

ЗАВИСИМОСТЬ РЕЗУЛЬТАТОВ ИЗМЕРЕНИЯ СПОСОБНОСТЕЙ УЧАЩИХСЯ ОТ СЛОЖНОСТИ ЗАДАНИЙ

Аннотация

Проведен анализ результатов ЕГЭ по физике для 10472 учащихся, а также моделирование на основе теории Раша с параметрами заданий, соответствующими результатам ЕГЭ. Предложена методика выявления формы зависимостей с помощью усреднения специальным образом отсортированных данных.

Показано, что результаты выполнения заданий ЕГЭ части А (тестов с выбором одного варианта из нескольких) могут быть объяснены с помощью теории Раша. Результаты выполнения заданий других типов (заданий частей В и С) качественно соответствуют предсказаниям теории Раша, но количественно отклоняются от соответствующих зависимостей.

При сравнении испытаний различной сложности очень важными факторами являются статистический разброс и нелинейность зависимостей. Показано, что способность выполнять простые задания соответствует очень широкому диапазону способности выполнения более сложных заданий, от низкого до высокого. Показано, что способность выполнять сложные задания гарантирует способность выполнять более простые задания.

Ключевые слова: ЕГЭ, Единый Государственный Экзамен, тестирование, теория Раша, IRT, образование, физика, моделирование, программное обеспечение.

1. ПОСТАНОВКА ПРОБЛЕМЫ

Единый Государственный Экзамен (ЕГЭ) уже несколько лет используется для отбора абитуриентов при поступлении в вузы. В других странах также имеются аналогичные виды испытаний. Выставление балла по результатам этих испытаний, как правило, считается достаточным. Однако в работе [1] нами было установлено, что способность отлично выполнять простые задания не гарантирует способности выполнять более сложные задания, и что, напротив, способность выполнять сложные задания гарантирует способность выполнять более простые задания – но, вообще говоря, не на высший балл.

© В.В. Монахов, 2011

В работе [2] автором было показано, что для ЕГЭ и олимпиад реализуется описанная выше ситуация «простых» и «сложных» заданий: самое сложное задание ЕГЭ 2010 года по физике соответствовало по сложности самому простому заданию очного тура интернет-олимпиады школьников по физике. В других олимпиадах школьников I и II уровней и на межрегиональном и заключительном этапах Всероссийской олимпиады школьников задания также обладают заметно более высоким уровнем сложности, чем задания ЕГЭ. Поэтому сравнение результатов испытаний в случае, когда сложности заданий заметно отличаются, требует специального исследования. Данная проблема имеет большую практическую значимость,

поскольку в настоящее время дипломантам олимпиад предоставляются льготы при поступлении в вузы, и олимпиады дополняют ЕГЭ, обеспечивая альтернативный механизм поиска талантливых учащихся.

У проблемы имеется и теоретический аспект. Разработка контрольно-измерительных материалов ЕГЭ и анализ результатов ЕГЭ основаны на теории Раша [3]. В соответствии с ней результаты измерения способностей не зависят от сложности заданий. В связи с этим возникает вопрос: можно ли объяснить наблюдавшиеся в работе [1] закономерности в рамках теории Раша, поскольку, на первый взгляд, данные результаты ей противоречат? Или необходим пересмотр представлений о применимости теории Раша и всех теорий группы IRT (Items Response Theory) в случае очень широкого диапазона изменения сложности заданий и способностей участников?

2. ЭКСПЕРИМЕНТАЛЬНЫЕ РЕЗУЛЬТАТЫ

В настоящей работе проанализированы результаты ЕГЭ по физике, проводившегося в Москве в 2010 году. Дегерсонифицированные результаты 10472 участников были получены от предметной комиссии ЕГЭ по

физике г. Москвы. Программное обеспечение для обработки данных и для моделирования было написано с использованием программного комплекса BARSIC [4].

Сначала было проведено разбиение 36 заданий ЕГЭ на две равные группы, к которым они относились в зависимости от сложности. Сложность определялась по проценту выполнения заданий, задания в группах указаны в порядке возрастания сложности.

«Простые» задания: А2, А9, А6, А24, А13, А1, А20, А4, А22, А8, А17, А15, А14, А18, А5, В2, А11, А10.

«Сложные» задания: А3, А21, А19, А7, В1, А16, А12, В4, В5, А25, С2, А23, С6, В3, С1, С3, С5, С4.

