

ЕГЭ: ОБЩИЕ ПРОБЛЕМЫ И ПЕРСПЕКТИВЫ РАЗВИТИЯ

Инструментарий

В.В. Монахов

Приказом от 09.04.2002 г. № 1360 Министерством образования и науки РФ было принято Положение о проведении единого экзамена (ЕГЭ). В 2004 году в регионах России начался массовый эксперимент по введению ЕГЭ, и с самого начала проведения не утихают споры о том, полезен ЕГЭ или вреден.

• педагогические измерения • тестирование • IRT • Items Response Theory • теория моделирования и параметризации тестов • модель Раша • модель Бирнбаума • Partial Credit Model • модель частично правильных ответов • образование • физика • способности • компьютерное тестирование

Единый экзамен: убедительные аргументы «за»

Сторонниками ЕГЭ выступают Минобрнауки и представители системы управления образованием, а также ряд экспертов: они выдвигают ряд аргументов, из которых, с точки зрения автора, только три являются убедительными.

Первый убедительный аргумент: ЕГЭ позволяет сохранить единое образовательное пространство. Можно даже сказать, продолжать восстанавливать, что наглядно показало честное проведение ЕГЭ в этом году.

Второй аргумент: ЕГЭ ликвидирует субъективность оценок и позволяет более адекватно оценить знания, умения и навыки выпускников. Этот аргумент подтверждает то, что ЕГЭ поддерживают наиболее «сильные» лицеи и гимназии — руководители, учителя, выпускники. Однако есть и доводы о неадекватности оценок ЕГЭ. В частности, наиболее «сильные» вузы, имеющие высокий международный рейтинг, предпочитают в первую очередь набирать дипломантов олимпиад, и только во вторую очередь берут «высокобалльников» ЕГЭ. Это означает, что в методике учёта баллов ЕГЭ есть принципиальный изъян, и этот аргумент

требует изучения того, когда результаты ЕГЭ применимы для приёма в вузы. Кроме того, массовые проблемы при проведении ЕГЭ в этом году вызывают подозрение, что часть проблем связана не только с низким уровнем подготовки учащихся, но и с самой процедурой ЕГЭ, ориентированной не на выпуск из школы, а на поступление в вуз.

Третий аргумент: ЕГЭ открыл возможность попасть в центральные вузы учащимся «глубинки». И действительно, многие «сильные» учащиеся и их родители из удаленных от крупных городов мест России приветствуют такую систему поступления в вузы.

Единый экзамен: сомнительные аргументы «за» и веские аргументы «против»

Помимо достаточно убедительных аргументов в пользу ЕГЭ часто звучат и другие.

Аргумент о том, что ЕГЭ способствует уменьшению коррупции, не выдерживает критики: очевидно, что коррупция перекочевала из вузов на места проведения ЕГЭ и в Интернет. В этом году была предпринята первая попытка честного проведения ЕГЭ на территории всей страны — и результаты показали реальное

состояние дел, сильно отличающееся от относительно благой картины прошлых лет. Оказалось, что примерно треть выпускников не получила бы аттестат, если бы не был резко снижен порог выставления положительной оценки по обоим обязательным предметам, русскому языку и математике. Честное проведение ЕГЭ во многих регионах стало возможно только после того, как из показателей эффективности работы губернаторов исключили пункт о показателях сдачи ЕГЭ в регионе. Однако трудно надеяться на то, что установка видеокamer может навсегда решить проблему коррупции на местах. Через два-три года произойдёт налаживание новых контактов и коррупционных связей, и нечестная сдача ЕГЭ возродится. Если, конечно, не ликвидировать первопричины проблемы — обязательность сдачи ЕГЭ и проведение ЕГЭ школьными учителями.

Второй не выдерживающий критики аргумент — что якобы лучше иметь один экзамен, который одновременно является и выпускным, и вступительным в вуз, да ещё ЕГЭ используется в качестве средства мониторинга системы образования. Проблема связана с тем, что попытка увязать три разные задачи

вместе приводит к невозможности повысить качество ЕГЭ. Требование к выпускному экзамену — подтверждение того, что ученик усвоил наиболее важные элементы основных тем. Это проверка типа «зачёт» без создания рейтинга. Такого рода проверки называют критериально ориентированными¹. Для них характерен максимально широкий охват тем, но с небольшой глубиной проверки — с простыми заданиями примерно одинакового уровня сложности, соответствующего порогу зачёта. Это должны быть задания типа простых заданий, применяемых в частях А и В ЕГЭ, а значит, они могут проходить в компьютерной форме, с мгновенным получением ответа после прохождения экзамена. Более того, в силу простоты всех заданий можно заметно сократить время прохождения экзамена до 30–40 минут. Помимо прочего, это минимизирует возможности списывания путем «звонка другу» при выходе в туалет, и т.п., а также упрощает процесс пересдачи. В перспективе выпускной экзамен можно проводить в самих школах.

