

Методология

ПРОБЛЕМА ДЕМАРКАЦИИ ПЕДАГОГИЧЕСКИХ ИЗМЕРЕНИЙ

Вадим Аванесов
testolog@mail.ru

Цель данной работы — разработка критериев демаркации (отграничения) педагогических измерений от псевдоизмерений.

Проведение Единого государственного экзамена (ЕГЭ) актуализировало проблему демаркации научных методов педагогического измерения от методов псевдонаучных. В сфере педагогических измерений общепринятым в мире методом является тест. В России был выбран иной путь: вместо тестов стали применять так называемые контрольно-измерительные материалы (КИМы). Их называли тестами, хотя имеющаяся по ним скудная информация не соответствует тестовым стандартам и критериям. Измерительные свойства этих материалов оказались неудовлетворительными. Президент РФ Д.А.Медведев недавно заявил о необходимости разобраться в сути проблем, возникающих с проведением ЕГЭ¹, и о корректировке² этого экзамена.

Главной проблемой этого экзамена являются некачественные КИМы, поэтому надо прекратить их производство. Аттестацию выпускников и приём в вузы лучше разделить, вместо ЕГЭ полезнее ввести добровольное негосударственное тестирование абитуриентов вузов. В образовательном процессе применение заданий в тестовой форме рекомендуется дополнять олимпиадами.

1

Дмитрий Медведев поддерживает идею о президентской комиссии по ЕГЭ. http://www.rian.ru/edu_egrus/20090810/180375075.html

2

А. Чёрных, Х. Дачаева, М. Юриша, А. Саргин
Президент поставил оценку системе ЕГЭ. 31.08.2009. <http://gzt.ru/topnews/business/258180.html>

ПЕД
измерения

3

Интерференция (от лат. inter – взаимно, между собой и ferio – ударяю, поражаю) – взаимоподавление одновременно осуществляющихся процессов, прежде всего относящихся к познавательной сфере.

4

Аванесов В.С.

Форма тестовых заданий. М.: Центр тестирования, 2005.156 с.

5

Факты расследования массовых фальсификаций и коррупции представлены в публикации Е. Мельник и О. Герасименко: «ЕГЭ получил хорошие расценки». http://www.gazeta.ru/2007/04/05/oa_235842.shtml

7

«Левада-центр» опубликовал 19.05.2009 результаты опроса общественного мнения по поводу Единого государственного экзамена (ЕГЭ). Социологический опрос показал отрицательное отношение к ЕГЭ большинства респондентов. 36% россиян категорически против ЕГЭ. Оказалось, что только 16% россиян отдадут предпочтение ЕГЭ по сравнению с обычными экзаменами.

4

*Какая-то подделка
под науку...
присвоила себе
название знания
и ставила возврату
последнего
почти неодолимые преграды.*
Ж.-Ж. Руссо

Проблемная ситуация

В сфере педагогических измерений бесспорным по качеству измерения и общепринятым методом является тест. Не случайно тестирование применяется в США, странах Латинской Америки, Европы и Азии. По сравнению с научно обоснованными тестами, все другие методы оценки знаний оказываются более субъективными, подверженными коррупции, менее точными и эффективными. Научной основой тестирования являются педагогические измерения.

На фоне общемирового процесса ускоренного развития теории и практики педагогических измерений в России возникли интерферентные³ тестам КИМы ЕГЭ, состоящие из трёх неподходящих форм – традиционных экзаменов с текстовыми задачами и двух форм, похожих на тестовые, но не соответствующие всем требованиям теории⁴.

Всенародную критику КИМов Министерство образования РФ все годы т.н. эксперимента по введению ЕГЭ в штатный ре-

жим игнорировало, а подлинные результаты скрывало. Тем не менее, коррупционность и некачественность этих материалов стали широко известными. В итоге на сегодняшний день сложилась ситуация заметного отторжения КИМов и ЕГЭ педагогической общественностью и населением страны в целом. Можно думать, что именно коррупционность⁵ и низкое качество безликих материалов послужили главной причиной того, что популярность ЕГЭ в массах стала снижаться год от года⁶.

Социологический опрос, проведённый фондом «Общественное мнение» 20–21 июня 2009 г., показал, что не одобряют введение ЕГЭ в школах теперь уже 47% опрошенных, а одобряют – только 21% россиян.

Хотя в КИМах ЕГЭ часто применяется лексика педагогических измерений, настоящих тестов там нет. Зато появилось множество бессмысленных слов и словосочетаний типа «тестотехник», «тесты ЕГЭ» и т.п., которые не имеют ни научного смысла, ни научного определения. Этот новояз создаёт видимость науки там, где её нет. Ввиду связи, существующей между качеством контроля и качеством образования, стало понятно, что без чётких разграничений педагогических измерений от прочих форм оценивания образование в России не выйдет на требуемый

уровень. Каков контроль, таково и образование.

Президент РФ Дмитрий Медведев поддержал идею создать комиссию по анализу результатов ЕГЭ (http://rian.ru/trend/exam_2009/). «На мой взгляд, есть весьма позитивные моменты, есть и негативные моменты. В целом я считаю, что в этой теме мы должны продолжать разбираться. Я поддерживаю идею создать отдельную, может быть, комиссию, которая проанализирует все плюсы и минусы ЕГЭ», — сказал он на встрече с лидерами думских фракций⁷. Про плюсы, минусы и о необходимости улучшения ранее говорил и председатель Правительства РФ В.В. Путин⁸.

Председатель Совета Федерации Федерального собрания РФ С.М. Миронов считает, что тема Единого государственного экзамена продолжает оставаться одной из самых острых и спорных и, наверное, одной из самых обсуждаемых в обществе. В этой связи он считает правильным обобщить итоги не только прошедшего в 2001–2008 гг. эксперимента, но и обобщить результаты единого государственного экзамена, проведённого в 2009 году. Нужно принципиально решать вопрос о будущем единого госэкзамена и об ответственности тех, кто его продвигал и организовывал. <http://viperson.ru/wind.php?ID=592021&soch=1>).

Первый заместитель председателя Правительства РФ И.И. Шувалов недавно заявил, что по мере выявления недостатков и элементов коррупции Правительство и Минобрнауки РФ будут корректировать механизм проведения ЕГЭ⁹.

Из приведённых мнений руководителей страны становится понятным наличие заметной политической составляющей в рассматриваемой проблемной ситуации.

Позиция автора относительно преодоления данной ситуации такова. Идея корректировки этого экзамена вряд ли станет эффективной без признания конкретных ошибок в планировании и проведении ЕГЭ. Ведь социуму непонятно и неизвестно, что и почему власть собирается корректировать, а без осознанной поддержки общественности никакая коррекция ЕГЭ не может.

Обнародовать ошибки и недостатки ЕГЭ важно ещё и потому, что сама идея коррекции трудно соотносится с имеющимися реалиями в стране и в сфере образования. Главные причины неудач видятся в ошибках формулирования целей ЕГЭ¹⁰ и в несоответствии КИМов этим целям. Несоответствие КИМов декларированным целям усугубилось их недопустимо низкими метрическими свойствами и нетехнологичностью, хотя эти КИМы изначально выдавались за сред-

Методология

36% отзываются о нём резко отрицательно, причём в основном это студенты и люди с высшим образованием. Тогда как ВЦИОМ ранее настаивал на том, что против ЕГЭ выступают исключительно двоечники, неврастеники и учителя-реформаторы. <http://www.dailyonline.ru/m/4627>

7

Дмитрий Медведев поддерживает идею президентской комиссии по ЕГЭ. http://www.rian.ru/edu_egrus/20090810/180375075.html

8

Путин В.В. Система ЕГЭ будет совершенствоваться. Правительство РФ будет совершенствовать систему единого госэкзамена (ЕГЭ). <http://www.eduhelp.ru/page.php?pageid=1043,15.02.08>.

9

Шувалов И.И. Говорить об эффективности или неэффективности ЕГЭ пока рано. <http://www.regnum.ru/news/1205987.html>, 16.09.2009.

ство педагогических измерений, не будучи таковыми по существу¹¹.

Конструкция КИМов ЕГЭ основана на ложной идее преимущественной значимости формы, а не трудности заданий, что противоречит сущности теории педагогических измерений. Об этом автор этой статьи писал с самого начала эпопеи с ЕГЭ¹². Используемые формы сильно подвержены коррупционным искажениям, что массово подтверждается на практике и имеет тенденцию к усилению. В этих условиях поддерживать идею корректировки КИМов и ЕГЭ при всём желании нет смысла. Частными корректировками эту ситуацию вряд ли удастся переломить. Нужны кардинальные решения.

Есть ли надежда, что со временем инструменты ЕГЭ будут обладать хотя бы частью свойств тестовой технологии? Сомнительно, если иметь в виду фактическое отсутствие теоретической основы разработки КИМов ЕГЭ, наличие множества явных и неявных целей и задач, эклектичность форм в их конструкции, а также, самое главное, реальный многолетний опыт их некачественного производства.

Ситуация с ЕГЭ очень похожа на положение с выпуском некачественных автомобилей. Это положение много лет пытались исправить непомерными затратами и несбывшимися надежда-

ми. Пока не поняли, что в наш прагматический век нужно проводить не коррекцию выпуска некачественных автомобилей, а решительное перепрофилирование автозаводов на выпуск нужной продукции.

В Белоруссии, Украине и в большинстве других стран бывшего СНГ, а также во вновь образовавшихся государствах — Абхазской, Нагорно-Карабахской, Приднестровской Молдавской и Юго-Осетинской республиках проявлялась тяга к внедрению аналога российского ЕГЭ, но в последнее время там стали понимать ошибочность замены полноценного тестирования КИМами ЕГЭ. Хотя многие по сей день полагают, что экзамен, КИМы и тест — это одно и то же, и что форма проверки не имеет принципиального значения. Но это — большая ошибка.

В России сейчас применяются следующие основные формы и методы проверки уровня подготовленности:

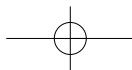
- КИМы ЕГЭ;
- олимпиады;
- традиционные экзамены;
- текущее и итоговое оценивание.

Эти формы, не являясь, в научном смысле, методами педагогических измерений, так или иначе противопоставлены тестированию, которое при должном уровне работы, напротив, обладает свойствами педагогических измерений. Но настоящие педагогические тесты в России

10
Цели ЕГЭ исследовались в статье: Аванесов В.С. Единый государственный экзамен в фокусе научного исследования // Педагогические измерения №1, 2006, стр. 3–31.