Максимальный первичный балл за каждое из заданий части А, а также задания В3, В4, В5 составлял 1 балл, за каждое из заданий В1 и В2 составлял 2 балла, за каждое из заданий части С составлял 3 балла. В связи с этим максимальный первичный балл за выполнение «простых» заданий составлял 19 баллов (задание В2 вошло в число «простых»), а за выполнение «сложных» заданий – 31 балл.

На рис. 1 приведена зависимость результатов выполнения «простых» заданий от «сложных». Для наглядности на рис. 2 при-

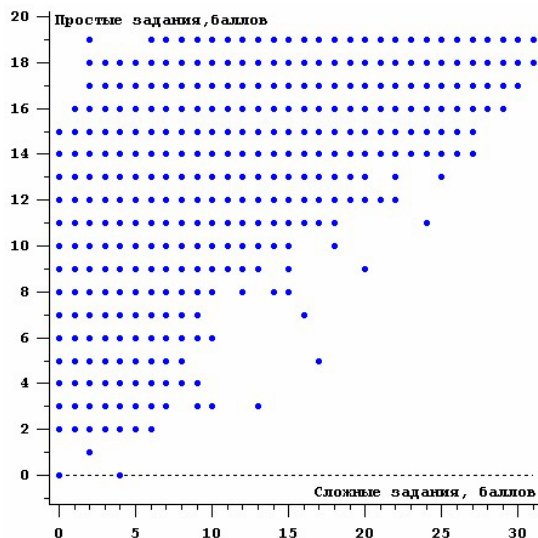


Рис. 1. Зависимость результатов выполнения 18 наиболее простых заданий ЕГЭ от результатов выполнения 18 наиболее сложных заданий

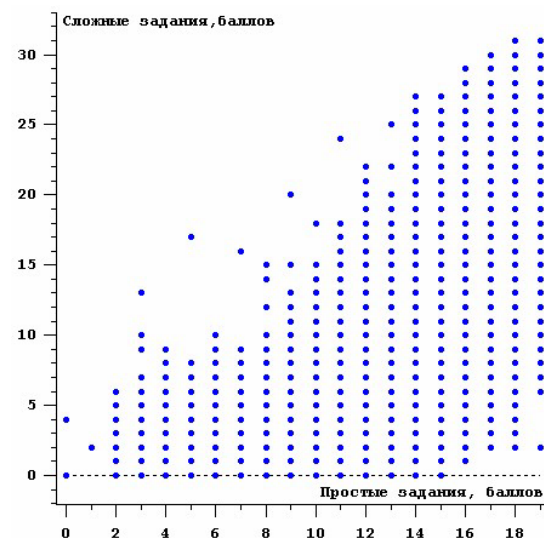


Рис. 2. Зависимость результатов выполнения 18 наиболее сложных заданий ЕГЭ от результатов выполнения 18 наиболее простых заданий

ведена зависимость результатов выполнения «сложных» заданий от «простых».

Видно, что зависимости представляют очень широкие полосы, фактически превращающиеся в треугольники, заполняющие левую верхнюю либо правую нижнюю половину области графика. При этом отчетливо наблюдается тенденция, отмеченная в работе [1]: способность отлично выполнять простые задания не гарантирует способности выполнять более сложные задания и, напротив, способность выполнять сложные задания гарантирует способность выполнять более простые задания – но, вообще говоря, не на высший балл.

Например, выполнение сложных заданий на половину от их максимального балла гарантирует выполнение простых заданий на половину от их максимального балла. Но с простыми заданиями все обстоит иначе: даже максимальный балл за выполнение простых заданий может соответствовать близкому к нулю результату за выполнение более сложных заданий – но может соответствовать и максимальному баллу. Баллы за выполнение сложных заданий оказываются очень слабо связанными с баллами, полученными за выполнение простых заданий, имеется единственное ограничение: у каждого участника доля решенных сложных заданий

не превосходит долю решенных простых заданий.

На рис. 3 приведены зависимости полных баллов за ЕГЭ от баллов, полученных за выполнение «простых» и «сложных» заданий.

Видно, что погрешность, связанная с разбросом, для зависимости от баллов за «простые» задания очень заметно возрастает по мере увеличения набранных баллов. А для зависимости от баллов за «сложные» задания, напротив, погрешность велика в области низких баллов и очень заметно убывает по мере увеличения набранных баллов.