Требования к экзамену для поступления в вуз совершенно другие — это должна быть так называемая нормативно-ориентированная проверка², предназначенная для создания рей-

тинга участников. Для подобной формы проверок характерно наличие линейно возрастающей сложности заданий — от простых, которые выполняют почти все участники, до очень сложных, которые не выполняет почти никто.

При попытке использования критериально-ориентированных проверок для ранжирования (т.е. в качестве нормативно-ориентированных) качество проверки будет низким из-за низкой надёжности результатов. При попытке использовать нормативно ориентированные проверки для получения зачёта (например, сдачи экзамена на окончание школы) не будет обеспечена проверка всех необходимых для получения зачёта знаний, умений и навыков, так как в этом случае задания сильно различаются по сложности, и основная часть учащихся не справляется с большим числом заданий.

Таким образом, попытка совместить выпускные экзамены и вступительные в вуз содержит принципиально несовместимые требования.

Третий сомнительный аргумент — что ЕГЭ обеспечивает реализацию государственных гарантий доступности и равных возможностей получения полноценного образования. Фактически этот аргумент состоит из трёх утверж-

Инструментарий

1

Контроль качества обучения при аттестации: компетентностный подход: учеб. пособие / В.И. Звонников, М.Б. Чельшкова. — М.: Университетская книга; Логос, 2009. — 272 с.

2

Там же.

ПЕД диагностика
ПЕД диагностика

3

Rhoades K., Madaus G.
Errors in Standardized Tests: A Systemic Problem. National Board on Educational Testing and Public Policy. Lynch School of Education, Boston College, 2003. 72 p. <http://www.bc.edu/research/nbeppp/statements/M1N4.pdf>;
Eva L. Baker, Paul E. Barton, Linda Darling-Hammond et al.
Problems with the Use of Student Test Scores to Evaluate Teachers // EPI Briefing Paper #27, 2010, p. 1–27. http://www.imagic.com/eLibrary/ARCHIVES/GENERAL/EPI_US/E100830B.pdf

4

Аванесов В.С.
Единый государственный экзамен в фокусе научного исследования // Педагогические измерения, — 2006. — №1. — С. 3–31. <http://testolog.narod.ru/Ege1.html>;
Аванесов В.С.
Являются ли КИМы ЕГЭ методом педагогических измерений? // Педагогические измерения. — 2009. — №1. — С. 3–26. <http://testolog.narod.ru/EGE26.html>

дений: одно относится к получению аттестата об окончании школы, второе — к обеспечению мониторинга системы школьного образования, третье — к поступлению в вузы. По первому из утверждений: мы уже видели, что форма организации заданий ЕГЭ препятствует качественным выпускным экзаменам зачётного типа. По второму: привлечение к проведению ЕГЭ тех, кто заинтересован в повышении формальных результатов (учителей, представителей учреждений управления образованием, местных властей), неизбежно приводит к искажению результатов. Особенно негативно было влияние результатов ЕГЭ на карьеру губернаторов, а также на результаты аттестации учителей или их зарплату. Подобного рода ошибки (влияние баллов, полученных учащимися, на зарплату учителей) в мире совершались неоднократно и всегда приводили к фиктивному росту успеваемости, с серьёзным расхождением результатов тестирования с реальностью³.

Помимо сторонников у ЕГЭ имеется множество противников. Они делятся на несколько категорий.

Во-первых, это та часть руководителей и преподавателей престижных вузов, которая считает, что ЕГЭ не способен

проверять те способности, которые необходимы студентам их вузов. Это сторонники экзаменов в обычной форме, письменной и устной, а также поступления по дипломам олимпиад из ежегодно пересматриваемого Перечня Российского совета олимпиад школьников (РСОШ).

Во-вторых, это та часть общества, которая считает, что тестирование, используемое в ЕГЭ, это либо «угадайка», либо натаскивание, и поэтому результаты ЕГЭ не имеют смысла. Эти люди считают, что обычные экзамены — более правильный способ выставления оценок. К этой категории противников ЕГЭ относится значительное количество учителей, родителей, преподавателей вузов.

В-третьих, это специалисты в области тестирования. Например, один из известных российских специалистов в области педагогических измерений, главный редактор журнала «Педагогические измерения» В.С. Аванесов считает, что ЕГЭ проводится с нарушением принципов тестирования, и поэтому Контрольно-измерительные материалы (КИМ) ЕГЭ не являются педагогическими измерениями⁴.

