

КОНСУЛЬТАЦИИ

УДК 37:519.2

Л. М. Нуриева

О ПРИМЕНЕНИИ ФОРМУЛЫ БЕРНУЛЛИ ДЛЯ АНАЛИЗА РЕЗУЛЬТАТОВ ЕГЭ

Аннотация. В статье рассматриваются возможности применения аппарата теории вероятностей для анализа статистики Единого государственного экзамена. Исследуется влияние на результаты испытаний закрытых заданий, которые пока, к сожалению, составляют большую часть контрольно-измерительных материалов для выпускников средних общеобразовательных учреждений. На основе произведенных расчетов убедительно доказывается, что чем больше в тестовой работе закрытых вопросов, тем больше шансов у сдающих экзамен элементарно угадать ответы и тем меньше вероятность получения низких оценок.

Предлагаемая автором формула Бернулли позволяет найти объяснения некоторым, на первый взгляд, непонятым аномалиям в распределении участников по набранным баллам, оценить качество экзаменационных материалов и системы начисления баллов, уточнить параметры границ освоения школьниками учебных дисциплин и даже проследить социальные эффекты экзамена, давая им количественную оценку. Автор надеется, что аналитики ЕГЭ при подведении его итогов будут активнее пользоваться возможностями описанной в статье методики.

Ключевые слова: Единый государственный экзамен, анализ результатов ЕГЭ.

Abstract. The paper looks at implementing the probability theory and mathematical statistics while analyzing the outcomes of the unified state examination (USE). The research is aimed at investigating the impact of closed questions that make the greater part of USE, on test results. The methodology is based on so called Bernoulli's trial. The research findings demonstrate the higher probability of incidental right answers to the closed questions compared with the open ones. The author makes a conclusion that the considerable number of closed questions in the test can misrepresent the final result which tends to improve.

The proposed method of statistic analysis can provide the explanation for the USE results anomalies, evaluate the quality of examination materials and scoring system, and give the quantified assessment of social implications.

Keywords: unified state examination, analysis of unified state examination results.

Последние несколько лет в профессиональных изданиях и Интернет стали появляться аналитические материалы, посвященные итогам Единого государственного экзамена (ЕГЭ) в различных субъектах Российской Федерации. Чаще всего их авторство принадлежит местным институтам повышения квалификации работников образования и методическим службам, реже – региональным центрам обработки информации. Образцом для подражания для аналитиков, как правило, служат ежегодные отчеты Федерального института педагогических измерений (ФИПИ). Содержание последних сводится к описанию контрольно-измерительных материалов, перечислению проверяемых элементов школьных курсов с указанием успешности их освоения и более или менее обоснованной интерпретацией полученных данных. В силу ограниченности данных об участниках экзамена [4] отчеты ФИПИ не дают ответа, как и почему достигнуты те или иные результаты. Сходная ситуация с региональными исследованиями итогов экзамена. Составить полную картину мешает не только специфика исходного статистического материала, но и ограниченность используемого аналитического инструментария.

В настоящей статье мы не беремся оценивать приемы, которыми пользуются специалисты при изучении результатов ЕГЭ, а хотим лишь обратить внимание на возможности применения аппарата теории вероятностей.

Уже не раз обсуждалась статистическая зависимость результатов экзамена от соотношения в контрольно-измерительных материалах (КИМах) вопросов открытого и закрытого типов. Большинство заданий ЕГЭ – закрытые, чаще всего с четырьмя (в иностранном языке – с тремя) готовыми альтернативными ответами¹. Вопросы, предполагающие краткую самостоятельную формулировку ответа (обычно в части *B*), на самом деле, как правило, также закрытые, хотя имеют множественный выбор. Иначе говоря, приходится в очередной раз констатировать, что открытых заданий в части *A* и *B* по всем предметам очень мало.

¹ Пока от закрытых вопросов отказались только в математике и литературе.

Алгоритм выполнения части А независимо от дисциплины – это последовательный перебор и исключение сомнительных вариантов ответа. Общий недостаток закрытых заданий – наличие подсказки в альтернативах. Дело даже не в том, что некоторые верные ответы видны сразу (как бывает даже в математике) [3]. Общим правилом является то, что большинство дистракторов (отвлекающих ответов в тестовых заданиях закрытого типа, похожих на правильные, но не являющиеся таковыми) откровенно неправдоподобны, что резко сужает круг поиска. Исключая их, можно получить правильное решение, даже не зная его.