11
Подробнее о педагогических измерениях см. в работах автора, опубликованных в журнале «Педагогические измерения» за 2004–2009 годы. А также на сайте автора <http://testolog.narod.ru>

12
Аванесов В.С. Мифология ЕГЭ. Газета «1 Сентября» №50/2001. <http://ps.1september.ru/article.php?ID=20010500>



сейчас практически не применяются. Усилиями чиновников и разработчиков ЕГЭ культура педагогических измерений оттеснена на периферию. Между тем она является теоретической основой разработки тестов. Подобные явления страна уже переживала в тридцатые годы XX столетия.

Результат — отставание по этой проблеме от многих стран мира. Полноценное возрождение тестового метода и педагогических измерений, как прикладной педагогической науки в России, так и не состоялось. Способы оттеснения — подмена педагогических измерений псевдоизмерениями, смешение лексики различных форм, келейность обсуждения значимых для страны вопросов, выдвижение на руководящие должности некомпетентность лиц, принимающих ошибочные решения. В общем, получилось так, как писали классики: в истории всё повторяется дважды: первый раз в виде трагедии, второй — в виде фарса.

Что такое демаркация?

Научной формой противостояния упомянутым формам псевдоизмерения может стать демаркация педагогических измерений от этих форм.

Демаркацией называется процесс разграничения научных теорий и методов от псевдонаучных и ненаучных. Необходимость в демаркации возникает

там, где научные теории, понятия и методы не различаются или намеренно подменяются. В России сейчас важно отграничить педагогические измерения от псевдоизмерений.

Псевдоизмерения можно определить как процесс производства и применения методов, позволяющих создать лишь внешнюю видимость педагогических измерений. Такого рода видимость чаще всего проявляется в ложных названиях, типа «тесты ЕГЭ», хотя там никаких тестов тоже никогда не было и нет. Не случайно КИМы не имеют научного определения и научно обоснованных свойств метода измерения. В сложившейся ситуации подменять тесты методами оценивания теперь уже больше нельзя. Именно на это автору данной статьи пришлось обращать внимание участников международной конференции в Харькове¹³.

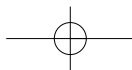
Ярким примером псевдоизмерений являются КИМы ЕГЭ. Там используются три формы заданий. Напомним, что в части «А» КИМов используются задания с выбором одного правильного ответа из четырёх-пяти ответов, предлагаемых на выбор. Из-за высокой вероятности угадывания правильных ответов в таких заданиях большинство испытуемых получают незаслуженно завышенные баллы. И такой метод, не моргнув глазом, чиновники называют «объектив-

Методология

Методология

13

Аванесов В.С.
Проблема становления
Национальной системы
педагогических измерений.
Харьков, 13–14 декабря,
2007.
[http://testolog.narod.ru/
Education19.html](http://testolog.narod.ru/Education19.html)



ПЕД
измерения

14

Завышенные результаты многих работ связаны, прежде всего, с качеством проверки местными экспертами, интересом последних к завышению результатов. См. статью <http://gzt.ru/search/?author=32>. «На 100 баллов ЕГЭ сдают только вундеркинды». 23.07.2009 <http://gzt.ru/topnews/education/250901.html>

15

Ковалёва Г.В.

Не впереди планеты всей». Народное образование. № 5. 1998.

16

Об этом и других преимуществах централизованного тестирования по сравнению с ЕГЭ см.: Аванесов В.С. Централизованное тестирование лучше Единого государственного экзамена. Тез. докл. Всерос. научно-методической конференции «Развитие тестовых технологий в России». /Под ред. Л.С. Гребнева. М.: Центр тестирования Мин. образования РФ. 2002. М.: 21–22 ноября 2002 г. с. 71–73. <http://testolog.narod.ru/Education27.html>

ным». Субъективное оценивание используется в третьей части КИМов ЕГЭ. Именно в этой части заметна необъективность выставляемых оценок¹⁴. Они недостаточно коррелируют с оценками по другим частям КИМов ЕГЭ.

Похоже, что для подмены педагогических измерений КИМами ЕГЭ условия создавались заранее. Вот цитата: «...Не следует забывать, — писала Г.В. Ковалёва, — что от «чистого тестирования», т.е. тестов, в которые включены только задания с выбором ответа, отказываются все страны мира. И очень жаль, что Центр тестирования и другие используют только традиционную форму контроля, навязывая стране старые методики. Причём в свои тесты они включают, как правило, задания на проверку знаний фактологического характера или расчётные задачи»¹⁵.

Здесь неправда оказалась смешанной с полуправдой. Качественные тесты применялись и продолжают применяться во всём мире, кроме России. Раскритикованные некачественные задания с выбором одного правильного ответа продолжают всё-таки применять во всём мире, и даже, как это ни удивительно, они используются в первой части КИМов ЕГЭ. Хотя это совсем не тесты, а лишь худшая форма заданий с выбором одного правильного ответа, где очень высокая вероятность угадыва-

ния правильного ответа. В работе Центра тестирования действительно было немало ошибок, критиковавшихся больше других автором этой статьи. Но не следует забывать главное: до прихода ЕГЭ в Россию тестирование было добровольным¹⁶, а теперь мы получили принудительный и коррупционный госэкзамен с некачественными КИМами.

Проблема демаркации получила развитие в XX веке, в трудах участников Венского кружка. Внимание философов было привлечено к вопросам построения системы эмпирических показателей (индикаторов), позволяющих подтвердить или опровергнуть правильность тех или иных теоретических построений (конструктов). Эти показатели участники кружка называли эмпирическими референтами теоретических концепций (конструктов). Демаркация позволит сохранить и развивать педагогические измерения, не смешивая их с КИМами, а также с аттестацией, сертификацией и т.п. формами оценивания уровня подготовленности личности.

Постановка проблемы

Применительно к педагогическим измерениям проблема демаркации, насколько известно, в явном виде, как в данной статье, ранее не ставилась и не обсуждалась. Исследование этой пробле-

мы касалось преимущественно вопросов обоснования надёжности и валидности тестовых результатов — этой важнейшей части демаркации измерений от прочих форм. Другие вопросы демаркации — это вопросы объективности¹⁷, справедливости и эффективности тестовых результатов¹⁸ — специально исследовались автором этой статьи.

Далее следовала серия статей, в которых были подвергнуты научной критике ЕГЭ¹⁹, КИМы²⁰ и была сделана попытка отграничить язык педагогических измерений от языка псевдоизмерений²¹. Научный анализ вопросов языка педагогических измерений входит в исследуемую проблему, поскольку каждое понятие позволяет отграничить тест от псевдотеста. Данная статья подводит итоги проделанной работы по демаркации педагогических измерений от псевдоизмерений. Проблема демаркации теории и практики педагогических измерений от различных форм оценивания стала особенно актуальной в России после решения о введении в стране ЕГЭ.

Автор этой статьи с первых дней эпопеи с ЕГЭ обращал внимание на отсутствие какой-либо внятной и кем-либо одобренной научной концепции или какого-либо проекта в основе ЕГЭ. Кроме того, обращалось внимание на отсутствие научной экспертизы самой идеи ЕГЭ и КИМов, имеющих результаты, публика-

ций по реальным результатам этого экзамена, на ненадёжность и несправедливость этого экзамена, на очень высокий уровень погрешности индивидуальных результатов испытуемых... Многократно говорилось, что по результатам таких КИМов принимать решения относительно социального и профессионального развития личности нельзя. Но всё было тщетно²².

Хотя некоторые дефекты были очевидны. Например, используемые формы допускают угадывание правильных ответов незнающими испытуемыми в первой части. В заданиях с выбором одного правильного ответа можно угадать, по вероятности, примерно четвертую или пятую часть всех правильных ответов.

В заданиях открытой формы возникает много трудностей с пониманием смысла заданий. Испытуемые разного уровня подготовленности по-разному понимают смыслы этих заданий, что вводит в предмет измерения такие дополнительные факторы, как уровень интеллектуальности, память, а это явный предмет не педагогических, а психологических измерений. Смыслы примерно четверти заданий второй части ЕГЭ не понимают даже учителя²³. А это логические дефекты, которых в КИМах ЕГЭ слишком много.

Задания третьей части КИМов за пределами трудные. В силу этого фактора они не стыку-

Методология

17

Аванесов В.С.
Проблема объективности педагогических измерений. Педагогические измерения, №3, 2008. С. 3–40.

18

Аванесов В.С.
Проблема эффективности педагогических измерений. Педагогические измерения, №4, 2008. С. 3–24.

19

Аванесов В.С.
«Единый государственный экзамен, или Эта вещь будет сильнее Фауста Гёте» // Учительская газета, № 49, 28 ноября 2000 года. Результаты оказались намного разрушительнее, чем можно было тогда себе представить. Аванесов В.С. Доживёт ли Единый государственный экзамен до 2009 года? <http://testolog.narod.ru>

20

Аванесов В.С.
Являются ли КИМы ЕГЭ методом педагогических измерений? Педагогические измерения. №1. 2009. С. 3–26.

21

Аванесов В.С.
Язык педагогических измерений. Педагогические измерения, №2, 2009.

ются по метрическим свойствам с заданиями первой и второй части. Кроме того, они включают субъективные компоненты оценивания результатов экзамена, что открывает возможности заметного искажения результатов ЕГЭ. Именно в этой части оказался наиболее высокий процент манипулирования результатами ЕГЭ на местах. Поэтому третья часть КИМов по форме, содержанию и метрическим требованиям не может быть элементом педагогических измерений. Пока в КИМах ЕГЭ есть третья, не тестовая часть с субъективными оценками, качественных педагогических измерений не будет.

Не случайно тестовые службы типа Educational Testing Service в массовых обследованиях используют только тестовые задания с выбором ответов, применяя далее автоматизированную считку результатов с бланков посредством оптических сканеров очень высокой производительности, от 5 до 50 тысяч бланков в час. Хороший пример успешного применения такой техники даёт Центр тестирования Республики Казахстан. Там результаты тестирования становятся известными испытуемым в тот же день.

Большинство же дефектов КИМов ЕГЭ — скрытые. Они проистекают из-за отсутствия открытых профессиональных дискуссий по общим и конкретным вопросам их содержания,

методов метрического анализа свойств каждого задания и материалов в целом. Отсутствие реальных публикаций по качеству КИМов, келейность в этой работе усугубили проблемную ситуацию, загнали её в тупик.

