученные с помощью разных оценочных процедур, при том что результаты тестирования (баллы) имеют разное распределение, а то и разные шкалы.

Во многих работах (см., например [Ковалева, 2005]) подчеркивается недопустимость сравнения результатов Единого го-

Рис. 1а. **Динамика среднего балла ЕГЭ по школам муниципалитета X**

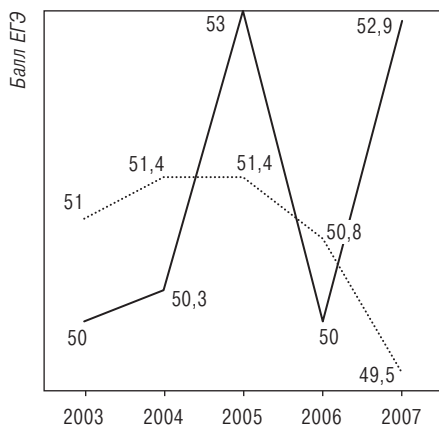
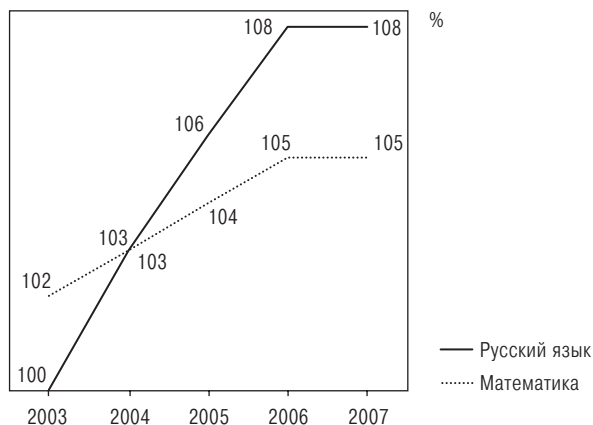


Рис. 1б. **Динамика отношения среднего балла ЕГЭ по школам муниципалитета X к среднему значению по региону, %**



сударственного экзамена за несколько лет хотя бы потому, что невозможно гарантировать одинаковую сложность текстов. Поэтому, в частности, результаты имеют разное распределение и их непосредственное сравнение некорректно.

До настоящего времени единственный более или менее корректный способ сравнения результатов тестирования за разные годы заключался в соотнесении этих результатов со средними значениями по некоторой совокупности, например в сопоставлении результатов школы со средним баллом по муниципалитету, региону или России [Агранович и др., 2006. С. 23]. Использование даже этого простого хода позволяет судить, насколько некорректно непосредственное сравнение результатов ЕГЭ за разные годы (рис. 1а, б.).

Если судить по рис. 1а, то приходится констатировать, что результаты учащихся по русскому языку год от года снижаются, а по математике очень неустойчивы, подвержены резким изменениям. На самом деле (рис. 1б) результаты учеников по обоим предметам плавно росли, но в последний год этот рост остановился.

Из-за разных характеристик распределения результатов нельзя сравнивать даже средние баллы ЕГЭ по разным предметам. Кривые распределения баллов ЕГЭ по разным предметам имеют разные параметры, это означает, что один и тот же балл ЕГЭ фиксирует разные уровни подготовки, о чем очень убедительно писал С. Боченков [2012]. Не вдаваясь в проблемы шкалирования, приведем официально установленные минимумы количества баллов, подтверждающие освоение программ

Таблица 1. Минимум количества баллов, подтверждающий освоение программ по общеобразовательным предметам

Предмет	Минимальное количество баллов
Русский язык	36
Математика	24
Физика	36
Химия	36
Информатика и ИКТ	40
Биология	36
История	32
География	37
Обществознание	39
Литература	32
Иностранные языки (английский, немецкий, французский, испанский)	20

Таблица 2. Значения граничных первичных и тестовых баллов в 2012 г.

