

ЕДИНЫЙ ГОСУДАРСТВЕННЫЙ ЭКЗАМЕН НУЖДАЕТСЯ В СУЩЕСТВЕННЫХ ИЗМЕНЕНИЯХ

Люция Нуриева,

liutsiya59@mail.ru

Сергей Киселёв,

Омский государственный педагогический университет,

ksg_sd@mail.ru

В статье рассматриваются некоторые стороны конкретной практики проведения Единого государственного экзамена, обращается внимание на ряд недостатков в технологии измерений учебных достижений школьников, проверки работ и анализа результатов ЕГЭ.

Ключевые слова: ЕГЭ, анализ результатов ЕГЭ, измерение учебных достижений, закрытые вопросы

Накануне 2013 года Президент РФ подписал закон «Об образовании», видимо поставившего точку в затянувшихся спорах «быть или не быть» Единому государственному экзамену. С неоднозначными результатами экзамен всё же состоялся. «Настал важный момент, когда должен произойти процесс нормальной «приватизации» результатов ЕГЭ субъектами Российской Федерации, — заявила в одном из интервью руководитель Рособнадзора (теперь уже бывший) Л. Глебова. — Пришло время задуматься над тем, что делать с результатами с точки зрения анализа итогов госэкзамена»¹.

Действительно, как показывает десятилетний опыт проекта, итоги ЕГЭ нуждаются в критическом осмыслении и инвентаризации. Отсутствие своевременного анализа ведёт к консервации застарелых ошибок и запоздалому реагированию на появление новых проблем, что оборачивается против самого же экзамена и его участников. В нашей статье мы хотим обратить внимание на практику проведения испытаний и подведения итогов, недостатки которых могут быть устранены.

Методология

Методология

1
[http://ria.ru/
interview/20100712/
253851237.html](http://ria.ru/interview/20100712/253851237.html)

Технология измерений

Одним из самых уязвимых мест в тестовой технологии измерений уровня знаний являются закрытые вопросы, позволяющие угадывать ответы. Их доля в первичном балле за всю работу по истории и обществознанию составляет почти 40%, по русскому языку, биологии, физике и химии — около 50%. Более того, вопросы с кратким ответом (обычно часть В) в основном тоже закрытые, хотя и предусматривают множественный выбор. Открытых вопросов в контрольно-измерительных материалах (КИМ) по всем предметам очень мало (как правило, только в части С).

Алгоритм выполнения заданий части А — это последовательный просмотр и отсев сомнительных вариантов ответа. Общий недостаток закрытых вопросов — подсказка в альтернативах. Дело даже не в том, что некоторые из них сразу видны как верные (такое случается даже в математике²). Общим правилом является то, что большинство неверных альтернатив откровенно неправдоподобно. Исключая их, можно получить правильный ответ, 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. Закрытый вопрос позволяет сравнивать полученный ответ с предложенными вариантами. Если совпадений нет, то задача решается до достижения успеха. Открытый вопрос сигнала для повторного поиска верного ответа уже не давал бы. При затруднении подходящий вариант выбирается приблизительно. И в любом случае сохраняется 25% вероятность угадать ответ, наобум поставив крестик.

Следы вопросов-угадаек в виде «горбов» в начале шкалы хорошо видны в распределении учащихся по баллам на обязательных экзаменах. Наиболее ярко это проявилось в период, когда охват испытаниями стал почти 100%-м и действовала норма, позволявшая в любом случае получить удовлетворительную оценку. Зная, что тройка им гарантирована, выпускники наобум проставляли крестики в бланке ответов и сдавали работы.

Например, в 2007–08 годах на ЕГЭ по математике почти каждый пятый школьник страны выполнял задания «методом тыка»³. На графике выбросы значений, соответствующие результативности простого гадания ответов, обведены кружком (рис. 1). В ежегодных отчетах Федерального института педагогических измерений (ФИПИ) знания детей, пока-

2

Киселев С.Г.,
Нуриева Л.М.

Единый государственный экзамен и проблемы его анализа, http://scepsis.net/library/id_2480.html.

3

Там же.

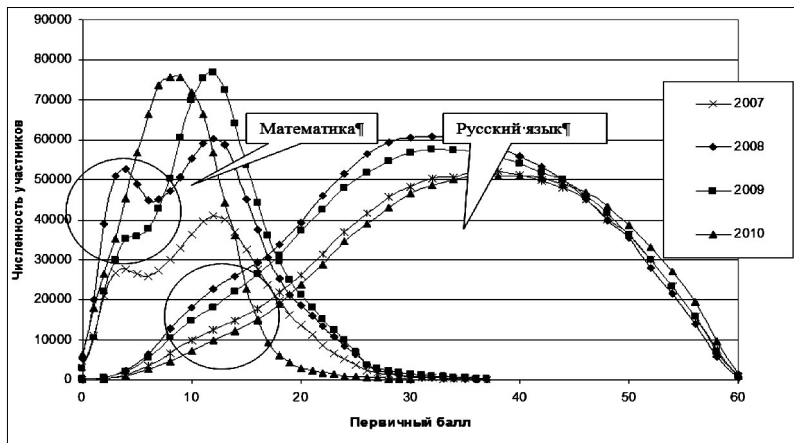


Рис. 1. Распределение участников ЕГЭ по первичному баллу по Российской Федерации (математика и русский язык)

завших минимальный их уровень, постоянно характеризовались как отрывочные и бессистемные. Хотя какая уж тут система?

Столь явное проявление в математике заданий-угадаек, выражающееся в распаде массива участников на две части, объясняется тем, что форма записи ответа (числа, выражения, числовые множества, уравнения) не позволяла выбирать альтернативы методом исключения неправдоподобного. Это стимулировало формальное отношение к выполнению работы: учащиеся стали наобум отмечать варианты ответов. В распределении по русскому языку, КИМы которого почти наполовину состоят из вопросов с альтернативами, содержащими подсказки, подъём линии графика той же приро-

ды сдвинут вправо и выражен менее.

На графиках прочих предметов всплесков, соответствующих результативности гадания, мы не найдем. На экзаменах по выбору дети стараются максимально хорошо выполнить работу и к гаданию не прибегают⁴. Однако следы вопросов-угадаек мы обнаружим и тут. Обратим внимание, что по всей стране и всем дисциплинам учащихся с 4–5 и менее первичными баллами практически нет. Чем больше в работе закрытых вопросов, тем выше вероятность что-то угадать, тем меньше вероятность получения низких результатов. Как только школьник начинает, даже глядя в потолок, заполнять бланк он непременно проставит верные ответы и что-то «заработает». График распределения по

Методология

4

Впрочем, если заставить выпускников сдавать все экзамены, вопросы-угадайки встанут «горбами» в начале шкалы по всем предметам аналогично тому, как это было в распределении по математике.

ПЕД
измерения

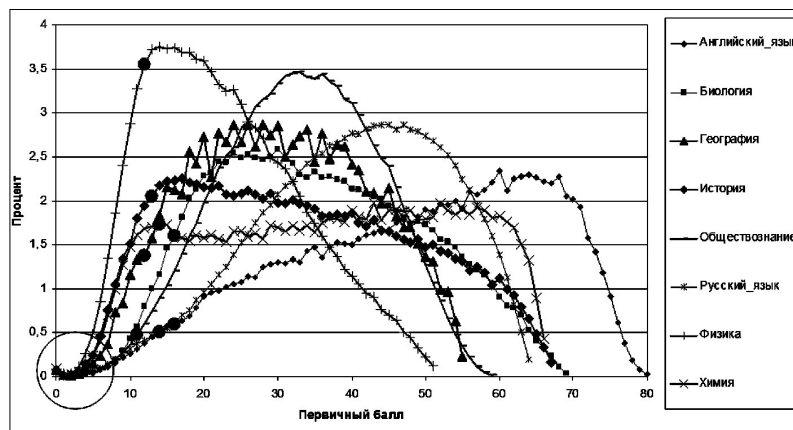


Рис. 2. Распределение участников ЕГЭ по первичному баллу, в процентах по Российской Федерации (2011 год)

баллам хорошо иллюстрирует этот факт (рис. 2).

Понимание того, что знания выпускников проверяются с помощью закрытых вопросов, в значительной мере определяющих результаты испытаний, позволяет избежать ошибок толкования последних. Так, например, уровень, обозначенный Рособрнадзором как минимальный, в значительной мере достигается элементарным угадыванием ответов. Поэтому при интерпретации итогов ЕГЭ точку отсчёта результативности следует считать не с нуля, а от значения, когда вероятность получить высокие баллы, набум отмечая варианты, становится допустимо малой (менее 1%). Для русского языка, например, она составляет 14 баллов, что хорошо согласуется с границей минимально-

го уровня усвоения курса, которую ежегодно устанавливает Рособрнадзор — 17 баллов (2013 г.). Для английского языка эти показатели насчитывают 16 баллов по формуле Бернулли (16 — Рособрнадзор), биологии — 17 (16), географии — 12 (14), истории — 13 (13), обществознания — 14 (15), физики — 12 (12), химии — 14 (14)⁵. Данный факт свидетельствует, насколько невысоки федеральные требования к подготовке учащихся. Ведь работу нужно выполнить немногим лучше, чем с известной долей везения наугад проставлять крестики в бланке ответов.

Проблема вопросов-угадаек уже поднималась во время работы Комиссии по совершенствованию ЕГЭ, созданной при Президенте РФ. В декабре 2010 года на встрече с прези-

5

На графике соответствующие точки обозначены маркером.

дентом министр образования заверил, что его ведомство неоднократно проводило исследования, доказывающие невозможность сдать ЕГЭ путем угадывания. Более того, из КИМов по математике закрытые вопросы вообще исключены. Все это действительно так. Обратим, однако, внимание на одну деталь: убраны закрытые вопросы из контрольно-измерительных материалов дисциплины, где они были представлены менее всего⁶. В 2009 году (до смены формата экзамена) их доля в первичном балле за всю работу составляла лишь 27% (для сравнения: в русском языке, биологии, физике, химии их около 50%). На наш взгляд, часть А из КИМов по математике удалена не столько потому, что её легко заменить открытыми вопросами с ответами в виде целого числа или десятичной дроби, хорошо читаемых машиной (ведь в физике и информатике часть А осталась), сколько потому, что распределение результатов наглядно и «предательски» показывало, как велико влияние закрытых вопросов на общий результат.

Казалось бы, стоит ли обращать внимание на такую мелочь, как возможность школьника что-то угадать? Таким путём высоких результатов на экзамене никогда не получить. К тому же всё равно все участни-

ки ЕГЭ находятся в одинаковых условиях.