3. РЕЗУЛЬТАТЫ МОДЕЛИРОВАНИЯ

Для выяснения того, можно ли объяснить наблюдавшиеся зависимости на основе теории Раппа, было проведено моделирование зависимости количества набранных баллов за сложные задания от баллов, набранных за простые задания. В рамках найденных для заданий ЕГЭ диапазонов сложности (от -1.1 до -0.34 логитов для «простых» и от -0.28 до 1.74 логитов для «сложных») модельные задания считались распределенными равномерно. Для результатов ЕГЭ были найдены следующие параметры распределения способностей участников, предполагавшегося

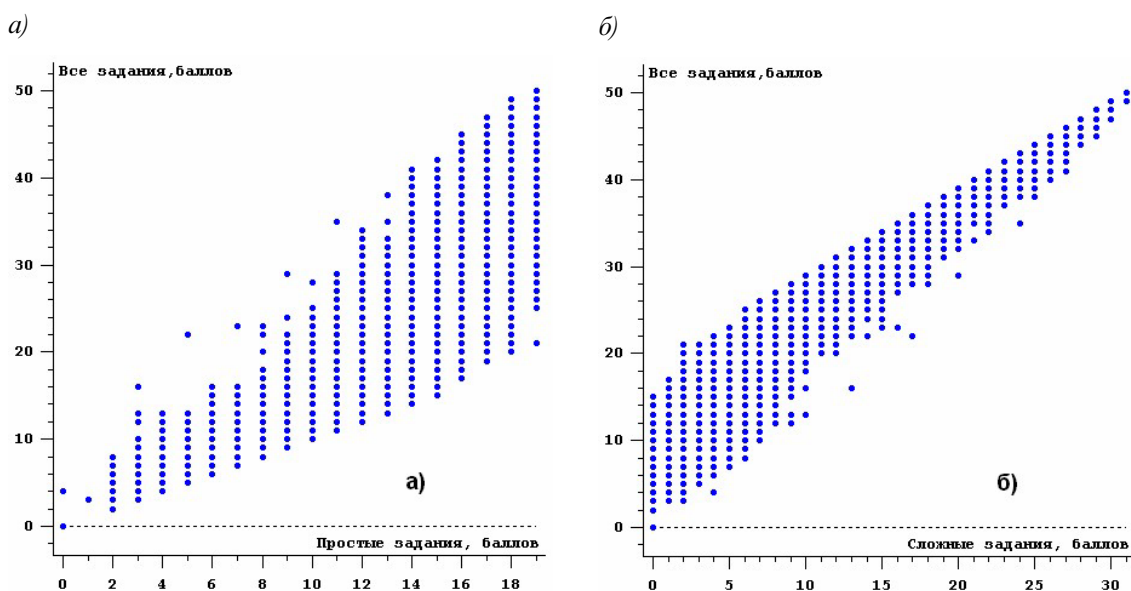


Рис. 3. Зависимость результатов выполнения всех заданий ЕГЭ от результатов выполнения: а) 18 наиболее простых заданий; б) 18 наиболее сложных заданий

нормальным: среднее значение способностей $I_0 = -0.5$ логитов, среднее квадратичное отклонение от этого значения $y = 0.8$ логитов. Эти параметры были заложены в модельное распределение.

В связи с тем, что максимальное количество баллов за задания В1 и В2 составляло 2 балла, каждое из этих заданий заменялось двумя модельными однобалльными заданиями, соответствующими требованиям теории Раша. Аналогично, каждое из заданий части С заменялось тремя модельными однобалльными заданиями, соответствующими требованиям теории Раша.

Результаты моделирования для 10 472 учащихся показаны на рис. 4. Точки – модельная зависимость результатов выполнения «сложных» заданий от результатов выполнения «простых» заданий. Сплошной линией показана зависимость, которая получилась бы при отсутствии разброса, связанного с вероятностным характером выполнения заданий в рамках теории Раша.

Видно, что модельная зависимость (диаграмма рассеяния) на рис. 4 очень похожа на экспериментально измеренную (рис. 2).

На рис. 5 приведены модельные зависимости, соответствующие показанным на рис. 3 экспериментальным. Поскольку эти зависимости также очень похожи, можно считать, что теория Раша на качественном уровне хорошо описывает результаты ЕГЭ.

4. СРАВНЕНИЕ СГЛАЖЕННЫХ РЕЗУЛЬТАТОВ МОДЕЛИРОВАНИЯ И ЕГЭ

На первый взгляд может показаться, что зависимость на рис. 4 представляет широкую линейную полосу, и по ее виду можно было бы предположить, что зависимость линейна. Однако на рис. 4 хорошо видно, что при отсутствии статистического разброса зависимость нелинейная (сплошная кривая), что связано с заметно отличающейся сложностью заданий [5]. Поэтому к линейной аппроксимации такого рода зависимостей необходимо подходить с большой осторожностью: предварительно требуется обосновать, что зависимость линейна.

