

Методология

ПРОБЛЕМА РАЗВИТИЯ ПЕДАГОГИЧЕСКИХ ИЗМЕРЕНИЙ

Вадим Аванесов

testolog@mail.ru

Каждая вещь известна настолько,
насколько её можно измерить.

У.Т. Кельвин

Хотеть совершить что-либо — значит хотеть всё то, что необходимо для достижения цели, в том числе хотеть приобрести те качества, которые нужны для воплощения замысла. Во всех прочих случаях вы не хотите, а просто стремитесь к чему-то, находитесь в плену собственной фантазии, сладостно упиваетесь задуманным¹.

В статье определены пять основных направлений развития теории и практики педагогических измерений. Первое — это демаркация (отграничение) педагогических измерений от контрольных материалов государственных экзаменов и от иных псевдометрических форм и методов, распространяемых сейчас в России под видом средств педагогических измерений. Второе направление — разработка вопросов формы тестовых заданий. Третье — содержание теста и тестовых заданий. Четвёртое направление — разработка критериев качества и эффективности педагогических измерений. И пятое

————— 1 —————

Ортега-И-Кассет.
Миссия университета.
М., 2010. С. 51.

направление — улучшение языка этой прикладной педагогической науки.

Развитию педагогических измерений препятствует проведение т.н. единого государственного экзамена, с его некачественными контрольными материалами. Эти материалы функционируют в России как бюрократическая подмена подлинных педагогических измерений. В статье обобщены ранее опубликованные факты и доказательства некачественности контрольных материалов российских государственных экзаменов.

Ключевые слова: педагогические измерения, демаркация, форма, содержание, язык, критерии качества результатов.

Введение

Педагогические измерения относятся к числу сравнительно малоисследованных направлений педагогической науки и образовательной практики. Они возникли в конце XIX — начале XX века за рубежом, вначале в Англии, США и Франции. В течение ста лет к этим трём странам добавились Австралия, Нидерланды, Израиль, Южная Корея, Япония. За прошедший век педагогические измерения превратились в одно из развивающихся направлений зарубежной педагогической науки.

В России сложилась иная ситуация. Здесь педагогические измерения заметного развития не получили. В XXI век Россия вошла с контрольными материалами государственного экзамена, созданными на основе политически привлекательной, но метрически ошибочной идеи — одновременной оценки выпускников школ и абитуриентов вузов общим (единым) набором заданий единого государственного экзамена (ЕГЭ). Качественно измерить заметно различающиеся уровни подготовленности выпускников российских школ и абитуриентов вузов бланковыми методами не удаётся в течение вот уже десяти лет. Контрольные материалы превратились, по сути, в антагонистов подлинных педагогических измерений.

Проблемная ситуация

Зримым поворотным моментом к ухудшению ситуации с педагогическими измерениями стало решение Правительства РФ о проведении в стране единого государственного экзамена (ЕГЭ)². В распоряжении Правительства РФ № 910-з- объяснялось, что ЕГЭ вводится с целью обеспечения *объективности* и унификации итоговой аттестации и вступительных испытаний в системе профессионального образования³.

2 —
Постановление Правительства РФ от 26 июля 2000 г. №1072-р. ОВД, № 18, 2000. С. 29.

3 —
<http://www.r-komitet.ru/obraz/EGE.htm>

Дело, казалось, нужное — обеспечение равного доступа граждан к высшему образованию и ликвидация разросшейся коррупции при приёме в вузы. Но идея унификации итоговой аттестации и вступительных испытаний в вузы не сочеталась с традициями российского образования, научными положениями социологии, педагогики, психологии и педагогических измерений, зато легко укладывалась в логику трансформации образовательного процесса в сторону усиления авторитарных бюрократических элементов.

Ответом на такую трансформацию стал ежегодный рост числа нарушений правил проведения ЕГЭ. Результаты фабрикуются в невиданных ранее масштабах, коррупция растёт, качество образования ухудшается. Информация о всех зафиксированных нарушениях в целом по стране и по отдельным территориям не публикуется; десять лет она оставалась скрытой образовательным ведомством.

Спорная изначально цель оценить, одновременно и качественно, разные по уровню подготовленности контингенты школьников и абитуриентов вузов не могла продуцировать позитивные результаты. Одновременно это делать можно, но качественно, бланковыми формами, нельзя. А потому не случайно в реальном исполнении контрольные материалы для дости-

жения этой спорной цели оказались некачественными.

Ошибочные цели введения единого государственного экзамена и некачественные инструменты достижения этих целей привели к тому, что в России педагогическое оценивание и педагогические измерения были заменены бюрократическим контролем посредством некачественных «материалов». Все исходные и подлинные результаты засекретили, после чего вместо разработки научно-обоснованных методов педагогических измерений и улучшения качества образования бюджетные средства стали тратиться на проведение государственного экзамена.

Сама идея отбирать абитуриентов одинаковыми «материалами» в разные вузы — тоже ошибочная и вредная. Если вузы разные и туда идут студенты с разной подготовкой, то и задания должны быть разные по уровню трудности, такие, которые позволяют каждому вузу точнее отобрать искомый контингент. При проверке всех одним и тем же набором «материалов» точность оценок резко ухудшается. Это особенно сказывается на правой стороне распределения исходных баллов, у наиболее подготовленных испытуемых.

Неоднократные обращения автора этой статьи и многих представителей научно-педаго-

| | |
|------------|------------------|
| ПЕД | |
| | измерения |

4

Каннети Э.
Масса и власть. М., 1997.

5

Хуторской А.В.
ЕГЭ — пример научно обоснованного нововведения. Фрагмент из книги: «Педагогическая инноватика». Уч. пос. 2 изд. М.: Изд. Центр «Академия», 2010. С. 97–101.
<http://testolog.narod.ru/Other20.html>. См. также Педагогические измерения. № 4. 2010. С. 17–21.

6

Хуторский А.В.
Дидактическая эвристика. Теория и технология креативного обучения. М.: Изд-во МГУ, 2003. 416 с.

7

Такой, например, алогизм: ВОССТАНИЕ СПАРТАКА БЫЛО:
• а) в 73 г. до н.э.
• б) в 73 г. н.э.
• в) в 1973 г.
• г) пока ещё не было
(Источник: Российская газета, 23 марта 2001 г.)

6

гической общественности к чиновникам всех уровней с предложениями прекратить производство некачественных материалов и проведение единого министерского контроля уровня подготовленности выпускников школ и абитуриентов различных вузов успеха не имели. Ответом было полное молчание. А это самый грубый способ защиты — притвориться глухим или непонимающим, как писал Э. Каннети. «Молчачий не раскрывается, но выглядит при этом опаснее, чем он есть на самом деле. В нём подозревают большее, нежели он скрывает. Он молчит, потому что ему есть что скрывать»⁴.

И действительно, с научно-педагогической точки зрения идея ЕГЭ оказалась уязвимой⁵. Преобладание внешней заданности в целях, содержании и технологии образования приводит к ослаблению внутренней мотивации учеников... к гиперболлизации формальных ценностей образования — получение отметки, сдачи экзаменов, поступление в вуз, его окончание⁶.

Ещё более идея ЕГЭ оказалась несостоятельной с точки зрения теории педагогических измерений. Похоже, что исполнители и чиновники не понимали сложности самой проблемы. Чтобы убедиться в некачественности заданий, с которых начинался так называемый эксперимент по ЕГЭ, достаточно взгля-

нуть на один пример контрольного «материала», не имеющего ничего общего с педагогическими измерениями⁷. Такой плохой старт стал одной из причин неудовлетворительного финиша. Некачественные контрольные «материалы» ЕГЭ дискредитировали благие замыслы Правительства РФ.