Один из очень серьёзных аргументов против ЕГЭ — утверждение, что по многим предметам, в основном гумани-

тарным, задания ЕГЭ не валидны. Валидностью называется степень соответствия того, что измеряет тест, и того, что он должен измерять. На примере демонстрационных заданий ЕГЭ на сайте ФИПИ (Федерального института педагогических измерений) <http://old.fipi.ru/view/sections/92/docs/> легко заметить, что многие задания ЕГЭ по литературе не проверяют способности к литературному творчеству (писать литературные тексты, сочинять стихи и т.п.) — это в значительной степени экзамен не по литературе, а по литературоведению: измеряется нечто, не имеющее отношения к тому, что необходимо измерять. Следовательно, у этого экзамена очень низкая валидность. Ситуация с проверкой валидности заданий ЕГЭ вообще очень плохая — никаких научных данных по этому вопросу нет.

С научным анализом ЕГЭ дело обстоит очень плохо. Все результаты ЕГЭ (даже обезличенные!) закрыты. Действует прямой запрет на разглашение любой информации по результатам ЕГЭ, выходящей за рамки официальных итоговых обобщённых данных. А эти данные не позволяют проводить сколько-нибудь серьёзный анализ качества заданий ЕГЭ и методики создания рейтинга участников ЕГЭ.

Таким образом, ситуация с ЕГЭ очень неоднозначна и требует прояснения. Автор уже несколько лет занимается анализом результатов ЕГЭ. Первоначально интерес объяснялся тем, что автор возглавляет разработку методики и программного обеспечения интернет-олимпиады школьников по физике⁵, и проводилось сравнение сложности заданий ЕГЭ и олимпиад, а также ограничений и проблем ЕГЭ по сравнению с олимпиадами⁶. Написание данной статьи связано с тем, что автор обнаружил ошибочность теоретических представлений о характеристиках заданий ЕГЭ у широкой категории людей, от которых зависит проведение ЕГЭ и его совершенствование: сотрудников ФИПИ (головной организации по подготовке КИМ ЕГЭ), представителей Рособнадзора, а также у критиков ЕГЭ — как политиков, так и представителей научного сообщества.

В связи с тем, что автор занимается конкурирующей с ЕГЭ методикой, интернет-олимпиадой школьников по физике, наличие положительных отзывов о тех или иных элементах ЕГЭ не могут быть приписаны ангажированности или личной заинтересованности автора.

Инструментарий

5

<http://distolymp2.spbu.ru/olymp/> — домашняя страница интернет-олимпиады школьников по физике. См. также Монахов В.В., Ханнанов Н.К. Сравнение интернет-олимпиады по физике с другими формами интеллектуальных состязаний // Дистанционное и виртуальное обучение. — 2011. — №4. — С. 4–19.

6

Монахов В.В. Анализ результатов ЕГЭ по математике и физике и интернет-олимпиады по физике // Компьютерные инструменты в образовании. — 2011. — №1. — С. 50–57. <http://is.ifmo.ru/works/2011/Monahov-KOI-EGE.pdf>;
Монахов В.В. Зависимость результатов изменения способностей учащихся от сложности заданий // Компьютерные инструменты в образовании. — 2011. — № 3. — С. 42–50;
Монахов В.В., Кожедуб А.В., Огинец О.В. Интернет-олимпиада школьников по физике, система образования и информационное общество // Народное образование. — 2014. — №7.

Задания ЕГЭ — тесты или не тесты? И можно ли проверять способности к творческому мышлению с помощью тестов?

Задания ЕГЭ состоят из частей А, В и С. Часто делается утверждение, что ЕГЭ — это тест, и столь же часто слышится возражение, что задания части С не тестовые, и что именно поэтому необходимо увеличивать их долю.

Задания частей А и В имеют классическую форму, использующуюся в тестах. Для части А это вопросы с заданными вариантами ответов, в которых необходимо выбрать один правильный вариант из четырёх имеющихся. В заданиях части В необходимо ввести ответ в виде числа или слова или же указать соответствие для предложенных вариантов.

Задания части С — это задания со свободным развёрнутым ответом. По физике, математике и ряду других предметов это письменный ответ с развёрнутым решением задач, по русскому языку и ряду других предметов это эссе (краткое сочинение на заданную тему).

Исторически сложилось, что под словом «тест» (test в переводе на русский язык — «проверка», «проверочная работа») подразумевается способ проверки знаний, умений и на-

выков за короткое время. С этой точки зрения выполнение теста с выбором одного варианта из нескольких, решение простой задачи, написание короткого сочинения (эссе) или экспресс-экзамен в устной форме — разные формы тестов.