Для выявления поведения сильно зашум-

ленных зависимостей был разработан метод сглаживания специальным образом отсортированных данных. Данные, представленные на диаграммах рассеяния на рис. 2 и рис. 4, усреднялись по 100 значениям – соответствующие усредненным данным зависимости приведены на рис. 6.

Усреднение проводилось для данных, отсортированных по увеличению суммы $kD_1 + D_2$, где D_1 – баллы за простые задания,

D_2 – баллы за сложные задания, $k = \frac{D_{2\max}}{D_{1\max}}$ –

весовой коэффициент, $D_{1\max}$ – максимальный балл за простые задания, $D_{2\max}$ – максимальный балл за сложные задания. Сортировка осуществлялась с помощью алгоритма быстрой сортировки [6].

Необходимость сортировки данных перед усреднением связана с тем, что усреднению должны подвергаться значения, наиболее близкие к заданной точке (сердине интервала на кривой, для которого производится усреднение). Предложенный подход позволяет реализовать данное требование в случае линейной зависимости. Для нелинейных зависимостей, наблюдающихся на рис. 6, результаты усреднения будут оптимальны только в средней области кривой. В

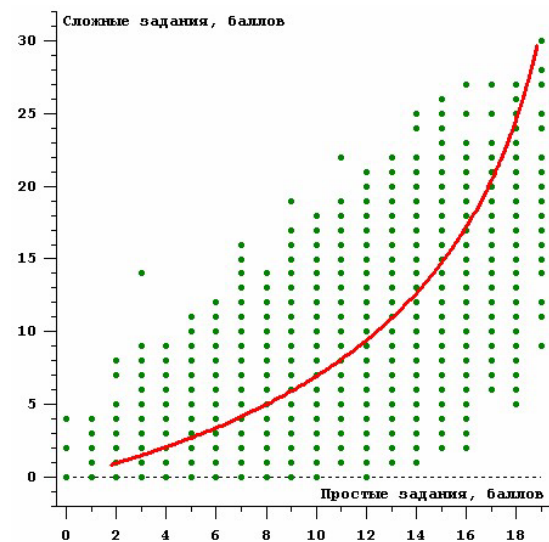


Рис. 4. Точки – модельная зависимость результатов выполнения «сложных» заданий от результатов выполнения «простых» заданий. Сплошная линия – модельная зависимость при отсутствии статистического разброса

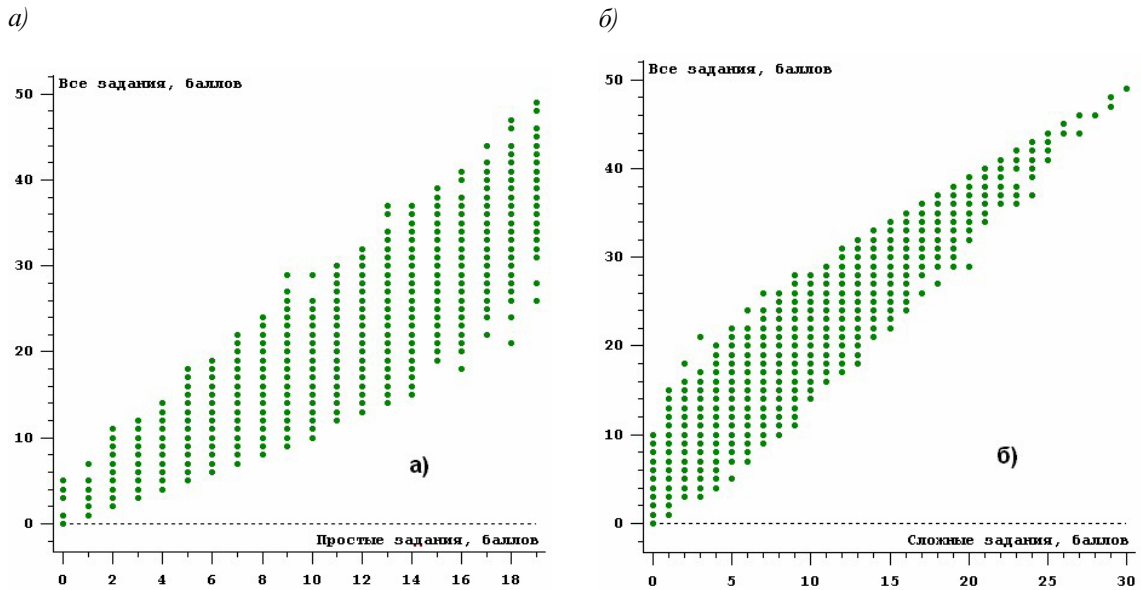


Рис. 5. Зависимость результатов выполнения всех заданий ЕГЭ от результатов выполнения: а) 18 наиболее простых заданий; б) 18 наиболее сложных заданий

области высоких баллов оптимальным будет усреднение по баллам сложных заданий, а в области малых баллов – по баллам простых заданий. Однако в связи с большим количеством участников возрастание погрешности усреднения в областях высоких и низких баллов оказалось незначительно: на

рис. 6 точки в этих областях лежат очень близко к зависимости, показанной сплошной линией.