Посмотрим, тем не менее, к чему такое допущение приводит на практике. Сравнение достижений учащихся производится не по первичным, а по так называемым тестовым баллам, система начисления которых равных условий уже не создаёт. Шкалы перевода предусматривают начисление максимальных тестовых баллов минимальным и максимальным первичным. В результате участники, проставляющие на экзамене кресты «от фонаря», тестовыми баллами МАКСИМАЛЬНО НАГРАЖДАЮТСЯ. Но это полбеды. Учащиеся, делающие лишь единичные ошибки, потерей тестовых баллов МАКСИМАЛЬНО НАКАЗЫВАЮТСЯ.

Например, за работу наугад на экзамене по русскому языку «непредвзятая диагностика ЕГЭ» одарит лоботряса 7–8 первичными или 15–17 тестовыми баллами. Сильных школьников, потерявших 7–8 первичных баллов из 80 возможных (10%), она обворует на пятую часть тестовых (20%). Отсюда понятной становится уже почти закономерная картина: неожиданно высокие результаты у детей, учившихся в школе на «3–4», и относительно низкие — у отличников, призеров олимпиад и различных конкурсов по предметам.

Методология

6

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

ПЕД
измерения

Именно нелепая система оценивания позволяет обвинять школы в «липовых» медалях, а олимпиады и конкурсы — в слабых победителях. Собственно в этом кривом зеркале ЕГЭ и отражается состояние школьной подготовки.

Отказ от закрытых вопросов и линейная шкала перевода баллов сделают систему оценивания более корректной. Основной причиной сохранения закрытых вопросов является технологичность их обработки. Между тем открытые вопросы тоже легко и быстро обрабатываются. Следует лишь изменить технологию: перейти от проверки символов, которыми записан ответ, к проверке его смысла. В действующей практике каждый бланк ответов верификаторы до рези в глазах, «до буквы», сравнивают с текстом, полученным при машинном распознавании. Таким способом открытые вопросы действительно проверяются очень трудно.

На наш взгляд, следует поступать иначе. Необходимо перевести результаты распознавания задания (пусть это будет, скажем, В1) в массив и отсортировать его по алфавиту (или по той или иной степени сходства). Поскольку ответы представляют собой однотипные записи с вариациями, вызванными ошибками распознавания, сходные ответы тут же сгруп-

пируются. В группах сразу будут видны отклонения в орфографии отдельных записей, не влияющие на их смысл. Ошибки исправляются приведением ответов к единой для группы форме одним щелчком «мыши». И так с каждой группой. После этого массив будет представлять собой несколько групп грамматически верных и одинаковых записей, соответствующих смыслу, который вкладывал школьник. Ответы, не попавшие в группы, будут единичны. Их легко просмотреть обычным способом. Аналогично далее проверяется задание В2 и т.д. Такая проверка может быть сделана очень быстро и качественно. Осталось лишь соотнести полученные варианты с верным ответом.

Применение архаичной системы проверки работ является причиной многих «болячек» экзамена. Так, например, удаление из контрольно-измерительных материалов по математике заданий в закрытой форме без адекватных изменений в технологии верификации повлекло обеднение содержательной части проверяемых элементов базового уровня. В угоду технологичности обработки результатов (ответ должен быть записан числом, легко читаемым машиной) были принесены в жертву такие вопросы курса алгебры и начала анализа средней школы базо-

вого уровня, как множество значений функции, графики основных элементарных функций, правила дифференцирования и таблица производных.

Не проверяется теперь на базовом уровне, и такой блок раздела «Уравнения и неравенства» как задачи на нахождение множества решений неравенств (дробно-линейного, показательного, логарифмического, иррационального). Задание В12 в некоторых вариантах только на первый взгляд требует решения неравенства. На самом деле в этих задачах достаточно найти корни соответствующего уравнения. В итоге умение решать неравенства (и их системы) проверяются только в заданиях повышенного (С3) или высокого (С5) уровня. По этой же причине не проверяются и умения решать тригонометрические уравнения. Более того, ограничения на форму записи ответа (целое число или конечная десятичная дробь) содержит обычную для закрытых вопросов подсказку в случае, если в результате вычислений, например, получается рациональное число $\frac{1}{3}$, которое нельзя представить в виде конечной десятичной дроби.

Наглядный пример влияния формы записи ответа на результат представляет собой задание типа: «Найдите наименьшее значение функции

$y = (x^2 - 7x + 7)e^{x-5}$ на отрезке $[4, 6]$ ». Очевидно, что только при $x = 5$ можно получить ответ, который не содержит числа e , а значит, и решать эту задачу нет смысла благодаря ограничениям на запись ответа.

Технология проверки работ

Проблемы точности и скорости проверки работ не ограничиваются недостатками программы верификации. Серьёзные изъяны имеют также инструментарий проверки и её организация.

За все время проведения ЕГЭ повышенное внимание организаторов уделялось качеству контрольно-измерительных материалов. И это правильно. Однако, абстрагируясь от их содержания, заметим, что сами КИМы баллов не начисляют. Баллы начисляются по бланкам ответов, содержимое которых знаний учащихся может не отражать. Связано это со сложностью заполнения бланков, оплошностями школьников при проставлении ответов, ошибками и низкой дисциплиной верификации.

Начнем с того, что бланк ответов, единый для всех предметов, в части А рассчитан на 60 вопросов. От вплотную наклеенных ячеек рябит в глазах. Между тем в КИМах всех дисциплин часть А насчитыва-

ПЕД	
	измерения

ет не более 30 заданий (кроме биологии, где их 36). Таким образом, второй блок клеток бланка почти не используется.

В 2011 году среди восьми сотен омских учащихся, чей результат по русскому языку составил свыше 80 тестовых баллов, в среднем за часть С потеряно 2 первичных балла из 23 возможных (или 8%), за часть В — 1 из 11 (или 9%). Зато за часть А потеряно 6 первичных баллов из 30 (или 20%). В сложных заданиях части В и С у сильных школьников ошибок вдвое меньше, чем в куче простейших угадаек. Объяснение этому может быть одно: *ошибки в части А у хорошо подготовленных выпускников связаны в основном с банальной путаницей в постановке крестиков* огромного числа заданий, не представляющих в действительности для этой группы экзаменуемых практически никакой трудности (апелляция, кстати, здесь не предусмотрена). В этих условиях большую роль при ранжировании участников по высоким баллам играют уже не столько собственно знания (а уровень подготовки таких учащихся понятен), сколько иные факторы: скрупулёзность в заполнении клеточек (теперь мы проверяем и такие характеристики выпускников) и качество распознавания и верификации ответов.

В иностранных языках закрытые вопросы вообще имеют только три альтернативы. Здесь бланк ответов не соответствует КИМам. Любой верификатор подтвердит, что школьники вместо третьего (последнего в КИМах) представляют четвёртый (последний в бланке) вариант. Это автоматически означает ошибку. Приведем замечательный фрагмент отчёта ФИПИ по английскому языку: «...количество ошибок при переносе ответов в бланки (запись лишних слов или символов, **ошибки в выборе позиции ответа, приводящие к полному отсутствию ответов в бланке ответов** по какому-либо заданию) (выделено нами. — *Авт.*) продолжает увеличиваться, что свидетельствует о снижении общей функциональной грамотности экзаменуемых, а также о том, что при подготовке к экзамену уделяется недостаточное внимание ознакомлению учащихся с бланком ответов и правилам работы с ним»⁷. Может ФИПИ стоит просто почистить бланк от мусора, разрядить клетки и привести его в соответствие КИМам?

На протяжении почти всех 10 лет проведения ЕГЭ верификация ответов осложнялась конструкцией КИМов. Закрытые вопросы, где ответ состоял из набора букв, проверялись с большим трудом. При распоз-

7

Итоговый аналитический отчёт о результатах ЕГЭ. <http://fipi.ru/view/sections/138/docs/624.html>.

навании текста машиной заглавная буква «А» путается с буквой «Д» и наоборот, «И» с буквой «Ч», «Н», «К» и т.д. Если школьник пишет прописью (а значительная часть участников экзамена игнорирует требование писать заглавными печатными буквами), то машина путает буквы «б», «в», «д». Даже верификатор иной раз затрудняется сразу определить, какую букву проставил школьник. Нередки случаи, когда вместе с кодом альтернативы пишется ее порядковый номер. При этом если для записи ответа использовались цифра 3 и буква «З», верификатор, обязанный оставить только допустимые символы, оставит оба этих знака, т.к. внешне они неразличимы. Верный ответ будет верифицирован как ошибочный. И это хорошо, если в условиях жесточайшего аврала оператор станет досконально разбираться во всех таких записях.

Между тем проблема решалась элементарно. Обозначения альтернатив следовало бы сразу сделать цифровыми. По свидетельству ФИПИ, замена в 2011 году в вопросе В5 по обществузнанию буквенного обозначения вариантов ответа на цифровое дало 20% рост его решаемости⁸. Иными словами, низкий показатель был вызван не слабыми знаниями учащихся, а ошибками верификации.

Сама программа верификации «TestReader» также нуждается в совершенствовании. Нацеленная на распознавание каждого знака, она составляет бессмысленные наборы символов из вполне читаемых слов. Между тем не следует делать упор на точности распознавания каждой буквы. Требуется правильно распознавать слоги. Процесс значительно ускорится, если программа сама будет подбирать в слоги наиболее вероятные сочетания букв, ибо практически вся верификация состоит из исправлений такой простейшей несурзацы.

Сомнительным представляется и использование единой для всех предметов системы распознавания. Действительно, программа «TestReader» позволяет работать с различным материалом (символы латиницы, кириллицы и цифры) и отличает один знак от другого. Но понять ответ целиком она не в состоянии. К тому же ответы состоят не только из слов. У каждой науки своя знаковая система: латинские и греческие буквы, математические символы, химические и физические формулы и т.д. Программы верификации должны «понимать», с каким материалом они имеют дело, в чем «TestReader», увы, абсолютно «не разбирается». Поэтому распознавать записи, представленные в виде формул, на ЕГЭ за десять лет так и не научились.

Методология

8
Итоговый аналитический отчет о результатах Единого государственного экзамена 2011 года. <http://www.fipi.ru/binaries/1189/1osnrez2011.pdf>.

ПЕД
измерения**9**

См., например, «Анализ результатов проведения единого государственного экзамена 2009 года в Новосибирской области», http://nscm.ru/docs/EGE_Analiz_2009.rar.

10

Автору довелось работать в отделе статистики РИАЦ, поэтому наблюдения здесь высказанные, касаются работы именно этого отдела.

11

О том, что центры обработки информации не обеспечены адекватным функционалом, Росбурнадзор знает и рекомендует нагружать их обработкой ведомственной статистики. Однако этот функционал тоже сезонный: приём статистических отчётов системы образования приходится, главным образом, на октябрь-ноябрь и объём этой работы даже на уровне области (края) сравнительно не велик.

Технологию проверки работ невозможно рассматривать также вне контекста её организационных вопросов. Острой проблемой экзамена остается дисциплина участников, прежде всего школьников. О недоработках организаторов если сообщается, то очень немного⁹. Между тем если взглянуть на работу региональных центров обработки информации (РЦОИ), можно обнаружить интересные вещи.

