

Статья опубликована в журнале «Компьютерные инструменты в образовании». 2011. № 1, с. 50 – 57.

Анализ результатов ЕГЭ по математике и физике и интернет-олимпиады по физике

Монахов В.В.

Санкт-Петербургский государственный университет, v.v.monahov@mail.ru

АННОТАЦИЯ

Проведено сравнение результатов ЕГЭ по физике и математике для относительно больших групп учащихся. Наблюдалась высокая корреляция результатов ЕГЭ по физике и математике (0,63-0,73). Среднеквадратичный статистический разброс результатов для каждого из этих экзаменов составлял 6-7 баллов, а зона статистически значимого разброса ± 13 баллов. Наличие этого разброса препятствует использованию ЕГЭ для приема учащихся в ведущие вузы.

Показано, что дистанционные туры интернет-олимпиады в имеющемся формате могут быть использованы в качестве независимого средства мониторинга системы образования. И что очный тур интернет-олимпиады позволяет проверять способности учащихся в диапазоне, наиболее интересном для ведущих вузов и недоступном для проверки с помощью ЕГЭ.

Ключевые слова: ЕГЭ, интернет-олимпиада, аппроксимация, МНК, TLS, тестирование, IRT, теория Раша, теория Бирнбаума, мониторинг, образование, физика, математика

Анализ результатов ЕГЭ по математике и физике и интернет-олимпиады по физике

Монахов Вадим Валериевич
Санкт-Петербургский государственный университет

В настоящее время прием в ведущие вузы осуществляется по двум основным механизмам: по результатам олимпиад школьников и по результатам ЕГЭ. Проводимая нами интернет-олимпиада школьников по физике [1 – 3] дает право на льготное поступление в вузы. В связи с этим встал вопрос о том, какие именно способности проверяются в ЕГЭ по физике, ЕГЭ по математике и в данной интернет-олимпиаде. В частности, в литературе отсутствуют достоверные данные о точности измерения способностей учащихся с помощью ЕГЭ и олимпиад.

1. Аппроксимация результатов ЕГЭ

Для сравнения результатов ЕГЭ по математике и физике строилась зависимость баллов одного испытания от баллов другого (диаграмма рассеяния) для одних и тех же учащихся (рис.1).