На рис. 6 приведена модельная зависимость при отсутствии статистического разброса (сплошная линия), а также результаты усреднения результатов моделирования

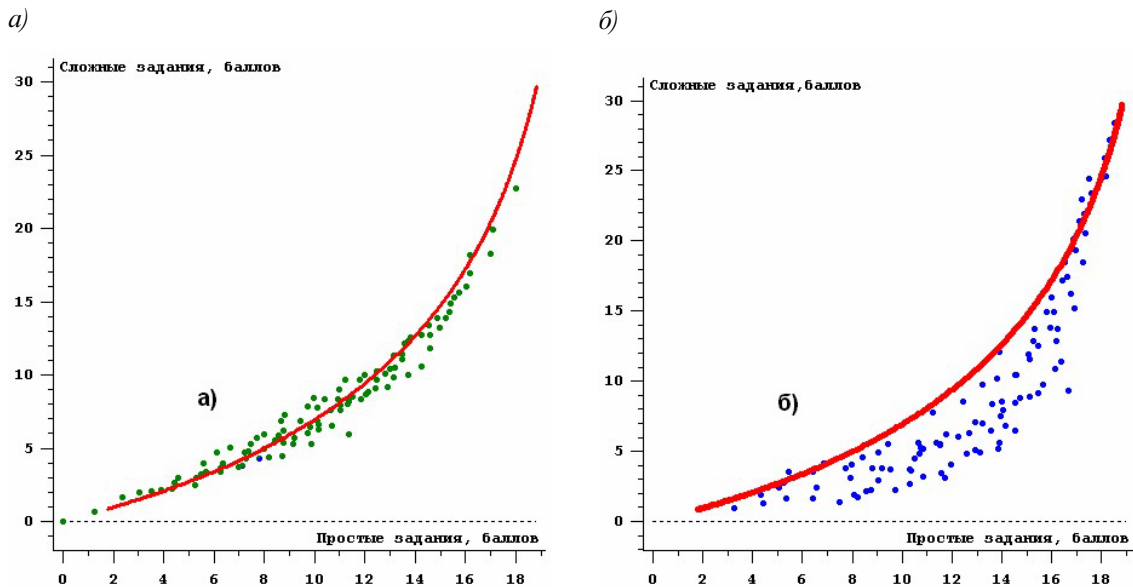


Рис. 6. Точки – зависимость усредненных результатов выполнения «сложных» заданий от усредненных результатов выполнения «простых» заданий:

а) результаты моделирования, б) результаты ЕГЭ.

Сплошная линия – модельная зависимость при отсутствии статистического разброса

(точки на рис. 6 а) и результаты усреднения результатов ЕГЭ (точки на рис. 6 б).

На рис. 6 б полученные в результате усреднения точки в области средних баллов, где выбранный вариант усреднения оптимален, заметно отклоняются от модельной зависимости при отсутствии статистического разброса, и это отклонение явно носит систематический характер. Поэтому можно констатировать, что экспериментальные данные качественно соответствуют результатам моделирования, но количественного соответствия нет.

Изменение параметров распределения участников по способностям показало, что они практически не влияют на характер зависимости, полученной в результате усреднения – он оказывается аналогичным показанному на рис. 6 а. Меняется только плотность расположения точек в тех или иных областях баллов, а сами точки располагаются вокруг сплошной модельной кривой. Таким образом, отклонения точек от модельной кривой на рис. 6 б могут быть связаны либо с заметными отклонениями распределения заданий по сложности от линейных, либо с невыполнением для заданий требований, предъявляемых теорией Раша.

На рис. 7 приведена зависимость сложности задания от его номера для простых и сложных заданий. Прямые – линейная аппроксимация методом наименьших квадратов (МНК) [2].

Линейность зависимостей может быть оценена как удовлетворительная. Моделирование показало, что неравномерность расположения сложности отдельных заданий слабо влияет на форму зависимости и не приводит к искажению ее формы в области средних баллов. Поэтому заметные отклонения точек на рис. 6б от модельной кривой, вероятнее всего, связаны с отклонением параметров заданий ЕГЭ от требований теории Раша. Для заданий частей В и С это вполне естественно, так как теория Раша разрабатывалась для тестов с выбором одного ответа из нескольких представленных.