В.С. Аванесов утверждает: «Есть один очень простой метод демаркации тестов от псевдотестов. Если время проверки знаний превышает примерно 40 минут, то этот признак свидетельствует о потере одного из самых существенных свойств теста — кратковременности процесса контроля знаний. Когда, например, говорят о четырёхчасовых «тестах» ЕГЭ по русскому языку или математике, легко видеть, что столь длительное время «тестирования» является убедительным признаком бюрократического выхолащивания самой сути тестового метода»⁷.

Хочется задать вопрос: а если время проверки составляет 41 минуту, что, задание перестанет быть тестом?

В области компьютерных технологий давно сложилась практика сдачи сертификационных экзаменов в виде тестов, основу которых составляют задания в виде выбора одного или нескольких вариантов из нескольких предложенных. Время прохождения таких экзаменов достаточно большое. Например, для сертификаци-

7

Аванесов В.С.

Item Response Theory: Основные понятия и положения. Статья первая // Педагогические измерения. — 2007. — № 2. — С. 3–28. <http://testolog.narod.ru/Theory59.html>

онных экзаменов по языку Java оно составляет 150 минут, т.е. 2 часа 30 мин⁸, а для сертификационного экзамена PRO: Windows Server 2008, Enterprise Administrator может достигать целых 240 минут, т.е. 4 часа⁹. Можно ли считать такие экзамены, проводимые в тестовой форме, тестами? Или же признаваемые во всем мире сертификационные экзамены — «псевдотесты»?

Вопрос о том, какова может быть длительность теста, может быть решен только на основе исследования зависимости результатов тестирования от времени. Если бы начиная с некоторой длительности тестирования наблюдались заметные изменения значений определяемых способностей участников или особенности в результатах, можно было бы считать, что это и есть порог, начиная с которого нельзя считать задания тестовыми.

Аналогично, если бы результаты, получаемые на основе типично тестовой формы проверки способностей, давали бы результаты, принципиально отличающиеся от результатов с заданиями с решениями со свободным развернутым ответом, можно было бы говорить о том, что эти формы заданий проверяют разные способности, и использование заданий в тестовой форме имеет очень ограниченную область применимости.

Автор изучил зависимость от времени результатов участников интернет-олимпиады школьников по физике и не обнаружил никаких отличий в закономерности выполнения заданий на интервале от 10 минут до 2 часов 30 минут. Более того, при снятии ограничений по времени при выполнении тренировочных заданий распределение по баллам мало отличалось от распределения при прохождении туров олимпиады с ограничением по времени. Это означает, что время проведения испытания не определяющий фактор, и что для определения способностей участников важна только валидность и точность измерения, а не время испытания.

Однако есть естественное ограничение: время, необходимое для проведения испытания (тестирования), должно быть больше суммы типичных времён выполнения каждого из заданий средним участником тестирования. При этом под выполнением подразумевается как выдача ответа (правильного или неправильного), так и отказ от выполнения задания. Если даётся меньшее время, участники не успевают дать ответы на все задания, что эквивалентно уменьшению числа заданий теста. Поскольку погрешность измерений увеличивается при уменьшении числа заданий N пропорционально,

Инструментарий

8

Сертификационные экзамены по языку Java <http://edu.softline.ru/certification/java-standard-edition-6-programmer-certified-professional-exam>, <http://edu.softline.ru/certification/oracle-certified-associate-java-se-7-programmer>, <http://edu.softline.ru/certification/java-se-7-programmer-ii>

9

Сертификационный экзамен PRO: Windows Server 2008, Enterprise Administrator <http://edu.softline.ru/certification/pro-windows-server-2008-enterprise-administrator>

ПЕД диагностика
ПЕД диагностика

10

Монахов В.В.

Зависимость результатов измерения способностей учащихся от сложности заданий // Компьютерные инструменты в образовании. — 2011. — № 3. — С. 42–50.

11

Сертификационные экзамены по языку Java <http://edu.softline.ru/certification/java-standard-edition-6-programmer-certified-professional-exam>, <http://edu.softline.ru/certification/oracle-certified-associate-java-se-7-programmer>, <http://edu.softline.ru/certification/java-se-7-programmer-ii>

такое уменьшение времени тестирования приводит к заметному возрастанию погрешности измерений. А время выполнения задания обычно очень сильно зависит от его сложности: чем задание сложнее, тем больше времени требуется для его выполнения. Для уменьшения погрешности измерений стараются увеличить число заданий N , и при 30–50 заданиях (минимальное число заданий, принятое при различных итоговых проверках) хотя бы шесть из них должны быть сложными, требующими в среднем по крайней мере по 10 минут на выполнение. Тогда время только для выполнения этих заданий заведомо должно превышать 1 час. Не случайно характерные времена прохождения сертификационных экзаменов (в тестовой форме!) в области компьютерных технологий составляют от 1 часа 30 мин до 4 часов.