Нет ясных ответов и консенсуса по вопросу — что измеряют КИМы ЕГЭ? Знания учебного предмета, интеллект, учебные компетенции, способности к освоению всех или конкретных программ высшего образования? Точный ответ на вопрос о предмете измерения остаётся неизвестным. А это затрудняет интерпретацию результатов ЕГЭ. Что мы и видим в реальности: данных много, качественных выводов, нужных для сферы управления, — очень мало.

Актуальность проблемы демаркации можно видеть на конкретном примере обоснования того, почему в России вместо тестов хотят использовать экзамены. Понадобится цитата. «Особое значение при создании национальной системы экзаменов имеет разработка требований к обеспечению качества при организации и проведении экзаменов, разработке контрольных измерительных материалов, на этапе обработки и шкалирования результатов и их использования, а также создание механизмов обеспечения качества системы. В США идёт активная разработка механизмов обеспечения качества экзаменов. С этой целью

22

Аванесов В.С.
Доживёт ли Единый государственный экзамен до 2009 года? Педагогические измерения. №1. 2006.

23

<http://www.newsru.com/russia/29aug2008/ege.html>

разрабатываются и внедряются стандарты по различным аспектам тестирования»²⁴.

Получается так: в США — тестирование, но оно оказывается «механизмом экзамена». Вследствие такой алогичной риторики, привязанной к другим бюрократическим «механизмам», вместо качественных тестов в России уже восьмой год используются некачественные КИМы. Теперь мы слышим пожелания коррекции, которая в сложившихся условиях может только продлить агонию КИМов и ЕГЭ, но вовсе не остановить её. Между тем, начинающаяся агония ЕГЭ — несомненный показатель кризиса образования.

Актуальность проблемы демаркации педагогических измерений вытекает также из обширности самой образовательной сферы и огромного числа участников образовательного процесса, а также из общего понимания весомости влияния качества измерений на состояние образовательной сферы вообще, на решение проблем профессионального и социального самоопределения личности.

Успешное решение проблемы демаркации возможно при росте интереса к разработке языка педагогических измерений²⁵ и уточнения системы понятий, позволяющих отчленить педагогические измерения от всех остальных форм педагогического оценивания, особенно субъек-

тивных и спорных. К последним можно отнести весьма спорную и затратную общероссийскую систему оценивания качества образования (ОСОКО).

Миссия ОСОКО предполагается ещё более обширной, чем ЕГЭ. Создание системы оценки качества образования призвано обеспечить обучающихся и их родителей, педагогические коллективы школ и преподавателей сузов/вузов, органы управления образованием всех уровней, институты гражданского общества, работодателей надёжной информацией о состоянии и развитии системы образования на муниципальном, региональном и федеральном уровнях²⁶. Поскольку в центр ОСОКО поставлен ЕГЭ, с некачественными КИМами, ОСОКО не может быть лучше ЕГЭ.

В случае с применением таких КИМов мы имеем дело с настоящей общественной проблемой, которая имеет тенденцию к противопоставлению общественных сил. Общественная проблема определяется в литературе как форма существования и выражения противоречия между уже назревшей необходимостью определённых общественных действий и недостаточными ещё условиями её реализации. Главная причина обострения данной проблемы — это отрыв методов управления образованием от реальности, игнорирование чиновниками мнения научно-педагогической общественности.

Методология

24

Развитие национальной системы экзаменов: опыт РОССИИ, СНГ И США. Материалы и тезисы докладов Международной конференции, 19–24 апреля 2003 года, Москва 2003. С. 6.

25

Аванесов В.С. Язык педагогических измерений. Педагогические измерения. №2. 2009.

26

<http://www.reitor.ru/ru/analytic/quality/index.php?id19=98>

ПЕД
измерения

Подмена культуры педагогических измерений субкультурой КИМов ЕГЭ приобрела в России необычайно широкий, можно сказать, государственный размах, по причине трудно объяснимой правительственной поддержки этого метода. Различные виды государственного оценивания стало щедро финансироваться из госбюджета и считаться синонимами педагогических измерений. Сложившиеся ментальные установки порождают в образовательной практике хаос, непрофессионализм и бессмысленные затраты денег на разработку оценок непонятно чего.

В методологии педагогических измерений это актуализировало проблему демаркации (разграничения) подлинных научных методов педагогического измерения от псевдонаучных. Единый подход к аттестации и к комплектованию студентов различных вузов посредством единого госэкзамена вреден и опасен своими отрицательными последствиями, о которых уже было сказано²⁷.

Сущность педагогических измерений

Демаркация педагогических измерений предполагает чёткое выделение главного объекта от иных объектов. Педагогическое измерение представляет собой *процесс* определения меры *инте-*

ресующего свойства личности испытуемых на *латентной интервальной шкале* посредством *качественного теста*, состоящего из *системы* заданий равномерно возрастающей трудности, позволяющего получать педагогически целесообразные результаты, отвечающие критериям *надёжности, валидности, объективности и эффективности*. В этом определении курсивом выделены основные термины, позволяющие отграничить признаки педагогических измерений, от прочих методов, научных и ненаучных.

Главное интересующее свойство личности, измеряемое в процессе педагогических измерений — это уровень *подготовленности испытуемых*, который рассматривается как латентное свойство. В определение этого свойства включаются знание учебной дисциплины, умения, навыки, представления и компетенции. Свойство — это внешнее выражение непосредственно ненаблюдаемого, латентного качества личности. Взятые вместе, все перечисленные свойства можно представить как обобщённую совокупность свойств, образующих уровень подготовленности.

Уровень подготовленности испытуемых — измеряемое свойство. Этот уровень измеряется посредством педагогического теста. Но в КИМах изучаемое свойство выражается не метрически, а

27

Вадим Аванесов.
Экзамены прошли, проблемы остались.
<http://viperson.ru/wind.php?ID=575957&soch=1&pay=ok>

счётно. Там нет процесса трансформации исходных баллов в линейную интервальную шкалу. А есть элементарный подсчёт условно назначенных (т.е. заведомо неточных) баллов за выполнение заданий каждой из трёх форм. Сложение таких баллов не корректно без специального обоснования. Нет, соответственно, и педагогического измерения²⁸.

В наше время исходные результаты тестирования уже не рассматриваются как педагогические измерения; это лишь первый этап сбора данных для последующего шкалирования. Можно сказать и так: тестирование — это незаконченное педагогическое измерение. В западной литературе, начиная с работ В.Д. Wright, только шкалированные результаты тестирования стали признаваться педагогическими измерениями²⁹.

Цель педагогических измерений — определить количество интересующего латентного свойства личности (меру интересующего признака), присущего данному испытуемому, определить значение интересующей латентной переменной у каждого испытуемого. Сама латентная переменная и значения этой переменной величины в КИМах ЕГЭ не представлены надлежащим образом. Там другие цели.

Уровни педагогических измерений — это наука и практика. Предмет науки — разработка те-

ории педагогических измерений, предмет практики — разработка и применение тестов. Это и есть определение тестирования, которое нередко отличается от педагогических измерений пренебрежением к науке (теории). Между тем, тестирование в отрыве от науки (педагогических измерений) приносит только вред. Критикой теории и конструктивным преобразованием практики призвана заниматься *методология педагогических измерений*.

Система педагогического тестирования — организация сбора и обработки данных посредством применения тестов на множестве испытуемых. Эта система может оказаться научно обоснованной, необоснованной и даже антинаучной. Научной основой системы тестирования является теория педагогических измерений. Этой системе в России противопоставлено традиционное педагогическое оценивание в процессе обучения, экзамены с их субъективными оценками, КИМы ЕГЭ.

Предметом практики педагогических измерений являются создание теста, проведение тестирования, шкалирование тестовых результатов, проведение рейтинга и мониторинга (в части разработки показателей текущих и итоговых учебных достижений). И кроме того, построение системы качественных показателей образовательной деятельности³⁰.

Методология

28

Хотя в российской литературе публикуются, время от времени, мнения противоположного толка.

29

Wright B.D.
Scores are not Linear Measures. Rasch Measurement Transactions, 1992, 6:1, 208. См. также раздел Raw Scores are NOT measures. In: Measurement for Social Science and Education. A history of social science measurement. <http://www.rasch.org/me62.htm>

30

Аванесов В.С.
Основы теории педагогических измерений // Педагогические измерения. №1. 2004. С. 15–21.

Педагогические измерения появляются в результате применения не КИМов, а тестов. Измерение какого-либо свойства испытуемых производится в предположении, что у каждого из них есть интересующее свойство личности, в каком-то количестве. Если выясняется, что у кого-то данного свойства нет, то это даёт основания для исключения данного испытуемого из предполагаемой выборки лиц, обладающих данным свойством.

Качественный педагогический тест образует система тестовых заданий равномерно возрастающей трудности, позволяющая получать объективные, надёжные и валидные результаты измерения уровня и структуры подготовленности испытуемых. В КИМах этого нет, что ещё раз подтверждает отсутствие у них свойств педагогических измерений.

Единицей педагогического измерения является логит уровня подготовленности испытуемого и логит уровня трудности задания. Исходное значение логита подготовленности испытуемого вычисляется как натуральный логарифм отношения долей правильных ответов испытуемого к доле неправильных ответов. Далее в процессе педагогического измерения логиты шкалируются. Среднее арифметическое шкалы принимается равным нулю, стандартное отклонение — равным одному логиту.

Идея возрастающей трудности тестовых заданий была впервые выражена в работе А. Бине и Т. Симона³¹ в начале XX века. Затем эта идея проходила во многих, если не во всех работах автора данной статьи по тестовой проблематике, начиная с 70-х годов XX века. Наиболее чёткое, аргументированное и системное выражение этой же идеи мы находим в классической работе Best Test Design³². Основу любого теста составляют содержательная педагогическая теория (концепция) подготовленности испытуемых по учебной дисциплине и математико-статистические характеристики как каждого задания в отдельности, так и теста в целом. Содержательная теория позволяет обосновать качество измерения на языке учебной дисциплины, математико-статистические характеристики позволяют обосновать качество каждого задания и теста в целом формальными и вычислительными методами.

Только тест позволяет получить результаты, соответствующие международным стандартам качества педагогической информации³³, эмпирически проверяемые по критериям надёжности, валидности, объективности и эффективности. Ни один другой метод в этом отношении сравниться с тестом не может. Столь строгие требования к обоснованию качества методов педагогических измерений позволяют

31

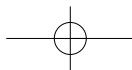
Binet A., Simon T.H.
The Development of
Intelligence in Young
Children. Vineland, N-J:
The Training School,
1916.