Для проверки этого утверждения было проведено исследование зависимости для заданий ЕГЭ части А. Из них было выбрано

18 «простых» заданий, которые были разбиты на две группы: 9 простейших заданий и 9 более сложных (наиболее сложных из «простых»). На рис. 8 а точками показаны результаты моделирования, а на рис. 8 б – результаты ЕГЭ. Видно, что в этом случае отклонение точек от зависимости при отсутствии статистического разброса лежит в пределах наблюдающейся на рис. 8 а погрешности, связанной с наличием статистического разброса. Небольшие систематические отклонения от этой зависимости могут быть связаны как с отклонением зависимости на рис. 7 а от линейной, так и с некоторыми отклонениями параметров заданий от требованиями параметрами заданий от требованиями теории Раша.

Таким образом, зависимость для простых заданий ЕГЭ части А по физике в целом можно считать соответствующей теории Раша, в отличие от зависимости для заданий ЕГЭ частей В и С.

5. ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Проведенное разбиение заданий на группы можно интерпретировать как имитацию проведения ЕГЭ в формате базового уровня («простые» задания) и профильного уровня («сложные» задания). Либо как имитацию результатов ЕГЭ («простые» задания) и

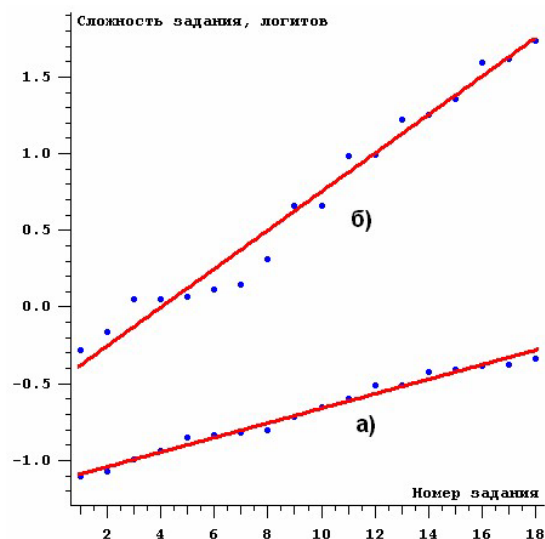


Рис. 7. Зависимость сложности задания от его номера: а) для простых заданий, б) для сложных заданий.

Прямые – линейная аппроксимация МНК

олимпиад («сложные» задания). В дальнейшем будем говорить о сравнении результатов ЕГЭ и олимпиад, подразумевая такую имитацию.

В теориях группы IRT и, в частности, теории Раша, являющейся теоретической основой ЕГЭ, считается, что набранные в результате тестирования баллы позволяют определить способности тестируемого независимо от сложности теста: при использовании теста с заданиями отличающейся сложности должно получаться такое же значение способностей. Однако не случайно при этом рекомендуется, чтобы диапазон сложности заданий примерно соответствовал диапазону способностей тестируемых. Полученные результаты показывают, что как при заметном превышении сложности заданий над способностями, так и при очень легких для тестируемых заданиях основной вклад в получаемые результаты вносит статистический разброс. Этот факт в теориях IRT хорошо известен, однако характер данного разброса и закономерности его поведения в зависимости от сложности заданий до сих пор не исследовались.

Как уже упоминалось, в [1] нами было обнаружено, что способность отлично вы-

полнять простые задания не гарантирует способности выполнять сложные задания, а способность выполнять сложные задания гарантирует способность выполнять простые – но, как правило, не на высший балл. В данном исследовании установлено, что эти особенности могут быть объяснены в рамках теорий IRT и, в частности, теории Раша. Они вызываются нелинейностью зависимости баллов, набранных за сложные задания, от баллов, набранных за простые, а также наличием статистического разброса и его зависимостью от сложности заданий и способностей тестируемых.

