



А.П. Иванов

КРИТИЧЕСКИЕ ЗАМЕЧАНИЯ ПО ЕГЭ (МАТЕМАТИКА): КАЧЕСТВО ТЕСТОВ И ТЕХНОЛОГИЯ ПРОВЕДЕНИЯ ЭКЗАМЕНА

Аннотация

В статье подробно обсуждаются проблемы ЕГЭ по математике: технологии проведения и соответствие используемых КИМов общеизвестным основным характеристикам эффективных тестов (техническим, дидактическим и методологическим). Приводятся результаты исследований по определению главных параметров теста — формата и длины (количества заданий), а также времени выполнения. Предлагаются технологии проведения тестирования, позволяющие изнутри проследить всю процедуру путем хронометрирования выполнения отдельных заданий; эта неоднократно апробированная технология выгодно отличается отсутствием сбоев при сканировании бланков, является экономичной и позволяет замерять объективный уровень знаний учащихся.

В соответствии с концепцией Единого государственного экзамена по математике ставится две задачи: проверка обязательного уровня усвоения школьного курса алгебры и начал анализа и отбор наиболее подготовленных учащихся для последующего обучения в вузе.

Однако проведение эксперимента по реализации программы ЕГЭ вызывает неоднозначные оценки в среде научной и педагогической общественности России. Это вполне объяснимо, так как затрагиваются интересы значительных социальных групп: педагогов, руководителей органов образования, ректоров вузов, учащихся и их родителей. Накопленный опыт проведения эксперимента показал, что ЕГЭ чаще всего встречает противодействие по политическим причинам. Единственный способ разрешить их состоит в том, чтобы примирить стороны, заставив их уважительно выслушать друг друга и принять такую структуру тестового экзамена, которая бы в наибольшей степени удовлетворяла максимальному числу требований и пожеланий всех участников процесса.

Следует отметить, что в начале проведения ЕГЭ многие конкретные вопросы по структуре, содержанию, числу тестовых зада-



ний, продолжительности проведения экзамена, технологии обработки результатов не были достаточно обоснованы. В последние годы постепенно ликвидируются указанные недостатки, в качестве примера следует отметить глубину идеологического обоснования общих подходов к разработке контрольно-измерительных материалов, предлагаемых Г.С. Ковалевой [11].

Общеизвестны основные характеристики эффективных тестов [14; 15].

1. Технические:

— согласованность времени (длительности) тестирования с количеством заданий в тесте;

— качество оформления теста.

2. Дидактические:

— задания теста соответствуют предметной области;

— результат тестирования позволяет оценить уровень и структуру знаний как одного тестируемого, так и целой группы;

— уровень трудности заданий отвечает целевому назначению теста.

3. Методологические:

— надежность;

— валидность (пригодность, обоснованность теста для достижения поставленной цели);

— дискриминативность (различительная способность задачи, т.е. способность отделять испытуемых с высоким общим баллом по тесту от тех, кто получил низкий балл);

— объективность (различные исследователи при измерении одного и того же признака приходят к одному и тому же результату).

Каждое из требований вносит свой вклад в эффективность теста. Их необходимо учитывать, чтобы тесты были не только правильно сконструированы, но и соответствующим образом использовались. Все требования, предъявляемые тестам, не случайны, а связаны с повышением точности измерения.

Конечно, не все можно измерить тестом (например, способность осуществить какой-либо проект), но при этом тестовая форма — самая эффективная. Противники тестов обычно приводят также и такой аргумент, что при тестировании проигрывают люди медлительные, с плохой памятью, но с развитым проективным мышлением, интуицией, а в жизни эти качества часто компенсируют недостаток знаний. Этого можно избежать построением четко продуманных тестовых заданий. Достаточно предложить задания, где можно что-то решить несколькими методами, тогда такие ученики не пострадают.

У нас большой опыт проведения тестов. В Пермском государственном университете тестовая форма экзамена используется с 1993г., более того, в течение 10 лет проводился для всех желающих конкурс SuperTest, по результатам которого победители могли быть зачислены на механико-математический факультет и на спе-



циальность «Математические методы в экономике» экономического факультета. Опыт проведения такого конкурса был признан положительным и вполне оправдал себя. Очевидно, что при получении искаженных результатов SuperTest'a, проводившегося 10 лет, они сразу же проявились бы в процессе обучения студентов, и факультеты отказались бы от их проведения.

Предлагаемые на конкурсе тесты представляют собой набор 30 пятиальтернативных заданий, время выполнения тестов не превышает 70 минут, части типа В и С не использовались. В последние 5 лет аналогичные конкурсы проводились в Пермском филиале Государственного университета Высшая школа экономики и головном университете ВШЭ (г. Москва), опыт был также признан удачным. Следует отметить еще одно положительное последствие этих конкурсов: в г. Перми и Пермской области имеет место значительное повышение мотивации обучения математике, о чем свидетельствуют результаты Централизованного тестирования по математике I, математике II и ЕГЭ 2003–2004 гг. — они выше средних по России.

Большой практический опыт позволяет нам увидеть некоторые недостатки по структуре и технологии ЕГЭ, которые повторяются из года в год при проведении эксперимента в России. Перечислим некоторые из них.

1. Отсутствует научное обоснование выбора одного из главных параметров теста — длины (количества заданий) и времени выполнения всех заданий. Они явно завышены. Остановимся на этом недостатке подробнее.

Очевидно, что короткий тест не в состоянии охватить достаточно широкий диапазон изучаемого материала. Чем тест длиннее, тем он в большей степени обеспечит качественное измерение обученности.

Второму параметру — времени выполнения теста и оптимальному его выбору — практически не уделяется никакого внимания, и чаще всего это время назначается сверху без каких-либо научных обоснований, в результате чего оказывается завышенным, хотя очевидно, что главным фактором в измерении уровня знаний является не временной, а фактор «знание или незнание». При излишках времени открываются широкие возможности искажения объективных результатов тестирования: выполнил свой тест, начал помогать соседу; при «плотной» посадке и недостаточном количестве вариантов появляется возможность списывания; сдал ранее окончания тестирования полупустой бланк ответов, и недобросовестные члены комиссии получают прямую возможность фальсифицировать результаты; при времени более 60с80 минут каждый тестируемый имеет возможность выхода из аудитории (и даже неоднократно), и, наконец, чисто физиологически человек не в состоянии заниматься очень интенсивной умственной деятельностью более 60 минут.

В литературе [14] обсуждались проблемы выбора длины теста и времени выполнения, где указывается, что время в тестировании



является фактором, который определяет качество всего инструментария и качество получаемых в результате тестирования результатов.

В.С. Аванесов [1] называет время системообразующим фактором при разработке и использовании тестов. Действительно, одна из главных особенностей теста состоит в том, что он является инструментом быстрого и достаточно точного оценивания знаний групп учащихся, экономия времени становится естественной, особенно в массовых тестированиях в образовании.

В работе А.Н. Майорова [14] достаточно подробно обсуждается проблема фактора времени в тестировании. Указывается, что каждый тест имеет оптимальное время тестирования, уменьшение или превышение которого снижает качественные показатели теста. Время в тестировании многоаспектно, учет временного фактора требуется на нескольких этапах создания и использования теста.

В настоящее время нет четких методик определения оптимального времени по многим причинам. Здесь играют роль многие аспекты и в первую очередь психологические и психофизиологические [2; 6; 9].