Последовало вынужденное признание В.В. Путиным справедливости критики ЕГЭ, сделанное, однако, вместе с предложением совершенствовать этот экзамен⁸. Но как это сделать? Конструктивных предложений по совершенствованию госэкзамена, созданного без науки об измерениях и мало-мальски обоснованного научно-педагогического проекта⁹, не было ни до, ни во время, ни после десятилетнего срока его проведения.

Между тем подлинные трудности с проведением ЕГЭ связаны не только, и не столько с проведением экзаменов, сколько с контрольными заданиями. Завеса секретности, незаконно возведённая вокруг исходных, реальных результатов, обернулась отрицательным образом для всей страны, в виде повсеместного ухудшения образовательной деятельности.

В органах управления образованием не считали нужным отвечать на обращения, критику или вступать в дискуссию со своими многочисленными оп-

понентами. Такого вызывающего отношения чиновников к своим оппонентам в истории России не было никогда. Это был новый стиль, утвердившийся в сфере управления образованием. Как отмечалось в литературе, особенно удручала полная глухота реформаторов к мнению педагогической науки и народа¹⁰.

В ситуации заинтересованности большинства сторон процесса ЕГЭ в повышении баллов испытуемых любыми способами, ЕГЭ начал саморазрушаться. В недавно опубликованном отчёте Общественной палаты РФ перечислено семь сторон социума, заинтересованных в завышении результатов ЕГЭ, общим числом миллион человек¹¹.

В результате справедливая, казалось бы, идея проведения в России ЕГЭ, но бездарно исполненная, стала с каждым годом терять своих сторонников. Надо подчеркнуть, что в западных странах ЕГЭ нет. Там проводится независимое, общественно-профессиональное тестирование, текущий и итоговый компьютерный адаптивный контроль в самих образовательных учреждениях, автоматизированный учёт всех решённых заданий, каждым учащимся или студентом, используется также рейтинг.

А потому ловчить невозможно и незачем, потому что

при приёме смотрят не только на бумаги и в кошелёк родителей, но и на реальные компетенции, которыми владеют поступающие в вузы. О подлинных учебных достижениях каждого абитуриента известно всё, и в любое время. Вот истинная польза открытости результатов и образовательных технологий для возможного свободного развития каждого гражданина, в зависимости от его способностей, прилежания и учебной мотивации.

Влияние ЕГЭ и контрольных материалов российского государственного экзамена на систему российского образования оказалось отрицательным. Давно известно, что вред от некачественных оценок в сфере образования существенно выше, чем кажущаяся от них польза¹².

Ещё раз подтвердилось на практике мнение выдающегося русского учёного и педагога Н.И. Пирогова. «Я почти ежедневно убеждаюсь из опыта, что экзаменационное направление в наших училищах не приносит никаких благих результатов. Оно вредно, оно возбуждает наклонность в учащихся учиться для экзамена, а не для науки¹³. Государственный экзамен довёл отмеченную наклонность до абсурда.

Станным образом верхняя власть дала исполнителям ЕГЭ право игнорировать всенародную, по существу, критику в

Методология

8

Путин В.В. назвал справедливой критику ЕГЭ. <http://www.gzt.ru/top-news/education/-putin-nazval-spravedlivoi-kritiku-ege-/362863.html?subscribenam=content&category=-putin-nazval-spravedlivoi-kritiku-ege-&date=310511>. 31.05.2011 в 14:21, обновлено 31.05.2011 в 14:49 GZTRU.

9

Аванесов В.С. Мифология ЕГЭ. <http://ps.1september.ru/2001/50/2-1.htm>.

10

Современные тенденции и проблемы модернизации российского образования. Донской гос. аграрный университет, с. 4. Пос. «Персиановский», 2010. 219 с.

11

Уроки проведения ЕГЭ-2010. Аналитический доклад. По материалам обращения граждан на горячую линию Общественной палаты Российской Федерации/ Под научной редакцией Л.Н. Духаниной. М.: Издательский дом Государственного университета Высшей школы экономики. М., 2010.

ПЕД
измерения

<http://www.oprf.ru/files/jkbhfhbd84729.pdf>. Ниже приводятся документированные выдержки из этого доклада.

12

DeVellis R.F.
Scale Development:
Theory and Application.
Applied Social Research
Methods Series, Vol. 26.
Sage Publications, 1991.

13

Пирогов Н.И.
Избр. пед. соч. М.,
1952. С. 219–220.

14

«Скажите, пожалуйста, что за беспредел у нас творится? Почему такое ощущение, что для детей введён концлагерь?» (запись № 269, Иркутская область); Цит. по книге: Уроки проведения ЕГЭ-2010. Аналитический доклад. По материалам обращения граждан на горячую линию Общественной палаты Российской Федерации/ Под научной редакцией Л.Н. Духаниной. Издательский дом Гос. унив. Высшей школы экономики. М., 2010.
<http://www.oprf.ru/files/jkbhfhbd84729.pdf>

их адрес, дала бюджетные средства на проведение асоциального бюрократического контроля¹⁴ и на производство некачественной¹⁵ продукции. И сверх того дала право не отчитываться должным образом о своей работе.

Тем не менее **В.В. Путин** не потерял ещё надежды на возможность совершенствования госэкзамена. Его понять можно. В 2000 году ЕГЭ казался важным фактором образовательной политики, утверждения идей усиления борьбы с коррупцией и установления справедливых форм приёма в вузы. Спустя 11 лет этот фактор обернулся ростом коррупции, снижения социальных гарантий обеспечения реального доступа граждан к высшему образованию. Сейчас госэкзамен стал похож на тяжёлый чемодан без ручки, который нести неудобно, а бросить, похоже, жалко.

Только в последнее время власть стала видеть глубину ямы под названием ЕГЭ, вырытой ошибочными целями, негодными инструментами (контрольными материалами) и бюрократическим режимом исполнения. Фактическое введение полицейских мер в образовательный процесс — незаконные обыски учащихся¹⁶, слежка за детьми в туалетах¹⁷ и т.п. — прямая дорога к разрушению не только ЕГЭ, но и нравственных основ российской педагогики и российского государства.

Государственный экзамен теперь становится отрицательным предвыборным фактором. Уже давно надо было начать поиск достойного способа его прекратить¹⁸, что могло бы дать некоторым избирателям новую надежду. Однако свой элемент неожиданности привнёс **А.А. Фурсенко**, заявивший, что этот экзамен — навечно! Неужели ему непонятно, что своеволие, которое позволяли себе чиновники все годы силового внедрения ЕГЭ, дошло до терпимых пределов. А потому его притязания увековечить бездарный ЕГЭ лишь ускорят созревание ситуации, о которой говорил известный классик: когда низы больше не хотят, а верхи уже не могут.

ЕГЭ провалился?

В критической литературе обнаруживается много мнений о провале ЕГЭ. В министерстве образования и науки, а также в Правительстве РФ так не считают. Хотя и об успехах тоже благоразумно перестали говорить. Видимо, играет свою роль периодическая замена кадров главных неудачников-толкачей ЕГЭ. Очевидно, вопрос в критериях — что считать провалом? Президент РФ Д.А. Медведев и Председатель Правительства РФ В.В. Путин считают, что у ЕГЭ есть плюсы

и минусы, но что в целом он эффективен¹⁹.

Никто из них, а также из министерства образования и науки, не назвали конкретные минусы ЕГЭ и что они считают нужным совершенствовать? Из чего можно понять, что их позиция имеет лишь общую, но не конкретно определённую направленность на сохранение этого экзамена. Но такая позиция не имеет положительной перспективы. Между тем конкретные недостатки ЕГЭ уже давно были выделены и названы в сотне научных работ многих авторов²⁰. На них не обращали внимания. Эта критика считалась маргинальной, а сами критики в кулуарных обсуждениях относились к маргиналам. На самом деле, маргинальной и ригидной оказалась позиция разработчиков ЕГЭ²¹ и их заказчиков.