Таким образом, имеются основания считать, что не только результаты выполнения тестов с выбором одного варианта из нескольких (задания ЕГЭ части А), но и выполнения заданий других типов (заданий ЕГЭ частей В и С по физике, заданий олимпиад) могут быть объяснены с помощью теорий группы IRT. При этом для зависимостей, основанных на результатах выполнения частей В и С ЕГЭ по физике, наблюдаются отклонения от предсказаний теории Раша, что вполне естественно, поскольку два балла, начисляемые за выполнение задания части В, не могут рассматриваться как баллы

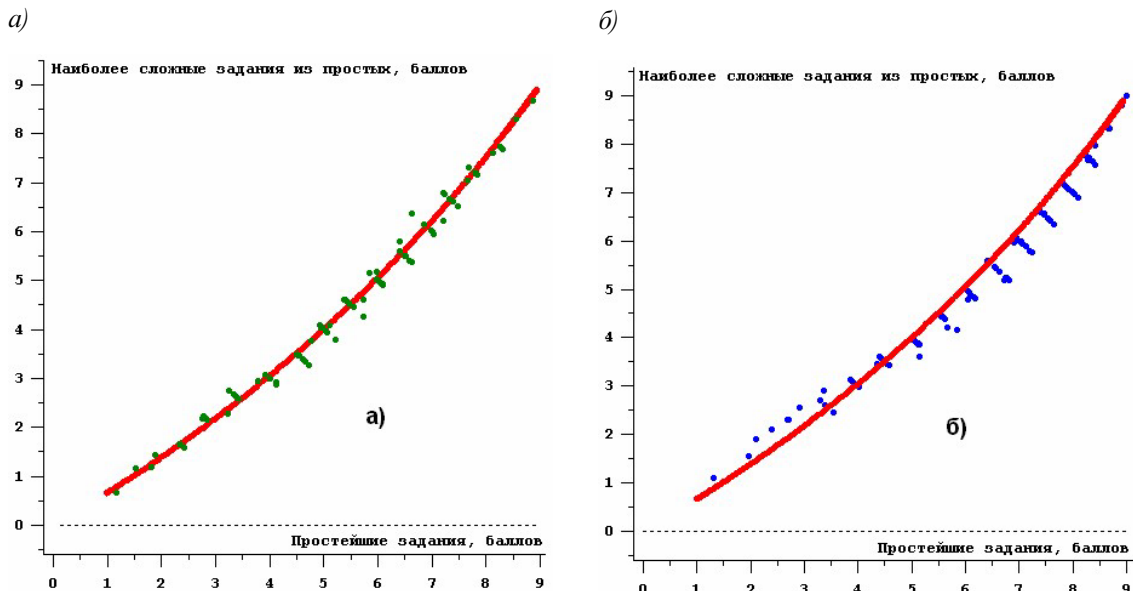


Рис. 8. Точки – зависимость усредненных результатов выполнения наиболее сложных из «простых» заданий от усредненных результатов выполнения простейших заданий:

а) результаты моделирования, *б)* результаты для части А ЕГЭ.

Сплошная линия – модельная зависимость при отсутствии статистического разброса

за два независимых задания, и три балла, начисляемые за выполнение задания части С, не могут рассматриваться как баллы за три независимых задания. Вопрос о теоретическом описании результатов выполнения частей В и С ЕГЭ по физике и заданий олимпиад требует специального рассмотрения.

Полученные результаты имеют важное практическое приложение: они позволяют принимать научно обоснованные решения в дискуссии о приеме в вузы по результатам ЕГЭ и олимпиад. Например, в этой дискуссии встречалось утверждение, что дипломантов олимпиад намного больше, чем число «стобалльников», и что поэтому следует проводить прием только по ЕГЭ. Очевидно, что данное утверждение ошибочно. Для ЕГЭ по физике полоса разброса в области высоких способностей составляет 24 тестовых балла [2]. Поэтому количество участников, получивших 100 баллов, заведомо меньше, чем количество тех, кто обладает равными с ними способностями, но получил меньшие баллы – оно в основном определяется случайностью. Более того, как было показано, выполнение заданий ЕГЭ на 100 баллов не гарантирует способности выполнять более сложные олимпиадные задания, и среди тех, кто получил менее 100 баллов, наверняка найдутся более талантливые учащиеся, чем «стобалльники». Хотя более высокие баллы корректно проведенного ЕГЭ по физике повышают вероятность того, что у учащегося более высокие способности в области физики.

Представляются возможными несколько вариантов предоставления льгот или начисления баллов, не противоречащих полученным результатам.

Во-первых, возможно сохранение имеющейся практики предоставления льгот дипломантам олимпиад – поступление в вуз без вступительных испытаний или зачет диплома олимпиады за 100 баллов ЕГЭ.

Во-вторых, возможно предоставление таких льгот только тем, кто набрал достаточно высокий балл ЕГЭ (например, 70 баллов и выше).

В-третьих, возможно добавление к баллам, полученным на ЕГЭ, баллов за диплом олимпиады. Например, 30 баллов за диплом

1 степени, 22 балла за диплом 2 степени и 15 баллов за диплом 3 степени.