В экспериментальной психологии и педагогике практически нет определенного ответа на вопрос об оптимальном времени и результативности умственной деятельности. Еще П.Ф. Каптерев [10] указывал, что вопрос об умственной работе учащегося есть один из важнейших вопросов экспериментальной дидактики. Умственная работа сопряжена с тратой сил и зависит от двух главных моментов, которые Каптерев назвал упражнением и утомлением. Утомляемость человека при выполнении конкретной работы тесно связана с тем, сколько он предварительно в ней «упражнялся». Работу, которой человек овладел вполне, он может выполнять почти автоматически; работа, трудная для кого-либо, в особенности мало соответствующая дарованиям, утомляет несравненно сильнее и быстрее.

А.Н. Майоров отмечает, что утомление характеризуется ухудшением продуктивности рабочих процессов, снижением темпа деятельности и ее качества, появлением характерных ошибок. Состояние утомления быстрее возникает при интенсивной или монотонной работе.

Время продуктивного (до момента утомления) выполнения учащимся тестовых заданий является ключевым при составлении теста.

Момент наступления утомления характеризуется появлением большого количества случайных ошибок, снижением скорости выполнения заданий. Интервал времени до момента наступления утомления вполне естественно считать оптимальным временем выполнения данного теста. Время наступления момента утомления зависит от многих причин. Это возраст тестируемых, уровень мотивации, монотонность выполняемой работы, индивидуальные особенности учащихся. Взаимосвязь мотивации и времени наступления утомления достаточно сложна. Слишком высокая мотивация, а также и слишком низкая приводят к уменьшению времени продуктивной работы.



Влияние монотонности можно значительно снизить за счет содержания теста, сделав его достаточно разнообразным и привлекательным для учащихся.

Разброс по времени наступления утомления в зависимости от индивидуальных особенностей учащихся может составлять от 20 до 100 минут рабочего времени [2]. На некоторые из перечисленных выше факторов можно повлиять при тестировании, но индивидуальные особенности, конечно же, учесть невозможно. Обязательным условием тестирования является устойчивая и положительная мотивация учащихся, которая достигается интересным содержанием теста и созданием благоприятных условий тестирования. Следует иметь в виду, что даже личность педагога, проводящего тестирование, его манеры и умение сформировать позитивную установку на выполнение тестовых заданий, отсутствие высокомерия, фамильярности, грубости могут оказать значительное влияние на формирование положительной мотивации и настроить на успех.

Все перечисленные выше факторы оценивают выбор времени тестирования только лишь качественно. Остановимся теперь на количественных параметрах выбора оптимального времени для тестирования.

Конечно, время тестирования можно выяснить при различных формах апробации готового теста, однако имеются и указания на ориентировочное время у некоторых авторов. Например, Н. Гронлунд [4] указывает, что в начальной школе тестирование не должно занимать более 30 минут, так как в этот промежуток времени внимание поддерживается мотивацией. Для средних школ и колледжей время может составлять 40–50–60 минут. Он отмечает: «Большинство тестов рассчитаны на 40–50 минут, потому что это продолжительность типичного урока». Реальные сроки уточняются в процессе апробации, где следует предусмотреть некоторый запас времени, который позволит оценить верхнюю границу приемлемого времени выполнения.

В литературе довольно часто дается практический совет — рекомендация А.С. Отиса: время теста должно быть таким, чтобы только 5% тестируемых могли справиться со всеми заданиями.

Автором данных замечаний также проводилось множество исследований по определению времени тестирования, результаты их достаточно подробно изложены в монографии [7]. Ниже приведем некоторые результаты этих исследований.

В.К. Гайда и А.П. Захаров [3] рекомендуют для определения времени тестирования дать испытуемым указание отмечать то задание теста, которое они выполняют в различные моменты времени в процессе тестирования. Но как указывает А.Н. Майоров, опыт апробации тестов свидетельствует о нерациональности этого пути: учащиеся отвлекаются или забывают отметить задание, даже для взрослых тестирующихся эта технология малопримемлема, а при работе с детьми лучше поручить фиксацию выполнения заданий наблюдателю за 5–6 учащимися, отмечая время на специальных бланках.



Автором рекомендуется следующая технология: оптимальное время определяется путем хронометрирования выполнения тестовых заданий на специальном бланке, вместо обычно используемого бланка при ручном тестировании.

Весь период выполнения теста в 80 минут (период может быть и другим, например 60 минут) был разбит на следующие временные интервалы (в минутах): [0;20], [20;30], [30;40], [40;50], [50;60], [60;80]. Длина первого промежутка (20 мин) в два раза превышает остальные промежутки для того, чтобы:

1) иметь резерв времени для включения в активную работу тестируемого;

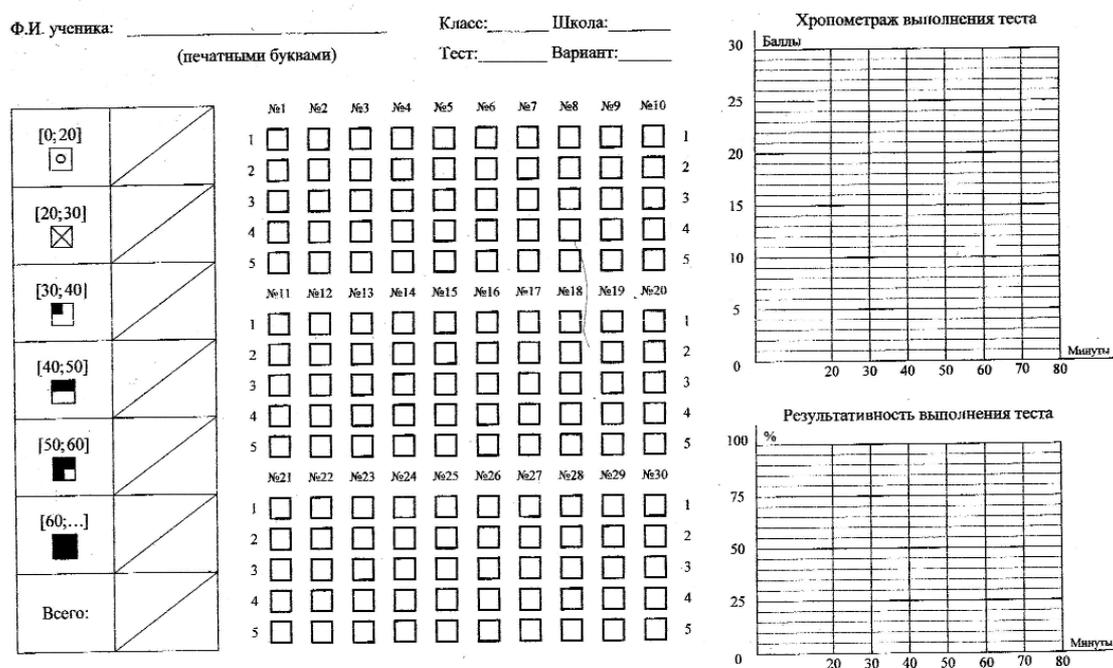
2) не отвлекать ученика в самое интенсивное время умственной деятельности, когда нет никаких признаков утомляемости и работа априори является максимально продуктивной.

Условные обозначения:

- на первом промежутке ответы фиксируются значком ;
- на промежутке $t \in [20; 30]$ — ;
- на промежутке $t \in [30; 40]$ — ;
- на промежутке $t \in [40; 50]$ — ;
- на промежутке $t \in [50; 60]$ — ;
- на промежутке $t \in [60; 80]$ — .

Время перехода от одного способа отметки ответов в бланке к другому способу (с указанием типа штриховки) объявлялось пре-

Рис. 1 Хронометраж выполнения теста (чистый бланк)





подавателем. Следует отметить, что даже учащиеся 4-х классов хорошо воспринимали эту процедуру хронометрирования выполнения тестовых заданий. Затем проводилась ручная обработка заполненных бланков (рис. 1), где значком «**V**» отмечены верно решенные задания. В результате в системе координат минуты-баллы получаются две линии: верхняя указывает на все указанные ответы, а нижняя — только верные ответы.