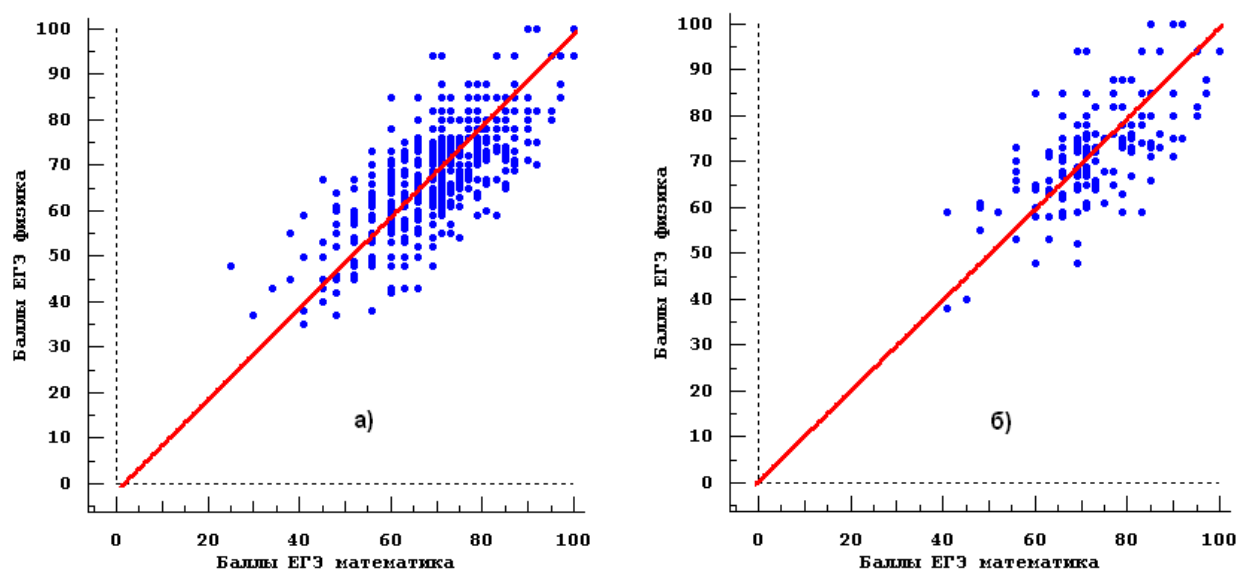


Рис.1. Зависимость баллов ЕГЭ по физике от баллов ЕГЭ по математике: а) для подавших заявление в вуз; б) для поступивших в вуз по результатам олимпиад и ЕГЭ.

На рис. 1,а приведена зависимость для участников, подавших заявление на факультет одного из ведущих вузов России (529 человек), а на рис. 1,б – для поступивших на этот факультет (150 человек). Из рис.1,а видно, что зависимость носит линейный характер. Была построена аппроксимирующая зависимость $y = ax + b$, наилучшим образом приближавшая экспериментальные точки x_i и y_i – набранные баллы по первому и второму испытанию. Аппроксимация проводилась методом наименьших прямоугольников (МНП,

часто называемого также методом RMA – Reduced Major Axis или OLP – Ordinary Least Products) [4, 5] и полных наименьших квадратов (Total Least Squares – TLS, часто называемого также методом MA – Major Axis) [4 – 6].

Для случая, показанного на рис.1, а:

аппроксимация МНП: $a = 1,00$, $b = -1,7$,

аппроксимация TLS: $a = 1,01$, $b = -1,8$,

то есть значение a равно 1 с точностью до процента, а значение b близко к нулю (составляет менее 2% от 100 баллов). Причем существенное изменение выборки не меняет эти значения – для случая, показанного на рис.1, б, получаются совершенно аналогичные результаты:

аппроксимация МНП: $a = 0,99$, $b = 0,2$,

аппроксимация TLS: $a = 0,984$, $b = 0,6$.

При этом значение b оказывается еще ближе к нулю.

Среднеквадратичный разброс точек от аппроксимирующей прямой вдоль каждой из осей составил 8.6 ± 0.4 баллов в первом случае и 9.7 ± 0.8 баллов во втором. Не наблюдалось зависимости величины разброса от значения аргумента, автокорреляция остатков отсутствовала – было выполнено условие гомоскедастичности. В такой ситуации совпадение результатов аппроксимаций МНП и TLS возникает только в случае, когда среднеквадратичные отклонения по оси x и по оси y равны [5]. Форма распределения отклонений была близка к нормальной. Это позволяет считать, что средний разброс баллов для ЕГЭ по физике и математике был в $\sqrt{2}$ раз меньше величины измеренного суммарного разброса и составлял $6,1 \pm 0,3$ балла в первом случае и $6,9 \pm 0,6$ баллов во втором.

Необходимо отметить, что наблюдалась очень высокая корреляция результатов ЕГЭ по физике и математике: для данных, представленных на рис. 1,а, коэффициент корреляции был равен $R = 0,73$. Для данных, представленных на рис. 1,б, он оказался чуть меньше, но все равно достаточно велик, $R = 0,63$.

2. Сравнение сложности заданий ЕГЭ и интернет-олимпиады школьников по физике

В настоящее время имеется ряд теорий, достаточно реалистично соответствующих экспериментальным результатам в области определения способностей учащихся. ЕГЭ основан на простейшем варианте теорий IRT (Items Response Theory), теории Раша [7]. В этой теории исход выполнения задания считается вероятностным. Сложность заданий и способности участников в рамках модели Раша рассматривают на одной шкале,