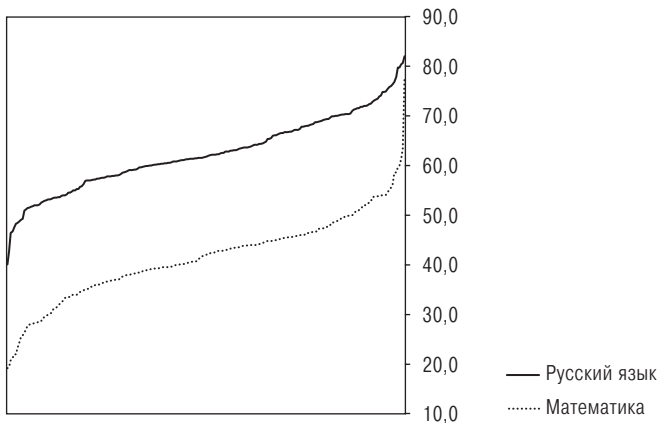
Предмет	ПБ1	ТБ1	ПБ2	ТБ2
Русский язык	17	36	54	73
Математика	5	24	15	63
Обществознание	15	39	48	72
История	13	32	46	72
Физика	12	39	33	62
Химия	14	36	58	80
Биология	17	36	60	79
География	14	37	44	69
Информатика	8	40	35	84
Иностранные языки	16	20	65	82
Литература	8	32	36	73

Источник: Таблица из Методики шкалирования результатов ЕГЭ, размещенной на официальном информационном портале Единого государственного экзамена <http://www.ege.edu.ru/ru/main/scaling/>

Примечание. ПБ1 — наименьший первичный балл, получение которого свидетельствует об усвоении участником экзамена основных понятий и методов по соответствующему общеобразовательному предмету, ТБ1 — балл по 100-балльной шкале.

ПБ2 — наименьший первичный 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, ТБ2 — балл по 100-балльной шкале.

Рис. 2. Сравнение распределения результатов ЕГЭ по всем школам региона X, 2010 г.



по общеобразовательным предметам [Об установлении минимального количества баллов...] (табл. 1).

Не только «граница отсечения», но и баллы, соответствующие высокому уровню освоения разных предметов, варьируют значительно (табл. 2).

Таким образом, часто используемый при оценке и рейтинговании школ показатель «средний балл ЕГЭ по русскому языку и математике» неадекватно отражает уровень достижений учащихся, поскольку один и тот же балл по этим предметам соответствует разным уровням знаний. Более того, средний балл ЕГЭ абитуриентов используется и для оценки вузов [Перечень показателей оценки...], что ставит технические и гуманитарные вузы в заведомо неравные условия. В частности, из табл. 1 и 2 видно, что в одном случае ученик, набравший в среднем по математике и русскому языку 35 баллов, — троечник (25 + 45), а в другом случае не получит аттестат из-за «неуда» по русскому языку (35 + 35).

Различия в результатах сдачи ЕГЭ по русскому языку и математике можно наглядно проиллюстрировать следующим графиком (рис. 2), на котором представлены упорядоченные по возрастанию результаты ЕГЭ по более чем 230 школам одного из регионов (каждая точка на графике — средний балл ЕГЭ одной из школ).

Еще более сложная задача — сравнение результатов Государственной итоговой аттестации и ЕГЭ. Здесь не только разные характеристики распределения, но и разные шкалы и разные процедуры, что влияет на достоверность результатов.

Возникает естественный вопрос: можно ли привести к сопоставимому виду результаты тестирования учащихся, когда показатели теста имеют разные параметры распределения или даже разные шкалы?

Приведение результатов ЕГЭ по математике и русскому языку к сопоставимому виду

Начнем с ЕГЭ. Я столкнулся с этой задачей при подготовке рейтинга школ повышенного уровня (лицеев, гимназий и школ с углубленным изучением отдельных предметов) для РИА Новости [Агранович, 2011]¹, где в качестве одного из показателей фигурировал как раз средний балл ЕГЭ по математике и русскому языку. В рейтинге участвовали более 900 «продвинутых» школ из 24 регионов. Распределение средних баллов по русскому и математике по этим школам представлено на рис. 3: каждая точка на графике соответствует значению среднего балла ЕГЭ по школам в порядке возрастания.

Поскольку один и тот же балл по русскому и математике означает разный уровень знаний, физико-математические и гуманитарные школы оказываются в заведомо неравных условиях: из рис. 3 очевидно, что получить определенный балл ЕГЭ по математике сложнее, чем по русскому языку. Поэтому встала задача привести данные о результатах ЕГЭ по этим предметам к сопоставимому виду.