32

Wright B.D., Stone M.H.
Best Test Design. MESA
PRESS. Chicago. 1979.

33

Аванесов В.С.
Вопросы методологии
педагогических измерений// Педагогические
измерения. №1. 2005.
С. 3–27.



обезопасить личность, общество и государство от некачественных оценок, повысить ответственность разработчиков тестов за результаты своей деятельности. В педагогическом тесте нет места заданиям, не имеющих приемлемых формальных эмпирических характеристик. В этой связи попытки создать методы, более научно обоснованные, чем тест, неизбежно становятся провальными. В число таковых попали и КИМы ЕГЭ.

Необходимость разграничения педагогических измерений от экзаменов

Экзамен — это традиционная нетехнологичная форма педагогического контроля уровня подготовленности, в процессе которой делаются попытки оценить у испытуемого, в основном, знания, умения и навыки. Экзамены применяются в России, в программах международного исследования PISA, на Тайване и в других странах Юго-Восточной Азии. Они обычно содержат вопросы и задачи. Задания в тестовой форме там иногда бывают, но тестов нет по причине антитестовых психологических установок участников исследования по программе PISA и ей подобным.

Экзамены имеют две основные формы — устную и письменную. Преимущества устного эк-

замена — возможность оценить речевые способности учащихся, провести углублённую проверку знаний по избранным темам, в небольших учебных группах. В условиях малых групп лучше обеспечивается связь между содержанием обучения и содержанием экзамена.

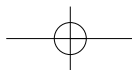
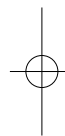
В наше время роль устных экзаменов недооценивается. Следствием такой позиции является ухудшение качества устной речи учащихся и студентов. Для массового процесса измерения уровня подготовленности учащихся и студентов устные экзамены не годятся. Результаты на них заметно зависят от речевых способностей испытуемых и от личности оценивающего педагога.

Письменные экзамены нетехнологичны, имеют повышенную трудоёмкость, оценки по ним не менее субъективны, хотя выглядят более аргументированными. При использовании экзаменов для приёма в вуз они сами нередко становятся инструментом коррупции. Поэтому экзаменационная форма для итогового контроля мало применяется в странах, имеющих развитые системы образования.

Недостатком экзамена является также несравнимость оценок, полученных студентами разных институтов, по одной и той же дисциплине, и отсюда принципиальная невозможность корректно установить, в каком вузе процесс обучения поставлен

Методология

Методология



ПЕД
измерения

лучше. Ответ на этот вопрос можно дать только в случае использования стандартной тестовой программы с применением технических средств контроля.

В наше время эффективное измерение подготовленности больших масс испытуемых посредством экзаменов экономически неэффективно. Этот вывод относится и к КИМа, представляющим смесь тестоподобных заданий и экзаменационных задач, предъявляемых испытуемым в форме «материалов» экзамена.

Результаты олимпиад не являются педагогическими измерениями

Олимпиада — это форма развития творческого мышления молодёжи и форма проверки уровня подготовленности в одинаковых для всех участников условиях. Чисто внешне различие между тестированием и олимпиадой несколько напоминает отличие между быстрой и обычной игрой в шахматы. В России сложилась традиция трёх-четырёхчасового проведения экзаменов и олимпиад, в странах Запада чаще используется примерно сорокаминутное тестирование. Если придерживаться аналогии с шахматами, то на результат олимпиад и тестирования влияет не только фактор мастерства, но и фактор лабильности мышления, специ-

альной тренировки по формированию алгоритмов быстрого выполнения заданий.

В системе образования олимпиадой называется соревновательная форма выявления наиболее подготовленных учащихся. Олимпиада является известной и весьма полезной формой педагогического контроля уровня подготовленности учащихся.

Цель олимпиад — выявление победителей соревнования. Как и в Олимпийских играх, при применении этой формы награждаются только победители, все остальные, независимо от уровня подготовленности, получают только личное удовлетворение от участия в таком празднике науки и одно общее название — участник олимпиады.

Конечно, должна быть положительная корреляция между результатами участия в олимпиадах и в тестировании, хотя такие данные автору пока не встречались в литературе. Корреляция не обязательно должна быть высокая, поскольку при ответах на задания по различным учебным дисциплинам требуется разная быстрота мышления. В общественных науках быстрее, в математике — дольше. Математика вообще требует иного подхода. При незначительной дисперсии олимпиадных баллов упомянутый коэффициент корреляции может оказаться искусственно заниженным.

При написании заданий для тестов авторы исходят из модели актуального знания: знать всё, что требуется, здесь и сейчас, уметь чётко и быстро дифференцировать приводимые ответы на один правильный, остальные — неправильные. Если в заданиях есть выбор только одного правильного ответа, испытуемые либо знают правильный ответ и готовы его быстро ввести в компьютерный бланк, либо не знают, после чего они пытаются угадать правильный ответ.

При тестировании надо уметь быстро и правильно отвечать на задания. Но в России ещё не сложилась традиция массового обучения отвечать на задания быстро. Нельзя проводить тестирование до тех пор, пока в школах не обучат основную массу детей быстро отвечать на задания. Детей и студентов, имеющих иной склад мышления, нельзя проверять тестовыми формами. Они заведомо покажут худшие результаты. В этом смысле ЕГЭ со своими тестоподобными заданиями в первой части КИМов не справедлив по отношению к детям с недостаточно высокой лабильностью мышления.

В России сотни лет существует традиция неспешного выполнения заданий и задач, а потом проверки и перепроверки своих ответов. В этом смысле проведение олимпиад больше соответствует российским традициям, а проведение быстрого тес-

тирования — западным и восточным традициям. Поскольку процент детей, не готовых быстро отвечать на задания теста, представляется автору достаточно высоким, таким испытуемым лучше подходят олимпиады. Отсюда и общественные приоритеты в пользу олимпиад и других традиционных форм неспешного выполнения заданий и субъективированного оценивания.

Скажем, в Японии, Южной Корее одним из приоритетных направлений развития учащихся становится формирование умения быстро и чётко отвечать на тестовые задания или поставленные вопросы. Эта способность формируется постепенно, и только там, где этому уделяют внимание. Внедрение ЕГЭ привлекло органов управления образованием к этому аспекту развития учащихся. Главным средством считаются тренировки в ответах на задания КИМов ЕГЭ.

Предмет соревнования в олимпиадах — уровень учебной или, по западной терминологии, академической подготовленности участников.

Оценка участникам олимпиады выставляется на основе числа решённых задач и качества проведённого обоснования решений. Каждая задача обычно ассоциируется с неким числом баллов, даваемых за её правильное решение. Чем труднее задача, тем больше баллов может получить участник олимпиады.

Методология

Положительная сторона — это равенство условий и документированность результатов участников. Каждый собственноручно пишет свои решения и обоснование решений. Недостаток олимпиад — нетехнологичность форм их проведения, возможности субъективных воздействий, нечастые утечки содержания олимпиадных заданий, субъективные методы оценивания результатов. Отсюда частые недовольства итогами олимпиад со стороны участников и наставников.

Содержание олимпиады включает обычно с десяток трудных и сверхтрудных задач, содержащих, как правило, интеллектуальный компонент. Иначе говоря, для победы на олимпиадах одних только знаний, имеющихся в учебнике, не хватает. Нужно проявить ещё высокую степень понимания сути задачи, способность рассуждать, логически обосновывать свои суждения и проверки обоснованности решения — в общем, проявить, по меньшей мере, три свойства личности: знания, интеллект и творчество. Заметим, что в педагогических измерениях, например, по модели Раша, предметом измерения всегда является одно, и только одно выделенное и хорошо определённое латентное свойство личности. Именно здесь видится главное отличие олимпиад от педагогических измерений.

Олимпиадные оценки не являются измерениями, поскольку

они, как и КИМы ЕГЭ, не обладают требуемыми для измерений свойствами. Среди уже упомянутых выше существенных свойств можно упомянуть недостаточную дифференцирующую способность олимпиадных оценок, элементы субъективизма, неизвестный уровень надёжности и валидности олимпиадных оценок, недостаточную сравнимость этих оценок с другими оценками, отсутствие шкалированных результатов. Несравнимы и ранги испытуемых, поскольку ранг каждого зависит не от трудности заданий, а от состава испытуемых.

Как сообщается в прессе, в минувшем 2008–2009 учебном году было проведено 120 олимпиад для школьников. В нынешнем году их подготовлено 175 (32 — первого уровня, 61 — второго, 82 — третьего), при этом в Российском союзе ректоров подчеркнули, что намерены добиться принципиально нового подхода к уровню олимпиад³⁴.

Цели проверки знаний как существенные признаки демаркации педагогических измерений от псевдоизмерений

Трудно понять сущность КИМов без ответа на вопрос: а зачем они создавались? Очевидно, для проведения ЕГЭ. Ранее автором этой статьи было установлено,

что Единый государственный экзамен замышлялся как механизм внедрения в стране с целью коммерциализации образования посредством т.н. ГИФО, государственных именных финансовых обязательств, а КИМы — это метод, посредством которого и должна была осуществляться коммерциализация образования — общего среднего и высшего.

Подлинные цели ЕГЭ Б.Виноградов усматривает в соединении с ваучерами (ГИФО). Это создание механизмов перехода государственных вузов в образовательные бизнес-центры. Уже в 2001 году были выбраны экспериментальные регионы, определена шкала стоимости ГИФО в зависимости от оценок по ЕГЭ³⁵.

Можно предположить, что этот план рухнул тогда из-за некачественных КИМов ЕГЭ. Но коммерческие замыслы сохранились и по сей день. Недавно было опубликовано мнение двух ректоров — Я. Кузьмина и Н. Кропачёва: государство должно оплачивать учёбу только победителей олимпиад и набравших высокие баллы; а середнячкам придётся учиться за свой счёт³⁶. Поскольку основная масса учащихся и абитуриентов имеет среднюю (а не среднячковую) подготовленность, то нетрудно догадаться, что такой ход мысли этих ректоров ведёт к практически полной коммерциализации высшего образования. Станет ли от предлагаемой ими коммерциализации

население России более образованным — вопрос, который упомянутые лица вряд ли обсуждали. Хотя ответ ясен: не станет!

С течением времени явно проявились четыре главные цели разработки и применения КИМов и ЕГЭ:

- создание равных условий доступа к качественному образованию;
- снижение уровня коррупции в сфере образования;
- побуждение учащихся к качественному образованию за счёт введения и ужесточения госконтроля вертикального типа;
- совмещение итоговой государственной аттестации выпускников школ и приёмных экзаменов во все вузы страны.