вероятность P правильного выполнения заданий зависит от разности сложности задания D и способностей участника I :

$$P = \frac{1}{1 + \exp(D - I)} . \quad (1)$$

В рамках данной модели обычно выбираются задания, равномерно отстоящие друг от друга по сложности, и всем заданиям присваиваются одинаковые баллы. Если задание составное, каждая часть рассматривается как отдельное задание. Основой определения способностей в рамках модели Раша является тест с выбором одного варианта из нескольких. В то же время имеются и другие варианты, более удачные с точки зрения проверки способностей (выбор нескольких вариантов из нескольких, ответ в виде числа). Вариант с выбором нескольких вариантов из нескольких используется в сертификационных экзаменах по программированию и в компьютерных тестах на знание английского языка *TOEFL*, и он хорошо себя зарекомендовал. Задания с компьютерной оценкой ответа, вводимого в виде числа, используется и в интернет-олимпиаде по физике, и также хорошо себя зарекомендовали.

Сложность D заданий ЕГЭ оценивалась по стандартной методике в области определения способностей [8] – по отношению числа q участников, не выполнивших задание, к числу p правильно выполнивших:

$$D = \ln \frac{q}{p} . \quad (2)$$

При этом сумма $q + p$ равна числу участников проверки способностей n .

В задачах и моделях интернет-олимпиады использовалась более сложная схема, чем предусмотренная в распространенных вариантах модели Раша: задания оценивались разными баллами в зависимости от сложности, баллы назначались в соответствии с экспертной оценкой. В одной модели или задаче, как правило, присутствовало несколько взаимозависимых заданий разной степени сложности. Причем при отсылке результатов на сервер участнику сообщалось, какие из заданий модели или задачи он выполнил правильно, а какие нет, и при желании он мог переделать неправильно выполненные задания, но с начислением штрафных баллов. Поэтому при проведении анализа результатов очного тура интернет-олимпиады по физике вместо числа правильно выполненных заданий учитывались полученные за задания баллы:

$$D = \ln \frac{B_{\max} - B}{B} , \quad (3)$$

где B_{\max} – максимально возможное количество баллов за задание, B – средний балл, набранный участниками за задание.

Сделанные таким образом оценки являются чисто качественными, однако они дают хорошее представление о том, какие задания оказались для учащихся простыми, а какие – сложными, и насколько сложными.

Значение сложности 0 для задания означает, что одна половина участников с ним справилась, а другая половина – не справилась. Такое задание считается имеющим среднюю сложность. Сложность -2,5 считается чрезвычайно низкой, с ещё менее сложными заданиями справляются все участники. Сложность 2,5 считается при проведении обычного тестирования чрезвычайно высокой, с более сложными заданиями может справиться малое число участников.

Сложность заданий ЕГЭ по физике 2010 года:

- Часть А: сложность от -2,5 (задание А6, чрезвычайно простое) до 0,62 (задание А23, сложность чуть выше средней)
- Часть В: сложность от -1,5 (задание В2, чрезвычайно простое) до 0,66 (задание В3, сложность чуть выше средней)
- Часть С: сложность от -0,06 (задание С2, очень простое) до 0,68 (задание С1, сложность чуть выше средней)

Таким образом, диапазон изменения сложности заданий ЕГЭ для данной группы составляет от -2,5 до 0,68. При этом распределение заданий в каждой части по сложности оказалось близким к равномерному.

Видно, что для рассмотренной группы участников задания ЕГЭ по физике оказались не очень сложными. Это и не удивительно, поскольку на очный тур приглашалось примерно 20% участников интернет-олимпиады, показавших наилучшие результаты в отборочном этапе олимпиады.

Сложность заданий отборочных туров для участников очного тура лежала в диапазоне от -1,1 до 2,3. Таким образом, наиболее простые задания дистанционных туров интернет-олимпиады по физике соответствовали примерно середине диапазона сложности заданий ЕГЭ по физике.

Сложность заданий очного тура интернет-олимпиады 2010 года менялась от 0,67 (задача «Определите скорость капли и скорость встречного ветра», сложность чуть выше средней) до 2,76 (модель «Движение частицы в магнитном поле», очень сложное задание). Таким образом, диапазон изменения сложности заданий интернет-олимпиады для участников очного тура составлял от 0,67 до 2,76 – самое простое задание очного тура интернет-олимпиады по уровню сложности соответствовало самому сложному заданию ЕГЭ по физике.