Одному из авторов статьи (С.К.) «посчастливилось» в течение двух лет работать в Омском региональном информационно-аналитическом центре системы образования (РИАЦ), выполняющего функции РЦОИ, участвовать в двух кампаниях по верификации школьных работ (в 2010 и 2011 годах), наблюдать организацию проведения экзамена изнутри. Что-то из увиденного является общей проблемой всех РЦОИ, что-то можно отнести к местной специфике¹⁰.

Главной особенностью функционирования РЦОИ является сезонный характер занятости. Работа здесь имеется только в период проведения экзамена (середина мая — июль), т.е. три месяца в году¹¹. В период массовой обработки экзаменационных материалов (июнь) нагрузка становится предельно высокой. Бланки ответов следует пропустить через сканер и программу распозна-

вания и установить соответствие их содержания символам, распознанным машиной. Установление такого соответствия осуществляется вручную операторами станций верификации. Несмотря на кажущуюся простоту, верификация является довольно трудной работой. В течение трёх-четырёх недель без выходных, по 8–10 часов не отрываясь от экрана, операторам необходимо сверять каракули школьников с текстом, который распознала машина. Своих сотрудников для этого в РЦОИ, как правило, недостаточно, поэтому они могут привлекать дополнительных работников по договору. На оплату верификации в регионы специально направляются деньги.

Проблему нехватки людей для верификации РЦОИ на местах, видимо, решают по-разному. В Омском РИАЦ, например, «под ружьё» ставят всех: не только программистов, методистов, аналитиков, но и бухгалтеров, секретаря, рабочих типографии (в Омске учителя школьные работы больше не проверяют). Наиболее стойкие сотрудники в дополнение к дневным сменам работают и по ночам: т.е. где-то 14–16 часов в сутки. Отказ от привлечения людей со стороны объясняется стремлением освоить присылаемые Москвой деньги узким кругом штатных сотрудников или «своих» людей.

Введение Государственной итоговой аттестации (ГИА) обострило проблему своевременной проверки работ. Вместо того чтобы привлечь людей со стороны на определённое время, Министерство образования Омской области пошло на увеличение штатов РИАЦ. В нем был создан отдел из шести человек, основной задачей которого стал сбор статистики по сфере образования (ранее этот функционал между делом выполнял всего один сотрудник). Решив проблему нехватки операторов, министерство усугубило проблему занятости. Основным времяпрепровождением новых работников в межэкзаменационный период стали пустые посиделки, блуждания в интернете, компьютерные игры и просмотр видеofilмов. К слову сказать, Омский РИАЦ собрал на своем сервере огромную видеотеку, изучению которой и стали посвящаться усилия нового отдела¹². Падение дисциплины отразилось и на верификации. Иные сотрудники, одновременно с проверкой работ, умудряются не то что слушать музыку (что запрещено инструкцией) — сериалы смотреть.

Бывший руководитель Рособнадзора В. Болотов, как-то посетовав на низкую дисциплину экзамена, предложил для её поддержания привлечь армию. Что касается Омской об-

ласти, то старшину в пору ставить в кабинетах, где проверяются работы учащихся¹³.

Насколько сказываются на качестве верификации недостатки её организации, вероятно, не скажет никто. Однако согласно статистическим исследованиям, проведённым по заказу Министерства образования и науки РФ группой специалистов Московского института открытого образования под руководством И.Яценко, в 2011 году до 7% участников экзамена получили оценку, которая не соответствует их знаниям¹⁴. В этот показатель вошли все погрешности измерений, в том числе ошибки верификации.

Анализ результатов

Одним из главных недостатков организации экзамена в настоящее время является система подведения итогов, а именно отсутствие их должного анализа. Материалы, посвящённые «разбору полётов» в субъектах Федерации, как правило, строятся по схеме, применяемой ФИПИ: характеристика контрольно-измерительных материалов и контингента участников, сведения о выполнении различного рода задач, выводы и рекомендации по совершенствованию преподавания предмета. Солидные по объёму, они почти одинаковы по содержа-

Методология

12

Как-то изучая результаты ЕГЭ по физике, автор столкнулся с необходимостью проверить, какие альтернативы отмечали сильные и слабые школьники на один из вопросов по молекулярно-кинетической теории. Однако выяснилось, что результаты сканирования школьных работ в РИАЦ были уничтожены уже спустя 3 месяца после окончания экзамена. Как следовало из пояснений заместителя директора, файлы изображений занимают на сервере много места. Зачем они нужны? Действительно, если сервер забит фильмами, места там, в самом деле, для ЕГЭ нет.

13

В июле 2011 г. в Омском РИАЦ грянул гром: против директора РИАЦ В.Федотова было возбуждено уголовное дело за фальсификацию результатов ЕГЭ (см. сайт Следственного комитета РФ по Омской области). Министерство образования Омской области, не желая скандала, тихо уволило Федотова и не стало разбирать состояние дел. Вместо того чтобы внимательно посмотреть на функционал и объёмы работ, оно принялось накачивать контору

ПЕД
измерения

деньгами и людьми. Сейчас в Омском РИАЦ работает свыше 50 чел. Для сравнения: РЦОИ Московской области насчитывает 16, Нижегородской — 10, Ставропольского края — 9, Челябинской области — 4 чел.

Число выпускников школ в Омской области 11 тыс., Московской — 34 тыс., Нижегородской — 15 тыс., Ставропольского края — 14 тыс., Челябинской области — 13 тыс.

14

Неожиданные итоги Единого госэкзамена 2011: Списывали меньше, чем в 2009-м!, Комсомольская правда, 17 января 2012, <http://www.kp.ru/daily/25818/2796538/>.

15

Данный выбор объясняется тем, что отчёты этих регионов мы нашли в интернете.

16

Результаты Единого государственного экзамена (май-июнь 2006 года) Аналитический отчёт, Москва, 2006.

17

Предметный анализ результатов единого государственного экзамена по математике в разрезе административно-терри-

нию. Значительную часть в них занимает пересказ нормативных документов. Из года в год они повторяют одни и те же выводы и рекомендации. Более того, справки ФИПИ, например, местами дословно переписаны из отчётов прошлых лет. В строгом смысле некоторые работы трудно назвать аналитическими, т.к. носят они преимущественно описательный характер. Большая часть содержащейся в них информации может быть представлена в табличной форме. Литературные дарования составителей в ряде случаев украшают эти документы, но не спасают положения. В них практически нет объяснений причин успехов и неудач школьников. Однако даже эта работа сопровождается систематическими ошибками и упущениями. Укажем на ряд из них на примере материалов ФИПИ, а также работ Московской, Мурманской, Новосибирской, Омской и Томской областей¹⁵.

Рассмотрим сначала **ошибки интерпретации** результатов ЕГЭ.

1. Подведение итогов без учёта особенностей состава участников и условий проведения испытаний. Довольно часто аналитики делают выводы о тенденциях в изменении уровня подготовки выпускников в отчётном году путем простого сопоставления результа-

тов с прошлогодним уровнем. Без учёта особенностей состава участников и условий проведения испытаний делать это нельзя.

К сожалению, одними из первых такой пример нам показывают специалисты ФИПИ. В 2006 году они приходят к заключению о повышении уровня математической подготовки выпускников по сравнению с предшествующим годом, проигнорировав при этом увеличение доли участников из крупных городов, где, по признанию самого ФИПИ, показатели всегда были выше¹⁶.

Омские аналитики, определяя равнотрудность заданий, принялись вычислять разницу между средним процентом выполнения «простого» и «сложного» вариантов, делая выводы о неравнозначности КИМов. Однако, как выяснилось, так называемые «простые» варианты были распространены среди более подготовленных городских школьников, а «сложные» — среди сравнительно слабых сельских¹⁷. Равнотрудность заданий и вариантов вычислялась и в других регионах¹⁸. Однако есть подозрения, что эти сравнения проводились без учёта особенностей состава участников, т.к. местные специалисты данную проблему не поднимали вовсе.

Длительное время распространённой ошибкой было не-

внимание к условиям проведения экзамена, напрямую определяющим его результаты. В 2004–2008 гг. происходило массовое подключение к ЕГЭ новых регионов. Обязательность экзамена вела к расширению состава участников за счёт нежелающих его сдавать. Чтобы побороть сопротивление эксперименту на местах, Рособрадзор установил правило «+1», гарантирующее выпускникам получение удовлетворительной отметки при неудовлетворительном результате. Но это стимулировало не участие в испытаниях, а его имитацию. На экзамене школьники стали наобум рисовать крестики в бланке ответов и сдавать работы. И такая форма «выполнения» заданий становилась всё популярнее.

Размеры саботажа обязательных экзаменов в те годы поражают. Например, в 2007–08 годах доля школьников, выполняющих задания по математике наугад, составляла до 20% по стране. Поражает и другое — близорукость специалистов, этот саботаж не замечающих. Более того, ФИПИ в своих отчётах преднамеренно замалчивал проблему, пряча неприятные факты в обтекаемых выражениях: «При интерпретации результатов ЕГЭ следует иметь в виду, что значительный процент учащихся, показавших неудовлетворитель-

ный или невысокий уровни подготовки, частично объясняется отсутствием положительной мотивации на овладение курсом математики у слабо подготовленных выпускников, которым известно заранее, что при любом даже неудовлетворительном выполнении работы им обеспечена хотя бы минимальная положительная аттестационная отметка «3» или они не планируют поступать в вузы, где надо сдавать математику»¹⁹.

«Холодным душем» для омского руководства образованием стали, например, результаты экзамена в 2005 году. Область оказалась в десятке худших регионов страны. Все чувствовали, что здесь что-то не так, но объяснить произошедшее не могли. Омский аналитический центр только сводил статистику и разводил руками. Между тем основной причиной низких результатов была бездумная политика руководителей разного уровня. С их подачи в школах широко развернулась агитация «за ЕГЭ»: «Дети! Не волнуйтесь. Меньше тройки вы не получите». Дети и не волновались...

Для того чтобы оценить масштабы халтуры, приведём распределение учащихся по тестовому баллу в период с 2004 по 2007 год (рис. 3), которое свидетельствует, что в первые три года участия Омской обла-

Методология

ториальных единиц (июнь 2006) Омск, 2006, с. 31.

18

Анализ результатов государственной итоговой аттестации выпускников 2010 года общеобразовательных учреждений Томской области, <http://coko.tomsk.ru/files/reports/analit-ege-2010.pdf>.

19

Результаты Единого государственного экзамена (май-июнь 2008 года), <http://www.fipi.ru/view/sections/138/docs/401.html>.