Важным объективным свидетельством провала ЕГЭ стал уже упоминавшийся выше документированный отчёт Общественной палаты РФ²². В нём, в общем, в очередной раз показана неприемлемость и абсурдность сложившейся бюрократически-режимной практики проведения ЕГЭ. Сделанные там выводы, ввиду документированности событий и обращений, никем не опровергаются, но и опять не принимаются чиновниками, как нужные и полезные для исправления положения

дел. В общем, для постсоветской России — это обычная практика. Как тогда совершенствовать ЕГЭ? В этих условиях он и начал разваливаться.

Ссылаясь на выводы независимой общественной комиссии, куда входил и автор этого доклада, **С.М. Миронов**, заявил, что эксперимент по введению единого госэкзамена в стране *провалился*²³. «ЕГЭ не устранил коррупцию при поступлении в вузы и не позволил объективно выставлять оценки выпускникам школ», — пояснил он. Судя по смещению его с должности Председателя Совета Федерации, последовавшим вслед за этим, эти его правильные слова пришло не ко двору.

В.В. Черников тоже считает, что попытки повышения качества обучения через введение ЕГЭ провалились. Как известно, считает он, в тонущем корабле никакие прогрессивные способы мытья стаканов на кухне не спасают корабль. И пока критерии работы системы образования не будут напрямую встроены в стратегию выживания социума, результат будет плохим²⁴.

Все попытки властей собственными силами улучшить ситуацию и остановить разрушительный процесс оказались неудачными; они принесли лишь новые, очередные трудности. Это следствие системных дефектов самого российского го-

Методология

15

Провал ЕГЭ: большинство оценок оказались фальшивыми. http://goryshulakov.ru/rossiya/news_2009-11-12-13-00-08-168.html и мн. др. См. аргументацию автора о некачественности «материалов госэкзамена в статьях, представленных по адресу <http://obrazovanie.viperson.ru/wind.php?ID=425098>

16

«Сотрудник на экзамене подошёл к моему ребёнку и заставил показать содержимое карманов. Имеет ли он на это право?» (запись № 00026847, Республика Башкортостан).

17

«Правомерно ли, что когда детей водят в туалет, преподаватель наблюдает за детьми в туалете?» (запись № 1390, Республика Бурятия, запись № 5259, Московская область).

18

Аванесов В.С. Единый государственный экзамен надо прекратить. <http://obrazovanie.viperson.ru/wind.php?ID=425098>

ПЕД
измерения**19**

Аванесов В.С.
Спорное решение кремлёвской комиссии. <http://obrazovanie.viperson.ru/wind.php?ID=425098>

20

Сайт <http://obrazovanie.viperson.ru/wind.php?ID=425098>

21

Помнится, примерно за десять лет до распада СССР так же называли критиков ошибок партийного аппарата КПСС.

22

Уроки проведения ЕГЭ-2010. Аналитический доклад. По материалам обращения граждан на горячую линию Общественной палаты Российской Федерации/ Под научной редакцией Л.Н. Духаниной. М.: Издательский дом Гос. унив. высшей школы экономики, 2010. <http://www.oprf.ru/files/jkbhfbhd84729.pdf>

23

Сергей Миронов вынес приговор ЕГЭ. 28.01.2011. Авторы Евгений Насыров, Александра Гавычева. <http://www.gzt.ru/topnews/education/-sergei-mironov-vynes-prigovor-ege/>

сударственного экзамена и используемых там материалов. На критику дефектов разработчики этих материалов не реагировали должным образом, как будто они жили в другом государстве. В таком асоциальном варианте ЕГЭ был, конечно же, обречён на провал²⁵.

Реакция на этот прогноз была неожиданной. Вместо привычного молчания в министерстве образования и науки РФ, без какого-либо признания ошибочности ЕГЭ, тихо и быстро отказались от идеи единых экзаменов для большинства школьных предметов, равно как и от единых контрольных материалов. И в нарушение закона переналадили большинство единых экзаменов только на приём абитуриентов в вузы. А норму статьи закона о *едином* экзамене сохранили, похоже, лишь для двух предметов — русского языка и математики. Генпрокуратура и министерство юстиции РФ заняли привычную отстранённую от ЕГЭ позицию.

К слову, оба упомянутых выше экзамена имеют наибольшие метрические погрешности. Остальные т.н. единые госэкзамены после такой операции фактически перестали быть едиными, а потому уже больше не применяются для итоговой государственной аттестации выпускников школ. Бюрократическая логика незатейлива: меньше экзаменов для выпускников

школ, меньше двоек, меньше станет и неприятностей.

Хотя буква «Е» в аббревиатуре ЕГЭ для большинства экзаменов стала обречённой на фактическое вымирание, во всех бумагах и в средствах массовой информации она сохранялась. Пропагандистский прессинг создавал в социуме ложное представление, что единый государственный экзамен жив и вечно будет жить. Случившаяся в министерстве образования и науки подмена большинства ранее *единых* экзаменов *не едиными* экзаменами осталась в Правительстве РФ и в Госдуме как бы незамеченной. А потому финансирование провальной затеи продолжается.

ЕГЭ провалили некачественные контрольные «материалы» ФИПИ

Для реализации правительственной идеи проведения единого государственного экзамена был создан Федеральным институтом педагогических измерений (ФИПИ). Ему и была поставлена задача создать новые, доселе неизвестные в мире методы унификации двух экзаменов. На наше предупреждение о ложности самой идеи и об ошибочности принятого правительственного распоряжения чиновники привычно не реагировали²⁶.

ФИПИ погрузился в бесплодный «эксперимент» на целых восемь лет, без опубликованной научной программы. Отчёта о результатах этого пролонгированного поиска нет и, видимо, никогда не будет. К 2009 году терпение оппонентов и власти иссякло, а потому было приказано «Пора»! После чего материалы ФИПИ для единого государственного экзамена объявили пригодными для масштабного применения на массах школьников и абитуриентов, в штатном, как было отмечено, «режиме». Смысл последнего слова был тогда явно не понят и не оценён должным образом. Никто не мог тогда предположить, в ФИПИ создавался не научно-педагогический метод, как это должно было быть, а новый контрольно-бюрократический режим проведения массового государственного экзамена.

Отсюда становятся понятными две главные причины провала ЕГЭ: авторитарный бюрократический режим, установленный для его проведения, и некачественный инструментарий, созданный в ФИПИ для реализации такого режима контроля знаний.

Сам ФИПИ был превращён в исполнительный придаток министерства образования и науки. В условиях зависимости от ошибочной воли чиновников институт не получил возможно-

сти стать подлинным научным учреждением. То же произошло с Центром тестирования, который стал, по той же воле, головной организацией Рособнадзора по организационно-технологическому обеспечению проведения ЕГЭ и по трансформации исходных результатов ЕГЭ в производные. Надо было так менять результаты ЕГЭ, чтобы мало кто мог понять, что происходит в реальности. После чего Централизованное тестирование, проходившее ранее, на добровольной основе, во многих регионах России, перестало существовать²⁷. Оно стало частью бюрократического режима контроля.

Настоящие научные статьи и научные монографии по педагогическим измерениям в ФИПИ не готовились и не издавались. Похоже, что писать их там было некому или некогда. И сейчас на сайте института нет таких изданий. Бюрократическому режиму контроля знаний это и не надо. По полезности и информативности публикуемой научной литературы по проблеме педагогических измерений этот сайт целого института заметно уступает сайтам отдельных авторов.

Быстро нашлось размытое и ложное название производимым там изделиям: *контрольно-измерительные материалы* (КИМы) ЕГЭ. Размытое, потому что под название «материалы» можно подвести даже кан-

Методология

345598.html?from=copiedlink.

24

Черников В.В.

Геодидактика: Основания общей теории обучения человека в глобализованной среде современного социума. М: ООО «Астра-Полиграфия», 2010. 459 с.