Технология выполнения части А по физике и информатике, где требуются вычисления, максимально приближена к обычной письменной работе. Однако и здесь присутствует элемент подсказки. Закрытое задание позволяет сравнивать результат решения с предложенными тестовыми вариантами. Если среди них полученного ответа не обнаружено, то задача может решаться до достижения успеха. Открытые же задания сигнала для повторного поиска не дают.

В закрытых заданиях при затруднении подходящий ответ может выбираться приблизительно. К тому же способ проверки бланков ЕГЭ требует, чтобы результат вычислений легко читался программами сканирования и распознавания чисел. При таком раскладе вероятность угадать ответ, наобум поставив крестик, сохраняется в 25%. Высоких оценок таким путем, конечно, добиться нельзя, но получить удовлетворительную оценку, как в свое время происходило, например, на экзамене по математике, вполне возможно [2].

Хорошим инструментом для демонстрации влияния закрытых вопросов на результаты экзамена является схема Бернулли. Она позволяет рассчитывать вероятность получения определенного результата в серии независимых испытаний. В нашем случае есть серия из n заданий части А, в каждом из которых четыре альтернативных ответа. Угадывание наступает с вероятностью 0,25 (при трех альтернативах – 0,33), а ошибка с вероятностью 0,75 (0,66).

В соответствии с формулой Бернулли

$$P_n(k) = C_n^k p^k q^{n-k}, \quad k = 0, 1, \dots, n,$$

где $P_n(k)$ есть вероятность того, что в n испытаниях произойдет ровно k успехов, вычисляем, с какой вероятностью можно полу-

читать баллы путем угадывания в серии из n заданий. Результаты вычислений сведены в табл. 1.

Таблица 1

Распределение вероятности получения первичных баллов за выполнение части А путем угадывания ответов

Первичный балл	Информатика	Обществознание	Физика и география	История	Иностранный язык	Русский язык и химия	Биология
0	0,56	0,18	0,08	0,04	0,00	0,02	0,00
1	3,38	1,31	0,63	0,38	0,02	0,18	0,04
2	9,58	4,58	2,51	1,65	0,11	0,86	0,22
3	17,04	10,17	6,41	4,59	0,48	2,69	0,84
4	21,30	16,11	11,75	9,17	1,50	6,04	2,31
5	19,88	19,33	16,45	14,06	3,60	10,47	4,93
6	14,36	18,26	18,28	17,19	6,91	14,55	8,49
7	8,20	13,91	16,54	17,19	10,85	16,62	12,13
8	3,76	8,69	12,41	14,32	14,25	15,93	14,66
9	1,39	4,51	7,81	10,08	15,83	12,98	15,20
10	0,42	1,95	4,17	6,05	15,04	9,09	13,68
11	0,10	0,71	1,89	3,12	12,30	5,51	10,78
12	0,02	0,22	0,74	1,38	8,71	2,91	7,49
13	с 13 по 18 ¹ – 0	0,06	0,25	0,53	5,36	1,34	4,61
14		0,01	0,07	0,18	2,87	0,54	2,52
15		с 15 по 22 – 0	0,02	0,05	1,34	0,19	1,23
16			с 16 по 25 – 0	0,01	0,54	0,06	0,54
17				с 17 по 27 – 0	0,19	0,02	0,21
18					0,06	с 18 по 30 – 0	0,07
19					0,02		0,02
20					с 20 по 28 – 0		0,01
21							с 21 по 36 – 0

¹ Здесь и далее указывается максимальный первичный балл за часть А.

На рис. 1 хорошо видно, как по мере увеличения в работе числа закрытых вопросов от наименьшего (информатика – 18) к наибольшему (биология – 36) растут шансы получения необходимых баллов, а распределение смещается вправо. Результаты по

иностранному языку несколько отличаются из-за того, что вопросы тестовых заданий по данной дисциплине сопровождаются тремя альтернативными ответами.