ПЕД
измерения

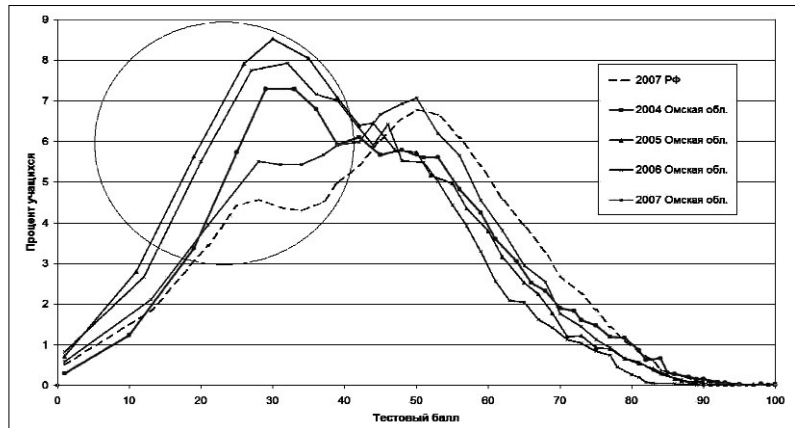


Рис. 3. Распределение омских учащихся по тестовому баллу (математика)

20

Этот приём полезен для любого специалиста, обладающего протоколами ЕГЭ. Работникам ФИПИ, например, он позволит оценить масштабы саботажа экзамена по субъектам Федерации. В 2009 году, рассказывая о методике обработки данных по русскому языку, ФИПИ продемонстрировал результаты нетипичных регионов, но не объяснил причин возникновения полученных аномалий. На графиках ФИПИ были зафиксированы смещения распределений участников к началу шкалы, что могло быть следствием как слабых знаний, так и формального участия школьников. Предлагаемый способ как раз и позволит это выяснить.

сти в ЕГЭ на экзамене по математике просто «от фонаря» поставили крестики выше трети выпускников — 34% (в 2007 г. — 26%)! (Значения, соответствующие результативности гадания, обведены кружком.) Обескураженно фиксируя низкие показатели, ни Омский аналитический центр, ни местное Министерство образования об этом даже не догадывались, а по итогам ЕГЭ региональные руководители отрасли из года в год беспомощно краснели перед губернатором.

Следы бойкота ЕГЭ мы найдем и в результатах по русскому языку. Они проявляются в виде левостороннего подъёма линии графика распределения по баллам, особенно хорошо заметного в 2008 году (т.е. до отмены правила «+1») (рис. 1). Масштабы саботажа

поддаются оценке. Для этого воспользуемся следующим приёмом: при подсчёте первичного балла уменьшим число вопросов части А до 10, как это было в математике²⁰. Таким образом мы получили, например, графики распределения омских выпускников 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 из них свидетельствует, что в 2008 году около 15% омских учащихся, не мудрствуя, отмечали варианты ответов наобум (рис. 4). Аналогичные результаты, правда, в несколько меньших размерах обнаружили и при исследовании результатов 2009 года. В 2010–2011 годах из-за отмены правила «+1» са-

ботажа ЕГЭ по русскому языку в омских школах сошел на нет.

Игнорировали условия проведения испытаний не только в Омске. Аналитики Московской области в 2008 году приводят таблицу зависимости тестового балла от количества часов в неделю изучения математики, из которой следует, что разница в успехах учащихся обучающихся 4 часа (гуманитарный класс), 5 часов (общеобразовательный класс) и 6 часов (профильный класс) невелика. Далее следует вывод: «Это говорит как о примерном равенстве контингента учащихся разных классов, так и о том, что многие учителя при большем числе учебных часов не столько расширяют и углубляют знания учащихся, сколько решают большее число однотипных заданий. Однако отработка навыков не даёт должного эффекта». О том, что расшире-

ние состава участников, правило «+1» и обязательность экзамена стимулировали его саботаж специалисты, видимо, не подозревают. А ведь именно последний и вел к нивелировке тестового балла у изучаемых групп²¹.

В 2009 году в поисках причин общих успехов специалисты Московской области последовательно перебирают и сравнивают с прошлогодними результатами выполнения отдельных заданий. Не найдя существенных изменений в редакции КИМов, они делают выводы об улучшении уровня подготовки, верно заметив при этом, что рост показателей произошел в заданиях с выбором ответа. Задания с кратким ответом остались в минусе. Поэтому последовало следующее заключение: «Так как содержание заданий ЕГЭ-2009 не пре-

Методология

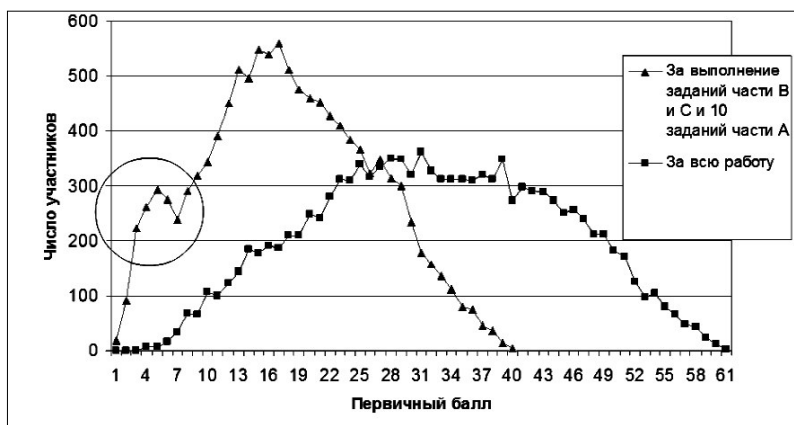


Рис. 4. Распределение омских учащихся (11 тыс. чел.) по первичному баллу в 2008 году (русский язык)

21

Анализ результатов Единого государственного экзамена в Московской области в 2008 году. Сборник информационно-методических материалов, <http://rcoi.net/statistika-i-analitika/105.html>.

ПЕД
измерения

22

Анализ результатов Единого государственного экзамена в Московской области в 2009 году. Сборник методических материалов, <http://rcoi.net/statistika-i-analitika/237-analitika-2009.html>

23

Анализ результатов проведения единого государственного экзамена 2009 года в Новосибирской области, Новосибирск, http://nscm.ru/docs/EGE_Analiz_2009.rar.

24

Результаты Единого государственного экзамена (май-июнь 2009 года), <http://www.fipi.ru/view/sections/138/docs/478.html>.

терпело существенных изменений по сравнению с прошлым годом, трудно оценить прогресс или регресс знаний учащихся по всей школьной программе»²². Проигнорировав изменения условий испытаний, исследователи зашли в тупик. Отмена гарантированной тройки снизила формальное участие школьников и гадание в части А, вызвав рост ее решаемости. По части В результаты, наоборот, ухудшились, т.к. гадать там было невозможно. Поэтому вывод по Московской области должен быть другим: в 2009 году работа выполнена хуже, чем в 2008-м.

На эти же «грабли» наступают в Новосибирске. Здесь так объясняют повышенные результаты в 2009 году: «В то же время существенное увеличение среднего балла по математике (на 22%) является следствием интенсификации процесса подготовки выпускников к ЕГЭ. При этом важно понимать, что в этот процесс включаются не только школьные учителя, но и подготовительные курсы вузов, репетиторы и пр. Немаловажную роль в получении такого прироста в величине среднего балла сыграли и штатный режим экзамена, и принятие педагогической общественностью его обязательности». О том, что из-за отмены правила «+1» школьники просто перестали гадать, ана-

литикам в голову не пришло. Далее они по той же причине делают еще один неверный вывод, объясняя повышение коэффициента корреляции результатов ЕГЭ и школьных оценок в 2009 году возросшей адекватностью оценивания учителем предметных достижений обучающихся²³.

Ошибки в оценке ситуации на местах отчасти лежат на совести ФИПИ, ибо бойкот экзамена он замалчивал. Графики 2004–2009 гг., опубликованные в отчетах института, содержат несложную уловку. Они выполнены в виде гистограмм, где значения, соответствующие угадыванию ответов, скрытаны в суммах интервалов шкалы и не бросаются в глаза, а потому не комментируются. В 2009 году повышение результата по математике ФИПИ, например, объясняют усиленной подготовкой, скрывая факт формального участия школьников на ЕГЭ-2008: «Наблюдаемая тенденция некоторого повышения уровня математической подготовки выпускников средней школы в 2009 году обусловлена в значительной степени тем, что они были мотивированы на продуктивную подготовку к выпускному экзамену в связи с отменой правила «+1» при оценке выполнения вариантов КИМ»²⁴.

2. Сопоставление средних баллов по дисциплинам с об-

щероссийскими показателями. Пожалуй, самой распространённой ошибкой являются сравнения региональных результатов с общероссийскими. Такие сопоставления с учётом ряда оговорок можно делать только по обязательным предметам. В дисциплинах по выбору необходимо обращать внимание на контингент участников, т.к. он может не отражать действительной структуры школьников всей страны. Например, состав учащихся, сдающих французский язык, наполовину представлен выпускниками Москвы, Московской области и Санкт-Петербурга, которые в значительной мере и формируют этот якобы «общероссийский» балл. Сравнение тут идет не со «среднероссийским», а со «среднестоличным» уровнем. И имеет ли смысл вычислять региональный балл, если участников экзамена можно «пересчитать по пальцам»?

При сравнении баллов по обязательным предметам необходимо также учитывать особенности социального развития каждой местности, в частности их урбанизацию, поскольку в регионах с повышенной долей городского населения показатели всегда будут выше. Видимо, межрегиональные сопоставления корректны среди территорий со сходными социально-экономическими условиями функционирования

образовательных учреждений: например, на уровне федеральных округов, результаты которых ФИПИ, к сожалению, не публикует.

3. Сопоставление среднего балла по дисциплинам. Простыми и наглядными, казалось бы, выглядят сравнения средних баллов по разным дисциплинам. Сами собой напрашиваются выводы о том, по какому предмету школьники лучше учатся. Так составители одного из отчётов пишут: «Представляется, что в целом по сравнению с результатами, продемонстрированными при сдаче других экзаменов, результаты по обществознанию выглядят достаточно неплохо, если учесть, что для данного экзамена был установлен наиболее высокий минимальный балл»²⁵.

Между тем представляется некорректным сравнивать средние баллы по разным дисциплинам. На итоги экзамена самым непосредственным образом влияют содержание, количество, характер и сложность заданий, которые при межпредметных сопоставлениях невозможно соотнести. Тот факт, что средний балл по русскому языку составил 57, а по математике 44, вовсе не означает, что русский язык школьники лучше знают. Один только формат КИМов, где преобладают задания, позволяющие угадывать ответы (русский

Анализ результатов государственной итоговой аттестации выпускников 2010 года общеобразовательных учреждений Томской области, <http://coko.tomsk.ru/files/reports/analit-ege-2010.pdf>.