Сравнение первичных баллов «тестовых» частей (А и В) с первичными баллами части С ЕГЭ по физике показало высокую корреляцию результатов — с точностью до разброса значений, связанного с вероятностным характером выполнения заданий (см. далее) и конечным числом заданий, а также нелинейности зависимости результатов из-за различия в сложности заданий¹⁰. Это означает, что в ЕГЭ по физике с по-

мощью заданий частей А и В проверяются те же способности, что и с помощью заданий части С, и часто встречающиеся в прессе слова про «разрушительность» тестовых технологий несостоятельны.

Точно так же несостоятельны утверждения о неспособности проверять с помощью тестов творческие способности — по крайней мере, в области физики, математики и компьютерных наук. Вряд ли кто сомневается, что квалифицированные программисты обладают высокими творческими способностями. И то, что сертификационные экзамены по программированию всегда проходят в форме компьютерного тестирования, однозначно свидетельствует о применимости тестовых технологий для проверки знаний, умений, навыков и способностей к творческому мышлению. Ну, а сомневающимся автор предлагает попробовать пройти сертификационные экзамены по языку Java¹¹, после чего утверждения о неспособности проверять с помощью тестов наличие творческих способностей будут восприниматься не иначе как с иронической улыбкой. Ведь написание сложной компьютерной программы — это классический образец творчества.

Таким образом, качественно составленные тесты позво-

ляют осуществлять проверку не только знаний, но и творческих способностей, а различие между тестовыми и «не тестовыми» типами заданий является условным, что будет подтверждено далее. Требование к короткому времени прохождения тестов также является условностью, поскольку типично тестовые технологии успешно применяются при времени проведения тестирования вплоть до 4 часов. При этом проблемы, связанные с применением в ЕГЭ тестовых технологий, кроются не в том, что сами эти технологии плохи, а в том, что они применяются неадекватно, или же в том, что качество тестов низкое. Кроме того, проблемы применения тестовых технологий в ЕГЭ по одному предмету (чаще всего по литературе) безосновательно переносятся на другие предметы — например, на физику и математику. Также недостатки реализации ЕГЭ (очень высокая погрешность измерений в области низких и высоких баллов) объявляются недостатками использования тестовых технологий. И, наконец, все отрицательные следствия имеющихся проблем в системе школьного образования (связанные со все возрастающей бюрократизацией системы управления образованием, низким социальным статусом учителей, подтасовками результатов и списыванием во

время прохождения экзаменов в значительном числе регионов РФ, и т.д.) начинают приписывать ЕГЭ и тестовым технологиям.

Теоретическая основа ЕГЭ — теория IRT и модели Раша

Теоретической основой¹² для выставления первичных и тестовых (итоговых, после процедуры шкалирования) баллов ЕГЭ является статистическая теория измерения способностей с помощью тестов, разработанная более полувека назад Героргом Рашем (G.Rasch)¹³. На деле под «моделью Раша» подразумевается три разных модели — однопараметрическая модель, двухпараметрическая и политомическая.

Модели Раша — варианты более общей теории — Items Response Theory (IRT)¹⁴.

Замечание: дословный на русский язык «Items Response Theory» — «теория ответов на задания». Оригинальное англоязычное название сложилось исторически, и его можно перевести как «теория моделирования и параметризации тестов», либо как «статистическая теория тестов», либо как «математическая теория тестов», либо как «стохастическая теория тестов». Общепринятое русскоязычное название

Инструментарий

12

Поляков М.А. Организационно-технологическое обеспечение проведения эксперимента по введению единого государственного экзамена в субъектах Российской Федерации // В материалах докладов международной конференции «Развитие национальной системы экзаменов: опыт России, СНГ и США». Москва. — 2003. — С. 49–53. http://stat.edu.ru/reforma/Publ_EGE.pdf

13

Аванесов В.С. Метрическая система Георга Раша // Педагогические измерения. — 2010. — № 2. — С. 57–80. <http://testolog.narod.ru/Theory68.html>

14

Аванесов В.С. Item Response Theory: Основные понятия и положения. Статья первая // Педагогические измерения. — 2007. — № 2. — С. 3–28. <http://testolog.narod.ru/Theory59.html>
Baker F.B. The Basics of Item Response Theory. — ERIC Clearinghouse on Assessment and Evaluation, 2001. — 172 p. <http://echo.edres.org:8080/irt/baker/final.pdf>

теории отсутствует, и её обычно называют по англоязычной аббревиатуре — IRT.