Дальнейшие рассуждения и расчеты исходят из предположения, что уровень знаний по двум предметам в рамках данного массива примерно одинаков и различия в оценках ЕГЭ и параметрах распределения этих оценок вызваны различиями используемого инструмента. Для того чтобы сравнить распределения, построим график, по одной оси которого будут результаты ЕГЭ по русскому языку, а по другой — по математике. Значения результатов ЕГЭ относятся к одной и той же совокупности школ и упорядочены по возрастанию. То есть точки на графике показывают, как соотносятся результаты ЕГЭ по русскому языку и математике с одинаковыми порядковыми номерами по возрастанию, и не являются парой «баллы ЕГЭ русский — математика» по конкретной школе. График представлен на рис. 4. Совокупность — 940 школ повышенного уровня из 24 регионов.

Мы видим, что распределение результатов ЕГЭ по двум предметам связано почти функционально (экспоненциальной) зависимостью с уровнем детерминации 0,98.

Подставляя в формулу

$$Y = 11,273 e^{0,0221x}$$

баллы ЕГЭ по русскому языку, мы можем привести их к сопоставимому виду для сравнения с результатами ЕГЭ по математике и корректного рейтингования школ по среднему баллу по обязательным предметам. Таким образом, с помощью формулы, полученной из анализа связей между двумя распределениями

¹ Последний рейтинг опубликован на сайте РИА Новости http://ria.ru/sn_edu/20130423/934054497.html

Рис. 3. Распределение баллов ЕГЭ по русскому языку и математике по школам повышенного уровня, 2012 г.

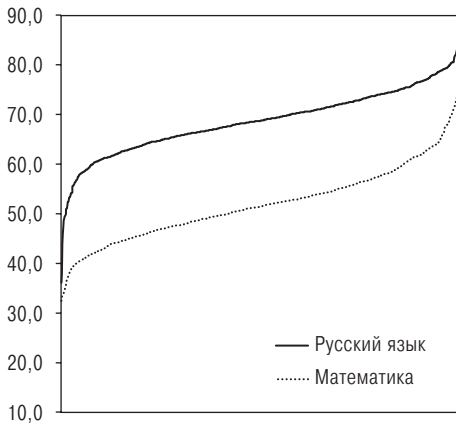
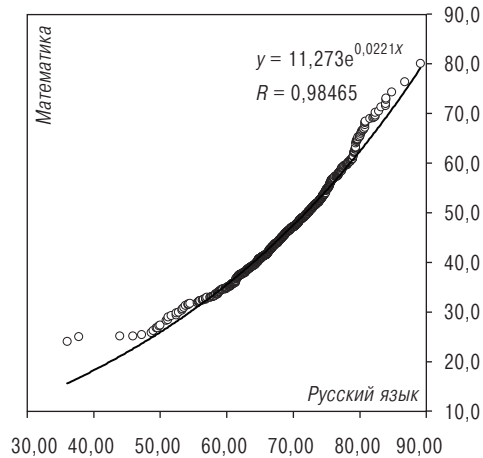


Рис. 4. Связь распределения результатов ЕГЭ по русскому языку и математике



по всей совокупности без привязки к отдельным объектам (в данном случае к школам), мы можем получить скорректированные значения среднего балла по русскому языку для каждого отдельного объекта и уже потом рассматривать пары значений, связанные с данным объектом.

В результате рейтинг школ по показателю «средний балл ЕГЭ по русскому языку и математике» меняется существенно. Как и следовало ожидать, в среднем школы физико-математического профиля улучшили свои показатели в рейтинге (табл. 3).

Иными словами, школы математического и физико-математического профиля поднялись в рейтинге в среднем на 12 позиций, а гуманитарные опустились на 6. Несовпадение значений изменений объясняется тем, что гуманитарных школ в рассматриваемой совокупности почти в 2 раза больше. Изменился состав групп и школ-лидеров, и школ-аутсайдеров.

Эта же схема была использована при анализе средних результатов ЕГЭ по школам одного из регионов и дала аналогичные результаты. Удалось подобрать функцию, обеспечивающую корректное сравнение результатов ЕГЭ по русскому языку и математике с коэффициентом детерминации выше 0,97. В результате пересчета среднего балла ЕГЭ по двум предметам позиции большинства школ в рейтинге изменились. В основном изменения были незначительны, но есть и случаи повышения или понижения ранга на 6–9 позиций.