ПЕД
измерения**26**

Результаты Единого государственного экзамена (май-июнь 2009 года), <http://www.fipi.ru/view/sections/138/docs/478.html>

27

Перестановка заданий внутри работы была вызвана желанием организаторов создать большее число оригинальных вариантов в условиях, когда исходный банк заданий был сравнительно невелик.

28

Не обходилось без курьезов. Например, в 2007 году на экзамене по математике школьникам предлагались разные задания, но под одним кодом и с одинаковым ответом! Если учащийся спросит соседа, каков, например, ответ на задачу А2, то получит верную подсказку, хотя оба решали разные задачи.

29

Предметный анализ результатов единого государственного экзамена по математике в разрезе административно-территориальных единиц (июнь 2006) Омск, 2006.

язык), даёт ощутимую фору его участникам перед школьниками, вынужденными давать краткий ответ (математика). В экзаменах по выбору к тому же имеются значительные отличия в составе участников, поэтому средний балл показывает уровень только их подготовки.

О том, что межпредметные сопоставления баллов некорректны, соглашаются и специалисты ФИПИ: «...на результаты ЕГЭ могут влиять различные факторы: состав участников, особенности КИМ, организационные особенности проведения экзамена и др. В этой связи нельзя однозначно сравнивать результаты по разным предметам, даже тестовые баллы, полученные по обязательным предметам и предметам по выбору»²⁶.

Ошибки содержательного анализа результатов испытаний

1. Аналитики не обращают внимания на тот факт, что под одинаковыми кодами содержатся задания разных типов. Широко распространённой ошибкой в течение первых пяти лет проведения экзамена было представление о том, что КИМы имеют одинаковую структуру. На самом деле первоначально они создавались на

основе нескольких планов, являвшихся модификацией общего плана, обозначенного спецификацией. Под одним кодом в разных вариантах скрывались задания разных типов²⁷. То, что такое возможно, организаторы не афишировали. Но после того, как на местах стали сводить несводимое, в спецификациях с 2007 года стали появляться предупреждения: «Порядок следования заданий в КИМ может быть изменен в разных вариантах». По мере накопления банка заданий составители стали отказываться от практики перестановки заданий внутри работы. По крайней мере, к 2010 году предупреждающее извещение осталось только в спецификациях по обществознанию, химии и информатике. К 2012 году оно сохранилось только в обществознании. Трудно сказать, насколько эта мера оказалась действенной в предотвращении списывания или обмена результатами во время испытаний²⁸. Однако она изначально ставила жирный крест на попытках содержательного анализа ЕГЭ на местах. Без приведения вариантов к единому порядку следования заданий статистика по типам задач представляла собой настоящую «кашу». Именно эту «кашу» и выкладывали РЦОИ в свои статистические сборники (так, в частности, было в Омске²⁹).

Приводя статистику выполнения заданий, обозначенных одним кодом, аналитики порой диву даются разбросу их решаемости³⁰, в то время как необходимо было посмотреть на тексты самих заданий. Многие вопросы отпали бы сами собой.

2. Аналитики не учитывают, что содержание или конструкция заданий иногда не соответствует спецификации. Хорошим примером этой ошибки является задание В1 по обществознанию, проверяемый элемент содержания ко-

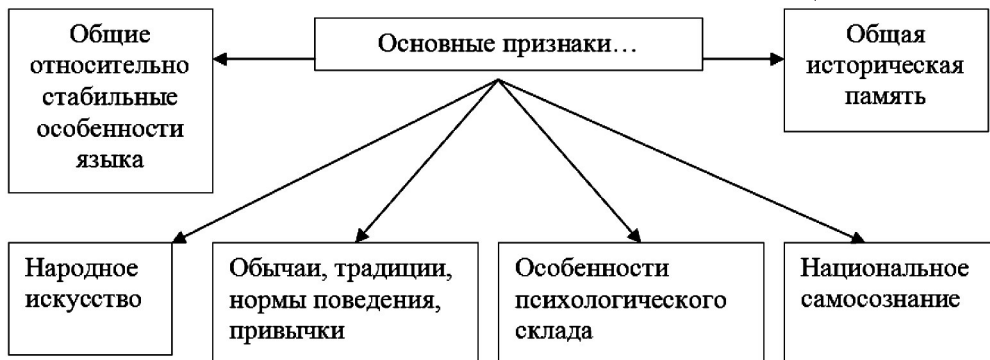
торого обозначен следующим образом: «Различное содержание в разных вариантах: выявление структурных элементов с помощью схем и таблиц». Здесь, согласно спецификации, процент выполнения демонстрирует уровень умений выявлять структурные элементы с помощью таблиц и схем, но не знаний в какой-либо области курса, т.к. эти задания построены на материале из различных его разделов. Задания выглядят следующим образом³¹:

Запишите пропущенное слово в таблице.

Признак выделения	Социальные группы
...	молодёжь
Отношение к власти	политическая элита

Ответ _____

Запишите слово, пропущенное в схеме.



Совершенно очевидно, что если знать ответ, записать его не составляет труда. Вот что действительно определяет успешность этой работы, так это

форма его записи. Отсутствие готовых альтернатив и элементов подсказки, присущих закрытым вопросам, ведёт к снижению результата. Именно от-

Методология

30

Анализ результатов Единого государственного экзамена по предметам по выбору на территории Московской области в 2011 году, <http://rcoi.net/statistika-i-analitika.html>

31

Примеры приведены из открытого банка заданий ФИПИ.

ПЕД
измерения

32

Если задание В1 действительно дифференцирует учащихся по умению выявлять структурные элементы с помощью схем и таблиц (ведь это вопрос открытого типа), то что же тогда измеряют все задания части А, напичканные подсказками?

33

Анализ результатов Единого государственного экзамена по предметам по выбору на территории Московской области в 2011 году, <http://rcoi.net/statistika-i-analitika.html>.

34

Итоговый аналитический отчёт о результатах проведения ЕГЭ в 2010 году (май-июнь 2010 года), http://www.fipi.ru/binaries/1069/OB_11_2010.pdf.

крытый характер вопроса (не смотря на то, что он относится к базовому уровню сложности) приводит к самым низким показателям. Так, в 2010 году в целом по стране с этим заданием справились только 51,2% участников. Лишь с частью С и заданиями В5 и В6, представляющими собой вопросы с выбором множественных альтернатив, школьники справляются хуже. А поскольку форма ответа определяет успешность выполнения задания В1, было бы ошибкой не только выдавать результат за оценку умения работать с информацией, представленной в различных знаковых системах (схема, таблица), но даже высчитывать среднее значение, ибо построено оно на материале из разных разделов курса. *В сопоставлении решаемостью иных заданий средний балл и процент выполнения задания В1 вообще лишены какого-либо содержательного смысла*³².

Поэтому ошибочными являются выводы, например, отчёта по Московской области: «Как показывает анализ результатов ЕГЭ по общественности, выпускники слабо владеют умениями... анализировать и классифицировать социальную информацию, представленную в различных знаковых системах, в том числе статистическую или графическую (напр. тексте, схеме, таблице, диаграм-

ме)... Это особенно наглядно проявилось при выполнении заданий... В1»³³.

То, что КИМы иногда могут не соответствовать спецификации, косвенно признают и специалисты ФИПИ: «Задания, предполагающие умения классифицировать понятия, явления, социальные объекты путем установления соответствия терминов и их определений, понятий и их признаков, выпускники выполнили в этом году менее успешно, чем в предыдущие годы (25–50%). Возможно, это связано, прежде всего, с тем, что *задания этого типа были построены на новом содержании* (выделено нами. — **Авт.**), которое оказалось сложным для усвоения (организационно-правовые формы предпринимательства)»³⁴. Иными словами, классифицировать школьники стали хуже не из-за снижения умений классифицировать, а из-за смены материала классификации. Хотя по спецификации оценивались именно умения.

3. Игнорирование формы вопроса. Как-то неловко говорить банальности о том, что результативность выполнения задания в значительной мере зависит от формы записи ответа. Между тем зачастую без комментариев, беззубо и беспомощно аналитические отчёты, в том числе ФИПИ, фиксируют факты успешного решения

задания повышенного уровня в сравнении с базовым. Так, например, ФИПИ в отчёте по общественнознанию сообщает: «В отдельных случаях результат выполнения заданий *базового уровня с кратким ответом ниже, чем результат выполнения заданий повышенного уровня с выбором ответа* (выделено нами. — *Авт.*). Так, задание В2 второй части работы по теме «Системное строение общества: элементы и подсистемы» (средний процент выполнения 35%) выполнялось правильно почти в два раза реже, чем задание повышенного уровня А3 (средний процент выполнения 58%)».³⁵ При этом работники ФИПИ не объясняют полученный результат. А ведь причина очень проста. Открытые вопросы всегда представляют наибольшую трудность — они не содержат подсказки.

Примеры влияния формы записи ответов на результат можно найти повсюду. В региональных отчётах и справках ФИПИ часто приводятся графики выполнения задач участниками с разной подготовкой. Решаемость заданий части А среди слабых школьников никогда не опускается ниже 10–15%. Теоретически этот показатель должен находиться в районе 25%. На практике альтернативы-провокации снижают его ниже вероятностной отметки. В качестве иллюстрации

приведем график ФИПИ по физике (рис. 5), который показывает, что в сравнении с другими заданиями задачу А23 менее всех провалили именно «двоечники» (задание отмечено стрелкой). Почему? Просто хуже выполнить его уже нельзя: форма записи ответа не позволяет это сделать.

Приведем ещё пример. Если сравнивать разницу решаемости заданий базового и повышенного уровня по разделам «Человек и общество» (А5 и А3), «Познание и духовная жизнь» (А4 и А6), «Социальные отношения» (А14 и А15), «Политика (А18 и А19), «Право» (А23 и А24) в курсе общественнознания, то выяснится что минимальные ее значения наблюдаются среди представителей самой слабой и самой сильной группы участников. И если в сильной группе это является свидетельством способности справиться с заданиями разной сложности, то в слабой — равной вероятности угадать верный ответ любого вопроса с 4 альтернативами. Только этим можно объяснить тот забавный факт, что по разделу «Познание и духовная жизнь» двоечники «выполнили» задание повышенного уровня сложности (А6) лучше, чем базового (А4)³⁶.

ФИПИ, конечно, понимает, что закрытые ответы угадываются, но признает это уклончи-

35

Там же.

36

Там же.

ПЕД
измерения

37

Аналитический отчёт
ФИПИ 2007 г.,
[http://www.fipi.ru/view/
sections/138/docs/
318.html](http://www.fipi.ru/view/sections/138/docs/318.html).

38

Не хотим обвинять в формализме всех подряд. Многие из просмотренных нами работ вызывают искреннее уважение. В них дотошно, вплоть до помарок школьников, разбираются решения и ошибки. Остро чувствуется желание специалистов разобраться, а не отписаться. Есть работы, где верно отмечается связь результативности учащихся с качеством КИМов.