Следовательно, интернет-олимпиада по физике проверяет способности учащихся в том диапазоне сложности, до которого ЕГЭ не доходит. Это не случайно, ведь олимпиады как раз и предназначены для выявления наиболее способных участников.

3. Анализ погрешности ЕГЭ в определении способностей учащихся

На основе приведенных данных о сложности заданий ЕГЭ автором была разработана программа, моделирующая в рамках теории Раша статистический разброс при прохождении ЕГЭ. Численное моделирование проводилось с помощью программы, написанной в среде BARSIC [9].

На рис. 2 показана модельная зависимость баллов ЕГЭ в зависимости от способностей участника. Шкала сложности смещена так, чтобы её левый край (сложность -2,5) соответствовал 0 баллам.

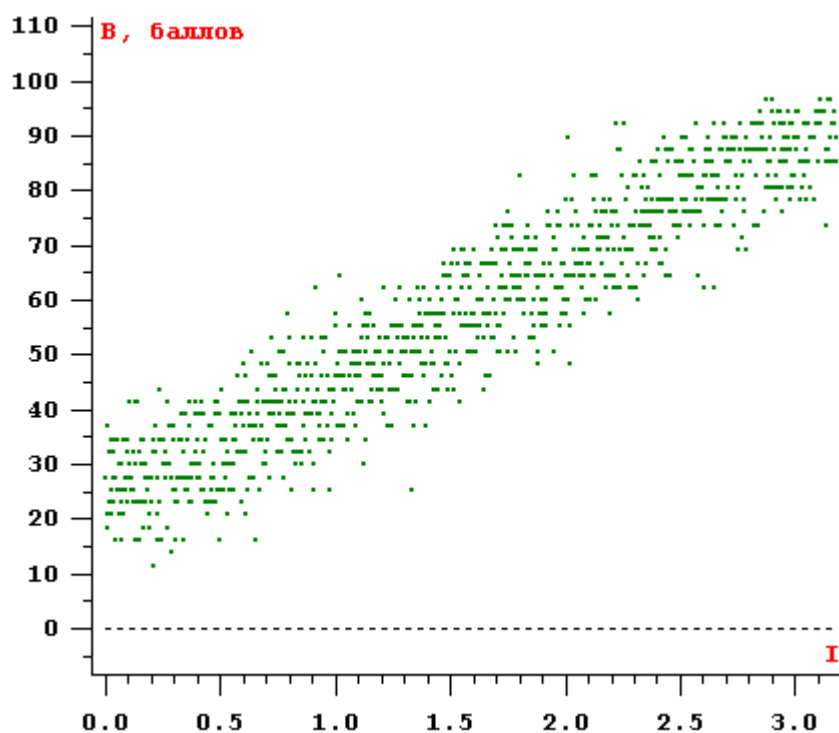


Рис.2. Модельная зависимость набранных баллов ЕГЭ от способностей участника.