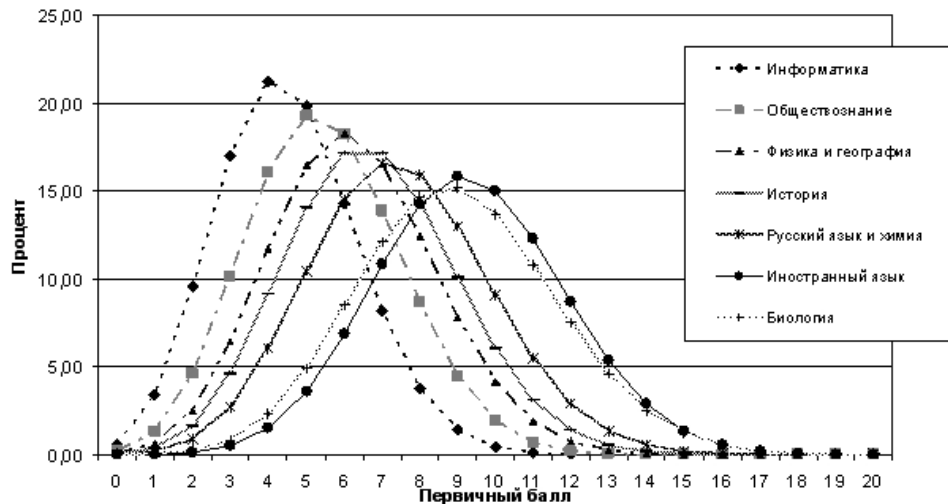


Рис. 1. Распределение вероятностей получения первичного балла по дисциплинам в случае угадывания ответов

Результативность выполнения части А на основе угадывания показана в табл. 2. Подсчеты по схеме Бернулли показывают, что ноль баллов на ЕГЭ получить практически невозможно. Хотя, согласно статистике, по всем предметам такую оценку получают больше учащихся, чем те, кто зарабатывает один-три балла, но происходит это потому, что первые просто сдают пустые (!) работы. Как только школьник начинает, даже глядя в потолок, заполнять бланк, он непременно проставит верные ответы и что-то «заработает».

Негативным примером использования закрытых вопросов служит опыт проведения ЕГЭ по математике в период, когда часть А по этому предмету еще не была отменена. В распределении учащихся по первичному баллу всегда присутствовал выброс значений, соответствующих эффекту угадывания ответов (рис. 2). Особенно это проявилось в 2008 г., когда охват испытаниями стал почти 100% и действовала норма «+1», позволявшая получить удовлетворительную оценку и в случае неудовлетворительного результата. Принудительное вовлечение регионов в экзамен и введение его обязательности привели

к расширению состава участников за счет нежелающих сдавать ЕГЭ. Такие выпускники выполняли экзаменационную работу формально, наобум проставляя крестики в бланке ответов. В 2007–2008 гг. почти каждый пятый школьник страны фактически игнорировал экзамен по математике, выполняя задания «методом тыка» [2]. Отмена «гарантированной тройки» в 2009 г. существенно снизила саботаж экзамена, но не устранила проблему закрытых вопросов. Поэтому выброс значений, равных 3–4 первичным баллам, все равно сохранился. И только в 2010 г. после отказа от части А¹ распределение участников приняло нормальный вид.

Таблица 2

Результативность выполнения части А на основе угадывания

Предмет	Число заданий с выбором ответа	Первичный балл, который в среднем можно получить путем гадания ответов	Интервал первичных баллов, которые в 90% случаев можно получить путем гадания ответов
Информатика и ИКТ	18	4,5	2–7
Обществознание	22	5,5	2–8
Физика	25	6,25	3–9
География	25	6,25	3–9
История	27	6,75	3–10
Русский язык	30	7,5	4–11
Химия	30	7,5	4–11
Биология	36	9	5–13
Английский язык	28	9,24	5–13

Почему закрытые вопросы столь явно обнаруживали себя в математике, хотя в КИМах по остальным предметам они составляют почти половину тестовой работы?² В нашем вопросе содержится ответ: из-за малого их числа общее распределение участников по баллу распадалось на две части. В дисциплинах, где доля «угадаек» выше, распределения выглядят более монолитными, так как значения, соответствующие гаданию, смещаются по шкале вправо. Однако и тут

¹ По заявлению ФИПИ, отказ от закрытых вопросов «отвечает традициям преподавания математики в российской школе» [1]. Упомянутые традиции не мешали, однако, организаторам ЕГЭ десять лет насаждать «тесты-угадайки», в том числе и в математике.

² В 2009 г. доля закрытых вопросов в первичном балле за работу по математике составляла 27%, в истории и обществознании – почти 40%, русском языке, биологии, физике и химии – около 50%.