Изложенный метод приведения результатов тестирования к сопоставимому виду ориентирован не только и не столько на ранжирование. Рейтинг и ранги используются в данном слу-

Таблица 3. **Сравнение среднего ранга школ разного профиля в рейтинге по показателю «средний балл ЕГЭ — русский и математика» до и после пересчета**

	Средний ранг (место в рейтинге)	
	До пересчета	После пересчета
Школы гуманитарного профиля	142	148
Школы физико-математического профиля	120	108

чае только для того, чтобы более наглядно продемонстрировать изменения, которые дает пересчет по предложенной схеме. Гораздо более важным, чем грамотный расчет рейтинга, является то, что разработанный метод позволяет сравнивать результаты тестирования с разными параметрами распределения и даже разными шкалами.

Сравнение результатов ГИА и ЕГЭ

Продemonстрируем применение разработанного подхода для сравнения результатов Государственной итоговой аттестации (ГИА) и ЕГЭ — задачи, которая носит вполне прикладной характер. Учителю важно понять, как соотносится багаж знаний, с которым ученики выходят из основной школы, с баллами ЕГЭ. Сравнение динамики результатов по разным школам позволит выделить наиболее успешные учебные заведения и попытаться распространить их опыт.

Для приведенного ниже расчета были использованы результаты ГИА и ЕГЭ с шагом в два года в школах одного из регионов. Расчет проводился на массиве, состоящем из 222 школ, в которых в 2010 г. в 9-х классах проводилась ГИА и в 2012 г. — ЕГЭ. ГИА и ЕГЭ не только проводятся с использованием разных инструментов, но и имеют разные шкалы (30-балльную и 100-балльную соответственно). Понятно, что для целей практического использования надо исключать при расчете среднего балла ГИА тех учеников, которые не продолжили обучение в данной школе, а из расчета среднего балла ЕГЭ — тех, кто не сдавал ГИА в этой школе, но в данном случае мы этих корректировок не проводили, поскольку для оценки применимости метода они принципиального значения не имеют.

Для расчета все результаты ГИА и ЕГЭ по каждому из предметов были упорядочены по возрастанию. Полученные результаты представлены на рис. 5 и 6: точки на графике соответствуют значениям средних баллов ЕГЭ и ГИА в порядке возрастания.

Для приведения результатов ГИА и ЕГЭ к сопоставимому виду подобраны функции, в одном случае логарифмическая, в другом экспоненциальная, с коэффициентами детерминации

Рис. 5. Сравнение средних баллов ГИА и ЕГЭ по школам региона X (математика)

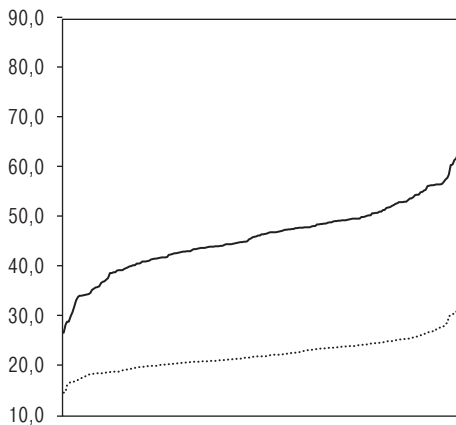
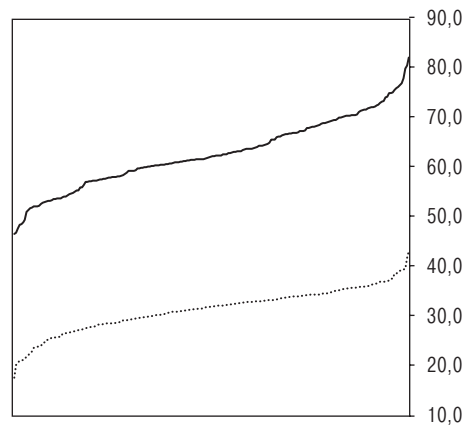


Рис. 6. Сравнение средних баллов ГИА и ЕГЭ по школам региона X (русский язык)