39

Анализ результатов государственной итоговой аттестации выпускников 2010 года общеобразовательных учреждений Томской области,
[http://coko.tomsk.ru/
files/reports/
analit-ege-2010.pdf](http://coko.tomsk.ru/files/reports/analit-ege-2010.pdf).

во: «...требование стандарта можно считать достигнутым..., если 65% и более учащихся... успешно выполнили задание с выбором ответа, проверявшее достижение данного требования. Для заданий с кратким или развёрнутым ответом установлена граница 50%»³⁷. Откуда такая разница? Ответ очевиден.

4. Формальный подход к анализу, игнорирование содержания заданий и альтернатив ответов. Главным недостатком аналитических работ является формальный подход к изучению результатов ЕГЭ³⁸. В них присутствует какая-то странная уверенность в том, что КИМы подобно деталям из-под штампа хороши и одинаковы, а коды спецификации отражают все их особенности. Между тем это не так.

В КИМах встречаются *задания, не имеющие решений*. Например, в 2010 году в задании А5 по обществознанию составители не видят разницы между абсолютным числом и процентом (материал за 5 класс). В задании приводится динамика охвата молодежи России программами профессионального образования в процентах, а предложенные альтернативы ответа оперируют понятием «число». Для того, чтобы найти верный ответ, школьникам следует не читать предлагаемый график, а заставить за вычисления. При этом

знать, как менялась численность населения страны в возрасте от 15 до 34 лет, чего в условии нет.

Встречаются задания, *имеющие несколько верных ответов*. Так в задании А3 по химии, где специалисты из Томска зафиксировали низкий процент решаемости, в качестве альтернатив было предложено два верных ответа: «...вопрос о водородной связи: Водородная связь не характерна для: 1) H₂O 2) CH₄ 3) HCl 4) CH₃OH.

Причиной неудовлетворительных ответов является наличие двух правильных ответов 2) и 3), так как и между молекулами метана и хлороводорода не возникают водородные связи»³⁹.

Встречаются *невразумительные задания*. Пример приводится в отчёте томских коллег. Составители его, фиксируя падение решаемости одного из заданий части С по обществознанию, пишут, что школьникам требовалось «привести три аргумента в подтверждение авторского вывода о том, что «демократия является постоянным усилием по созиданию сферы гражданского состояния и обеспечению того, чтобы частные убеждения и групповая принадлежность могли сойтись и соединиться в ней в атмосфере взаимного уважения». Отвечающие, как правило, про-

сто не поняли смысла задания, что и определило уровень ответов»⁴⁰. Немудрено. Кто же разберет такую галиматью?⁴¹ Интересно, насколько единодушно было бы решение этого задания у самих экспертов? Дифференцирующая способность задания — нулевая.

Встречаются задания с *обратной дифференцирующей способностью*. Показательной являлась ситуация с задачей А9 по физике на ЕГЭ-2010, обозначенной спецификацией как задание по молекулярно-кинетической теории. По результатам испытаний оказалось, что его одинаково хорошо (плохо) решали как слабые учащиеся, так и отличники, что было вызвано неверным подбором альтернатив. В своем отчёте ФИПИ приводит диаграмму решаемости задач учащимися с

различным уровнем подготовки (рис. 5)⁴², но аномальные результаты задания А9 (обведены кружком) обходит молчанием. Между тем в 20% поступивших в Омск вариантов школьники с минимальным уровнем подготовки решили это задание лучше отличников. Таким образом, дифференцирующая способность её — отрицательная. Причина провала — применение альтернатив-провокаций. Составители настолько намудрили с альтернативами ответа, что запутали сильных учащихся. Слабые выпускники, полагаясь лишь на собственный бытовой опыт (речь шла о закипании воды в кастрюле), отмечали самый простой ответ, который и оказался верным.

О некорректности КИМов сообщают специалисты Санкт-

Методология

40

Там же.

41

ФИПИ, конечно, может спрятаться за авторитет французского социолога, автора этих строк. Но вырванная из контекста и вне стили первоисточника, фраза А.Турена в качестве задания явно не годится. ФИПИ следовало бы здесь подыскать что-то более внятное, пусть и не такого маститого автора.

42

Итоговый аналитический отчёт о результатах проведения ЕГЭ в 2010 году, http://www.fipi.ru/binaries/1070/FI_11_2010.pdf.

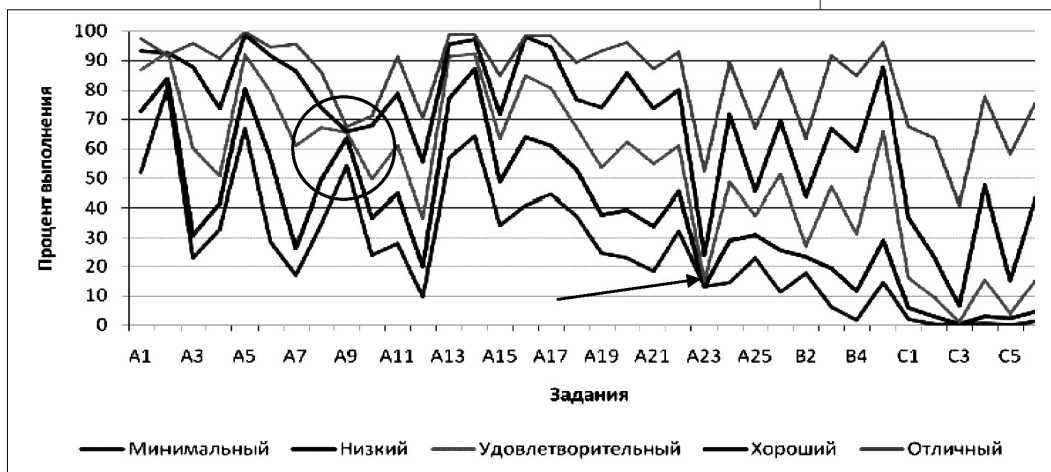


Рис. 5. График из отчёта ФИПИ по физике в 2010 году

ПЕД
измерения

43

Основные итоги Единого государственного экзамена по немецкому языку в 2012 году в Санкт-Петербурге. Аналитический отчёт предметной комиссии, <http://podelise.ru/docs/index-2122.html>

44

Анализ результатов государственной итоговой аттестации выпускников 2010 года общеобразовательных учреждений Томской области, <http://coko.tomsk.ru/files/reports/analit-ege-2010.pdf>.

45

Единый государственный экзамен в 2010 году, Новосибирск, <http://nscm.ru/index.php?p=0&year=2010>.

46

Итоговый аналитический отчёт о результатах проведения ЕГЭ в 2010 году (май-июнь 2010 года), http://www.fipi.ru/binaries/1085/1_gazdel_11_21092010.pdf.

47

Нет в учебниках? Ну что ж? Тем хуже для учебников. Вот только детям что делать? — Правильно! Приобретать пособия ФИПИ...

Петербурга, анализировавшие результаты по немецкому языку⁴³, историки Томска⁴⁴, обществоведы Новосибирска⁴⁵ и др.

Все это лишний раз подтверждает, что **успех решаемости заданий в значительной мере определяется не столько сложностью изучаемого материала, сколько формулировкой вопроса, формой ответа и подбором альтернатив**, т.е. качеством инструмента проверки. Результат уже заложен в КИМах альтернативами-провокациями и альтернативами-подсказками. **Поэтому обязательным этапом изучения результатов испытаний на местах должен стать анализ используемых в данном регионе контрольно-измерительных материалов.**

Не нужно также забывать, что содержание КИМов не всегда соответствует содержанию учебников. В этом признаются сами составители КИМов, в частности, по общественнознанию. Оказывается «...не все содержательные линии, представленные в ЕГЭ, раскрыты в учебниках с необходимой степенью полноты...»⁴⁶. Так ФИПИ осторожно сообщает, что берётся проверять материал, которого в учебниках нет⁴⁷.

Томские аналитики приводят пример применения задачи С4 по физике, «связанной с малыми колебаниями одного заряженного тела в поле другого.

В критериях предложено решение, в котором используется малая величина второго порядка малости, которой пренебрегают, так как она не вносит существенной поправки в искомую величину. Такие задачи объясняются только в вузе, а в школьной программе ни по математике, ни по физике подобное не проходят. Естественно, что это задание даже никто и не приступал решать»⁴⁸.

Показательным примером, характеризующим соответствие КИМов школьной программе, может послужить история, произошедшая в 2007 году на пробном экзамене по математике. Оценку в 100 баллов в Омске не получил никто, в том числе абсолютный чемпион Международной математической олимпиады школьников в Ханое⁴⁹, выпускник омского лицея №66 Константин Матвеев. На штатном экзамене максимальный балл получили два омика. Один из них — чемпион мира. Какие измерения ведёт ФИПИ?..

Поверхностный анализ ведёт не только к ошибочным выводам, но и не позволяет своевременно корректировать инструмент оценки. Так, например, на протяжении нескольких лет специалисты ФИПИ игнорировали существенные расхождения в результатах испытаний по иностранным языкам. Разница среднего

балла по английскому и немецкому языкам в 2008 г. составила 8,5, в 2009 г. — 15,1, в 2010 г. — 14, 2011 г. — 10,9. В 2012 году она вдруг резко сократилась до 3,2 балла. Как получилось, что в 2012 году средний тестовый балл по немецкому языку взлетел на 8 пунктов, а по английскому вырос только на 0,3?