Слева во второй колонке бланка (рис. 1) проставлялось количество всех отмеченных на данном интервале ответов (в знаменателе) и верно решенных в этом временном интервале тестовых заданий в числителе. Полученную информацию очень удобно изобразить графически в виде двух ломаных линий в системе координат (минуты — по оси абсцисс, баллы — по оси ординат). Таким образом, на каждого испытуемого составляется самая полная временная картина выполнения теста. Имея результаты по каждому тестируемому, можно получить полную информацию о результативности работы каждого ученика в отдельности; группы учащихся (например, «двоечники» и «отличники»); всех учащихся. Эту информацию можно получить по отдельному тесту; по серии тестов и т.д. Ниже приводятся результаты трех тестирований: два теста **A** и **B** из 30 заданий, на выполнение которых отводилось 80 минут, и один тест **B** из 15 заданий, предлагавшийся в качестве вступительного экзамена в одном из Пермских вузов, на выполнение которого отводилось 180 мин. Тест A выполнялся классом, уровень подготовки которого выше уровня класса, выполнявшего тест **B**.

Проанализируем результаты по всем трем тестированиям. Для удобства их рассмотрения, составим три таблицы.

Табл. 1 **Результативность выполнения теста A**

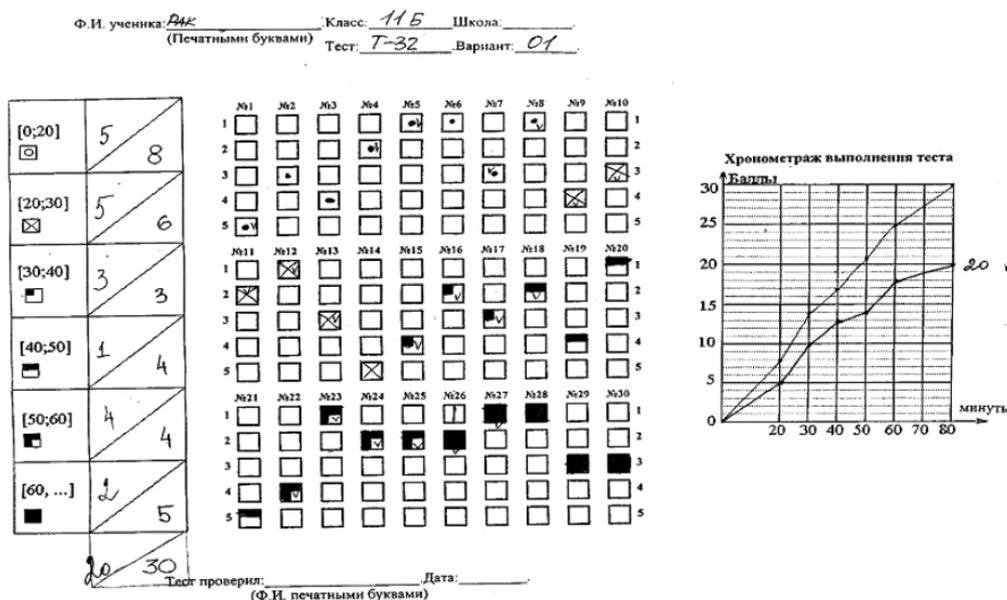
Интервалы времени	Процент испытуемых, допустивших не более 2 ошибок	Среднее число выполненных заданий			Среднее число верно выполненных заданий		
		среди всех участников	среди «отличников»	среди «двоечников»	среди всех участников	среди «отличников»	среди «двоечников»
[0;20]	77,8	8	10	6	7	9	3
[20;30]	85,2	5	6	5	3	5	2
[30;40]	66,7	4	4	4	2	3	2
[40;50]	92,6	3	3	1	2	3	0
[50;60]	92,6	3	3	3	2	2	1
[60;...]	37	7	4	11	3	2	0

Заметим, что 14,8% испытуемых закончили работу по истечении часа. Причем 75% из них имеют отличные результаты.

Из табл. 1 видно, что с увеличением времени уменьшается количество решенных заданий и резко падает количество верно ре-



Рис. 2 Хронометраж выполнения теста



шенных заданий (скачок на последнем этапе объясняется тем, что принимаются решения относительно вызывавших сомнения, но обдуманых ранее тестовых заданий и производится угадывание ответов в «безнадежных» ситуациях).

Построим три графика:

1. Средний результат по всем испытуемым.
2. Средний результат «отличников».
3. Средний результат «двоечников».

Рис. 3 Хронометраж выполнения теста А среди всех участников тестирования

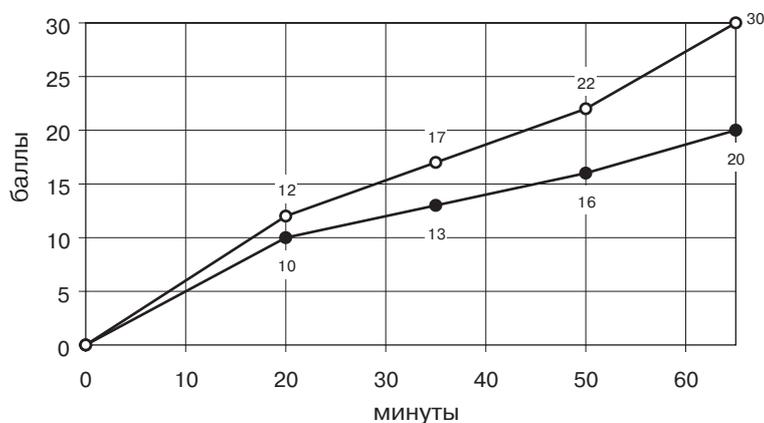




Рис. 4 Хронометраж выполнения теста А среди «отличников»

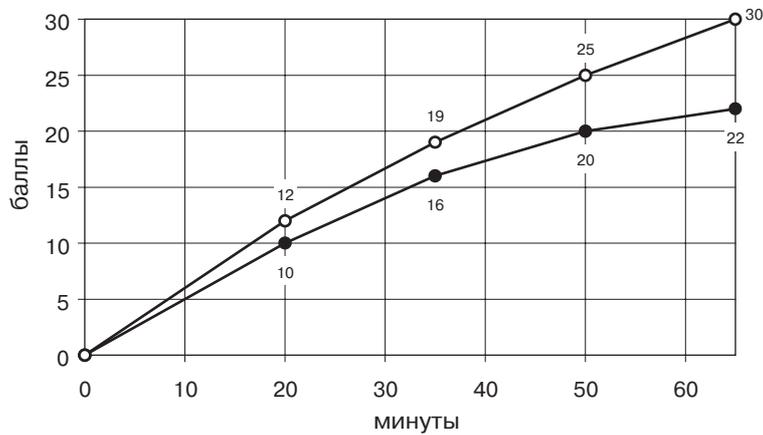
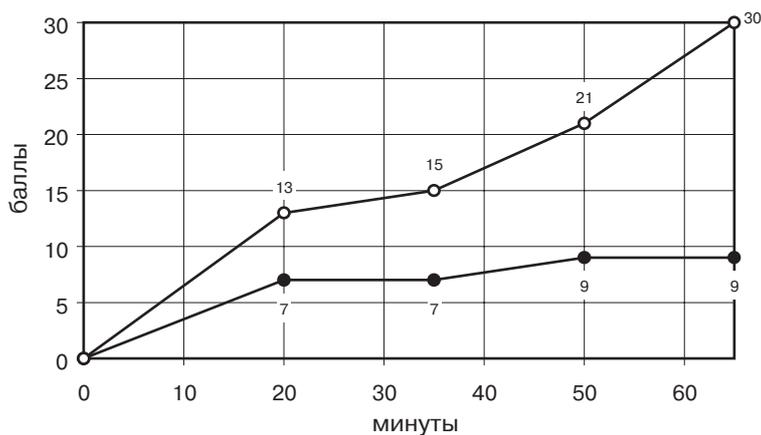


Рис. 5 Хронометраж выполнения теста А среди «двоечников»



Из полученных графиков наглядно видно, что площадь, заключаемая между кривыми общего числа решенных заданий и числа верно решенных заданий минимальна у тех испытуемых, которые удачно справились с тестом (в идеале кривые должны совпадать), максимальна у не справившихся с работой и имеет среднее значение среди всех участников тестирования. Увеличение рассматриваемой площади связано с увеличением ошибок тестируемых, так как верхняя ломаная показывает общее число решенных задач, а нижняя — число верно решенных тестовых заданий.