25

Аванесов В.С.

Доживёт ли Единый государственный экзамен до 2009 года?
http://www.socpolitika.ru/rus/social_policy_research/analytics/document8205.shtm.

26

Аванесов В.С.

Единый государственный экзамен, или «Эта вещь будет посильнее Фауста Гёте». Учит. Газета. № 49. 2000 год.
<http://www.ug.ru/00.49/t25.htm>.

27

Михалёва Т.Г.

Двадцать лет централизованному тестированию в России. Что дальше? Педагогические измерения. № 1. 2011. С. 88–105.

ПЕД
измерения

28

Аванесов В.С.
Ни тест, ни экзамен.
Ж. «Университет и школа» №1–2. 2002. С. 26–27.

29

Обоснование этого вывода смотрите в работе автора: Являются ли КИМы ЕГЭ методом педагогических измерений? Педагогические измерения. № 1. 2009. С. 3–26. Вторая редакция — 25 мая 2009 г. и др. работы
<http://viperson.ru/wind.php?ID=425098>

30

Аванесов В.С.
Доживёт ли Единый государственный экзамен до 2009 года?
http://www.socpolitika.ru/rus/social_policy_research_analytics/document8205.shtml

31

В МВД отметили двойной рост коррупции после введения ЕГЭ.
<http://www.gzt.ru/top-news/accidents/-mbpch-prizyvaet-chinovnikov-sfery-obrazovaniya-307628.html?subscribenam=content&category=-mbpch-prizyvaet-chinovnikov-sfery-obrazovaniya-&date=270510>

целярский чих, попавший на служебную бумагу. Ложное, потому что эти КИМы не являются ни тестами, ни обычными экзаменами²⁸, ни педагогическими измерениями²⁹.

В ситуации отсутствия объективной статистики и вакуума эмпирических данных, общественное обсуждение проблем ЕГЭ было искусственно направлено исключительно в бездоказательный мир мнений. Это позволяло инициаторам и лоббистам ЕГЭ несколько лет держать оборону. Факты качества материалов они представить не смогли, а факты некачественности — не хотели, пришлось скрывать.

В данной статье сделана попытка собрать некоторые факты, указывающие на непригодность контрольных материалов ЕГЭ, в предположении, что это поможет, наконец, перевести общественный дискурс из сферы мнений в область бесспорных, эмпирически установленных фактов.

Начнём с того факта, что большинство единых госэкзаменов уже не дожили до прогнозированного автором этой статьи срока в своём прежнем, подчеркнём, *едином* виде³⁰. Разве это не признак провала прежнего ошибочного замысла проведения *единого* государственного экзамена?

А кто, кроме чиновников, сейчас считает, что новый вари-

ант проведения ЕГЭ, ориентированного преимущественно на приём в вузы, лучше? Никто. Сама идея отбирать абитуриентов, одинаковыми «материалами», в разные вузы — такая же ошибочная и вредная, какой была и идея единого экзамена. Если вузы разные и туда идут студенты с разной подготовкой, то в соответствии с требованиями постулата соответствия теории педагогических измерений, задания должны быть разные по уровню трудности, такие, которые позволяют, соответственно, каждому вузу точнее отобрать искомый контингент. При проверке всех одним и тем же набором «материалов» точность оценок резко ухудшается. Это особенно сказывается на правой стороне распределения исходных баллов, у наиболее подготовленных испытуемых, что видно на приводимых далее гистограммах.

Коррупция не уменьшилась, а увеличилась, теперь уже вокруг ЕГЭ. И это уже не мнение, а профессионально установленный точный факт³¹. А значит, это, одновременно, и факт провала одной из главных целей введения ЕГЭ — снижения коррупции. В министерстве образования и науки все годы т.н. эксперимента по ЕГЭ ублаживали себя тем, что этот факт их не касается, что борьба с коррупцией вокруг ЕГЭ — дело других ведомств. Никто в Пра-

вительстве РФ их не поправил. Хотя без шума меняли, время от времени. Именно такая негласная позиция и привела к всплеску коррупции.

Факты непригодности «материалов» ЕГЭ

На рис. 1 представлено трансформированное распределение т.н. тестовых баллов ЕГЭ по русскому языку, полученное благодаря запросу комиссии при Совете Федерации. Реальное распределение так и не удалось получить даже этому представительному органу.

Представленное распределение похоже на гребёнку, вызванную, скорее всего, искусственным нарушением ранее выбранных интервалов построения гистограммы. Что само по

себе, при отсутствии реальных результатов, может свидетельствовать о стремлении исполнителей скрыть настоящую картину — непригодность исходных «материалов» и полученных по ним результатов ЕГЭ. Реальное распределение было бы много полезнее для улучшения дел. Но и это слышать не хотели.

Здесь неработающей оказалась правая часть, связанная с использованием части «С». Разве это не признак провала некачественного контрольного «материала» и самого ЕГЭ? Справа на рис. 1 два всплеска, в районе девяносто четырёх и ста баллов, могут интерпретироваться как коррупционные, в принципе не вытекающие из логики распределения результатов. А это тоже вполне статистически доказательный признак провала.

Методология

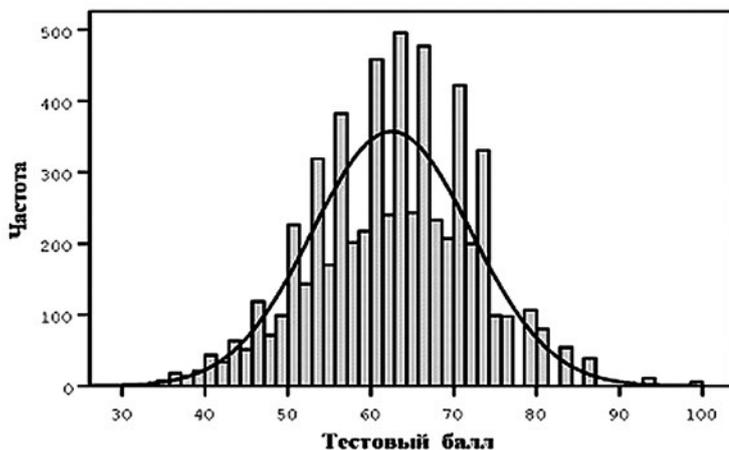


Рис. 1. Распределение тестового балла ЕГЭ по русскому языку в Вологодской области в 2008 году

ПЕД
измерения

32

Русский язык. 2.2.1.
Характеристика контрольных измерительных материалов по русскому языку 2010 г. Отчёт ФИПИ. 2010.

33

Уроки проведения ЕГЭ-2010. Аналитический доклад. По материалам обращения граждан на горячую линию Общественной палаты Российской Федерации / Под научной редакцией Л.Н. Духаниной. М.: Издательский дом Государственного университета Высшей школы экономики, 2010.
<http://www.oprf.ru/files/jkbhfhbd84729.pdf>

Другой факт искусственно сжатого (справа) трансформированного распределения результатов по русскому языку, но теперь уже во всероссийском масштабе представлен на рис. 2³². Этот т.н. тест, сделан избыточно лёгким. Таким смещённым, асимметричным тест не позволяется делать даже студентам. По приведённым данным, только 3,3% испытуемых не смогли преодолеть минимально определённый в кабинетах Рособнадзора порог, который устанавливали, в нарушение существующих на Западе этических норм, не до, а после проведения экзамена.

Слева, на рис. 2, возник своеобразный «проходной двор» — почти все испытуемые (96,7%) успешно сдали требования единого государственного экзамена к владению русским языком. Это очевидный артефакт, вызванный завышенной лёгкостью, высокой вероятностью угадывания правильных ответов в

первой части и массовыми нарушениями процесса проведения государственного экзамена.