— ГИА 2100
..... ЕГЭ 2012

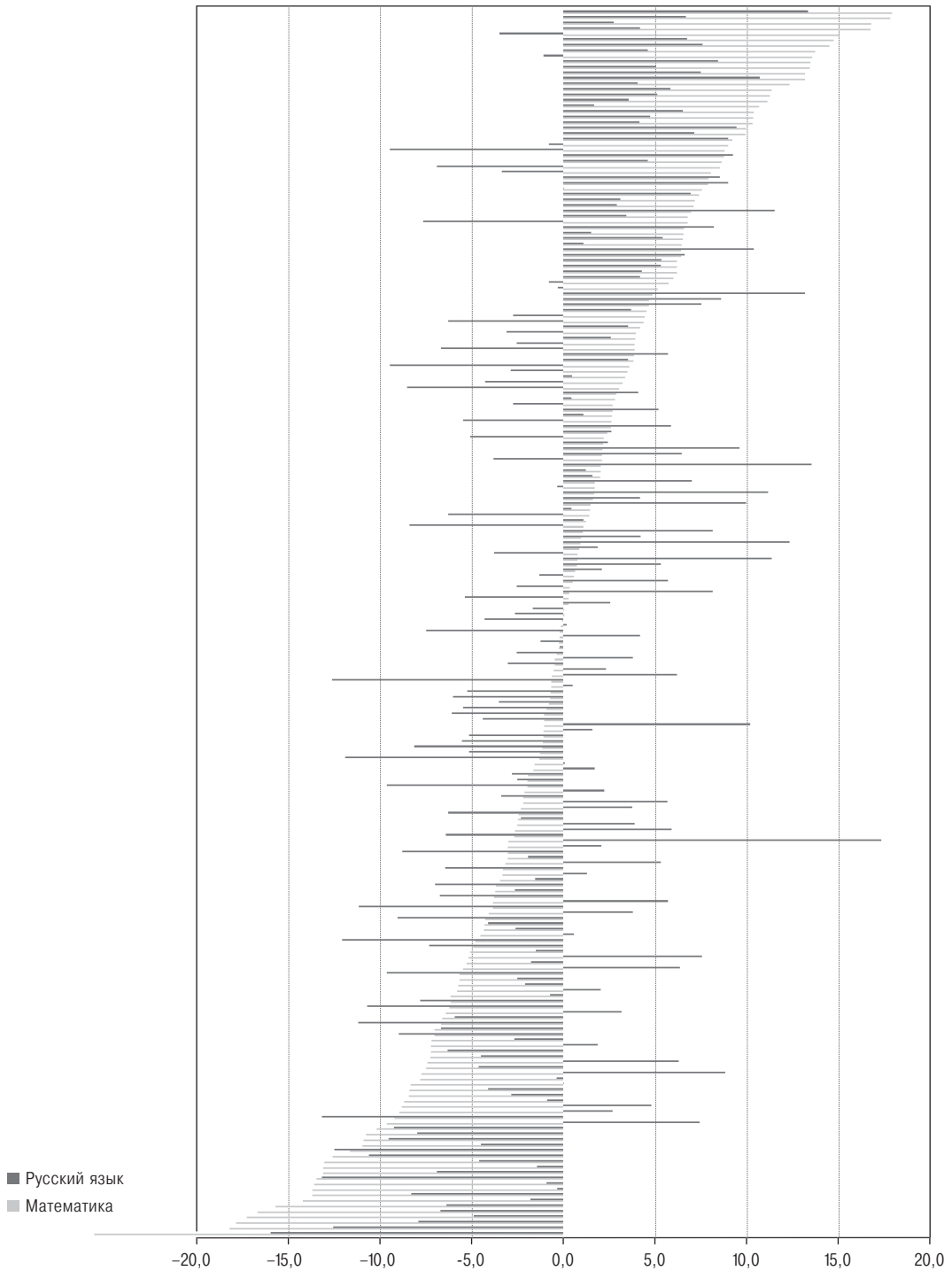
0,99 и 0,98 соответственно. То есть между двумя распределениями была определена практически функциональная зависимость. Результаты ГИА по каждой школе были пересчитаны и затем сопоставлены с результатами ЕГЭ. Итоги сопоставления представлены на рис. 7.

Анализ результатов этого сравнения позволяет сделать следующие выводы.