Заметим, что среди перечисленного нигде не видно цели создать метод педагогических измерений.

Создание равных условий доступа граждан к качественному образованию — очень важная цель. Она достигается совокупностью существенных политических, экономических, социальных и образовательных условий. Посредством КИМов ЕГЭ создать действительно равные условия доступа к качественному образованию невозможно, поскольку эта цель более высокого, социально-политического, а не только образовательного уровня.

Вот показательный пример. Рособрнадзор начал кампанию

Методология

35

Виноградов Б.
Единый государственный эксперимент. Власти признают провал ЕГЭ.
<http://www.kasparov.ru/material.php?id=4A0AAE22DB28E>
14.05.09

36

Ректоры предлагают отказать двоечникам в высшем образовании.
<http://gzt.ru/topnews/education/256534.html>

ПЕД
измерения

37

В России закроют 200 филиалов вузов. 01.07.2009 в 13:32 GZT.RU <http://gzt.ru/topnews/education/246780.html>. В 2009-2010 годах Рособразование планирует закрыть до 200 филиалов, а также объединить те филиалы, которые находятся в одном населённом пункте». В ведении Рособразования в настоящее время находятся 345 вузов и 1177 филиалов, где учатся 4,6 миллиона студентов, из них 2,1 миллиона — за счёт бюджета. В 2008 году ведомство закрыло 36 филиалов подведомственных ему вузов, а в 2009 — 50 филиалов. Кроме того, было закрыто более 400 представительств российских вузов.

38

Юлия Котова.
Президенту доложили об успехе ЕГЭ. <http://gzt.ru/topnews/education/247336.html>. 03.07.2009.

39

<http://gzt.ru/topnews/education/249524.html>.

40

Четверть работ отличников по ЕГЭ оказалась фальшивками. 23.07.2009 в 14:10

по ликвидации филиалов вузов. Их в стране сейчас насчитывается 1177³⁷. Высшее образование станет доступнее для граждан страны, живущих на огромной территории, нередко в труднодоступных местах? Нет, конечно. А как это отразится на размещении трудовых ресурсов, структуре населения? Отрицательно.

Из поставленных вопросов и ответов на них легко понять, что вопрос создания равных условий доступа к качественному высшему образованию — не ведомственный, а общегосударственный, посредством КИМов ЕГЭ и закрытия филиалов не решаемый. Рособрнадзор со своими методами борьбы за качество образования может больше навредить, чем улучшить положение дел с образованием населения. Закрыть легко, труднее наладить там настоящий учебный процесс. А потому вектор деятельности Министерства образования и науки должен был быть другим — надо меньше закрывать, а больше заниматься улучшением качества образования, особенно в тех вузах и филиалах, которые ещё в состоянии дать качественное образование, если их оснастить современным оборудованием и помочь им методически.

Не менее важна и вторая цель — снижение уровня коррупции. Хотя эта цель считается частично достигнутой, но это мнение нуждается в доказательной статистике, которой пока что

нет. Коррупция не исчезла, а переместилась из вузов в пункты проведения ЕГЭ.

Пример с Дагестаном, Ингушетией и другими территориями показывает, что уровень коррупции не только не снизился, но многократно вырос³⁸. А в Адыгее на ЕГЭ по биологии и истории 50% результатов оказались неподтверждёнными³⁹. Официальное заключение об огромном количестве «липовых» отличников по ЕГЭ и информация о стоимости сертификата — весомые аргументы⁴⁰ для такого вывода. Эти и многие другие официально установленные факты, появившиеся в печати, убедительно говорят о необъективности и несправедливости результатов и о полной непригодности КИМов для проведения ЕГЭ.

Как справедливо отмечает О.Н. Смолин, Единый государственный экзамен — не менее коррупционная схема, чем обычные экзамены, а может быть, даже и более. И система взятков при ЕГЭ совсем не исчезла. По его оценкам, «хороший» ЕГЭ в Москве или на Кавказе стоил около 60 тыс. рублей⁴¹. Таким образом, по результатам применения КИМов ЕГЭ похоже, что и вторая цель оказалась не достигнутой. Мы всё больше убеждаемся, что КИМы ЕГЭ коррупционны не менее иных методов, а возможно, и более.

Третья цель противоречит самой сущности образователь-

ной деятельности. Ещё в древности было известно, что... «ни одну науку не следует изучать рабски», так как «...насильственно внедренное в душу знание не прочно». И действительно, восемь лет проведения этого экзамена изменили учебный процесс, подчинили его сдаче ЕГЭ. Последний превратился в универсальный показатель, главный критерий успеха учащихся, учителей, школ, городов и субъектов РФ. Собственно образовательную деятельность заменило стремление хорошо сдать ЕГЭ. Известно, что негодный показатель порождает негодную работу.

Бесплодные попытки совмещения несовместимого

Если первые три цели — социального и управленческого толка, то четвёртая цель ЕГЭ, совмещение итоговой государственной аттестации выпускников школ и приёмных экзаменов во все вузы страны представляется метрическим нонсенсом.

Корни идеи совмещения обнаруживаются в публикации, где делается попытка дать научное определение КИМа ЕГЭ. «КИМы — это стандартизованная экзаменационная работа, создаваемая в соответствии с требованиями теории педагогических измерений, позволяющая с достаточной объективностью и надёжностью провести независимую госу-

дарственную аттестацию выпускников образовательных учреждений и отбор абитуриентов вузов»⁴².

Интерпретация текста приведённого фрагмента такова. Автор, по всей видимости, считает КИМы ЕГЭ экзаменационной работой, которую он надеялся стандартизовать. Весь мировой предыдущий исторический опыт и последующая российская практика проведения ЕГЭ показали, что стандартизация экзаменационных работ влетает государству в огромную сумму. Это тупиковый путь, интересующий только лоббистов российского ЕГЭ, имеющих свой интерес к дележу государственных средств.

Очевидно, что цитированный автор так же, как и некоторые его коллеги, верит в возможность чудесного превращения *экзаменационных работ* в качественное средство педагогических измерений. Критерии качества тестовых результатов были произвольно приспособлены этим автором для оценки качества экзаменационных работ. Как будто такое возможно. Автор также верит в возможность качественного совмещения в КИМах ЕГЭ выпускных экзаменов в школе и вступительных экзаменов в вузе. В общем, вместо науки — сплошная вера в удобные схемы.

Под влиянием таких ошибочных схем бывшее Правительство РФ под руководством

Методология

GZTRU.
<http://gzt.ru/topnews/society/250851.html>.

41

<http://gzt.ru/topnews/education/250952.html>

42

Тамур А.О.
«Контрольно-измерительные материалы для Единого государственного экзамена. Высшее образование сегодня». №7–8. С. 28–33. 2002.

ПЕД
измерения

М.Касьянова не критически восприняло предложение бывшего Министерства образования РФ о проведении в стране ЕГЭ. В Постановлении Правительства РФ №119 «Об организации эксперимента по введению единого государственного экзамена», от 15.03.2001 г. № 1033 было написано: «Принять предложение Министерства образования Российской Федерации о проведении эксперимента по введению единого государственного экзамена, обеспечивающего совмещение государственной (итоговой) аттестации выпускников XI (XII) классов общеобразовательных учреждений и вступительных испытаний для поступления в образовательные учреждения высшего профессионального образования (<http://www.ege.ru/prikaz119.html>).

Так возник миф о том, что используемые в ЕГЭ т.н. контрольно-измерительные материалы (КИМы) позволяют совместить государственную итоговую аттестацию выпускников общеобразовательных учреждений со вступительными испытаниями в учреждения среднего и высшего профессионального образования.

«В результате эксперимента, — написано в документе Рособнадзора, — доказана принципиальная возможность совмещения государственной (итоговой) аттестации выпускников XI (XII) классов общеобразовательных учреждений и вступитель-

ных испытаний в ссузы и вузы в форме ЕГЭ, а также эффективность выбранной организационно-технологической схемы проведения ЕГЭ⁴³.

Многие авторы с первых дней затеи с КИМаи ЕГЭ утверждали о невозможности качественного совмещения этих двух экзаменов, поскольку это противоречит аксиомам педагогических измерений. Различны цели и педагогическое содержание этих экзаменов, уровни подготовленности испытуемых, не говоря уже о критериях оценки, которые отличаются как небо от земли. Но министерскими чиновниками и руководителями ФИПИ это мнение игнорировалось и опровергалось ложными доказательствами.

Далее идея совмещения двух экзаменов в одном перекочевала в те положения «Закона об образовании», которые связаны с ЕГЭ. Правительство РФ и доминирующая в Госдуме партия «Единая Россия» восемь лет пытались доказать за счёт госбюджета возможность проведения единого совмещающего экзамена, но ничего путного из этого не получилось.

В 2009 году, когда статью Закона об образовании относительно единого совмещённого госэкзамена надо было выполнять в обязательном порядке, единый экзамен неожиданно для всех раздвоился. Невозможность совмещения стала, наконец, понят-

43

Федеральная служба по надзору в образовании и науке. Стратегические цели и тактические задачи Федеральной службы по надзору в сфере образования и науки. <http://www.obrnadzor.gov.ru/strategy>. 23.03.06.

ной и министерским чиновникам. Единые государственные экзамены быстро разделили на аттестационные, нацеленные на выпускников школ (русский язык, математика), и на приёмные экзамены в вузы по остальным учебным дисциплинам.

Таким образом, как единый, экзамен, в настоящем смысле выделенного слова, в 2009 году реально исчез, оставшись номинально. В вузах к этому добавились собственные экзамены или собеседования. Иначе говоря, вопреки принятому Закону, вместо единых экзаменов в 2009 г. возникли удвоенные и утроенные экзамены. Положения Закона «Об образовании» относительно совмещающего ЕГЭ после таких изменений повисли теперь в воздухе.

Всё пошло так, как изложено в учебниках логики относительно соотношения объекта и его имени, замещающего в сознании объект. Несмотря на то, что сам объект (ЕГЭ) существенно изменился, название осталось, деньги выделялись на проведение единого экзамена, работа кипела, но единая сущность экзамена испарилась. Получилось, как предсказывал автор этой статьи в работе «Доживёт ли единый государственный экзамен до 2009 года? <http://news.babr.ru/?IDE=36646>. ЕГЭ начал умирать после первого же года реальной, а не мифической жизни.