Ответ мы найдем в КИМах. Если посмотреть, где более всего отстают «немцы» от «англичан», то окажется, что это задание В2 базового уровня, перегруженное призовыми баллами (за его выполнение предусмотрено 7 баллов). Именно из-за высоких баллов за выполнение задания В2 и сравнительной лёгкости его в КИМах по ан-

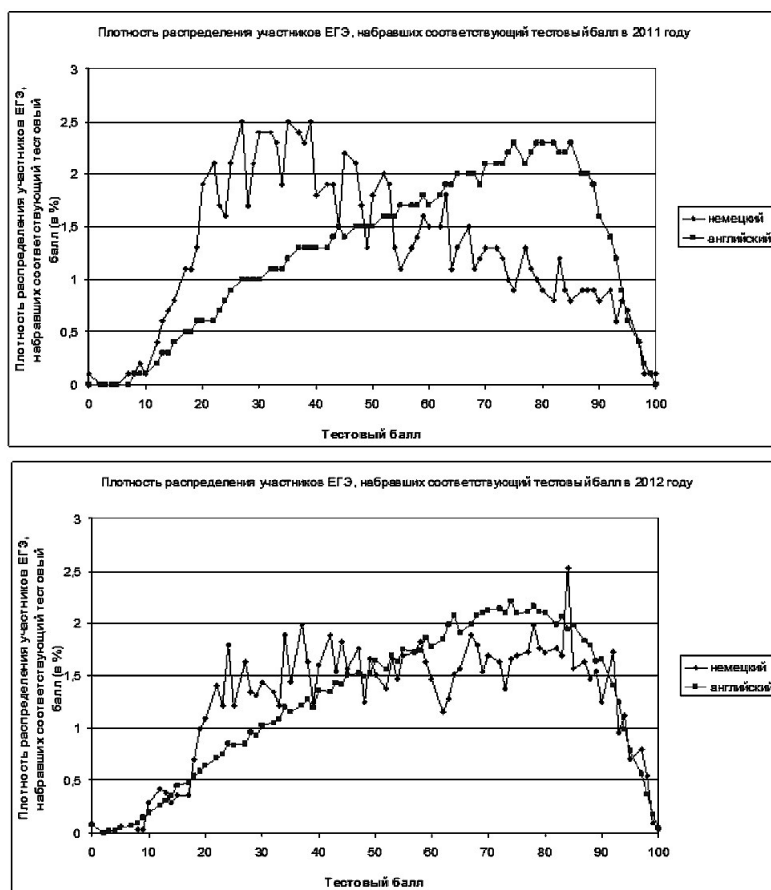


Рис. 6. Распределение учащихся по тестовому баллу в 2011 и 2012 гг. (немецкий и английский языки)

Методология

Показательным примером коммерческого интереса ФИПИ в проекте ЕГЭ являются размещенные на его сайте призывы не покупать литературу конкурентов: «Уважаемые коллеги! В настоящее время в ряде публикаций появилась информация об использовании в учебных пособиях для подготовки к единому государственному экзамену «реальных заданий ЕГЭ 2010 года». Считаю необходимым разъяснить, что ни одно учебное пособие не содержит заданий, которые будут использованы в экзаменационных материалах ЕГЭ 2010 года. С другой стороны, учебные пособия, подготовленные авторскими коллективами ФИПИ в 2009 г., включают как реальные задания, использовавшиеся в ЕГЭ предыдущих лет и размещенные в Открытом сегменте федерального банка тестовых материалов, так и задания в формате ЕГЭ, разработанные в строгом соответствии с демонстрационными вариантами КИМ ЕГЭ 2010 года. Директор ФИПИ А.Г. Ершов. Директор издательства «Экзамен» Д.В. Яновский. Директор издательства «Астрель» Ю.В. Дейкало.

И еще: «ФИПИ рекомендует к использованию в качестве учебных

ПЕД
измерения

пособий для подготовки к ЕГЭ и ГИА как издания с грифом «Допущено ФИПИ к использованию в учебном процессе в образовательных учреждениях» на титульном листе, так и издания, на обложке которых изображены логотип или наименование ФИПИ или знак печати «Разработано ФИПИ», а на обороте титульного листа — знак охраны авторских прав (копирайт) ФИПИ.

48

Анализ результатов государственной итоговой аттестации выпускников 2010 года общеобразовательных учреждений Томской области, <http://coko.tomsk.ru/files/reports/analit-ege-2010.pdf>.

49

Международная математическая олимпиада (англ. IMO, International Mathematical Olympiad) — старейшая из международных научных олимпиад школьников. Одними из самых сильных являются команды Китая, России, Вьетнам, США и Болгарии. В 2007 году впервые за несколько лет Россия заняла в неофициальном рейтинге первое место, опередив на одну позицию Китай, а ученик омского лицея № 66

лийскому и французскому языкам распределение результатов всегда было смещено вправо (типичный график — 2011 год), чего нет в немецком языке, где задание В2 сложнее (см. рис. 6), так как здесь, по свидетельству специалистов из Санкт-Петербурга, «тематика экзаменационных текстов и заданий зачастую не соответствует социально-личностному опыту выпускников средней школы»⁵⁰.

Стоило лишь поправить задание В2⁵¹ и произошло чудо, которое, судя по реакции ФИПИ, требовало неприятных объяснений. Видимо поэтому ФИПИ предпочел не заметить его. В своем отчёте он вскользь объяснил некоторые успехи «немцев» 10%-м ростом выполнения задания С2, относящегося к разделу «Письмо»⁵². Между тем, из графиков ФИПИ (рис. 7) следует, что основной вклад в повышение результата внес раздел «Чтение», к которому и относится задание В2. Именно здесь учащиеся со средней подготовкой (с первичным баллом от 20 до 60) чаще всего набирали баллы.

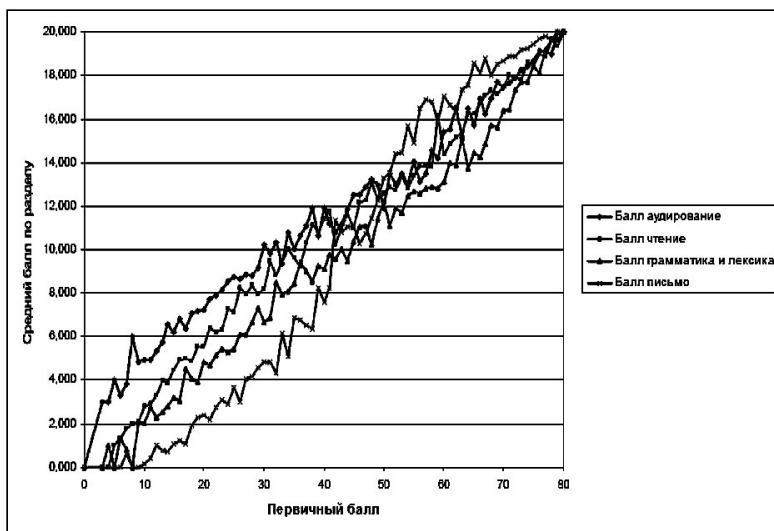
Странно, что проверяя умение школьников читать информацию, представленную в различных знаковых схемах (график, таблица), ФИПИ не может прочесть графики в собственных отчётах. Неравноценные задания по иностранным

языкам не применялись бы так долго, проведи ФИПИ своевременный анализ указанных расхождений. Свои ошибки институт исправил лишь после того, как о них уже заговорили с мест, в частности питерские эксперты⁵³.

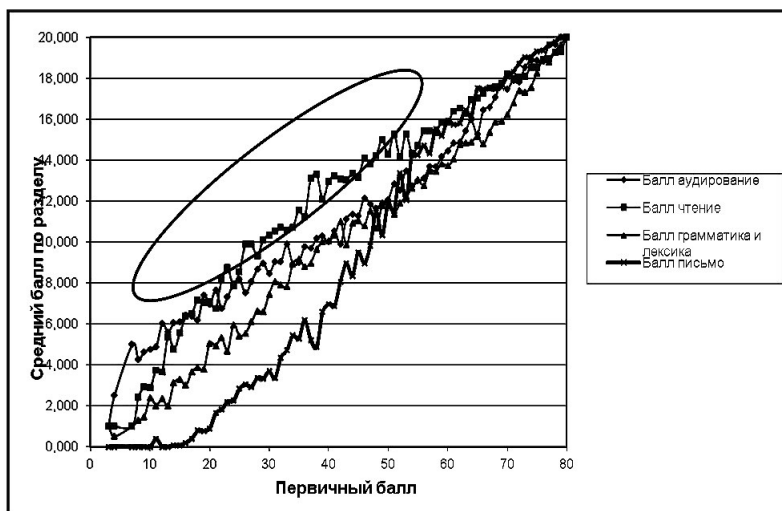
Сопутствующей проблемой является уровень специальной подготовки работников, анализирующих результаты ЕГЭ.

После окончания экзаменов РЦОИ всей страны подводят итоги. В основном результаты испытаний сводятся в статистические сборники⁵⁴. Аналитику сотрудники РЦОИ почти не пишут. Пишут её преподаватели и методисты местных институтов повышения квалификации. Однако для РЦОИ это не оправдание. Экзамен закончен, а понимания того, что сделано, нет. Чем же занят персонал после экзамена?

Что касается Омской области, то здесь даже статистику грамотно свести не могут. Результаты по региону без комментариев и объяснений из года в год сваливались в огромные и пустые таблицы, представляя читателям самимковыряться в цифрах, зачастую сомнительных и несопоставимых. Так, омские сборники по математике в течение нескольких лет приводили данные о решаемости задач по спецификации 2004 года, в то время как



2011 год



2012 год

Рис. 7. Графики ФИПИ «Зависимость общего первичного балла от результата выполнения разных разделов экзамена (немецкий язык)»

Методология

Константин Матвеев стал абсолютным победителем.

50

Анализ результатов ЕГЭ–2011 и методические рекомендации при подготовке ЕГЭ–2012. Санкт-Петербург, <http://podelise.ru/docs/index-2122.html>.

51

Отчёт Регионального центра оценки качества образования (Санкт-Петербург) за 2012 год свидетельствует: «Следует отметить, что тематика текстов и заданий соответствует социальному опыту учащихся и не создаёт дополнительные трудности в понимании, не требует дополнительных затрат времени».

52

Итоговый аналитический отчёт о результатах ЕГЭ, Москва, <http://fipi.ru/view/sections/138/docs/624.html>.

53

Кстати, после таких исправлений средний балл по немецкому языку в Санкт-Петербурге подскочил на 13 пунктов. Это самый большой рост среди территорий РФ.

ПЕД
измерения**54**

Как правило, статистические сборники по итогам ЕГЭ выглядят очень внушительно: огромные таблицы, масса цифр, «широжки» и «столбики»... Казалось бы, столько труда... Откроем секрет: по опыту авторов вся статистика и ее графическое оформление, встречающаяся в региональных отчётах, легко делаются в таблицах Excel одним человеком за два-три дня.

55

Журнал «Фундаментальные исследования» №5 за 2008 г. (http://www.rae.ru/fs/?section=content&op=show_article&article_id=7780873)

56

Эффективность РИАЦ как организации видна в сопоставлении с затратами: расходы областного бюджета только на оплату труда работников центра составляют свыше 9 млн. руб. в год, что превышает расходы на зарплату учителям любой общеобразовательной школы региона. Учитывая данный факт, трудно утверждать, что ЕГЭ это система, которая позволяет экономить средства на организации вступительных испытаний (по крайней мере, в Омске).

контрольно-измерительные материалы постоянно менялись. «Творения» РИАЦ вводили в заблуждение не только омских учителей. Из них ошибки стали переключиваться и в профессиональные издания⁵⁵. Публикация недостоверных сведений по Омской области стала возможной по причине отсутствия анализа результатов ЕГЭ. Сотрудники омского аналитического центра не в силах не то что объяснить, но даже просто рассказать о них. Причины — низкая квалификация, неумение работать со специализированными статистическими и офисными программами, отсутствие навыков аналитики и, главное, интереса к экзамену.