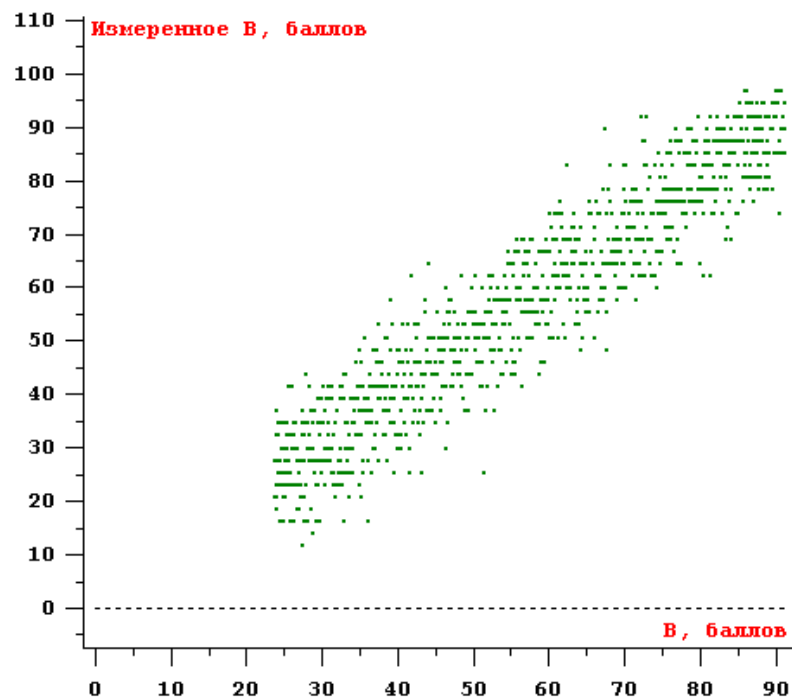


Рис.3. Модельная зависимость набранных баллов ЕГЭ от баллов, которые были бы получены при отсутствии статистического разброса.

На рис.3 показана модельная зависимость баллов ЕГЭ в зависимости от баллов, которые были бы получены участниками с разными способностями при отсутствии статистического разброса (либо от баллов, полученных при усреднении по неограниченно большому количеству испытаний для каждого участника).

На рис. 4 показана ошибка измерения баллов ЕГЭ, возникающая в рамках теории Раша из-за наличия статистического разброса, связанного с вероятностным характером выполнения заданий.

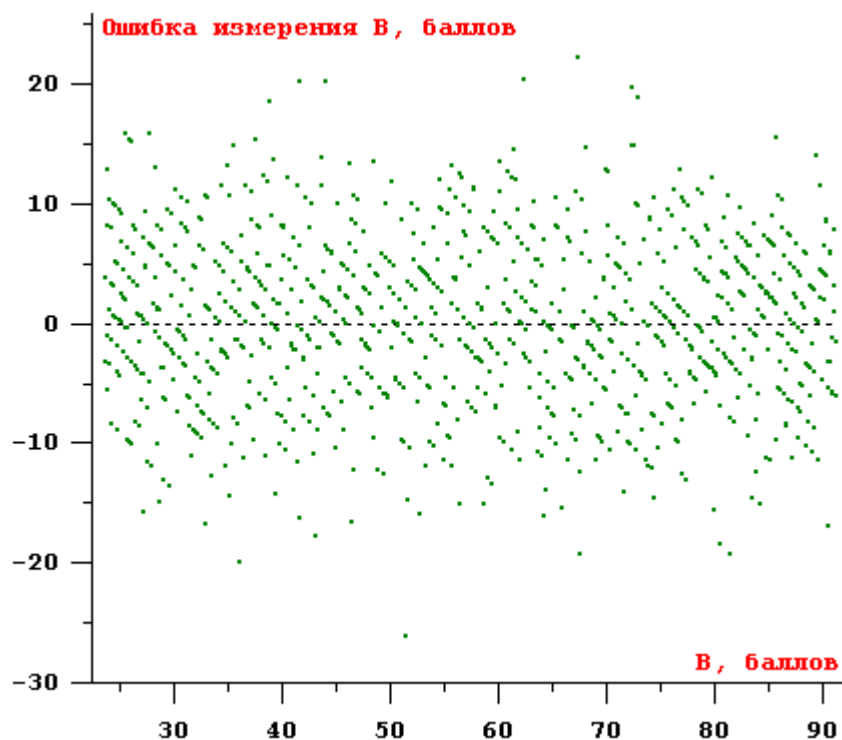


Рис.4. Модельная зависимость погрешности определения баллов ЕГЭ из-за статистического разброса.

При этом среднеквадратичный разброс составляет 7,0 баллов, а статистически значимый разброс 14 баллов (с вероятностью попадания в зону разброса 95%). При заданном диапазоне изменения сложности заданий и равномерном распределении заданий по сложности величина разброса в теории Раша зависит исключительно от числа заданий: чем больше число заданий, тем меньше разброс. Можно считать, что он обратно пропорционален квадратному корню от числа заданий.