Данные табл. 2 этого тестирования подтверждают выводы, полученные при анализе предыдущего: наибольшее число решенных (в том числе верно) заданий приходится на первые 30 минут работы.

Построим графики хронометража результатов среди всех участников, а также среди хорошо справившихся и не справившихся с работой.



Табл. 2 **Результативность выполнения теста Б**

Интервалы времени	Процент испытуемых, допустивших не более 2 ошибок	Среднее число выполненных заданий			Среднее число верно выполненных заданий		
		среди всех участников	среди «отличников»	среди «двоечников»	среди всех участников	среди «отличников»	среди «двоечников»
[0;20]	60,3	10	9	8	7	8	5
[20;30]	77,6	7	6	5	4	4	1
[30;40]	87,9	4	4	5	3	4	2
[40;50]	82,8	4	4	2	2	3	1
[50;60]	82,8	2	2	3	1	1	0
[60;...]	56,9	5	5	5	1	2	1

Рис. 6 **Хронометраж выполнения теста Б среди всех участников тестирования**

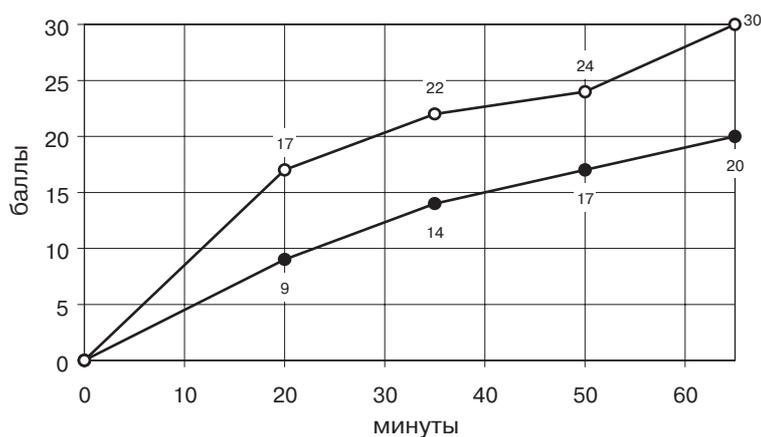


Рис. 7 **Хронометраж выполнения теста Б среди «отличников»**

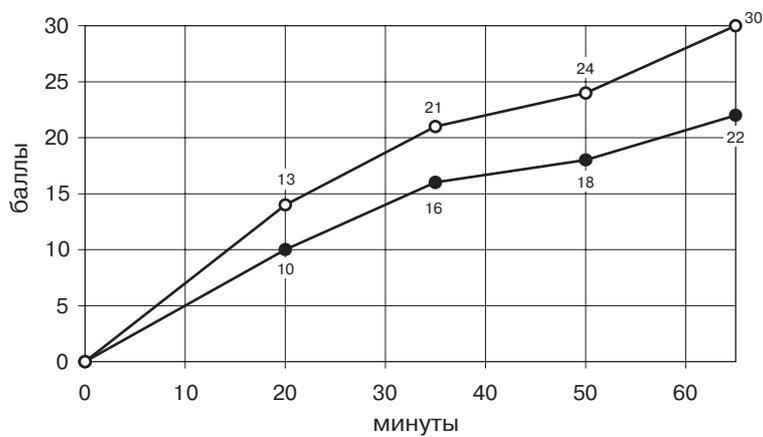
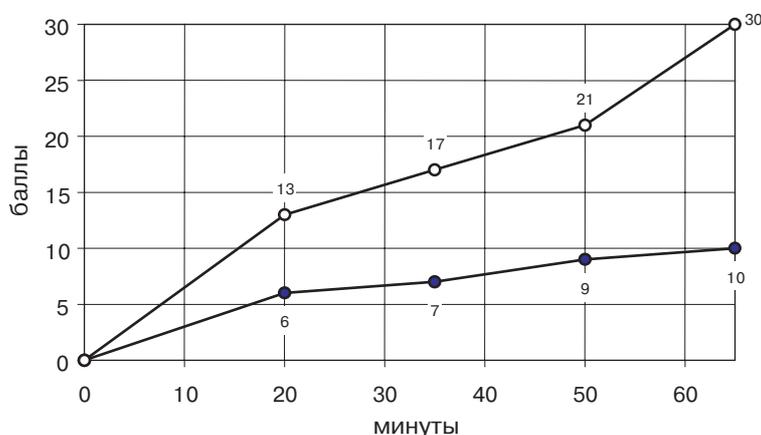




Рис. 8 Хронометраж выполнения теста Б среди «двоечников»



Общая структура графиков, построенных к тесту **Б**, аналогична структуре графиков, отображающих хронометраж выполнения теста **А**: наибольший разброс кривых характерен для «двоечников», наименьший — для отличников.

Табл. 3 Результативность выполнения теста **В**

Интервалы времени	Процент испытуемых, допустивших не более 2 ошибок	Среднее число выполненных заданий			Среднее число верно выполненных заданий		
		среди всех участников	среди «отличников»	среди «двоечников»	среди всех участников	среди «отличников»	среди «двоечников»
[0;20]	66,6	6	7	5	5	7	2
[20;30]	61,1	3	3	3	2	3	1
[30;40]	88,8	2	2	3	2	2	1
[40;50]	77,7	2	1	2	1	1	1
[50;60]	83,3	1	1	1	1	0	1
[60;...]	27,7	1	1	1	0	1	0

Наиболее интересным при проведении данного тестирования явилось то, что ни один из его участников не затратил на выполнение теста более 70 минут. 11,1% тестируемых справились с заданиями за 40 мин, 5,5% участникам понадобилось 50 мин, за час закончили работу 11,1% испытуемых, показав при этом достаточно высокие, а порой и максимальные результаты.

Особенностью данного тестирования явилась также невозможность качественно ранжировать его участников, так как очень многие получили одинаковые результаты. Причиной этого стало, прежде всего, недостаточное количество тестовых заданий. В связи с этим создание «укороченного» теста высокой сложности, предполагающего длительное время выполнения, себя не оправдало.