мы найдем следы закрытых вопросов. Так, в распределении учащихся по баллам по русскому языку наблюдается левосторонний подъем линии графика, что в условиях саботажа испытаний особенно хорошо видно на основе результатов 2008 г. (рис. 2). Масштабы формального участия здесь также поддаются измерению. Для этого мы воспользовались следующим приемом: при подсчете первичного балла уменьшили количество вопросов части А до 10, как это было в математике. Таким образом получили графики распределения омских выпускников (свыше 11 тыс. чел.) по первичным баллам за всю работу (нормальное распределение) и по сумме баллов за части В и С и любых 10 заданий части А (бимодальное распределение). Второй график наглядно показывает, что в 2008 г. около 15% омских школьников, не мудрствуя при выполнении части А, наобум отмечали варианты ответов (рис. 3). Аналогичные результаты, правда, в несколько меньших масштабах обнаружались и при исследовании результатов 2009 г. В 2010–11 гг. из-за отмены правила «+1» саботаж экзамена по русскому языку в омских школах сошел на нет.

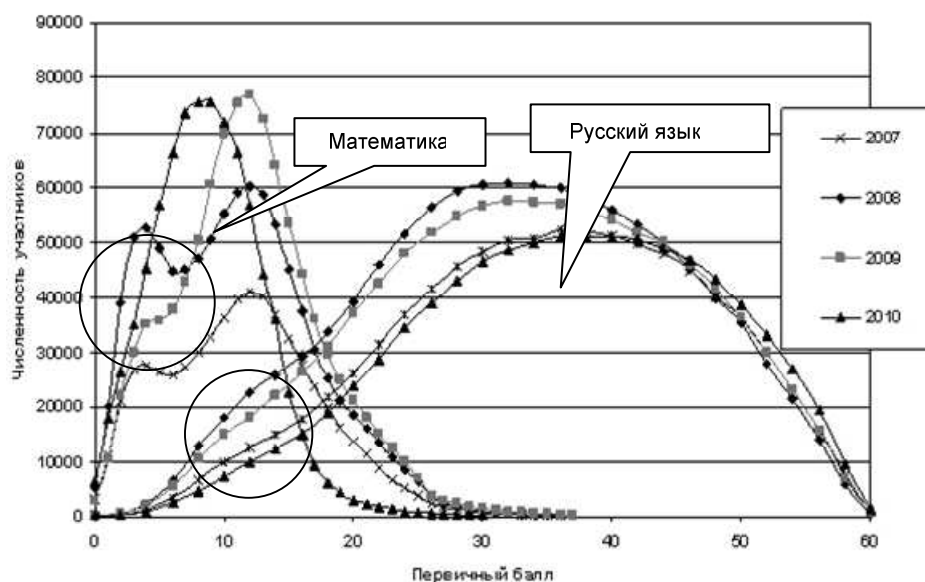


Рис. 2. Распределение участников ЕГЭ по первичному баллу по Российской Федерации (математика и русский язык)

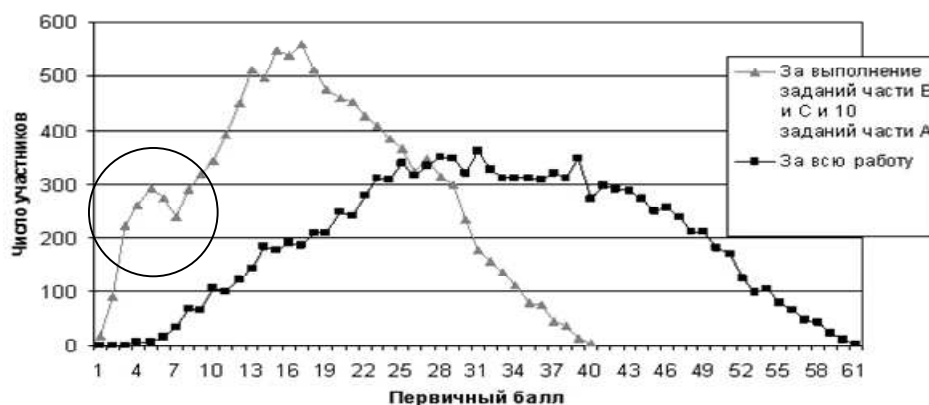


Рис. 3. Распределение омских учащихся по первичному баллу в 2008 г. (русский язык)