Реально дети говорят на русском языке, а особенно пишут, не лучше, а хуже, чем это было до введения ЕГЭ. Похоже, что именно про такой случай разработки контрольных «материалов» прочувственно говорила пожилая учительница из Самарской области, позвонившая в Общественную палату РФ: *«Я хотела бы передать благодарность тому, кто придумал ЕГЭ. ЕГЭ построен таким образом, что почти невозможно сдать на двойку!»* (документированная запись № 2128, Самарская область)³³. Разве эта благодарность опытного педагога, довольного таким удачным исходом государственного экзамена, — не убедительный пример некачественности используемого тем же государством контрольного «материала» ЕГЭ?

Справа на рис. 2 распределение сжато так, что коррупцион-

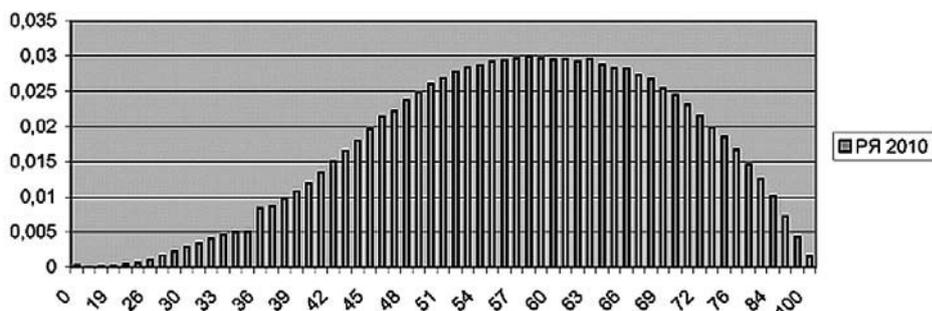


Рис. 2. Распределение участников экзамена по полученным тестовым баллам в 2010 г.

ные всплески исчезли совсем, появилось немало отличников. Таким ЕГЭ тоже оказались довольными большинство испытуемых, часть родителей. Довольны и чиновники министерства образования, создавшие видимость большого скачка в работе министерства, хотя на самом деле, по общему мнению, образование в стране деградирует.

Напрашиваются вопросы. Сколько времени всё это шоу будет ещё продолжаться? Так же «вечно», как и ЕГЭ? Тогда надо быть готовым к тому, что ЕГЭ, независимо от желания толкачей этого экзамена, закончится очень скоро! Похожая на ЕГЭ система была в Древнем Китае, но она рухнула под напором коррупции³⁴.

Шансы закончить среднюю школу с документом об успешном окончании школы даёт большинству экзаменуемых и т.н. КИМ по математике (рис.3). Но здесь распределение исходных баллов оказалось асимметричным в противоположную сторону.

Распределение в левой части шкалы исходных результатов по математике похоже на нормальное, но это характерно лишь для слабой части испытуемых, с модальными значениями 8 и 9 баллов, близких к минимально «проходным». Иначе говоря, массовое математическое образование у нас достигло минимального одобренного порога.

Из-за стремления разработчиков этого «материала» выполнить министерские установки «единого» экзамена, точность оценок у хорошо подготовленных выпускников школ на российском государственном экзамене оказалась недопустимо низкой. В зоне чрезмерно высокого риска получения ошибочных оценок оказались самые подготовленные выпускники школ. Разве это не метрический аргумент провала ЕГЭ?

При сравнении результатов получается, что математику в российских школах дети знают *много* хуже, чем русский язык. Именно на такой вывод наталкивает единственный в истории проведения ЕГЭ случай публикации распределения исходных результатов ЕГЭ по математике в 2010 г. (рис. 3 (рис. 1.1 второй главы отчёта (с. 5)³⁵). Если это считать фактом, то он указывает на провал образовательной политики. А если это артефакт, вызванный неправильной разработкой контрольных «материалов», то это — очередное свидетельство непригодности контрольных «материалов» проведения единых государственных экзаменов.

В классической (статистической) теории педагогических измерений уже давно сложилась традиция считать результаты тестов с асимметричным распределением баллов как относительно невалидные по двум

Методология

34

Аванесов В.С.
Тесты: история и теория. Из глубины веков. Управление школой. № 12, март, 1999 г. С. 14.

35

Результаты единого государственного экзамена (май-июнь 2010 года). Москва, 2010, Общее руководство — Ершов А.Г.
http://www.fipi.ru/binaries/1085/1_razdel_11.pdf

ПЕД
измерения

возможным причинам: либо трудность большинства заданий не соответствует уровню подготовленности большинства испытуемых, либо, наоборот, уровень подготовленности большинства испытуемых не соответствует уровню трудности большинства заданий. И то, и другое снижает качество измерений и практическую полезность.

Посмотрим ещё на один факт³⁶, представленный в табл. 1.

Анализ таблицы позволяет доказательно утверждать, что в этом материале, созданном для проверки знаний по физике, нет ничего похожего на тест как систему заданий возрастающей трудности, имеющих параллельные варианты. Они совсем не параллельны по уровню трудности. В таблице видна не-

допустимо большая вариация заданий по трудности, между вариантами одного и того же номера контрольного «материала». Из-за чего ЕГЭ превратился не в средство педагогического измерения, а в форму проведения лотереи.

Уже по одной этой причине единый экзамен, как государственный, не имеет права на применение, поскольку этим некачественным экзаменом нарушаются права испытуемых и их родителей на объективную и справедливую оценку уровня подготовленности. Баллы испытуемых сильно зависят от попавшегося варианта экзамена. И это ещё одно доказательство некачественности контрольных материалов государственных экзаменов. Разве и это не провал ЕГЭ?

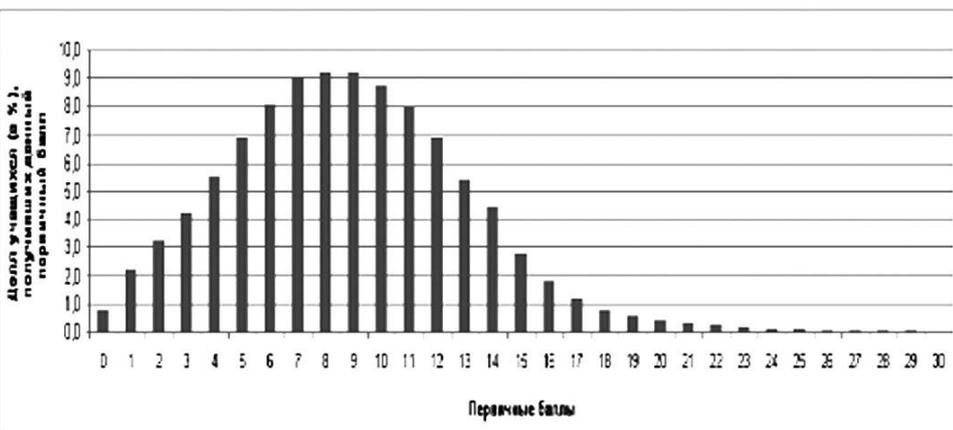


Рис. 3. Распределение исходных результатов по ЕГЭ (математика-2010)

36
Источник: Государственное учреждение «Областной центр мониторинга качества образования». Единый государственный экзамен 2007. Физика. Сборник аналитических материалов. Кемерово, 2007.

Таблица 1

**Решаемость заданий ЕГЭ по физике в 2007 году,
по вариантам (1 волна)**

| 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 |

Методология

Продолжение табл. 1

| 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 |
| 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 |

| | |
|-----|-----------|
| ПЕД | |
| | измерения |

Три вопроса

В связи с представленными здесь фактами могут возникнуть три вопроса и три вероятных ответа.