Рис. 9 Хронометраж выполнения теста В среди всех участников тестирования

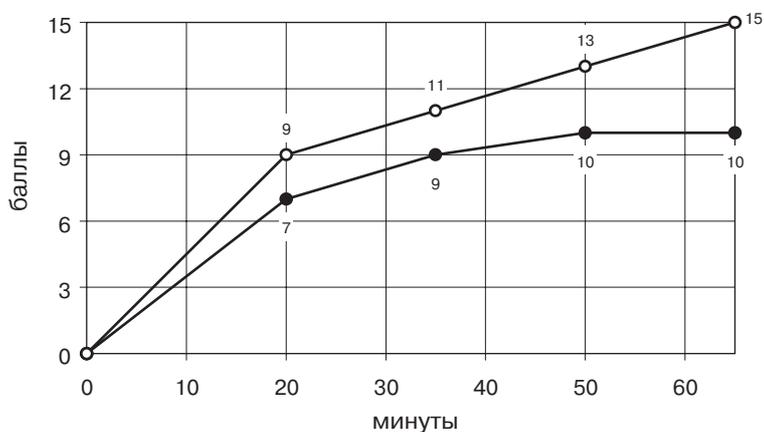


Рис. 10 Хронометраж выполнения теста В среди «отличников»

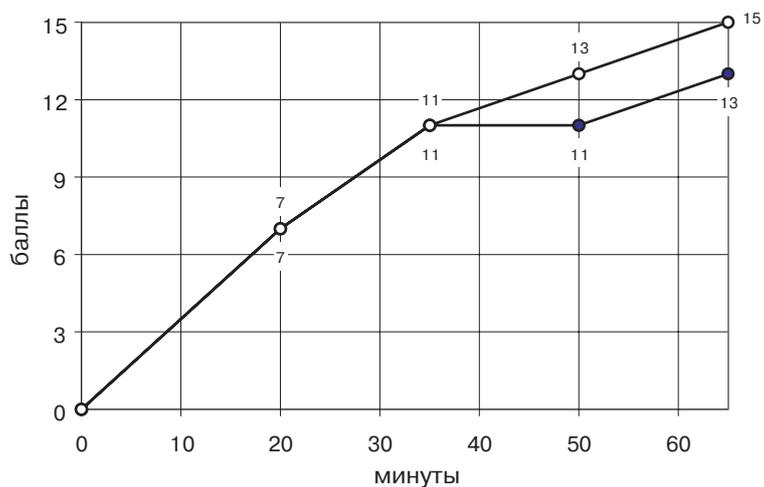
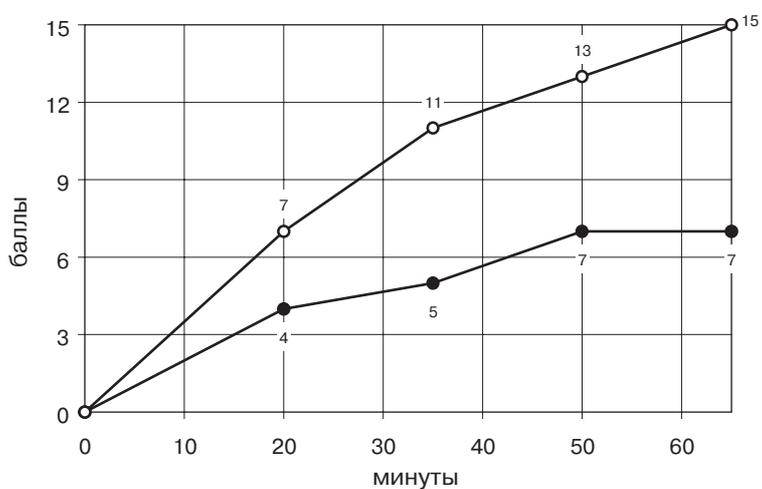


Рис. 11 Хронометраж выполнения теста В среди «двоечников»





Параметры изменения количества решаемых задач с течением времени не изменились по сравнению с результатами предыдущих таблиц.

Таким образом, по результатам полученных таблиц и графиков, можно заключить следующее:

1. Максимальное число заданий теста испытуемые выполняют в первые 30 минут тестирования.

2. Резкое отличие результатов хорошо подготовленных и плохо подготовленных участников наблюдается уже в первые 20–30 минут.

3. На последнем этапе работы производится угадывание ответов тех заданий, которые не удается выполнить, и принимается окончательное решение при выборе ответов заданий, вызывавших сомнения ранее. Процент совершения ошибок максимален.

Источником ошибок часто становится утомление и невозможность сосредоточиться. Известно, что максимально активизировать умственную деятельность человеку удастся лишь в течение часа. Проведенное исследование подтверждает этот вывод: по истечении часа испытуемые стараются завершить работу над тестом, даже если часть заданий осталась невыполненной. Хорошо подготовленные ученики выполняют работу за 10–15 минут до конца отведенного времени, а затем осуществляют проверку и сдают бланки. Преждевременное окончание работы имеет свои существенные недостатки:

1. В оставшееся время тестируемые пытаются сверить результаты.

2. Более сильные могут помогать другим в выполнении теста, что неизбежно ведет к искажению истинного результата.

3. Заинтересованные представители из числа тестирующих получают время и возможность улучшить результаты учеников.

4. Двухлетний опыт проведения вступительного экзамена в вузе в форме теста из 15 заданий с продолжительностью выполнения в 180 минут показал очевидные искажения объективных результатов: учителя констатировали получение высших баллов их слабоуспевающими в школе учащимися.

В связи с этим следует отводить на выполнение теста только необходимое время, то есть подбирать такое количество заданий определенной сложности, чтобы на их решение не требовалось более часа. Известно, что чем тест длиннее, тем он надежнее. Поэтому нецелесообразно уменьшать количество заданий, увеличивая их сложность.

«Хороший» тест имеет оптимальную длину (в математических тестах это, как правило, 30 заданий), оптимальную сложность, которая варьируется относительно поставленных тестирующим целей (контроль знаний, отбор лучших, выявление самых способных и др.) и требует на выполнение не более часа.

Отметим также, что влияние продления времени тестирования на результативность легко проверить путем проведения одного и того же теста в двух или трех равных по уровню группах учащихся



с приращением времени минут. Такие опыты автором проводились неоднократно и показали, что прирост верно решенных заданий после 60 минут не превышает 1,2–1,6 баллов.

Нами неоднократно проводились замеры среди учащихся двух основных групп:

1. Учащиеся с высокой мотивацией обучения и способностью сосредоточиваться на достаточно длительном интервале времени («отличники»);

2. Учащиеся с низкой мотивацией обучения (условно «двоечники»). Результаты предоставлены в виде графиков: верхняя кривая показывает общее число выполненных заданий, а нижняя — число верно решенных (см. рис. 12, 13, 14).

Рис. 12 Хронометраж выполнения теста всеми участниками тестирования

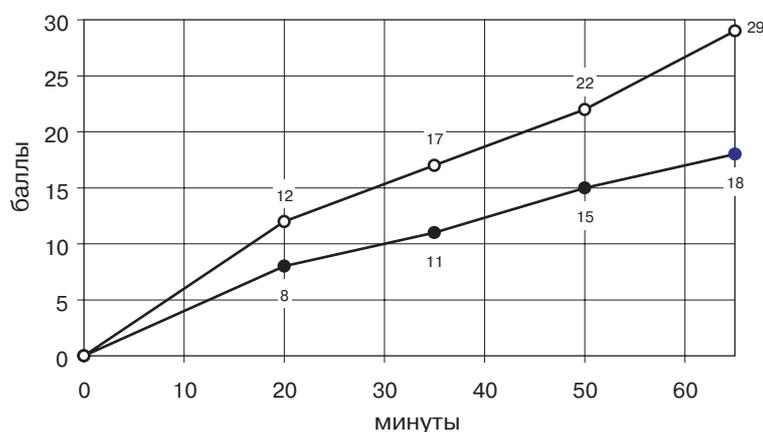
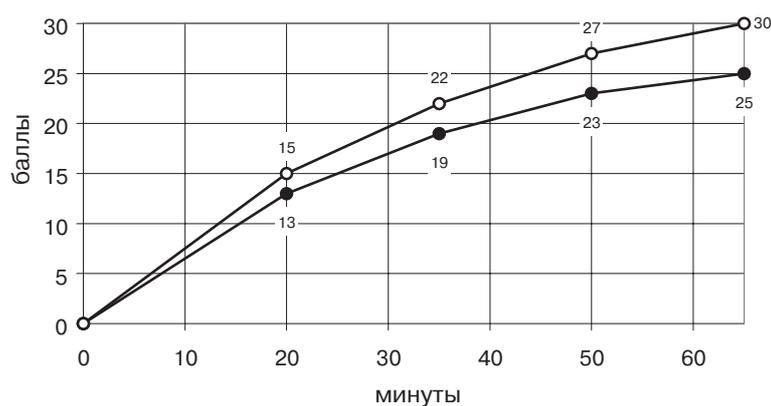