В дисциплинах, КИМы которых в значительной мере состоят из «угадаек», последние обнаруживают себя практически полным отсутствием участников с минимальными баллами. Чем больше в работе закрытых вопросов, тем больше шансов что-то угадать, тем меньше вероятность получения низких результатов¹. Поэтому при интерпретации итогов испытаний точку отсчета результативности следует относить вправо по шкале до уровня, когда возможность получить высокие баллы, не утруждая себя, становится допустимо малой. Для русского языка, например, по формуле Бернулли эта точка соответствует 14 первичным баллам (вероятность получения 14 и более первичных баллов путем гадания составляет около 1%). Границы минимального уровня усвоения курса для русского языка Рособрнадзором установлены в 0–16 первичных баллов. Для английского языка эти показатели составляют 16 баллов по формуле Бернулли (здесь и далее в скобках: у Рособрнадзора – 0–15), для биологии – 16 (0–16), географии – 12 (0–15), истории – 13 (0–13), обществознания – 11 (0–14), физики – 12 (0–11), химии – 14 (0–14). Таким образом, уровень, обозначенный Рособрнадзором как минимальный, на самом деле достигается элементарным гаданием ответов, что позволяют сделать закрытые задания. Это демонст-

¹ Учащихся с 4–5 и менее первичными баллами по всем дисциплинам (кроме предметов, где нет части А, и информатики, где число заданий с выбором ответа невелико) в стране практически нет.

рирует, насколько невысоки федеральные требования к уровню подготовки школьников. Ведь выпускники должны выполнить работу немногим лучше, чем просто с известной долей везения спонтанно проставлять крестики в бланке ответов.

Рисование крестиков наугад не смущает организаторов и не мешает выдавать такое «участие» в ЕГЭ за чистую монету. В ежегодных отчетах ФИПИ знания детей, показавших минимальный результат, постоянно характеризуются как отрывочные и бессистемные. Хотя какая уж тут система...

Все выше изложенное имеет еще один аспект. Оценивание учебных достижений с помощью ЕГЭ производится некорректно и несправедливо. Шкалы перевода предусматривают начисление максимальных тестовых баллов минимальным и максимальным первичными. При этом получается, что слабые школьники, формально участвующие в испытаниях, тестовыми баллами максимально награждаются. Сильные же учащиеся, делающие лишь единичные ошибки, потерей тестовых баллов максимально наказываются. Например, только за то, что выпускник явился и наугад «налепил» крестиков в бланке, ЕГЭ по русскому языку «наградит» его в среднем 7–8 первичными или 15–17 тестовыми баллами. А сильные школьники, потерявшие 7–8 первичных баллов из 80 возможных (10%), будут лишены пятой части (20%) тестовых баллов.

Осознание того, что уровень знаний школьников пока, к сожалению, проверяется прежде всего заданиями закрытого типа, в значительной мере определяющими результат испытаний, и применение аппарата теории вероятностей помогают выявить многие ошибки в толковании результатов ЕГЭ. Формула Бернулли позволяет найти объяснения некоторым, на первый взгляд, непонятным аномалиям в распределении участников по набранным баллам, оценить качество контрольно-измерительных материалов и системы начисления баллов, уточнить параметры границ освоения школьниками учебных дисциплин и даже проследить социальные эффекты экзамена, давая им количественную оценку. Хочется надеяться, что аналитики Единого государственного экзамена при подведении его итогов будут активнее пользоваться возможностями данной схемы.

Литература

1. Итоговый аналитический отчет о результатах проведения ЕГЭ в 2010 году (май – июнь 2010 г.). [Электрон. ресурс]. Режим доступа: <http://www.fipi.ru/binaries/1084/mat11.pdf>.

2. Нуриева Л., Киселев С. ЕГЭ и анализ качества обучения математике // Образование и наука. Изв. УрО РАО. 2008. № 5. С. 11–25.

3. Нуриева Л., Киселев С. Единый государственный экзамен и проблемы его анализа. [Электрон. ресурс]. Режим доступа: http://scepsis.ru/library/id_2480.html.

4. Нуриева Л., Киселев С. Является ли ЕГЭ инструментом анализа качества образования? // Математика (приложение к газете «1 сентября»). 2009. № 4. С. 3–7.