Интересен вопрос: а какая наука является основой разработки КИМов ЕГЭ? Поскольку КИМы не тесты и не педагогические измерения, то автор склоняется к мысли, что это может быть нечто вроде ЕГЭведения или КИМологии. Объединение в одном КИМе заданий нескольких форм, сопровождаемое ненаучным шкалированием исходных баллов по признаку формы, а не по уровню трудности заданий, — верный путь к провалу. Итогом этих и множества других ошибок стал некачественный метод — ни тест, ни экзамен⁴⁴.

Как следует из публикации Федерального института педагогических измерений (ФИПИ)⁴⁵, там уже в самом начале эпопеи с КИМаи ЕГЭ возникли противоречия по вопросу соотношения КИМов ЕГЭ и тестов. Одни сотрудники и авторы, вслед за руководителями Министерства образования и Рособнадзора, называли КИМы ЕГЭ тестами. Другие проявляли осторожность, считая, что «говорить о ЕГЭ, как тесте, основываясь на результатах первого года эксперимента, конечно, рано» (с. 34). Похоже, что тоже лелеяли надежду на чудесное превращение КИМов в тесты.

В основе определения КИМов ставится слово «материал», аналогом которого в науке является, вероятно «материя», вряд ли в философском смысле слова. Под ненаучное понятие «материалы» может подпадать всё что угодно.

Методология

44

Аванесов В.С.
И не тест, и не экзамен // Университет и школа. №1–2, май, 2002. С. 26.
<http://www.testolog.narod.ru/Ege4.html>

45

«Единый государственный экзамен. Научные основы, методология и практика организации эксперимента» / Под ред. В.А. Болотова. М.: Логос, 2002. 208 с.

ПЕД
измерения

Следовательно, при столь широком объёме содержание этого понятия оказалось слишком узкое, для научного анализа абсолютно непригодное. Этот вывод вытекает из логического закона обратного отношения между объёмом и содержанием понятий. А потому в логическом смысле КИМы обречены оставаться неконкретным названием не определённых наукой объектов. К тому же если на секунду допустить, что слово «материалы» является частью научного понятия «контрольно-измерительные материалы», то объём такого понятия оказался бы слишком широк для того, чтобы оно было в состоянии выразить что-то существенное или конкретное.

Кроме того, в словосочетании «контрольно-измерительные материалы» закладывалась, вероятно, ещё одна странная, если не сказать опасная, идея использования материалов и для контроля, и для измерения. Дело в том, что если для текущего педагогического контроля в учебном процессе какие-то учебные материалы использовать можно, то для итогового педагогического тестирования эти материалы, не имеющие обычно никакого статистического обоснования, не пригодны. В тесте нужны тестовые задания, объединённые в систему посредством научных методов. Если тестовые задания тоже назвать материалами, то возникнет ещё одна логическая ошибка: точное

понятие меньшего объёма заменили неточным понятием большего объёма. В общем, даже само название «контрольно-измерительные материалы» — алогизм.

При культурном и качественном тестировании каждое задание теста должно обязательно иметь параллельные варианты. Параллельными называются варианты задания, имеющие сходное содержание из одной и той же общей дидактической единицы, примерно совпадающие меры трудности и вариации результатов испытуемых.

И действительно, все варианты к каждому заданию, по каждому предмету, должны иметь общее содержание, примерное равенство средних арифметических показателей, а также показателей вариации по каждому заданию. Только тогда обеспечивается справедливость и объективность выставляемых оценок⁴⁶.

В КИМах ЕГЭ задания не параллельны по трудности, что подтверждается фрагментом приводимого здесь Кемеровского отчёта 2007 г. Непараллельность вариантов заданий, явно видимая из таблицы, делает КИМы ЕГЭ принципиально несправедливыми при оценке испытуемых. В таблице видны различия в процентах решаемости каждого варианта КИМа ЕГЭ по физике⁴⁷. Из-за этого несравнимы результаты испытуемых вообще, не говоря уже о сравнении результатов по регионам, где ис-

46

Аванесов В.С.
Проблема объективности педагогических измерений // Педагогические измерения. №3. 2008.

47

Государственное учреждение «Областной центр мониторинга качества образования». Единый государственный экзамен 2007. Физика. Сборник аналитических материалов. Кемерово, 2007.

пользуются не один и тот же т.н. Решаемость заданий ЕГЭ по
«КИМ», а его варианты сильно физике в 2007 году по вариантам
различающейся меры трудности. (1 волна). Задания группы «А, В»

Методология

Таблица

Номер варианта	A1	A2	A3	A4	A5	A6	A7
1	79,71%	60,87%	43,48%	55,07%	72,46%	81,16%	40,58%
2	52,17%	53,62%	60,87%	62,32%	65,22%	78,26%	30,43%
3	70,67%	58,67%	49,33%	50,67%	69,33%	68,00%	32,00%
4	72,22%	52,78%	48,61%	66,67%	66,67%	37,50%	58,33%
5	45,83%	47,22%	58,33%	40,28%	54,17%	38,89%	83,33%
6	55,41%	20,27%	55,41%	89,19%	83,78%	27,03%	66,22%
7	50,00%	25,00%	56,94%	84,72%	86,11%	69,44%	76,39%
8	76,81%	56,52%	47,83%	50,72%	69,57%	78,26%	34,78%
9	58,57%	47,14%	38,57%	72,86%	68,57%	71,43%	41,43%
10	77,46%	50,70%	56,34%	67,61%	64,79%	73,24%	43,66%
11	72,06%	66,18%	64,71%	94,12%	82,35%	48,53%	75,00%
12	54,55%	36,36%	57,58%	86,36%	59,09%	34,85%	71,21%
13	54,69%	32,81%	65,63%	50,00%	84,38%	54,69%	60,94%
14	50,00%	51,52%	71,21%	89,39%	83,33%	56,06%	86,36%
15	В69,49%	22,03%	64,41%	81,36%	69,49%	32,20%	74,58%
Среднее	62,64%	45,66%	55,69%	69,21%	71,91%	56,85%	58,01%

Окончание таблицы

Номер варианта	A8	A9	A10	A11	A12	A13	A14	A15
1	23,19%	31,88%	34,78%	60,87%	50,72%	56,52%	34,78%	44,93%
2	24,64%	37,68%	39,13%	59,42%	59,42%	63,77%	50,72%	17,39%
3	42,67%	36,00%	37,33%	45,33%	56,00%	69,33%	48,00%	14,67%
4	36,11%	16,67%	34,72%	41,67%	63,89%	79,17%	47,22%	19,44%
5	45,83%	11,11%	31,94%	48,61%	37,50%	58,33%	70,83%	40,28%
6	32,43%	54,05%	56,76%	18,92%	29,73%	64,86%	41,89%	18,92%

ПЕД
измерения

Номер варианта	A8	A9	A10	A11	A12	A13	A14	A15
7	62,50%	12,50%	43,06%	50,00%	38,89%	63,89%	29,17%	27,78%
8	33,33%	36,23%	56,52%	52,17%	63,77%	62,32%	33,33%	15,94%
9	45,71%	35,71%	57,14%	50,00%	62,86%	70,00%	52,86%	57,14%
10	43,66%	47,89%	46,48%	69,01%	61,97%	66,20%	32,39%	23,94%
11	63,24%	22,06%	51,47%	29,41%	42,65%	83,82%	63,24%	29,41%
12	51,52%	13,64%	65,15%	48,48%	78,79%	68,18%	42,42%	19,70%
13	50,00%	57,81%	43,75%	57,81%	26,56%	62,50%	34,38%	32,81%
14	36,36%	25,76%	53,03%	53,03%	42,42%	71,21%	56,06%	43,94%
15	62,71%	15,25%	55,93%	57,63%	28,81%	86,44%	55,93%	37,29%
Среднее	43,34%	30,41%	46,91%	49,23%	49,81%	68,24%	46,14%	29,34%

О какой объективности оценок КИМов ЕГЭ можно говорить в свете представленного факта — реальной таблицы результатов? Можно ли утверждать, что оценки и по всем остальным учебным дисциплинам не объективны? Для этого Правительство РФ должно дать указание рассекретить все подобные таблицы результатов по КИМах ЕГЭ, а виновных за незаконное засекречивание важных для личности, общества и государства оценок, свидетельствующих о некачественности методов, — привлечь к ответственности.

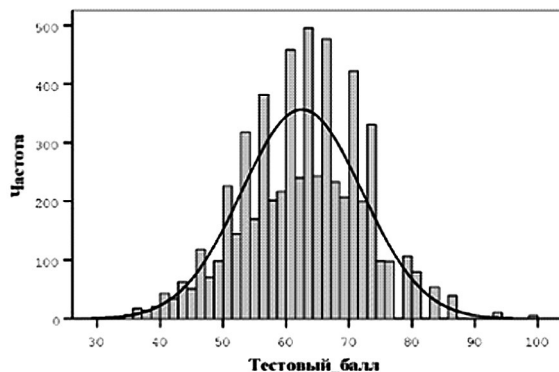
В России КИМы ЕГЭ часто называли (и продолжают называть) тестами. Но в таких случаях лучше говорить о псевдотестах. Если под педагогическим тестом понимать систему тестовых

заданий равномерно возрастающей трудности, позволяющую качественно выявить структуру и измерить уровень подготовленности испытуемых с определённой степенью надёжности и валидности результатов, то псевдотест можно определить так. Это совокупность вопросов и задач неопределённой формы и неопределённого уровня трудности, с неизвестными статистическими характеристиками, не пригодных для качественного выявления уровня и структуры подготовленности испытуемых. Все доступные гистограммы оказались искусственно сжатыми по краям исходной шкалы, чтобы не были видны зияющие дефекты разработки «тестов ЕГЭ».

Пример такой гистограммы с очень странным распределением

«т.н. «тестовых баллов ЕГЭ» благодаря содействию Общественной комиссии Совета Федерации. Этот отчёт был получен

Распределение тестового балла ЕГЭ по русскому языку в Вологодской области в 2008 году



Псевдоизмерения характеризуются похожими распределениями результатов, отсутствием науки и научных отчётов при разработке тестов. Досконально качество КИМов ЕГЭ не может сейчас проверить никто, поскольку по ним нет ни опубликованной документации, ни реальных исходных результатов, ни показателей качества.