Основой IRT является утверждение о том, что выполнение задания имеет вероятностный характер, и вероятность P правильного выполнения задания зависит от двух характеристик — сложности d задания и способностей u (иногда говорят об уровне подготовленности) участника, проходящего тестирование.

Когда говорится про способности участника, имеется в виду исключительно способность выполнять задания теста. То, какое отношение это имеет к способности учиться, творческим способностям и т.д., требует отдельного изучения. Как уже говорилось, степень соответствия того, что измеряет тест, и того, что он должен измерять, называется валидностью.

Исследования автора показывают, что задания ЕГЭ по физике и математике имеют высокую валидность. Как уже упоминалось, многие задания ЕГЭ по литературе не валидны. Судя по многочисленным замечаниям специалистов, содержат ошибки (а значит, не валидны) многие задания по биологии. Вероятно, не валидны многие задания по обществознанию и истории. Впрочем, есть шанс, что данные претензии необоснованны или гораз-

до менее обоснованны, чем это утверждается в прессе или отдельными специалистами. Однако из-за закрытости, непрозрачности для общества и научного сообщества процедуры анализа результатов ЕГЭ, приходится слепо верить разработчикам КИМ в валидность заданий — что не может считаться научным методом. К тому же голословным утверждениям разработчиков КИМ доверять нельзя, так как даже теоретические основы КИМ ЕГЭ, как будет показано далее, не соответствуют действительности.

Вероятностный характер выполнения задания означает две вещи:

а) если взять очень большое количество заданий одинаковой сложности, то доля решенных заданий конкретным участником будет равна P ;

б) если взять очень большое количество участников с одинаковыми способностями, то доля участников, которые справятся с данным конкретным заданием, будет равна P .

Таким образом, $P = P(d, u)$ — функция, выражающая зависимость вероятности правильного выполнения задания со сложностью d участниками способностями u . Для конкретного задания значение d фиксировано, и функция $P(d, u)$ превращается в зависимость P от u , которая называется ха-

рактической функцией задания. График этой функции называется характеристикеской кривой задания.

В настоящее время IRT получила множество экспериментальных подтверждений и является наиболее активно развивающимся направлением в области теорий педагогических и психологических измерений. Модели Раша — простейшие из моделей IRT.

Способы борьбы с «угадайкой» в ЕГЭ

Как уже упоминалось, одна из главных претензий к ЕГЭ в том, что ЕГЭ является «угадайкой». При этом Рособрандзор под давлением общественности вынужден идти на постепенный отказ от части А¹⁵. Однако подобный отказ от заданий части А обусловлен исключительно голословными утверждениями, не основанными на фактах, и, похоже, специалистам всё-таки удастся отстоять задания части А, хотя и с уменьшением доли этих вопросов¹⁶.

В данной работе были проанализированы задания ЕГЭ на наличие угадывания ответов. В силу специфики заданий части С в них угадывание ответов невозможно в принципе, так что интерес представляли задания частей А и В. Ни в од-

ном задании части В попыток угадывания не обнаружено, хотя в принципе они могли бы быть. С заданиями части А обнаружены ненулевые параметры угадывания для заданий А19, А23 и А25.

Очень большой разброс в области угадывания (способности участников и <0) говорит о том, что двухпараметрическая модель Раша лишь очень грубо описывает ситуацию: часть участников проходят тест честно, а часть занимается угадыванием. В то же время без учёта коэффициента угадывания характеристическая кривая даёт сильно заниженное значение дифференцирующей способности задания (0,82 вместо 1,75). Также наблюдается заметная ошибка в значении сложности задания.

В простых заданиях части А «угадайки» не наблюдается — как, например, в задании А2, имеющем сложность —0,63 логитов. В самом сложном из заданий части А, А23, имеющем сложность 0,95 логитов, угадывание отчётливо выражено. В следующих по сложности заданиях части А, А25 (0,54 логитов) и А19 (0,36 логитов), угадывание менее заметно, но присутствует.

Таким образом, «угадайку» можно устранить очень простым способом — не использовать в части А задания со сложностью $d_j > 0$.