В 2011 году РИАЦ, имея в распоряжении пятнадцать аналитиков, семь методистов и уйму времени, по итогам ЕГЭ не написал ни строчки⁵⁶. В 2012 году анализ по Омской области уместился в страничку безграмотного текста, обильно сдобренного рисунками и таблицами для увеличения объёма. К примеру, работники РИАЦ пришли к странному выводу об улучшении в 2012 году качества подготовки омских школьников по литературе, в то время как средний балл по предмету снизился, а также увеличилась доля неудовлетворительных оценок. Практически весь сборник составляют таб-

лицы. Допускаем, что составители приложили массу усилий. Но чтобы оценить их, скажем, что для подготовки аналогичной статистики по региону за 2011 год автору (С.К.) понадобилось всего четыре часа.

Ошибки специалистов проявляются не только в интерпретации результатов ЕГЭ. Можно только посочувствовать, например, томскому гуманитарии, видимо не знающему базовых понятий статистики. Вместо того чтобы просто указать, что распределение имеет нормальный вид, он взялся его описывать: «Похоже, что общая успеваемость учащихся по общественнознанию стремится к усреднению, сужаясь по обоим краям: с одной стороны, уменьшается количество получающих низкие баллы, но в то же время сокращается количество набирающих наиболее высокие баллы»⁵⁷.

Оригинальное ноу-хау демонстрируют омские специалисты по оценке качества образования. Они рассчитали средний балл по всем предметам, вычли из него прошлогодний результат и представили полученную разность как тенденцию состояния образования в регионе: «Общий средний балл по всем предметам в сравнении с 2011 годом снизился на 2,4 балла»⁵⁸. Остается загадкой, как рассчитывался этот «общий средний балл». Если как

среднее арифметическое, то получается, что наравне с результатами по русскому языку и математике в подсчётах использовался балл по французскому и испанскому языкам, где в испытаниях участвовало по 2 человека. Если показатель является частным от деления суммы всех баллов на число человеко-экзаменов, то итог будет определяться результатами двух самых массовых испытаний (математика и русский язык), выдавать который за балл по остальным предметам было бы неверным. При этом горе-аналитиков не интересуют ни изменения в составе участников и условиях проведения экзамена, ни изменения в КИМах.

Далее РИАЦ пишет: «В 2012 г., так же как и в 2011 г., 3–4 экзамена сдают более 70% участников ЕГЭ. Однако наибольший общий средний балл по всем выбранным предметам показали выпускники текущего года, сдававшие пять предметов»⁵⁹. Затем следует пассаж: «В 2011 г. в сельской местности по 8 из 13 предметов средний балл был выше, чем в населенных пунктах городского типа, а успеваемость по всем предметам была выше в селе. Таким образом, можно говорить, что по сравнению с 2011 г. результаты ЕГЭ в 2012 г. в населенных пунктах городского типа улучшились относительно

сельской местности». Уважаемые читатели! Вы что-нибудь поняли?

Невысокое качество публикуемых материалов по итогам ЕГЭ объясняется тем, что анализ проводят сотрудники, не имеющие не только специальной подготовки, но даже профильного образования. Сразу вспоминается известное изречение: «Об образовании, как и о футболе, может судить любой».

Подведём итоги

1. Несмотря на десятилетнюю практику, серьёзной проблемой ЕГЭ остается слабость аналитической работы. Последствия самые плачевные: неверная оценка уровня подготовки выпускников, бойкот экзаменов, неравные условия проведения испытаний (в частности среди учащихся, сдающих иностранные языки), неоправданные сложности при проверке работ и т.д. Отсутствие практики отправки региональных отчётов в ФИПИ, специалисты которого могли бы увидеть свои промахи и вовремя на них отреагировать, является причиной многих застарелых ошибок. А судя по унылому однообразию публикуемой ФИПИ аналитики, институт остро нуждается в свежем взгляде на экзамен со стороны.

Методология

57

Анализ результатов государственной итоговой аттестации выпускников 2010 года общеобразовательных учреждений Томской области, <http://coko.tomsk.ru/files/reports/analit-ege-2010.pdf>.

58

Сборник статистики результатов Единого государственного экзамена (ЕГЭ), государственной (итоговой) аттестации (ГИА) обучающихся, освоивших образовательные программы основного общего образования, с участием территориальной экзаменационной комиссии в Омской области в 2012 году, <http://ege55.ru/%d1%81%d1%82%d0%b0%d1%82%d0%b8%d1%81%d1%82%d0%b8%d0%ba%d0%b0/>

59

Да... В огороде бузина... Омские «аналитики» вряд ли сдали бы ЕГЭ по русскому языку. Даже название их опуса говорит о том же: «Сборник статистики результатов Единого государственного экзамена (ЕГЭ), государственной (итоговой) аттестации (ГИА) обучающихся, освоивших образовательные программы основного общего образования, с участием тер

ПЕД
измерения

риториальной экзаменационной комиссии в Омской области в 2012 году». Удивительно, но под грифом министерства ОБРАЗОВАНИЯ в Омске рождаются официальные документы, безграмотные не только по содержанию, но имеющие орфографические, синтаксические или стилистические ошибки почти в каждом абзаце текста объёмом в одну страничку.

К не менее печальным результатам приводит отсутствие анализа итогов ЕГЭ на местах. Именно по этой причине, в Омской области, например, в первые пять лет процветал саботаж обязательных экзаменов.

2. Слабой стороной экзамена остается зависимость конструкции КИМов от технологии проверки работ: до сих пор широко распространены вопросы-угадайки и из контроля исчезают разделы математики, где невозможно дать ответ целым числом или конечной десятичной дробью. Делается это все ради того, чтобы работу можно было легко проверить. Но как? Силами программиста, методиста, бухгалтера и секретаря, на корявых бланках и немощной программой распознавания? Не слишком ли высокая цена? Ведь сложности верификации связаны не с большими объёмами и сжатыми сроками исполнения работ, а с организацией труда и несовершенной программой распознавания. Пока штатные сотрудники РЦОИ будут получать деньги за верификацию, они будут стремиться освоить их собственными силами, следствием чего неизбежно станут штурмовщина, халтура и срывы сроков. Пока верификация будет вестись с помощью несовершенных программ, распознающих не смысл ответа, а буквы, его составляющие, проверка

вопросов, где требуется записать что-либо иное кроме крестика, всегда будет требовать огромных трудозатрат.

3. Применяемая в настоящее время система перевода первичных баллов в тестовые крайне несправедлива. Всемерно награждая за мизерные успехи, достигаемые зачастую лишь гаданием ответов, она предельно наказывает учащихся за единичные ошибки. Это вносит значительные искажения в измерения учебных достижений: если первичные баллы дифференцируют школьников по уровню подготовки, то тестовые, наоборот, нивелируют их результаты. Стоит ли цепляться за непропорциональную шкалу начисления баллов, учитывая большой разброс вариантов по трудности? Межгодовая разница в сложности заданий, которую якобы сглаживает нелинейная шкала, вряд ли превысит этот разброс. Конечно специалистам, владеющим полной информацией, следует еще раз все посчитать. Но с несуразной системой перевода баллов в любом случае нужно что-то делать.

4. Одной из причин невысокого качества контрольно-измерительных материалов является стремление организаторов быстрее нарастить действующий банк заданий. Между тем применение экспериментально не испытанных, сырых

и сомнительных задач вредит измерениям больше, чем использование пусть ограниченного, но проверенного их набора. Зачастую достижения выпускников зависят не столько от сложности материала, сколько от корректности инструмента контроля. Результаты ЕГЭ уже во многом запрограммированы технологией проверки работ и корректностью КИМов, внешняя экспертиза которых крайне затруднена их излишней закрытостью. По итогам испытаний специалистам оставлена возможность обсуждать лишь убогий демовариант, не изменяемый в некоторых дисциплинах годами. Закрытость КИМов после окончания испытаний не позволяет своевременно выявлять и устранять ошибки и консервирует их, что оборачивается против экзамена и его участников.

5. Проблемой организации ЕГЭ является низкая эффективность центров обработки информации как следствие сезонности работ. Передача им задач по сбору ведомственной статистики проблему не решает. Оптимальный вариант их функционирования — это деятельность в качестве структурных подразделений институтов повышения квалификации работников образования. В рамках института легче решаются вопросы загруженности работников РЦОИ в межэкзамена-

ционный период. Всегда есть возможность привлечь преподавателей в период экзамена и тем самым решить проблему нехватки людей. Одновременно специалисты профильных кафедр будут обладать полной информацией об экзамене и могут давать квалифицированный анализ результатов испытаний.

Предложения

Что можно предложить в условиях, когда аттестация выпускников школ в форме ЕГЭ законодательно закреплена? — Только лишь меры, направленные на её совершенствование:

1. Полностью открыть федеральный банк тестовых заданий. Контрольно-измерительные материалы накануне экзамена должны формироваться из него случайным образом, отдельно для каждого региона.
2. Перейти на линейную шкалу перевода первичных баллов в тестовые.
3. Усовершенствовать инструментарий: убрать лишние клетки в бланке ответов и разрядить остальные, отдельно печатать бланки для иностранных языков (три альтернативы).
4. Изменить способ верификации: перейти от проверки знаков к проверке смысла ответа. Необходимы специализиро-

ПЕД
измерения

ванные «интеллектуальные» программы распознавания ответов отдельно для каждой из дисциплин, использующих свою символику записи. По мере решения этой проблемы следует последовательно отказываться от закрытых вопросов. Уже сейчас это можно сделать в КИМах по физике и информатике.

5. Регламентировать верификацию:

- станции верификации не должны иметь иного программного обеспечения кроме программы верификации, должны быть отключены от Интернета и мультимедийных средств;
- исключить участие в проверке работ случайных людей: список привлекаемых к верификации лиц должен утверждаться приказом руководителя областного органа управления образованием;
- работа должна быть организована по сменам (в две, три, четыре), но не более 6 часов в смену, сотрудники РЦОИ должны выполнять обязанности старшего верификатора в сменах, также не более 6 часов;
- верификацию проводить в отдельно отведённых для этого помещениях;
- запретить штатным работникам РЦОИ получать деньги за верификацию (верифицируют

они в рабочее время, за что и получают зарплату).

6. Перевести РЦОИ в подчинение местных институтов повышения квалификации работников образования или аналогичных им структур. Ввести в обязательный функционал анализ результатов ЕГЭ (а не простой свод статистики). Аналитический отчёт по региону выставлять на сайте органа управления образованием субъекта Федерации за подписью его руководителя, а также направлять в ФИПИ. Возможно, ФИПИ следует организовать конкурс отчётов, которые позволили бы тиражировать по стране опыт успешных регионов.

7. ФИПИ обеспечить методически и информационно проведение анализа результатов ЕГЭ на местах. Для корректных межрегиональных сопоставлений разработать типологию регионов со сходными условиями функционирования образовательных учреждений. Ежегодно публиковать итоги испытаний по субъектам Федерации и (или) федеральным округам с необходимыми пояснениями о пределах допустимости сравнения результатов. Разработать для регионов алгоритм изучения итогов ЕГЭ. Провести обучение специалистов на местах приёмам статистического анализа.