4. Обсуждение результатов

В разд. 1 было показано, что коэффициент корреляции между результатами ЕГЭ по физике и ЕГЭ по математике достаточно велик и составляет 0.63 – 0.73. При этом наблюдается среднеквадратичный разброс 6.1 – 6.9 баллов для результатов ЕГЭ по математике и 6.1 – 6.9 баллов для результатов ЕГЭ по физике.

Разброс, приводящий к отличию коэффициента корреляции от 1, может быть связан с наличием трех причин:

- 1) разброс для параллельных вариантов из-за отличия в сложности заданий в этих вариантах;
- 2) разброс из-за неидеальной воспроизводимости результатов для одного и того же варианта;

3) разброс из-за различия предметного содержания ЕГЭ по физике и ЕГЭ по математике.

Полученные результаты показывают, что вклад от статистического разброса, равный 7,0 баллам ЕГЭ, в точности соответствует экспериментально измеренному. Поэтому можно предполагать, что разброс в основном носит статистический характер, а разброс, связанный с различием сложности заданий в параллельных вариантах, а также связанный с различием предметного содержания ЕГЭ по физике и ЕГЭ по математике, мал.

Таким образом, ЕГЭ по математике и ЕГЭ по физике в очень значительной степени проверяют одни и те же способности, несмотря на совершенно различное предметное наполнение заданий. То есть проверяются некоторые метапредметные способности – по-видимому, способности к логическому мышлению и решению задач с использованием математических формул. Не случайно существует термин “физико-математические науки”.

Кроме того, можно говорить о хорошей параллельности вариантов данных экзаменов и правильности методики шкалирования баллов ЕГЭ (при использовании первичных данных коэффициент корреляции был меньше, а результаты аппроксимации были заметно хуже – отсутствовала прямая пропорциональность, и параметры аппроксимации для двух выборок заметно различались).

В то же время полученные экспериментальные и модельные результаты означают, что в рамках ЕГЭ невозможно отличить способности участника с 87 баллами ЕГЭ от способностей участников с 74 баллами и со 100 баллами, поскольку зона статистически значимого разброса составляет ± 13 баллов (с вероятностью попадания результатов в зону разброса 95%). То есть диапазон баллов ЕГЭ, применяемый для приема в ведущие вузы (от 70 до 100 баллов), находится практически полностью в зоне этого разброса. Однако ЕГЭ по физике и математике вполне пригодны для мониторинга системы школьного образования (за исключением проверки в области нижней части шкалы способностей) и для поступления по их результатам в вузы, не претендующие на лидерство, поскольку имеющийся разброс обеспечивает относительно точное различение способностей в области, ширина которой превышает 26 баллов.

При этом олимпиады обеспечивают измерения в том диапазоне способностей, который наиболее интересен для ведущих вузов и которые не могут быть обеспечены с помощью ЕГЭ. Сочетание ЕГЭ и олимпиад перекрывает основной диапазон сложности заданий, который представляет интерес с точки зрения массовой проверки способностей учащихся.

Проведенный анализ основан на двух предположениях: что результаты ЕГЭ описываются теорией Раша, и что сложность заданий интернет-олимпиады может быть оценена по формуле (3). Оба эти предположения нуждаются в проверке.

В частности, предварительные исследования показали, что зависимость вероятности выполнения заданий ЕГЭ и интернет-олимпиады от способностей участников описывается не однопараметрической теорией Раша [7], а несколько более сложной двухпараметрической теорией Бирнбаума [10], и что эти задания обладают разной дифференцирующей способностью [8, 10]. Однако качественная оценка сложности заданий по формулам (2) и (3) мало чувствительна к выбору модели, поскольку характер сложности задания (очень простое, простое, средней сложности, сложное или очень сложное) и не может измениться при смене модели. Поэтому выводы, связанные с качественной оценкой диапазонов сложности заданий ЕГЭ и интернет-олимпиады, не зависят от применяемой для анализа модели.

В то же время оценка статистического разброса результатов, проведенная в разделе 3, напрямую опирается на модель Раша, и переход к теории Бирнбаума может вызвать изменение оцениваемой по модели ширины разброса до двух раз. Поэтому из совпадения модельного статистического разброса с экспериментально наблюдаемым нельзя делать вывод о том, что вклад в экспериментально наблюдаемый разброс полностью определяется статистическим разбросом, связанным с вероятностным характером выполнения заданий участниками. Можно только делать вывод о том, что этот вклад значителен.