ВЫВОДЫ

Как из экспериментальных результатов, так и из результатов моделирования на основе теории Раша следует, что:

- Очень важным фактором, влияющим на полученные результаты, является статистический разброс. Измерение способностей с помощью ЕГЭ и олимпиад школьников без учета имеющегося разброса не имеет смысла, так как величина разброса сильно зависит от сложности заданий и способностей участников.

- Способность выполнять простые задания соответствует очень широкому диапазону способности выполнения сложных заданий, от низкого до высокого. То есть способность отлично выполнять простые задания никоим образом не гарантирует способности выполнять более сложные задания.

- Способность выполнять сложные задания гарантирует способность выполнять простые задания – но из-за наличия статистического разброса, как правило, не на высший балл.

- Маленькие баллы за ЕГЭ гарантируют неспособность выполнять более сложные задания. Но чем большие баллы учащийся набрал за ЕГЭ, тем больше неопределенность в информации о его способности или неспособности выполнять более сложные задания.

- Высокие баллы ЕГЭ позволяют судить о способности выполнять относительно простые задания, но не дают никакой информации о способностях учащихся выполнять более сложные задания олимпиадного уровня, требующие творческих способностей. Только высокие баллы за олимпиаду дают гарантию способности выполнения сложных заданий. При этом баллы ЕГЭ будут высокими, но в большинстве случаев не самыми высокими.

- Хорошие результаты олимпиад, в отличие от результатов ЕГЭ, достаточны для вывода о способности учащегося хорошо справляться как со сложными, так и с более простыми заданиями. Поэтому имеющаяся

практика льгот дипломантам олимпиад (поступление без вступительных испытаний или зачет диплома за 100 баллов ЕГЭ по соответствующему предмету) вполне оправданна. Но только в случае, если сложность заданий олимпиады превышает сложность заданий ЕГЭ, либо если на олимпиаде проверяются те способности учащегося, которые на ЕГЭ не проверить.

- Результаты выполнения заданий части А ЕГЭ (тестов с выбором одного варианта из нескольких) могут быть объяснены с помощью теории Раша.

- Результаты выполнения заданий других типов (заданий частей В и С) качественно соответствуют предсказаниям теории Раша, но количественно отклоняются от соответствующих зависимостей. Для построения теории, дающей правильные количественные предсказания, требуется проведение дополнительных исследований.

Данные выводы следует учитывать в практике применения результатов ЕГЭ и олимпиад при приеме в вузы, а также при организации профильного ЕГЭ.

Литература

1. Монахов В.В., Монахова С.В. Определение способностей учащихся с помощью олимпиад, тестов и компьютерных моделей // Физическое образование в вузах, 2008. Т. 14, № 3. С. 75–86.
2. Монахов В.В. Анализ результатов ЕГЭ по математике и физике и интернет-олимпиады по физике // Компьютерные инструменты в образовании, 2011. № 1. С. 50–57.
3. Rasch G. On general laws and the meaning of measurement in psychology // in Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, IV. Berkeley, California: University of California Press, 1961. P. 321–333.
4. Монахов В.В. и др. BARSIC: программный комплекс, ориентированный на физика-исследователя // Программирование, 2005. № 3. С. 68–80.
5. Wright B., Stone M. Measurement Essentials. 2nd Edition / Wide Range, Inc., Wilmington, Delaware, 1999. P. 159.
6. Кнут Д. Искусство программирования. Т. 3. Сортировка и поиск. 2-е изд. М.: Вильямс, 2007.

Abstract

The analysis of results of Russian Unified State Examination in physics for 10472 pupils was done as well as modeling on the basis of Rush's theory. The technique of revealing of the form of dependences by means of averaging of in special way sorted data is offered.

It is shown that only results for part A of the examination can be explained by means of Rush's theory. Results for the other types of tasks (parts B and C) agree qualitatively with the predictions of the Rush's theory, but quantitatively deviate from the corresponding dependencies.

Statistical scatter and non-linearity of dependences are very important factors in a comparison of results of tests of different difficulty. It is shown that ability to perform simple tasks corresponds to very wide range of ability to perform more difficult tasks, from low to the high. It is shown that ability to perform difficult tasks guarantees ability to perform more simple tasks.

Keywords: Unified State Examination, assessment, Rush theory, IRT, education, physics, mathematical simulation, software.



Наши авторы, 2011.
Our authors, 2011.

*Монахов Вадим Валериевич,
кандидат физико-математических
наук, доцент кафедры
вычислительной физики
физического факультета СПбГУ,
v.v.monahov@mail.ru*