Инструментарий

15

«В Рособрандзоре готовы к постепенному отказу от тестовой части ЕГЭ»

<http://www.rosbalt.ru/main/2014/04/09/1254531.html>;
<http://www.rg.ru/2014/04/09/ege-site.html>

16

«Тестовая часть в ЕГЭ упразднена не будет, сообщили разработчики» <http://ria.ru/society/20140717/1016397512.html>
«Медведев: ЕГЭ должен стать более гуманным, содержательным и понятным» <http://itar-tass.com/obschestvo/1375494>

Выводы

- Теоретическая основа ЕГЭ на основе модели Раша абсолютно не соответствует действительности.
- Реальные параметры всех заданий ЕГЭ по физике 2010 года находятся в очень хорошем соответствии с моделями Бирнбаума (двухпараметрической, трёхпараметрической и Partial Credit Model).
- Распространённые взгляды о недостатках моделей Бирнбаума по сравнению с моделями Раша ошибочны и вытекают из неверных теоретических представлений (основанных на анализе характеристик одного задания, а не на анализе погрешности измерений для всего набора заданий).
- Не наблюдается принципиальной разницы между закономерностями, описывающими результаты выполнения заданий «тестовых» частей А и В ЕГЭ по физике и «не тестовой» части С (в виде письменного решения задачи). Они проверяют одни и те же способности. Единственное принципиальное различие между заданиями частей А, В, С заключается в сложности заданий.
- Мнение о том, что «тесты не способны проверить знания и умения учащихся» ошибочно, по крайней мере, в области физики, математики, компьютерных технологий. Основные про-

блемы подготовки заданий тестов — достижение высокой валидности и, при высокой валидности, обеспечение малой погрешности измерений. Однако на валидность заданий ЕГЭ по ряду предметов разработчики заданий, к сожалению, не обращают внимания. Причиной критики тестовых технологий часто выступает низкая валидность заданий (например, по литературе), а не тестовая технология сама по себе. Однако критики переносят ошибочность подходов разработчиков тестов на саму тестовую технологию.

- Требование к короткому времени прохождения тестов условность. Никакой принципиальной разницы между проведением тестирования в интервале времени от 10 мин до, по крайней мере, 150 мин (2 часа 30 мин) нет. Время, необходимое для тестирования, связано с числом заданий и их сложностью: оно должно превышать сумму средних времён выполнения каждого из заданий средним участником. Если время тестирования меньше этого типичного времени, многие участники не успевают выполнить все задания, и заметно растёт погрешность измерений. Чем сложнее задания и чем больше их число, тем больше времени необходимо отводить на их выполнение. Если давать только очень лёгкие за-

дания, на тестирование можно отводить очень короткое время, но при этом будет катастрофически высокая погрешность измерения в области высоких способностей участников.

- Вопреки несостоятельным утверждениям сторонников модели Раша для повышения точности измерений следует выбирать задания с максимальной дифференцирующей способностью, эти задания не следует отбраковывать.

- Практическая реализация оценивания результатов участников ЕГЭ не использует модель Раша, хотя и была разработана на основе этой модели. Но эта реализация очень хорошо описывается в рамках моделей Бирнбаума и имеет большие преимущества перед предлагавшимися ранее схемами оценивания в рамках моделей Бирнбаума. Таким образом, для адекватного анализа результатов ЕГЭ следует использовать модели Бирнбаума и не бояться, что такой переход вызовет необходимость пересмотра методики подсчёта баллов ЕГЭ.

- Попыток угадывать ответы в заданиях части В ЕГЭ не наблюдалось (хотя они в принципе возможны), а в части А они имеются только для сложных заданий. Поэтому для того, чтобы не было попыток угадывания, у заданий части А должна быть низкая сложность. У

заданий части А большое преимущество — они позволяют проверить знания в очень широкой области при очень малой затрате времени, что позволяет заметно повысить точность измерений в области малой сложности заданий, а также обеспечить широкий охват по темам.

- Имеющийся вариант ЕГЭ имеет очень хорошие метрологические характеристики для участников со средними способностями (от $-1,0$ до $1,5$ логитов), но для «слабых» и «сильных» участников точность измерений катастрофически снижается. Вне области сложности, в которой расположены задания, наблюдается резкий рост погрешности измерений. Чем выше дифференцирующая способность заданий, тем ниже погрешность измерений в области расположения заданий, и тем резче рост погрешности при удалении способности участника от данной области.

- По указанной причине ЕГЭ плохо подходит как в качестве выпускного экзамена для школы, так и в качестве вступительного экзамена в ведущие вузы. А вот в качестве вступительного экзамена в обычные вузы ЕГЭ по физике подходит очень хорошо.

- Распространённая методика подсчёта числа стобалльников ЕГЭ принципиально непра-

ПЕД диагностика
ПЕД диагностика

17

Монахов В.В.

Анализ результатов ЕГЭ по математике и физике и интернет-олимпиады по физике // Компьютерные инструменты в образовании. — 2011. — №1. — С. 50–57.

<http://is.ifmo.ru/works/2011/Monahov-KOI-EGE.pdf>

Монахов В.В.

Зависимость результатов измерения способностей учащихся от сложности заданий // Компьютерные инструменты в образовании. — 2011. — №3. — С. 42–50.