Выводы

- Наблюдается высокая корреляция результатов ЕГЭ по физике и математике – эти экзамены в очень большой степени проверяют одни и те же способности, несмотря на совершенно различное предметное наполнение. По-видимому, это способности к логическому мышлению и решению задач с использованием математических формул.
- Среднеквадратичный разброс результатов ЕГЭ по физике и ЕГЭ по математике составляет 6 – 7 баллов, зона статистически значимого разброса составляет ± 20 баллов.
- Результаты моделирования в рамках теории Раша показывают, что основной вклад в разброс, наблюдавшийся на диаграмме с зависимостью баллов, полученных учащимися на ЕГЭ по физике, от баллов, полученных на ЕГЭ по математике, вносит статистический разброс, связанный с вероятностным характером выполнения заданий. При заданном диапазоне сложности заданий он может быть уменьшен только путем

значительного увеличения числа заданий, что нереалистично в рамках экзамена, сдаваемого за один раз.

- Экспериментально наблюдаемая зависимость результатов ЕГЭ по физике от результатов ЕГЭ по математике очень близка к идеальной, предсказываемой в рамках теории Раша. Что позволяет говорить о хорошей параллельности вариантов и правильности процедуры шкалирования баллов, однако для получения достоверных численных оценок требуется проведение анализа параметров заданий ЕГЭ в рамках теории Бирнбаума.
- Из-за наличия большого статистического разброса результатов ЕГЭ по этим результатам невозможно достоверно отличать способности учащихся в диапазоне, наиболее интересном для приема в ведущие вузы.
- Наиболее простые задания дистанционных туров интернет-олимпиады по физике соответствуют примерно середине диапазона сложности заданий ЕГЭ по физике. Дистанционные туры интернет-олимпиады в имеющемся формате могут быть использованы в качестве независимого средства мониторинга системы образования, но только для проверки хорошей и отличной подготовки учащихся.
- Задания очного тура интернет-олимпиады по физике лежат в диапазоне сложности, начинающемся с верхней границы, достигаемой в заданиях ЕГЭ по физике. Причем этот уровень сложности заданий позволяет проверять способности учащихся в диапазоне, наиболее интересном для приема в ведущие вузы.

Литература

1. Монахов В.В. и др. Проведение дистанционных экспериментальных туров олимпиад по физике с использованием программного комплекса BARSIC // Компьютерные инструменты в образовании. 2005 № 2, с. 5 – 15.
2. Монахов В.В. и др. Назначение и опыт проведения интернет-олимпиад по физике // Физическое образование в вузах, 2007, т.13, № 4, с. 53 – 63.
3. 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.
4. Kermack K. A., Haldane J. B. S. Organic correlation and allometry // Biometrika. 1950. V37. No. 1/2, pp. 30 – 41.
5. Ludbrook J. Comparing methods of measurement // Clinical and Experimental Pharmacology and Physiology. 1997. V.24. No 2, pp.193 – 203.

6. *Jolicoeur P.* Linear regressions in fishery research: some comments // *J. of the Fisheries Research Board of Canada.* 1975. 32, pp. 1491 – 1494.
7. *Rasch G.* On general laws and the meaning of measurement in psychology /*Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, IV.* Berkeley, California: University of California Press. 1961, pp. 321 – 333.
8. Нейман Ю. М., Хлебников В. А. Введение в теорию моделирования и параметризации педагогических тестов. М.: 2000. – 168 с.
9. Монахов В. В. и др. BARSIC: программный комплекс, ориентированный на физика-исследователя // Программирование. 2005. № 3, с. 68 – 80.
10. *Birnbaum A.* Some Latent Trait Models and Their Use in Inferring an Examinee's Ability /*Statistical Theories of Mental Test Scores.* Reading Mass.: Addison-Wesley. 1968. Ch.17 – 20, pp. 397 – 479.