Первый вопрос: могли бы данные 2011 года стать свободными от недостатков или стать лучше ранее полученных результатов, представленных в этой статье? Ответ прост: из-за режима секретности результатов ЕГЭ науке это неизвестно. Надо открыть исходные распределения прошлых лет и новые данные. Эти данные не конфиденциальны по смыслу, но превратились в конфиденциальные по антиобразовательному замыслу и исполнению. Только после этого можно узнать — новые контрольные «материалы» стали лучше или, скорее всего, ещё хуже? Неужели гражданам страны и Правительству РФ неинтересно знать ответ на такой вопрос?

Второй вопрос: а сколько нужно ещё привести фактов некачественности контрольных материалов государственных экзаменов, чтобы Министерство образования и науки, а также Правительство РФ приняли наконец, решение о прекращении проведения единых государственных экзаменов для выпускников школ и абитуриентов различных вузов?

До выборов 2012 года мы вряд ли получим ответ и на этот

вопрос. Пока только видно, что власть, похоже, стремится сохранить «навечно» асоциальный вариант экзамена и хочет засекречивать некачественные материалы далее. А после 2012 года ответ будет зависеть от результатов выборов и от исхода возможных послевыборных событий.

Третий вопрос — а что, если всё продолжится так же, как сейчас? Тогда можно предположить, что государственные экзамены развалятся, и окончательно. Никакой бюрократический режим их проведения и никакие контрольные «материалы» их не спасут. А образование в стране начнёт деградировать ещё сильнее. Россия, с её природными ресурсами и огромными территориями, останется без качественно образованного населения. Как писал О. Шпенглер, сегодняшняя мощь «наций» есть следствие всего-навсего прошлой школьной политики. Пусть каждый читатель решит для себя — правильно ли будет сказать, что возможная завтрашняя немощь страны станет результатом сегодняшней образовательной политики?

Спорное решение Государственной Думы

Развитию педагогических изменений в России мешают кон-

трольные материалы для проведения государственных экзаменов. Теперь они станут мешать ещё больше. Наметившаяся тенденция законодательного усиления ответственности за нарушения режима проведения единого государственного экзамена — это ещё один шаг к стороне разрушения образования в стране.

Госдума приняла законопроект, в котором написано, что контрольные измерительные материалы (КИМы) ЕГЭ признаются конфиденциальной информацией³⁷. Это было сделано опять-таки, казалось бы, из благих побуждений — сохранить упомянутые материалы от несанкционированного доступа до начала экзаменов, обеспечивая, тем самым, равный доступ граждан к высшему образованию.

Факты нарушения сохранности используемых в ЕГЭ заданий возникают ежегодно, во многих местах и в массовом порядке, что, очевидно, указывает на несовершенство идейных основ, проектной схемы, процедуры и самой организации этого экзамена. Достаточно привести один самый массовый пример. В 2011 году во время ЕГЭ по математике многие выпускники получали ответы на мобильные телефоны из популярной социальной сети «В Контакте». В специально созданной в этой сети группе, которая существует уже несколько лет и в которую

вступили около 300 тысяч человек, размещались варианты ответов ЕГЭ, причём информацию предлагали купить за деньги³⁸.

По замыслу депутатов Госдумы, придание материалам юридического статуса конфиденциальности позволит удерживать ЕГЭ от дальнейшего интенсивного саморазрушения. Но это очень спорный, если не ошибочный, замысел. Он укладывается в логику бюрократического режима проверки знаний, но противоречит логике научной организации системы педагогического контроля, с педагогическими принципами этой системы³⁹.

Само слово «конфиденциальный» происходит от латинского *confidentia* — доверие, и в современном русском языке означает «доверительный, не подлежащий огласке, тайный, секретный». Получается, с точки зрения языка, что понятия «конфиденциальная информация», «тайна» и «секрет» могут восприниматься как равнозначные. Однако в юридической литературе эти понятия не признаются равнозначными.

Согласно п. 2 ст. 10 ФЗ «Об информации...» документированная информация с ограниченным доступом по условиям её правового режима подразделяется на информацию, отнесенную к государственной тайне, и на конфиденциальную информацию. Отсюда можно вы-

Методология

37

Госдума одобрила поправки по совершенствованию процедуры ЕГЭ. РИА «Новости». http://www.proforientator.ru/index.php?option=com_content&view=article&id=462:2010-09-10-20-29-49&catid=2:5

38

Римский В.Л.
ЕГЭ способствует росту коррупции в системе образования. Педагогические измерения. № 2. 2011.

39

Аванесов В.С.
Основы научной организации педагогического контроля в высшей школе: Уч. пособие для слушателей Учебного центра. М.: МИСиС, 1987. 107 с.

ПЕД
измерения

вести, что КИМы ЕГЭ законом «Об информации...» не могут быть отнесены ни к государственным секретам, ни к государственной тайне. Кроме того, понятие «тайна» означает, как видно из п. 2 упомянутого закона, ещё и наличие правового режима сохранности информации. Судя по всему, конфиденциальную информацию о некачественных материалах ЕГЭ, не имеющих никакого юридического статуса, в Госдуме не сопроводили наличием такого правового режима. Да и как это сделать?

К тому же применение данной юридической новации в практике усложняется тем, что в настоящее время нет чёткой и единой классификации видов конфиденциальной информации. Действующими нормативными актами установлено свыше 30 разновидностей конфиденциальной информации. Какую разновидность конфиденциальности педагогических заданий держат в уме думские законодатели, науке так же не известно.

И уже совсем не может быть юридического термина «контрольные измерительные материалы», которым сегодня оперирует послушное правительство большинство Госдумы. Инициаторам конфиденциальности и каждому голосовавшему за предложение правительства полезно задать один-единственный вопрос: где они видели из-

мерительные свойства пресловутых «материалов»?

Важно, к тому же, понять, что в решении Госдумы речь идёт, по сути, не об информации, вполне юридическом термине, а о педагогических «материалах», имеющих неконкретный, скорее мифологический смысл⁴⁰, относительно которых режим конфиденциальности и правового режима если и возможен, то в гротескном или абсурдном виде. Хотя бы задумались над тем, что педагогические задания ещё нигде и никогда в мире в законодательном органе не засекречивали.

В словосочетании «контрольно-измерительные материалы» слово «измерительные» не имеет никакого реального смысла. «Материалы» также нигде и никогда не были средством измерения. Это неумный вымысел российских министерских умельцев, размноженный услужливой пропагандой.

В других странах к секретной информации, ограниченной коротким временем, относят не сами задания, а только коды ответов на варианты используемых заданий. Так это установлено, например, законодательным органом Казахстана. И это правильно.

В большинстве демократических стран тестированием занимается не государство, а специальные и независимые центры, которые по уставу не имеют

40

Аванесов В.С.
Мифология ЕГЭ. Первое сентября. № 50,
2001 г. <http://ps.1september.ru/2001/50/2-1.htm>

права засекречивать задания и иные «материалы». Если там последние вдруг обнаружатся. Напротив, они *обязаны* заблаговременно знакомить испытуемых со всеми типами заданий, которые потенциально могут быть (а могут и не быть) использованы в процессе конкретного предстоящего контроля. Здесь возникает вопрос стратифицированной выборки заданий теста из генеральной совокупности заданий.

Если какой-то абитуриент в процессе специальной подготовки сумеет найти правильные ответы на большинство заданий по интересующему предмету, имеющихся в базе, то это может означать только одно — такой испытуемый имеет высокий уровень подготовленности. И ему нечего бояться и ловчить. Он готов к тестированию или экзамену. Легко видеть, что открытость заданий в таком варианте даёт положительные образовательные результаты. Принятая же Госдумой РФ конфиденциальность контрольных «материалов» ведёт к отрицательным результатам.

В разных странах мира не было случаев запрета на ознакомление учащихся с примерами заданий, содержащихся в базе данных. Эти задания, наоборот, активно используются для тренировочного контроля уровня подготовленности испытуемых в процессе самообразова-

ния. Этим затрагивается другой ключевой вопрос эффективной образовательной деятельности — использование обучающего потенциала заданий в тестовой форме для обучения. В этом аспекте Россия слишком отстала от Запада и Востока. Очевидно, мы имеем дело с проявлением ретроградной позиции министерства образования и науки.