1. Есть четкие группы лидеров и аутсайдеров, примерно по 20% от общего количества школ, у всех школ первой группы выросли результаты по обоим предметам, у всех школ второй группы результаты по обоим предметам снизились.
2. Имеется значимая, около 0,5, положительная корреляция между результатами сдачи ЕГЭ и приростом, т. е. школы, показывающие высокие результаты на ЕГЭ, обеспечивают и положительный прирост по сравнению с результатами ГИА.

Причин прироста результатов ЕГЭ по сравнению с ГИА может быть несколько: например, он мог быть достигнут благодаря учителям данной школы, или за счет высокого отсева после 9-го класса, или благодаря наличию налаженной системы подготовки к ЕГЭ на курсах вне школы в том или ином населенном пункте, развитой системы репетиторства и т. д. Чтобы определить, какая именно причина сработала в каждом отдельном случае, требуется большой объем конкретной информации. Наша задача была показать, что можно грамотно идентифицировать школы, которые дают прирост знаний (или просто лучше готовят к ЕГЭ) на ступени старшей школы, т. е. сделать корректную оценку и выбрать объекты анализа в такой важной для школ и родителей сфере.

Рис. 7. Прирост/снижение среднего по школам балла ЕГЭ-2012 по сравнению с результатами ГИА-2010 (скорректированными) региона X в баллах по 100-балльной шкале



Итак, мы видели, что непосредственное сравнение показателей, рассчитанных по результатам разных тестовых испытаний, по меньшей мере некорректно. При этом, если предположить, что уровень знаний по разным предметам достаточно большой группы учащихся примерно равный, а различия в параметрах распределения результатов тестирования объясняются применением нетождественных инструментов, то, используя методы математической статистики, можно привести эти результаты к сопоставимому виду. Предложенный способ приведения к сопоставимому виду показателей, рассчитанных по результатам разных тестовых испытаний, когда эти результаты имеют разные параметры распределения и даже разные шкалы, позволяет сравнивать результаты ЕГЭ по разным предметам и за разные годы, а также результаты ГИА и ЕГЭ.

В использовании предложенного метода, безусловно, существуют ограничения. Он применим в следующих случаях:

- 1) когда используются массивы значений, полученных в результате различных тестовых испытаний для одних и тех же объектов;
- 2) когда массив (количество объектов) достаточно велик для использования инструментов математической статистики, применяемых для оценки тесноты связи.

1. Агранович М. Л., Заир-Бек Е. С., Кожевникова О. Н., Поневаж В. П., Рачевский Е. Л. Методические рекомендации по разработке, публикации и распространению аналитических докладов о состоянии и развитии систем образования национального, регионального и субрегионального уровней на основе статистики. М.: Аспект Пресс, 2006.
2. Агранович М. Л. Рейтинг школ повышенного уровня: зачем он нужен и как он строится // Директор школы. 2011. № 6. С. 18–22.
3. Боченков С. А. Анализ и интерпретация результатов ЕГЭ-2012 // Управление образованием: теория и практика. 2012. № 3. С. 9–34 <http://www.iuorao.ru/20120930/216-----2012>
4. Ковалева Г. С. ЕГЭ: можно ли качество образования измерить в баллах? // Первое сентября. 2005. № 66.
5. Об установлении минимального количества баллов Единого государственного экзамена по общеобразовательным предметам, подтверждающего освоение основных общеобразовательных программ среднего (полного) общего образования в соответствии с требованиями федерального государственного образовательного стандарта среднего (полного) общего образования. Распоряжение Рособнадзора от 29.08.2012 г. № 3499–10 <http://www.referent.ru/1/205385>
6. Перечень показателей оценки эффективности деятельности федеральных государственных образовательных учреждений высшего профессионального образования и их филиалов, утвержденный Минобрнауки 09.08.2012 г. № АК-11/05вн <http://www.muiв.ru/upload/iblock/4f2/4f24c9eb1f091852d8ca3d2ae9dd3088.pdf>

Литература

Is it Possible to Compare USE and SFA: A Comparison of Indicators Calculated on the Basis of Different Tests

Author **Mark Agranovich**

Head, The Centre for Monitoring & Statistics of Education, Federal Institute of Education Development; Executive Director, The Interregional Association for Monitoring & Statistics of Education. Address: 9, Chernyakhovskogo str., Moscow, 129319, Russian Federation. E-mail: magran@firo.ru