Рис. 13 Хронометраж выполнения теста «отличниками»

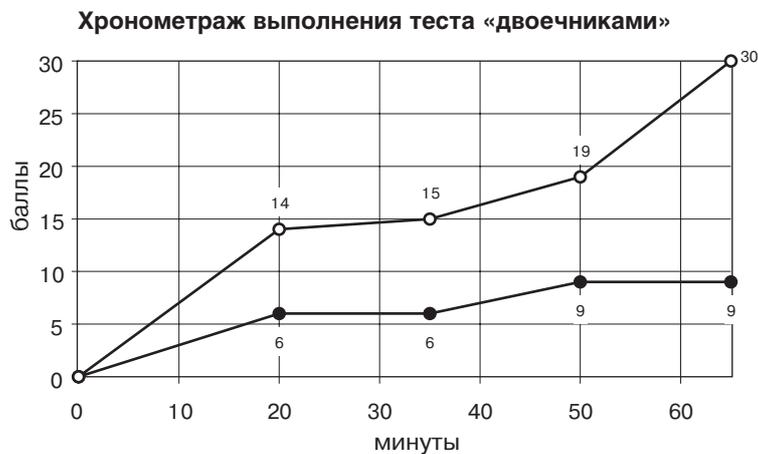


Из полученных графиков можно делать следующие выводы:

1. «Отличники» и «двоечники» дифференцируются уже на первых 20 минутах тестирования: у «отличника» из 10 выполненных заданий верно 9, у «двоечника» — из 6 верно только 3;



Рис. 14



2. Отчетливо видно, что результативность у «двоечника» практически нулевая уже после 40 минут;

3. Наглядно видно, как увеличивается «вилка» между верхней и нижней кривой: у «отличника» она минимальна, у «двоечника» она значительна;

4. Площадь этой «вилки» можно использовать для штрафования любителей угадывать.

Процедуру хронометрирования выполнения тестовых заданий можно существенно модернизировать с помощью компьютерного тестирования.

Специально для самоконтроля усвоения знаний была разработана программа «Компьютерный тренинг». С помощью этой прог-

Рис. 15

Временная диаграмма результативности тестирования

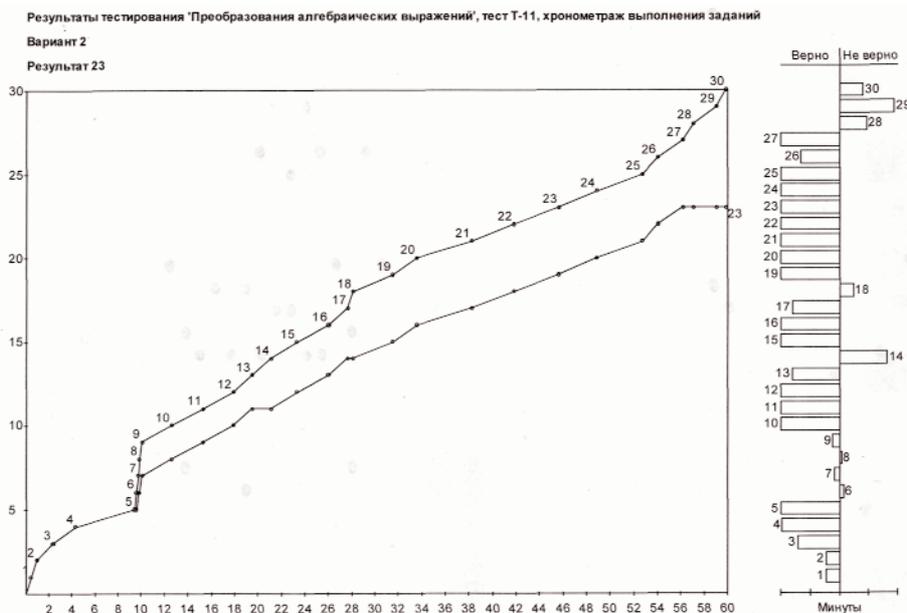
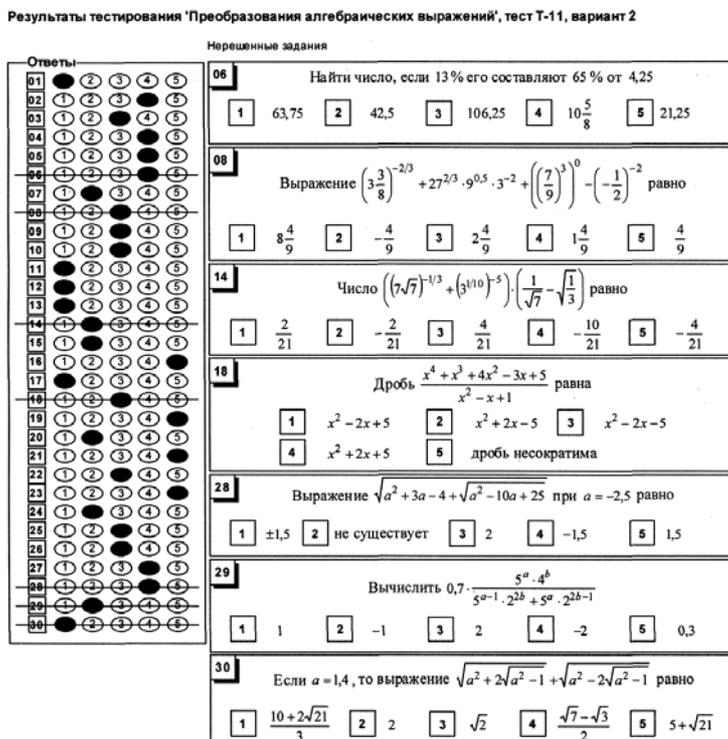




Рис. 16 Подробные результаты компьютерного тестирования



раммы учащиеся могут самостоятельно тестироваться, а родители учащегося получают *реальное* представление о состоянии знаний по математике своего ребенка.

Программа позволяет тестироваться по тематическим тестам и итоговым. Используя механизм генерации тестов, можно получить огромное количество различных итоговых тестов.

Основное отличие от бланкового тестирования — ответы сразу же заносятся в память компьютера, также автоматически ведется хронометраж выполнения заданий. После окончания тестирования программа выдает результаты тестирования, а также всю остальную, полученную за время тестирования информацию в наглядном виде (рис. 15 16). Рассмотрим подробнее выдаваемые программой отчеты.

Временная диаграмма представляет собой две кривые. В каждый момент времени (ось времени по горизонтали) они отображают: верхняя — количество отмеченных ответов, нижняя — количество верных ответов. Цифра рядом с каждым перегибом верхней ломаной означает номер задания, на которое был дан ответ. Например: на четвертой минуте тестирования было правильно решено задание № 4 за 2 минуты, так как предыдущий ответ был отмечен на второй минуте (задание №3). Таким образом, по виду диаграммы можно легко определить, решал ли учащийся задания теста или нашел совпадающие задания в пособии [8] и списал ответ.



Второй отчет (рис. 16) представляет собой задания, которые были решены неверно либо пропущены. С помощью этого отчета удобно проводить работу над ошибками.

Перейдем к обсуждению следующих существенных недостатков технологии проведения ЕГЭ.

2. Отсутствует обоснование введения 4-х альтернатив ответа, хотя оптимальным считается число их, равное 5.