Вопрос, следовательно, там стоит не о запрете на ознакомление учащихся с заданиями вообще, а об ограничении доступа к ознакомлению с ответами на конкретные варианты заданий, отобранные специальными методами педагогических измерений для применения *только* во время проведения контроля.

Например, настоящее тестирование абитуриентов и школьников проводится в течение примерно сорока минут. За это время никто из испытуемых не успеет ни скопировать, ни передать тексты всех доставшихся лично ему вариантов заданий. И затем получить решённые другими лицами задания; на решение трудных заданий и на проверку ответов тоже нужно немало времени. И потом ещё надо будет вписать правильно полученные ответы в специальный бланк. Гораздо эффективнее в таких случаях тактика использования отведённого времени на быстрое решение тех заданий, которые ис-

ПЕД
измерения

пытуемый может решить самостоятельно.

Разгадку смысла думского решения о признании КИМов ЕГЭ «конфиденциальной» информацией можно найти в работе **А.В. Ефремова** «Бюрократия и бюрократизм». Там говорится об использовании общих приёмов искусственной служебной конфиденциальности, помогающей бюрократии уйти от общественного контроля, закрытия доступа к информации, о действиях аппарата, недопущения действительной гласности⁴¹ и много чего ещё интересного. К ЕГЭ это относится на все сто процентов.

Своим сомнительным решением Госдума фактически вводит запрет на ознакомление педагогов и общественности с некачественными, надо подчеркнуть, контрольными заданиями и на реальные распределения результатов ЕГЭ. Между тем в любой демократически развитой стране граждане имеют право знакомиться с заданиями, которые могут быть использованы в процессе проверки знаний. Из обозримой совокупности каждому испытуемому достанется выборочная совокупность из вариантов таких заданий.

В разумно управляемой стране научные работники обязаны знать реальные распределения результатов и объективную статистику, чтобы подсказать властям, где и что можно улуч-

шать. Особенно по областям и территориям России. Но торжествующее ныне маргинальное бюрократическое мышление не позволяет говорить правду и что-то улучшать в этой сфере. А потому засекречивание «материалов» год от году крепчает.

Любопытна ещё одна сторона признания конфиденциальными контрольных материалов государственного экзамена. Своим решением, вольно или невольно, Госдума РФ вводит, по существу, запрет на научную критику этих самых «материалов». Потому что невозможно или очень рискованно аргументированно критиковать конфиденциальную информацию бюрократического режима проверки знаний учащихся.

И не появятся ли, вслед за решением о конфиденциальности, желание и возможность привлечь к ответственности за критику тех, кто выступает против применения некачественных, но теперь уже «конфиденциальных контрольных материалов» российского государственного экзамена? По сути, заготовлена юридическая дубинка, которая может соблазнить неумные головы использовать её против критиков ошибочных и вредных для страны решений. Однако пришла пора образумиться и остановиться, не множить далее число глупостей, которых и так накопилось много в системе образования.

41

Ефремов А.В.
Бюрократия и бюрократизм. Образование: проблемы, поиски, решения / Сб. научно-методических работ. Вып. № 4, Ханты-Мансийский ГУИПП, «Полиграфист», 2002.

Наконец, в свете изложенного не совсем понятно — каким образом законодательная инициатива Правительства РФ о введении конфиденциальности «материалов ЕГЭ», одобренная Госдумой, сочетается с заявлением главы того же самого Правительства В.В. Путина⁴² о его поддержке справедливой критики ЕГЭ⁴³? Конструктивная критика невозможна без открытых публикаций «материалов» ЕГЭ? Нет ли здесь противоречия? Если есть, то эти противоречия хорошо бы устранить самому Правительству РФ.

Ранний этап развития педагогических измерений

Начало развития теории и практики психологических и педагогических измерений можно проследить, начиная с работ Ф. Гальтона, Э. Торндайка, А. Бине и других зарубежных авторов, работавших в конце XIX — начале XX века⁴⁴. Практика измерений существовала тогда в форме применения вопросов и заданий, с последующим суммированием баллов испытуемых за каждое задание. Эти баллы нередко назывались тестовыми, и они воспринимались тогда как результаты педагогических измерений. Такая практика перевода качественных явлений в количественные

оценочные показатели имела глубокие исторические корни⁴⁵.

Позже, во второй половине XX века, возникла культура педагогических измерений, включающая в себя, как тогда писали, «теорию тестов», культуру формулирования заданий⁴⁶, а затем шкалирования исходных баллов испытуемых и уровня трудности заданий. А это полностью изменило прежние представления о сущности педагогических измерений. Теперь измерения понимаются как объединённый процесс разработки теорий, методик тестирования испытуемых и компьютерного шкалирования получаемых результатов.

К счастью для педагогических измерений, необходимость соотносить теоретические соображения с проверкой реальных результатов была осознана довольно рано. В начале XX века А. Бине и Т. Симон проводили эмпирическую проверку каждого задания, которое предполагалось включить в тест. Для оценки пригодности заданий для измерения уровня развития детей разного возраста авторы использовали два основных критерия: 1) меру трудности заданий, которую они определяли по доле правильных ответов в группах детей разного возраста⁴⁷ и 2) информацию о степени совпадения результатов теста с мнением преподавателей. А. Бине и Т. Симону принадле-

Методология

42

В.В. Путин назвал справедливой критику ЕГЭ. <http://www.gzt.ru/topnews/education/-putin-nazvaspravedlivoi-kritiku-ege-/362863.html?subscribe=content&category=-putin-nazvaspravedlivoi-kritiku-ege-&date=310511>. 31.05.2011 в 14:21, обновлено 31.05.2011 в 14:49 GZT.RU.

43

Аванесов В.С. Владимир Путин поддерживает критику ЕГЭ. <http://obrazovanie.viperson.ru/wind.php?ID=639869&soch=1>

44

Аванесов В.С. Возникновение и развитие тестового контроля. Гл. 2 дис. уч. степени д-ра пед. наук «Методологические и теоретические основы тестового педагогического контроля». С-Пб: СПбГУ, 1994. 339 с.

45

Аванесов В.С. Тесты: история и теория. Из глубины веков. Управление школой. № 12, март, 1999. С. 14.

ПЕД
измерения

46

*Haladyna, T.M.,
Downing, S.M.,
Rodriguez, M.C.*

A Review of Multiple-Choice Item-Writing Guidelines for Classroom Assessment. *Applied measurement in education*, 15(3), 309–334.

47

В наше время в качестве одной из наиболее распространённых мер трудности задания принимается не доля правильных ответов, а, наоборот, доля неправильных ответов. Ведь чем труднее задание, тем больше неправильных ответов.

48

«...it will be noticed that our tests are well arranged in a real order of increasing difficulty». Binet A. & Simon T. (1916). The development of intelligence in children. (Translations of articles in *L'Annee Psychologique*, 1905, 1908, and 1911). 1905, p. 185. Vineland, NJ: Vineland Training School.

жит также важная идея расположения заданий теста в зависимости от возрастания меры их трудности⁴⁸. Этими критериями проверки пригодности заданий для разработки теста как системы заданий возрастающей трудности тестологи пользуются до сего дня.

В России, после революционных потрясений начала XX века, возрождавшееся народное образование открылось практически всем известным тогда в мире педагогическим новациям. Появились комплексный и исследовательский методы обучения, Дальгон план, метод проектов, система полного усвоения знаний, коллективного обучения и многое другое. Среди этих новаций были попытки создания теории и методики педагогических измерений. Период развития был, однако, недолог. В 1936 г. Постановлением Совета Народных Комиссаров СССР тесты были запрещены⁴⁹. В 1937 году репрессиям подверглись нарком просвещения союза и всех республик, их заместители, начальники отделов, более 300 учёных-педагогов и педологов.