18

Нейман Ю.М.

Основные принципы шкалирования результатов ЕГЭ // В материалах докладов международной конференции «Развитие национальной системы экзаменов: опыт России, СНГ и США». Москва. — 2003. — С. 53–57.

http://stat.edu.ru/reforma/Publ_EGE.pdf

19

Монахов В.В.

Анализ результатов ЕГЭ по математике и физике и интернет-олимпиады по физике // Компьютерные инструменты в образовании. — 2011. — №1. — С. 50–57.

<http://is.ifmo.ru/works/2011/Monahov-KOI-EGE.pdf>

вильна, так как практически все «стобалльники» получают высший балл благодаря статистическому разбросу. Практически все получившие 50 первичных баллов учащиеся имеют такие же способности, как и получившие 48–49 первичных баллов (88–94 тестовых), а иногда и как получившие 46–47 первичных баллов (82–85 тестовых). Поэтому следует говорить о «высокобалльниками», а не о «стобалльниками», и рассматривать диапазон 80–100 тестовых баллов.

- По этой причине для поступления в ведущие вузы ЕГЭ необходимо дополнять вариантом поступления для дипломантов олимпиад РСОШ¹⁷.

- Шкала тестовых баллов ЕГЭ принципиально отличается от шкалы способностей, и, по идее, наиболее правильный вариант оценивания способностей — выражение их в логитах¹⁸. Однако есть произвол как в выборе начала координат на шкале способностей, так и масштабных коэффициентов шкал способностей (из-за нестойкости модели Раша). Поэтому на имеющемся этапе развития теории и практики измерения способностей используемый вариант шкалирования (перевод из первичных баллов в тестовые) представляется наиболее адекватным решением. Тем более что его корректность подтвержда-

ется экспериментальными данными¹⁹.

Рекомендации по изменению ЕГЭ

В связи с изложенным в статье автор считает, что необходимо:

- Разделить выпускной экзамен и вступительный в вузы, сделав сдачу ЕГЭ на профильном уровне по всем предметам добровольной. Каждый вуз должен сам принимать решение о том, какие предметы профильные для его образовательных направлений и программ. При поступлении должно учитываться наличие аттестата об окончании школы и баллы по профильным предметам, а также льготы дипломантам олимпиад. Выпускной экзамен для получения аттестата о среднем образовании должен обеспечивать проверку на базовом уровне.
- Обеспечить создание открытых банков заданий ЕГЭ по всем предметам. Провести работу по калибровке этих заданий по сложности и дифференцирующим способностям.
- Обеспечить открытость обезличенных результатов ЕГЭ на уровне регионов РФ. Это позволит выявлять как проблемные регионы, так и обеспечивать проверку качества заданий ЕГЭ.
- Необходимо организовать научную проверку валидности

заданий ЕГЭ по всем предметам. После проведения этой проверки для каждого предмета либо будет доказано, что для него форма и содержание ЕГЭ адекватны, либо будет доказано, что для данного предмета необходимо вносить изменение формы и содержания ЕГЭ. При этом не следует бездоказательно обобщать проблемы ЕГЭ, имеющиеся для какого-либо предмета, и переносить их на другие предметы, по каждому предмету должен проводиться независимый анализ.

- Создать круглогодично работающие центры по сдаче профильного ЕГЭ по разным предметам, оборудованные видеоканерами, с отправкой видеосъемки лица тестируемого и процесса тестирования в центральный архив, без возможности сотрудников центра влиять на то, что отсылается в архив.

- Проводить профильное тестирование по частям А и В в компьютерной форме с выдачей результатов сразу после тестирования (от частей А и В отказываться нельзя, так как это заметно снизит точность измерений, но в части А не должно быть сложных заданий).

- Проверка заданий части С должна проводиться круглого-

дично специально подготовленными высокопрофессиональными специалистами, с независимой проверкой каждой работы не менее чем тремя специалистами. Также должна проводиться автоматическая оценка качества работы этих специалистов по корреляции результатов их проверки с результатами специалистов, параллельно им проверяющих те же работы, а также корреляции с результатами выполнения заданий частей А и В.

- Часть из предложенных мер уже выполняется, это движение в правильном направлении. Однако ЕГЭ не является самоцелью и не может заменить школьное обучение. При отсутствии в старших классах сельских школ и малых населенных пунктов профильного обучения шансы на успешную сдачу учащимися профильного экзамена становятся призрачными. Поэтому одно из важнейших условий для того, чтобы введение двух уровней сдачи ЕГЭ не привело к деградации старшей школы — обеспечение в каждом населенном пункте (желательно — в каждой школе) профильного обучения для всех, кто хочет после окончания школы поступать в вуз.

Инструментарий