Abstract The article discusses the possibilities of different tests results comparisons. The issue of correct compare of different tests results raises in relation of analysis of tests results for different years and in a number of other cases. These results could not be compared directly, without some adjustments. The article presents the approach to make tests results with different parameters of distribution comparable. The suggested approach based on the assumption that the level of students' knowledge in different subjects are more or less equal and differences in parameters of tests results distributions arise due to instruments and differences in testing scaling. Suggested scheme of using statistical methods allow to eliminate these differences in testing instruments and make results of different tests comparable. Applications of the suggested method is shown on an example of comparison of results of Unified State Exam (upper secondary graduation tests) in different subjects and for different years, and also comparison of Unified State Exam and State Final Attestation (low secondary graduation tests) results. The sphere of application for the suggested method is limited by the following conditions: the number of objects should be sufficient to use mathematical statistics tools applied to evaluate the closeness of relationship; the paper compares the sets of values obtained in different tests for the same objects.

Key words Unified State Exam, State Final Attestation, distribution of tests results, comparison of different tests results.

- References** Agranovich M. (2011) Reiting shkol povyshennogo urovnya: zachem on nuzhen i kak on stroitsya [The Rating of Advanced Schools: What it is for and how it is formed]. *Director Shkoly*, no 6, pp. 18–22.
- Agranovich M., Zair-Bek E., Kozhevnikova O., Ponevazh V., Rachevsky E. (2006) *Metodicheskie rekomendatsii po razrabotke, publikatsii i rasprostraneniyu analiticheskikh dokladov o sostoyanii i razvitii sistem obrazovaniya natsionalnogo, regionalnogo i subregionalnogo urovney na osnove statistiki* [Methodical Recommendations on Preparing and Dissemination of Statistics Based Analytical Reports on State and Development of Education Systems of National, Regional and Sub-Regional Levels]. Moscow: Aspekt Press.
- Bochenkov S. (2012) Analiz i interpretatsiya rezultatov EGE-2012 [Analysis and Interpretation of USE-2012 Results]. *Upravlenie obrazovaniem: teoriya i praktika*, no 3, pp. 9–34. Available at: <http://www.iuorao.ru/20120930/216> (accessed 10 February 2014).
- Kovaleva G. (2011) EGE: mozno li kachestvo obrazovaniya izmerit v ballakh? [USE: Is it Possible to Measure in Points Quality of Education?]. *Pervoe Sentyabrya*, no 66.
- Ministerstvo obrazovaniya i nauki (2012) *Perechen pokazateley otsenki effektivnosti deyatelnosti federalnykh gosudarstvennykh obrazovatelnykh uchrezhdeniy vysshego professionalnogo obrazovaniya i ikh filialov, utverzhdeniy Minobrnauki 09.09.2012 № AK-11/05BH* [The List of Indicators to Measure the

Efficiency of the Federal Public Educational Institutions of Higher Education and Their Branches. Approved by the Ministry of Education and Science 09.08.12 № AK-11/05vn]. Available at: <http://www.muiv.ru/upload/iblock/4f2/4f24c9eb1f091852d8ca3d2ae9dd3088.pdf> (accessed 10 February 2014).

Rosobrnadzor (2012) *Ob ustanovlenii minimalnogo kolichestva ballov edinogo gosudarstvennogo ekzamina po obshcheobrazovatelnyim predmetam, podtverzhdayushchego osvoenie osnovnykh obshcheobrazovatelnykh program srednego (polnogo) obshchego obrazovaniya v sootvetstvii s trebovaniyami federalnogo gosudarstvennogo obrazovatel'nogo standarta srednego (polnogo) obshchego obrazovaniya. Rasporyazhenie ot 29.08.2012 № 3499-10* [On Establishing the Minimum Number of Points of the Unified State Examination in General Subjects, Confirming the Development of Basic Educational Programs of Secondary (Complete) General Education in Accordance with the Requirements of the Federal State Educational Standards of Secondary (Complete) General Education. The Order of Rosobrnadzor dated from 29.08.2012 № 3499-10. Available at: <http://www.referent.ru/1/205385> (accessed 10 February 2014).