3. Нельзя признать идеальной форму бланка автоматизированного считывания информации сканером. Полученное при этом изображение обрабатывается программой FineReader. Теоретически это выглядит достаточно хорошо, в работе [5] нет никаких сведений о проценте учащихся, которые заполняли бланки таким образом, что считывание осуществлялось с погрешностями. А они обязательно появлялись, и процент их, скорее всего, значителен (в подтверждение этого тезиса достаточно вспомнить долголетнюю путаницу при написании номеров почтовых индексов, та же проблема возникает и у учащихся, когда они пишут, например, цифры 1 и 7). Почему бы не использовать положительный опыт штриховки овалов, используемый на протяжении нескольких лет в международных математических соревнованиях школьников «Кенгуру»?

В соответствии с приказом №130 от 01.02.2005 руководителя Федеральной службы по надзору в сфере образования и науки В.А. Болотова «Об утверждении форм-бланков ответов участника Единого государственного экзамена, проводимого с использованием автоматизированной информационной системы «Экзамен», в 2005 году» вводится следующая форма бланка ответов АВ, где указываются образцы нанесения метки: , отмены нанесенной метки: и восстановления отмененной метки: (при этом) количество отмен и восстановлений не регламентируется. Очевидно, что такая форма заполнения бланка позволит заполнять и исправлять результаты на любом этапе проведения экзамена и проверки (причем даже без участия учащегося).

«Гордостью» тестовых материалов Центра тестирования является наличие в тестах части В, где требуется вписывание количественного результата решения тестового задания, и части С, где требуется приведение полного решения. Беру на себя смелость утверждать, что введение этих частей призвано снизить конфиденциальность процедуры обработки по многим причинам. Перечислим эти неприятные последствия:

а) неверное считывание результата В приведет к появлению необходимости апелляции, а следовательно, попадания бланка ответа в руки конкретного человека.

б) необходимость ручной проверки решения заданий части С: получается, что «от чего уходили (от письменного экзамена), к тому и пришли».

В 2004 г. появились еще две проблемы: сложности заданий части С были совершенно различными в разных вариантах:



а) задания С1 вариантов 151–180 значительно проще С1 вариантов 91–120; задача С3 вариантов 91–120 значительно проще С3 вариантов 121–180.

б) в задачах С3 вариантов 153, 156, 158, 161, 167, 171 были допущены описки, которые не позволили даже выполнить чертеж к задачам. Это привело к недополучению баллов многими сильными учащимися. Некоторые ученики представить себе не могли, что возможны ошибки в условиях, и потратили много времени на бесполезные попытки построить чертеж, а это, в свою очередь, отразилось на их психологическом состоянии и появлении стресса, из-за которого они не смогли качественно выполнить и другие задания.

Такие «промахи» составителей дискредитируют всю идею ЕГЭ. Любое решение, направленное на исправление этой ситуации при расчете баллов (например, добавление всем четырех баллов) нельзя признать справедливым.

Эти проблемы с частью С еще раз подтверждают необходимость широкого обсуждения и проведения многочисленных экспериментов, с помощью которых можно было оценить целесообразность включения этой части. Эксперты, проверяющие часть С, отмечают, что из приступавших к этим задачам учащихся (а таковых обычно бывает менее 50%) в среднем по протоколам на 1 школьника приходится по 1–1,5 первичных баллов.

Считаю, что приведенные задачи и способы их проверки убедительно показывают, что каждое из заданий части С лучше заменить на 2–3 обычных тестовых задания по проверяемым с их помощью интересующим темам. Самым неприятным во введении части С является необходимость попадания работ в руки конкретных людей, необходимость подготовки заранее вариантов и решений, секретность содержания которых в России сохранить невозможно (есть печальный опыт передачи по Интернету в 2001 г. всех вариантов выпускных письменных работ как для обычных школ, так и для школ с углубленным изучением математики накануне экзамена). В связи с проблемой сохранения секретности экзаменационных материалов возникает естественный вопрос: как обеспечить секретность материалов тестирования при заготовке даже 40 вариантов тестов Центра тестирования, почему бы не использовать процедуру автоматической генерации тестов и передачи их в регионы за самое минимально возможное время до начала экзаменов?

Вышеперечисленные замечания в большей мере относятся к технологии проведения ЕГЭ по математике, которую никак нельзя признать удовлетворительной. Кроме несовершенства технологии значительную часть его противников не устраивает и низкое качество конкретных заданий КИМов по математике. Эти замечания зачастую носят очень общий характер, и не все они справедливы. Тем не менее, следует указать на такие недостатки заданий (на примере вариантов 2003 г.):



- многие задания можно вообще не решать, а просто угадывать отбрасыванием совершенно непригодных и абсурдных ответов;
- излишне много очень простых заданий во второй половине группы заданий части А (А10, А12, А13, А14, А15);

Замечания по части В:

- все задания достаточно интересны, но не выдержан принцип составления теста: от простого к сложному;

Замечания по части С:

- задания С1 и С3 явно не удовлетворяют продекларированным требованиям, предъявляемым к заданиям части С.

Представляется, что благие намерения введения части С в ЕГЭ «для проверки состояния более сложных интеллектуальных и предметных умений — анализировать ситуацию, разрабатывать способ решения, проводить логически и математически грамотные рассуждения, обоснования, доказательства своих действий и грамотно записывать их» не могут быть практически реализованы по многим причинам. Перечислим некоторые из них.

В научно-методической литературе очень много внимания уделяется проблеме выработки единых подходов к оценке решения задач, выходящих за рамки минимальных требований (см. например, статью Г.В. Дорофеева в журнале «Математика в школе», 2000г.). В процессе решения таких задач ученик может проявить более высокий уровень самостоятельности, использовать знания и умения, приобретенные им не только в школе. И здесь начинают сталкиваться разные методические представления и взгляды на критерии оценки, возникает извечный вопрос об оформлении решения (как должно быть записано решение, чтобы его можно было признать логически полным и грамотным).

Г.В. Дорофеев указывает, что «ни математическая наука, ни методика обучения математике не могут дать на этот вопрос объективного ответа». Этим и объясняются разногласия по поводу критериев оценивания заданий части С. Используемый измеритель части С не обладает высокой надежностью, а погрешность в 1с2 первичных балла ведет к искажению результата 100-балльной шкалы в 3с6 баллов, что является значительной погрешностью. Попытки сформулировать общие критерии проверки к оценке заданий части С и приведение образцов их решения нельзя признать удачными. Таким образом, имеет смысл обсудить вопрос о целесообразности введения части С вообще.

Выполнение последних двух заданий части С требует чаще всего «олимпиадного» мышления, особой смекалки и нетривиальных подходов. Проверить справедливость этого утверждения достаточно просто (чисто экспериментально): попросить выполнить эти задания членов какой-либо математической кафедры вуза или учителей самой высокой квалификации. Предсказуемый результат следующий: задания будут решены некоторыми участниками эксперимента, но для решения их потребуется никак не один-два лис-



та такой бумаги, а значительно больше. Количество требуемой бумаги легко установить, прописав «идеальные» решения, выдаваемые членам экспертных комиссий, проверяющим часть С, в виде критериев оценивания. Суть этих требований хорошо известна из опыта работы медальных комиссий. Учителя-практики хорошо знают, сколько недочетов можно «накопать» в любом задании. Каждый эксперт будет трактовать необходимый уровень подробностей в объяснениях по-своему, одна и та же работа у разных проверяющих может заслужить разные первичные оценки, отличающиеся на 1, 2, 3 балла, а при переводе в 100-балльную шкалу эти различия надо умножать на коэффициент не меньше трех. О какой точности измерений с помощью заданий части С можно говорить?!