Напротив, тест, как научный метод, интересен тем, что при его разработке обязательно нужно использовать методы обоснования качества получаемых результатов. Все остальные методы этой методологической нормы не имеют. А потому любое название, вроде КИМов, освобождаёт их творцов от использования обязательных при разработке тестов методов

обоснования качества информации. Привлекательная, казалось бы, ниша оказалась, как и следовало ожидать, методологическим тупиком.

Полная система псевдоизмерений включает в себя ещё и так называемое оценивание, аттестацию, сертификацию и т.п. Все они основаны на использовании элементарного подсчёта исходных баллов испытуемых. Это счёт, а не измерение.

Отсутствие этапа трансформации счётных данных в шкалы с равными интервалами в наши дни можно назвать решающим критерием отсутствия педагогических измерений. В педагогических измерениях наибольший интерес представляют две шкалы — уровень подготовленности испытуемых и уровень трудности заданий.

Отличие педагогических измерений от педагогического оценивания

Ещё одна форма псевдоизмерений — педагогическое оценивание, которое можно определить как выражение числами уровня проявления интересующего признака, основанное на содержательном анализе качества выполнения каких-либо заданий, письменных работ, устных ответов и др.

В зарубежной педагогической литературе широко используется понятие «evaluation», что можно перевести как «оценивание», имея в виду не только конечный результат, но и процесс формирования оценки. Соответственно выделяются два основных вида оценивания — формирующее и итоговое⁴⁸. Цель первого — оказывать, посредством оценивания, формирующее влияние на текущий процесс обучения, в смысле его улучшения, за счёт установления обратной связи от студента к преподавателю. Цель второго — получить итоговые результаты обучения.

Английское educational evaluation можно рассматривать как одну из начальных форм численного оценивания, в которой, однако, нет результатов измерений. На Западе понятие «evaluation» нередко включает в себя все традиционные педагогические оценки и экспертные оценки⁴⁹.

Там в это понятие раньше иногда включали и измерение. Но в последние тридцать лет это уже не делается из-за необходимости дифференцировать оценивание от более наукоёмкого процесса педагогического измерения. В России делают наоборот.

Педагогическое оценивание надо отличать и от педагогического тестирования. В педагогическом тестировании применяются тесты, состоящие из тестовых заданий, решающих правил квантования результатов решения таких заданий, а также из инструкций для испытуемых. Инструкции даются в зависимости от применяемых тестовых форм и вариантов этих форм.

Первоначальным результатом педагогического тестирования являются исходные тестовые баллы испытуемых. Это ещё не измерения. Исходные тестовые баллы далее подвергаются шкалированию, что позволяет получать меру подготовленности испытуемых на интервальной шкале. Только такие шкалированные результаты тестирования могут называться педагогическими измерениями.

Можно сказать и так: измерение — самый высокий уровень численного оценивания, где к оценкам добавляется ещё и мера. Все остальные формы оценивания, включая учительские оценки, это применение числа, но не меры. Рассмотрение их как равноценные понятия низводит из-

48

Bloom B.S. a. o.
Handbook on Formative
and Summative
Evaluation of Student
Learning. N-Y., McGraw-
Hill, 1971. 923 pp.

49

Аванесов В.С.
Вопросы методологии
педагогических измере-
ний. Педагогические из-
мерения. № 1. 2005.
С. 3–27.

мерение до уровня оценивания. Итак, существенным критерием демаркации педагогических измерений от педагогического оценивания является признак наличия меры, точнее, интервальной шкалы.

Сейчас уже можно говорить о системе оценивания, которая пока не внедрена, но уже широко обсуждается на различных совещаниях. Она называется общероссийской системой оценки качества образования (ОСОКО) и мыслится её создателями как не имеющая аналогов в мире. Фактически система ОСОКО призвана установить не только полный контроль уровня подготовленности всех учащихся и студентов, но и тотальный контроль за работой всех образовательных учреждений России, во всех краях, областях и в автономных республиках, что противоречит принципам подлинного федерализма и соответствует, скорее, принципам унитарного государства. Это надо понимать.

Критерии демаркации педагогических измерений

1. *Наличие в стране процесса педагогических измерений.* Среди возможных критериев отличия педагогических измерений от псевдоизмерений главным критерием представляется наличие в стране самого процесса измерений. Без такого процесса надеж-

ды на появление качественных тестов тщетны. Сейчас такого процесса нет.

Отмеченный процесс включает абстрагирование интересующего свойства, выбор математической модели измерения, разработку заданий в тестовой форме, проверку и публикацию их тестовых свойств, поиск наилучших методов композиции тестовых заданий в систему заданий возрастающей трудности. При этом каждое задание должно иметь множество параллельных заданий со своими статистическими характеристиками, график зависимости вероятности правильных ответов испытуемых в зависимости от уровня их подготовленности, статистику соответствия характеристик задания требованиям модели измерения. Далее идут требования к форме и содержанию заданий, а также требования к тесту в целом⁵⁰.

2. *Научная обоснованность теста.* Это обеспечивается обязательной публикацией концепции теста, применением содержательных и метрических теорий, на основе которых готовится тест.

Повышению научной обоснованности теста способствует определённость предмета педагогического измерения. Наиболее распространённым предметом педагогических измерений является уровень подготовленности испытуемых. Предметом могут быть и другие свойства

личности, и каждое требует концептуализации, операционализации, структурирования, оснащения индикаторами, эмпирическим обоснованием выделенного для педагогического измерения свойства личности. Таких работ в стране практически нет. И это тоже критерий оценки состояния дел с педагогическим измерением в стране.

Уровень подготовленности испытуемых — измеряемое свойство. Этот уровень измеряется посредством педагогического теста. Но в КИМах изучаемое свойство выражается не метрически, а счётно. Там нет процесса трансформации исходных баллов в линейную интервальную шкалу. А есть элементарный подсчёт условно назначенных (т.е. заведомо неточных) баллов за выполнение заданий каждой из трёх форм. Сложение таких баллов не корректно без специального обоснования. Нет, соответственно, и педагогического измерения⁵¹.

В наше время исходные результаты тестирования теперь уже не рассматриваются как педагогические измерения; это лишь первый этап сбора данных для последующего логарифмического шкалирования. Можно сказать и так: тестирование — это незаконченное педагогическое измерение. В западной литературе, начиная с работ В.Д. Райта, только шкалированные результаты тестирования

стали признаваться педагогическими измерениями⁵².

Каждое из названных условий является достаточным для признания применяемых сейчас форм, вроде КИМов ЕГЭ, неприемлемым методом.

3. *Обязательность эмпирической апробации всех заданий, включаемых в тест.* Это требование является обязательным не только для педагогических измерений, но и для любого метода оценки личности. Именно в процессе апробации появляется возможность отделить тестовые задания от заданий в тестовой форме и элементарных вопросов. Делается это посредством проведения статистического анализа характеристик каждого задания.

Эмпирическая апробация позволяет провести тщательный подбор заданий теста, соответствующих конкретной цели педагогического измерения. Подбор заданий выполняется на основе специальных компьютерных программ. Нет никаких следов того, что при разработке КИМов ЕГЭ эти программы действительно и по-настоящему использовались.

В процессе производства КИМов ЕГЭ результаты эмпирической апробации остаются неизвестными, хотя такая информация очень интересна для методистов — представителей неустраиваемой в нынешней России профессии. Вследствие этого мы не можем знать, имеем

51

Хотя в российской литературе публикуются, время от времени, мнения противоположного толка.

52

Wright B.D.
Scores are not Linear Measures. Rasch Measurement Transactions, 1992, 6: 1, 208.

ли мы дело с методом педагогических измерений или с примитивным набором заданий и вопросов.

Судя по отдельным просочившимся в Интернет данным, нет в КИМах и расположения заданий по принципу возрастающей трудности заданий, что возможно только по итогам предварительной эмпирической апробации заданий. Похоже, что расположение заданий по принципу равномерно возрастающей трудности в КИМах никогда не выполнялось. Тут возможны две версии реальных событий. Либо разработчики КИМов ЕГЭ отвергают это принцип, либо они оказываются не в состоянии что-либо сделать. В отчётах ФИПИ нет данных по предварительной апробации заданий, включаемых в КИМы. Об этой важной части тестового процесса нет ни слова.

4. Открытость результатов педагогических измерений. В свете имеющегося опыта секретности, возведённой вокруг некачественных КИМов, этот критерий призван предотвратить подобные явления. Здесь особо важно наличие открыто опубликованного научного отчёта по созданию теста. Не создатель теста, а читатель отчёта решает — тест это или не тест?

Наличие качественного отчёта является существенным признаком демаркации тестов от псевдотестов и, если можно так

назвать, нетестов. С точки зрения требования иметь опубликованный и доступный для научной критики отчёт по разработке теста в России тестов нет.

Вся система признаков педагогических измерений должна излагаться в отчёте по разработке каждого теста. Это особенно важно в отношении упомянутых ранее критериев эффективности и качества тестовых результатов — надёжности, валидности и объективности.

Отчёт по разработке теста подлежит обязательной публикации. Без подробного отчёта по разработке, а также пособия по применению теста, как системы заданий равномерно возрастающей трудности, не бывает. Важно напомнить, что ничего похожего в разработке КИМов ЕГЭ нет. Отсутствие требуемых отчётов является важным критерием разграничения педагогических тестов от псевдотестов и нетестов.

Этические нормы тестологической службы США, Великобритании, Австралии и многих других стран Запада и Востока запрещают применять тесты, если нет опубликованной информации об их качестве. В ещё большей мере этот запрет относится к иным методам. Таким образом, в этих странах личность защищена от негодных методов. В России нет ни таких служб, ни упомянутых норм, ни этических запретов. А потому вместо тестов применяются псевдотестовые

методы оценивания и различные материалы сомнительного качества.

Для тестового метода обязательны публикация подлинных статистических данных по каждому заданию и по тесту в целом, необходимость построения графических образов каждого задания теста, обязательность проведения т.н. дистракторного анализа, построение графиков информационных функций каждого задания и теста в целом, проверка адекватности данных избранной модели педагогических измерений. Список обязательных требований, относящихся к мировым стандартам разработки и применения настоящих тестов, довольно обширен⁵³. КИМы ЕГЭ оказались очень далеки от соответствия этим требованиям.

Важно заметить, что в описательно-методических отчётах Федерального института педагогических измерений о результатах проведения ЕГЭ ничего из упомянутого научного арсенала нет. Объяснение двоякое. Первое: КИМы ЕГЭ — это не тесты, а потому тестовые стандарты качества измерений к ним не применимы. А стандартов качества КИМов нет. Но когда выгодно, говорят что КИМы — это тесты. Это как две шапки: какую нужно, ту и надевают. Методическая литература и газеты полны тестоподобной фразеологией.

5. *Наличие результатов дистракторного анализа как демар-*

кационный признак. В заданиях с выбором одного или нескольких правильных ответов дистракторами называют неправильные, но правдоподобные ответы в заданиях с выбором одного или нескольких правильных ответов, привлекательные для испытуемого, не знающего точный правильный ответ. Это название происходит от английского глагола to distract, что можно перевести словом «отвлекать». При создании теста обязателен процесс проведения дистракторного анализа качества заданий в тестовой форме.

С точки зрения психологии дистракторы выполняют интерферентную функцию. Они вторгаются в процесс мышления испытуемых, побуждают его сравнивать разные ответы к одному и тому же заданию, анализировать их, искать аргументы в пользу правильности (или неправильности) каждого ответа, выбирать правильный или иногда самый правильный ответ, если это написано в инструкции для испытуемых. Необязательно иметь в каждом задании одинаковое число дистракторов.

Качество дистрактора обычно оценивается несколькими методами. Чаще всего используется подсчёт долей (процентов) испытуемых, которые выбирают тот или иной дистрактор при ответе на задание теста. Чем выше процент испытуемых, выбравших данный неправильный от-

вет, тем выше его привлекательность для незнающих студентов.

Необходимость проведения дистракторного анализа вытекает из логики организации качественного процесса педагогического тестирования. Не бывает теста без качественных тестовых заданий, как не бывает качественных тестовых заданий без качественных дистракторов. Поэтому без проведения дистракторного анализа и публикации информации о качестве каждого использованного задания и ответов к ним, утверждения о «качестве тестов», и тем более т.н. КИМов-ЕГЭ, носят декларативный характер.

6. Наличие теории или теорий, позволяющих обосновать качество методов педагогических измерений. Разработка настоящих методов педагогических измерений является наукоёмкой сферой; без научных концепций и научных методов создать качественные тесты невозможно. Логика обоснования истинности одной или нескольких концепций посредством построения системы эмпирических показателей (индикаторов) активно используется при разработке (композиции) педагогического теста.

А что положено в основу композиции КИМов ЕГЭ? Идея эклектического соединения трёх форм заданий и наивная, необоснованная надежда на работоспособность такого соединения. Если кто-то считает, что есть науки, позволяющие обосновать качест-

во КИМов ЕГЭ, то было бы справедливо назвать их ЕГЭведением и КИМологией.

7. Технологичность теста. Это означает, что в тесте недопустимы нетехнологичные и субъективные оценочные формы, которые используются сейчас во второй и третьей части КИМов ЕГЭ. В наше время для массового контроля уровня подготовленности испытуемых традиционные экзамены уже не применяются ввиду их субъективности и нетехнологичности. Используются только тестовые технологии. Яркий пример эффективного применения тестовых технологий даёт всемирно известная служба Educational Testing Service⁵⁴.

8. Наличие стандартов. Каждое задание настоящего педагогического теста имеет содержательную и формальную аргументацию, необходимость которой специально оговорена в зарубежных стандартах на разработку тестов⁵⁵. Отсутствие подобных стандартов на русском языке является объективным препятствием для создания тестов с проверяемым качеством измерения. Полезно напомнить, что на разработку КИМов ЕГЭ стандарты не существуют. Без разработки общественно профессиональных стандартов настоящую культуру педагогических измерений поднять будет трудно.

Стандартами предусмотрена обязательная проверка результа-

Методология

54

Educational Testing Service. Princeton, N-J. <http://www.ets.org/portal/site/ets/menuitem>

55

Am. Educ. Research Ass., The Am. Psychological Ass., National Council on Measurement in Education. Standards for Educational and Psychological Testing. 1999. V. 1–3.

тов теста по критериям надёжности и валидности. Это идеи точности результатов испытуемых и адекватности метода измерения поставленной цели. Одной цели. Могут ли результаты теста быть валидными относительно другой цели? Могут, но только отчасти, и в меньшей обычно мере. Не бывает теста, который одинаково пригоден для разных целей.

Перед разработчиками КИМов было поставлено несколько различающихся целей, и они попытались создать метод, решающий многие проблемы российского образования. Пренебрежение к теории имеет своим результатом неудачу. Всегда.

А потому результаты ЕГЭ оказались ни надёжными, ни валидными, ни объективными, ни эффективными относительно каждой отдельно поставленной цели. Например, в борьбе с коррупцией вообще недопустимо применение форм, используемых сейчас в КИМах ЕГЭ. Об этом автор писал неоднократно, но все соображения такого рода не замечались в течение всех лет псевдонаучного экспериментирования с негодными материалами.

9. *Сочетаемость тестовой технологии с другими современными образовательными технологиями.* Вторая и третья формы КИМов не технологичны, а вопросы образовательных технологий, на фоне проведения масштабного и нетехнологичного ЕГЭ, в Министерстве образова-

ния и науки отодвинуты, похоже, в сторону. Так образовательная политика ведёт страну к технологическому отставанию от стран Запада и Востока, где технологии уже внедрены в образовательную практику и показывают хорошие результаты.

Используемые сейчас в России методы педагогического оценивания далеки от требований данного критерия. Можно определённо утверждать, что в стране нет процесса педагогических измерений. Есть процессы производства КИМов, проведения ЕГЭ и разного рода оценивания. Но давно требуемого для развития образования подлинного процесса педагогических измерений определённо нет.

10. *Наличие программно-вычислительного обеспечения.* Для разработки настоящих тестов используются специальные компьютерные программы типа RUMM 200 и WINSTEPS, а также ряда других, которые позволяют трансформировать счётные данные в измерения, получать графические образы каждого задания и создать систему заданий с требуемыми свойствами теста. Если графический образ удовлетворяет требованиям настоящего теста, то задание с таким графиком включается в тест, если не удовлетворяет, то в тесте такому заданию места нет. Без таких программ качественных педагогических измерений не бывает.

Таким образом, ещё одним критерием демаркации педагогических измерений от псевдоизмерений и неизмерений — это использование специальных компьютерных программ, которые позволяют получать настоящие педагогические измерения.

11. *Наличие статистических характеристик каждого задания.* Компьютерные программы позволяют получить статистические характеристики каждого задания теста. А это является важным критерием разграничения тестов от псевдотестов и не тестов. Есть ли такие характеристики у КИМов ЕГЭ? Это науке неизвестно

12. *Единица педагогического измерения.* Единицей педагогического измерения является логит уровня подготовленности испытуемого и логит уровня трудности задания. Исходное значение логита подготовленности испытуемого вычисляется как натуральный логарифм отношения долей правильных ответов испытуемого к доле неправильных ответов. Далее в процессе педагогического измерения логиты шкалируются. Среднее арифметическое шкалы принимается равным нулю, стандартное отклонение — равным одному логиту.

В псевдоизмерениях в качестве единицы рассматривается балл, получаемый как результат ответа испытуемого на задание. Иногда используются взвешен-

ные баллы. В КИМах используются субъективно взвешенные баллы, научно никак не обоснованные.

13. *Время проверки.* Есть и другой простой метод демаркации тестов от псевдотестовых материалов. Если время проверки уровня подготовленности 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 по математике, легко видеть, что столь длительное время является убедительным признаком отсутствия теста как кратковременного метода педагогического измерения и выхолащивания самой сути тестового метода.

14. *Наличие системы решающих правил и адекватность интерпретации результатов.* После проведения государственного экзамена выпускников школ в Рособнадзоре определили допустимый процент тех, кого можно лишить аттестата зрелости. 28863 выпускникам 2009 года — а это примерно 3% — на основании некачественных оценок по КИМах ЕГЭ не выдали аттестаты зрелости. Но так не делается ни в одной стране мира. Из этических соображений критерии оценки и системы решающих правил в массовых процессах всегда определяются не по-

сле проверки, а до неё. По министерским критериям, гений русской литературы Александр Пушкин, не знавший математики, живи он в 2009 году, был бы однозначно лишён аттестата зрелости. Страна потеряла бы ещё одного гения. К счастью, тогда Россия имела Министерство народного просвещения.

Разные цели потребуют разные системы заданий, а вместе с этим и разных решающих правил и интерпретаций. Поскольку КИМы ЕГЭ создавались для достижения сразу нескольких целей, то реально это означало непригодность ни для одной цели по существу. Именно поэтому в отчётах ФИПИ фактически нет интерпретации результатов, адекватной для каждой поставленной цели.

Например, коррупция уменьшилась или увеличилась? Об этом мы можем узнать только по публикациям в независимых СМИ. Или, например, другой вопрос: восемь лет применяли КИМы ЕГЭ. Образование в России, в каждой отдельной области, за это время улучшилось или ухудшилось? Ответов на эти злободневные вопросы в отчётах нет.

Выводы

1. Демаркация педагогических измерений — важное направление работы по защите тестов от псевдотестовых и нетестовых методов. Полноценным и объек-

тивным методом педагогического измерения КИМы ЕГЭ не станут, если их не перепроектировать в тесты. Поскольку эта идея имеет мало шансов быть принятой Министерством образования и науки, а также частью педагогической общественности, надежды высших руководителей страны на корректировку КИМов и ЕГЭ могут оказаться неэффективными. Восемь лет бесплодного «экспериментирования» с КИМаи ЕГЭ подтверждают правильность такого мнения.

2. Пришло время уяснить, наконец, что единым этот экзамен не может быть по метрическим причинам, государственным — по соображениям необходимости выбора демократического пути развития образования. Государственный экзамен может представить Россию как полицейско-бюрократическое государство, стремящееся к контролю над системой образования и своими гражданами. Для сравнения: на Западе есть национальное общественно-профессиональное тестирование, независимое от государства, которое на добровольной основе помогает гражданам выявить свои способности к овладению той или иной образовательной программой.

3. Вместо предложенной корректировки КИМов надо вводить в стране систему педагогических измерений⁵⁶, настоящее тестирование и доброволь-

ность участия испытуемых в таком тестировании. Только тогда страна получит шанс полноценно развивать образование.

4. Для аттестации выпускников школ достаточно иметь систему сравнительно лёгких заданий, определяющих овладение допустимым минимумом подго-

товленности. Для приёма в вузы нужно применять тесты, различающиеся по содержанию и по уровню трудности, в зависимости от требований самого вуза. В учебных заведениях, там, где нужно выявить уровень речевой и логической культуры, полезен традиционный экзамен.

Методология

Методология