По поводу последних («олимпиадных») заданий части С уместно напомнить следующее: для решения «олимпиадных» задач времени требуется намного больше, но и «олимпиадники» порой оказываются неспособными к продолжительной продуктивной деятельности. Б.К. Делоне считал, как указывал А.Н. Колмогоров, что большое научное открытие отличается от «олимпиадной» задачи только тем, что для решения задачи требуется 5 часов, а получение крупного научного результата требует 5000 часов [12; 13]. Академик П.С. Александров считает, что если бы во времена его юности были олимпиады, он, возможно, вообще не стал бы математиком: его главные достижения в математике являлись не плодом быстро работающей изобретательности, а итогом длительного и углубленного созерцания.

Приведенные рассуждения свидетельствуют, что тесты должны удовлетворять особым требованиям, условия их использования должны быть тщательно обоснованы.

Сформулируем основные выводы, вытекающие из всего вышесказанного по поводу экзамена по математике.

1. Несомненно, что старая форма проведения экзамена по математике по экзаменационным пакетам из письменной работы в 5сб заданий или открытым текстам устарела, так как весь экзамен превратился в фарс. Поэтому новая форма проведения экзамена в виде теста неизбежна, но требует значительной доработки (как содержание теста, так и технология проведения).

2. Так как основная идея проводимого экзамена — единый для школ и вузов, то очевидно, следует совершенствовать его прежде всего в направлении сближения интересов и школ, и вузов.

3. Нельзя признать, что тест составлен методически грамотно в соответствии с основными положениями тестологии: следует максимально исключить возможность угадывания ответов путем элементарного отбрасывания ненужных.

4. Формулировку некоторых заданий нельзя признать идеальной. Следует исключать задания, неверное или неполное решение которых приводит к правильным числовым ответам (например, задание С4 вар. 163, 2004 г.). Такие задания приводят к апелляциям.



5. Непонятен отказ от формирования 5-альтернативных заданий части А (как это было в самом начале проведения эксперимента по ЕГЭ) и переход к 4-альтернативным.

6. Представляется целесообразным путем проведения экспериментов выбрать оптимальные параметры тестов: длину (количество заданий и типы заданий), время выполнения теста (4 часа — явно завышенное время) и трудность.

В заключение перечислим прочие недостатки технологии проведения ЕГЭ по математике и возможные способы их устранения.

1. Часть С не выполняет своей основной функции — четкого выявления наиболее сильных учащихся, особенно если в заданиях имеются ошибки. В 2004 году искажение результатов можно оценить в 10–12%. Имеет смысл обсудить целесообразность выполнения части С всеми учащимися, так как эта часть экзамена является самой затратной в финансовом плане и очень трудоемкой как для составителей, так и для экспертов на местах.

2. Нужно исключить ситуацию, когда школьники, набравшие одинаковое количество баллов, получают различную оценку в пятибалльной шкале.

3. Недостаточно строгое следование инструкции по выполнению теста в пунктах проведения экзамена может приводить к использованию средств связи для передачи решений (это уже наблюдается на практике).

4. Необходима возможность переэкзаменовки в сомнительных случаях («2»–«3» в школе и более 74 баллов по ЕГЭ). В случаях аннулирования результатов нужно продумать меры наказания и профилактики нарушений в будущем. Известны примеры, когда «чистые двоечники» получили в 2004 г. по ЕГЭ 88–92 балла.

5. Необходимо исключить описки и ошибки в условиях. В задачах С они приводят к искажению оценки на 1с4 первичных балла, а при переводе в 100-балльную шкалу погрешность достигает 10–12%.

6. В Государственном университете Высшей школе экономики вступительный экзамен по математике с 2000 года проходит в виде теста, содержащего только часть А с 5-альтернативными вариантами с временем исполнения 60–90 мин. В Перми в классическом университете вступительный экзамен в такой форме проходит уже в течение 11 лет; таким образом, имеется опыт использования технологий, исключающих или минимизирующих многие вышеперечисленные недостатки. Представляется, что без особых затрат можно провести анализ качества отбора абитуриентов, проследив результаты их успеваемости за время обучения в вузе (достаточно ограничиться одними предметами математического цикла).

7. Следует продумать мероприятия по передаче экзаменационных материалов в регионы России (с десятью часовыми поясами), не в виде пакетов, а по компьютерным сетям с использованием технологий информационной безопасности, причем желатель-



но эти тесты автоматически генерировать за очень небольшой промежуток времени до начала экзамена и распечатывать непосредственно перед экзаменом. Эти мероприятия исключили бы утечку информации и появление в Интернете ответов.

8. Провести в течение учебного года эксперименты в школах по использованию различных технологий и тестов.

9. Избежать «подтасовок» результатов вполне возможно, проводя 2–3 замера уровня знаний в форме теста (30 заданий, 60 минут) в 10-х–11-х классах в течение учебного года. Очевидно, что банк личных достижений учащегося — лучшее средство от взятки. Все эти достижения, заносимые в электронную базу данных, переводятся в какую-либо шкалу (лучше всего в 100-балльную), каждый школьник получает идентификационную электронную карту (или просто номер) для того, чтобы иметь возможность проверить свой рейтинг, узнать свое место в своей параллели по всем школам района, города или другого сообщества общеобразовательных учреждений.

10. Отказаться от использования программы FineReader, дающей множество сбоев при считывании результатов части В. Вместо этой системы можно использовать технологию ПФ ГУ-ВШЭ.

11. В части В с кратким ответом подбирать только задания, где ответом является натуральное число, и результат отмечать в бланке, как остаток от деления на 5.

12. Проводить хронометрирование выполнения теста путем использования трех бланков. В Перми (ПФ ГУ-ВШЭ) такая технология апробирована, она малозатратна.

1. Аванесов В.С. Композиция тестовых заданий. М.: Адент, 1998.
2. Волочков А.А. Субъект активности и развития индивидуальности подростка (теория, практика, диагностика). Пермь: Изд-во ПОИПКРО, 2002.
3. Гайда В.К., Захаров А.П. Психологическое тестирование. Л., 1982.
4. Гронлунд Норман. Тесты достижений в конструировании. Лондон; Прентис-Холл, 1982.
5. Единый государственный экзамен: материалы Центра тестирования Минобразования России на конкурс контрольно-измерительных материалов и технологий проведения эксперимента в 2001 г. М.: Век книги, 2001.
6. Зак А.З. Различия в мышлении детей. М.: РОУ, 1992.
7. Иванов А.П., Систематизация знаний по математике в профильных классах с использованием тестов, М.: Изд-во Физматкнига, 2004.
8. Иванов А.П., Иванов А.А. Тематические тесты для систематизации знаний по математике: Учеб. пособие. Изд. 3-е, испр. и доп. М.: Физматкнига, 2004. Ч.1.
9. Ильин Е.П. Мотивация и мотивы. СПб: Питер, 2000.
10. Каптерев П.Ф. Избранные педагогические сочинения / Под ред. А.М. Арсеньева. М.: Педагогика, 1982.
11. Ковалева Г.С. Подходы к разработке контрольных измерительных материалов для Единого государственного экзамена: Материалы и тезисы докладов Международной конференции «Развитие национальной системы экзаменов: опыт России, СНГ и США». М., 2003.

Литература



12. Колмогоров А.Н. Математика — наука и профессия. М.: Наука, 1988.
 13. Колмогоров А.Н. О профессии математика. М.: Наука, 1973.
 14. Майоров А.Н. Теория и практика создания тестов для системы образования. М., 2000.
 15. Майоров А.Н. Тесты школьных достижений. СПб.: Образование и культура, 1997.
- На пути к Единому экзамену: идеология и технология. Пермский опыт: Сб. статей / Сост. и отв. ред. В.П. Севрук. Пермь: Социнком, 2001.