Роль авторитаризма и эгалитаризма

В начале 30-х годов в СССР начал усиливаться авторитаризм, вначале в политике, а затем и в

сфере образования. Он препятствовал развитию объективных методов контроля знаний, ибо в нём усматривалась угроза существованию авторитарной системы, с её тщательно культивируемым отрицательным подбором кадров из близкого окружения, из-за чего каждое последующее поколение управленческих кадров оказывается хуже предыдущего. Решающим критерием подбора кадров становятся не конкурс способностей и уровня образованности, а исполнительность в выполнении приказов вышестоящих лиц.

Одновременно с авторитаризмом в те же годы начал распространяться и эгалитаризм, который проявлялся в виде идеи всеобщего равенства и искусственного выравнивания, даже там, где неравенство неизбежно. Например, неравенство способностей разных лиц к различным видам деятельности. Эгалитаризм проявлялся в виде псевдонаучных суждений о равенстве способностей, одинаковой обучаемости, о возможности каждым человеком овладеть любой профессией, а также в виде одинаковой зарплаты преподавателям, независимо от фактических знаний тех учащихся, которых они учат. Эгалитаризм проповедует принцип: — «за равный труд — равную зарплату». Главный упор делается на равном труде, с неизбежным отсюда нормирова-

нием, как главным средством повышения общей производительности труда.

Под влиянием авторитарных и эгалитарных социально-психологических и политических установок в педагогической среде актуализировались афоризмы типа «нет плохих учеников, есть плохие учителя», «незаменимых людей нет» и т.п. В обществе, устроенном по эгалитарному принципу, все люди рассматриваются как взаимозаменяемые «винтики» одной огромной государственной машины, управляемой незаменимым лидером. Такие установки уже нанесли стране неисчислимый вред.

Воцарившаяся после запрета ситуация отрицательного отношения к тестам не давала кому-либо практической возможности опубликовать что-либо в их защиту. И эта ситуация продолжалась в течение длительного времени. Авторитарный стиль легко утверждается, но трудно и долго искореняется.

Сейчас в России вместо тестов используют контрольные «материалы» государственного экзамена. Вместе с тем растёт и число тех, кто путает тесты с контрольными материалами ЕГЭ, кто рассматривает тест как элементарный перевод с английского языка слов «проба, испытание, проверка». Без указания на существенные признаки, выводящие тест за пределы элементарного пере-

вода, как того требует научная логика.

Как показывает опыт издания российского научно-методического журнала «Педагогические измерения», к настоящему времени в стране оказалось слишком мало авторов, способных писать качественные тексты по проблеме педагогических измерений.

Период активного развития педагогических измерений

На Западе периодом активного развития педагогических измерений стало время между 1950 годом и серединой 80-х годов. Этот период отмечен научно-методическими достижениями D.C. Adkins⁵⁰, N.E. Glonlund⁵¹, C.M. Lindvall & A.J. Nitko⁵², W.A. Mehrens & Lehmann⁵³, A.J. Nitko⁵⁴, G. Sax⁵⁵, A.G. Wesman⁵⁶ и мн. др.

В девяностые годы XX века в педагогических измерениях США наметился явный успех, связанный с появлением работ R.E. Bennet & W. Ward⁵⁷, S.B. Carlson⁵⁸, N. Frederiksen⁵⁹, E.E. Freeman⁶⁰, R.J. Mislevy⁶¹, S.J. Osterlind⁶², W. Wiersma & S.G. Jurs⁶³ и мн. др. Ближе к созданию теории педагогических измерений подошёл в США R. Ebel⁶⁴.

В СССР и России проблемой развития педагогических

Методология

49

Аванесов В.С.

Проблема психологических тестов. Вопросы психологии, 1978, № 5. С. 97–107; Из истории психологических тестов (с. 23–34) в учебном пособии «Общая психодиагностика» / Под ред. А.А. Бодалева и В.В. Столина // М.: МГУ, 1987. 304 с.; СПб. Изд-во «Речь». 440 с. 2006. Общая психодиагностика. <http://medbookaide.ru/books/fold1002/book2002/p1.php>

50

Adkins, Dorothy C.

Test construction: Development and Interpretation of Achievement Tests. 2 ed. Columbus, Ohio, Ch.E. Merrill Publ. Co..., 1974. 164 pp.

51

Gronlund N.E.

Measurement and Evaluation in Teaching. The MacMillan Co. N-Y., 1965.

52

Lindvall C.M., Nitko A.J.

Measuring pupil achievement and aptitude. 2nd ed. Harcourt Brace Jovanovich, Inc., 1975.

ПЕД
измерения**53**

*Mehrens W.A.,
Lehmann I.J.*
Measurement and
Evaluation in Education
and Psychology. N-Y.,
Holt, Rinehart and
Winstin Inc., 1978.

54

Nitko A.J.
Educational Tests and
Measurement: An
Introduction. Harcourt
Brace Jovanovich, Inc.,
N-Y., 1983. 674 pp.

55

Sax G.
Principles of Educational
Measurement and
Evaluation. Belmont, Ca:
Wardsworth Publ. Co.,
Inc., 1980.

56

Wesman A.G.
Writing the Test Item //
In: Thorndike R.L. (Ed.)
Educational
Measurement. 2nd ed.
Am. Council on
Education. 1971. 768 pp.

57

*Bennet R.E.,
Ward W. (Eds.)*
Construction vs. Choice
in Cognitive
Measurement. Hillsdale,
N-J. 1993.

измерений занимался, среди других, автор этой статьи. За период с начала 80-х годов XX века до начала второго десятилетия XXI века удалось подготовить и опубликовать десятки работ по ключевым вопросам развития педагогических измерений. Почти все они представлены на сайте автора⁶⁵. Ранее этого периода были опубликованы труды по психологическим⁶⁶ и социологическим⁶⁷ измерениям.

В качестве основы теории педагогических измерений была разработана система понятий педагогических измерений, принципы разработки формы⁶⁸ и содержания тестовых заданий⁶⁹, концепция пяти этапов тестового педагогического процесса⁷⁰. В этой концепции бюрократическому режиму проведения контроля посредством контрольных материалов места нет.

Уровневое функциональное педагогических измерений

Первый, самый распространённый уровень функционирования педагогических измерений связан с практической деятельностью, похожей на тестирование. Но только внешне. Как, например, в контрольных материалах российского государственного экзамена есть задания,

внешне похожие на тестовые, но их действительные свойства далеко не тестовые. Засекречивание помогает сохранить этот дефект режимного метода.

Действительно, получаемые в ЕГЭ баллы слишком часто ошибочно называются тестовыми, хотя сущностных признаков теста там нет. Именно в таких случаях на Западе говорят, что тестовые баллы без процесса специального шкалирования ещё не образуют измерений⁷¹. А иные баллы, полученные посредством контрольных материалов, вне культуры метрического процесса, вообще далеки от педагогических измерений. Именно этот случай подпадает под известное определение классика: если бы внешняя видимость вещей и их суть совпадали бы, то всякая наука была бы излишней.

Первый уровень существует стихийно, без видимой организации. Носители такого низкого уровня функционирования педагогических измерений редко контактируют между собой, поскольку нет реальной содержательной основы для их совместной деятельности. Но их иногда собирали для осуждения неугодных лиц. Тогда этот уровень принимал форму антинаучной, по сути, деятельности.

Например, при министерстве образования в 90-х годах был организован т.н. Научный совет по тестированию. Ничего науч-

ного после себя этот совет не оставил, но ущерб педагогическим измерениям, науке он нанёс немалый. Главное — он «научно» одобрил проведение ЕГЭ. Недостаточно качественно проводившееся централизованное тестирование и массовое применение тестоподобных форм в большинстве государственных и вузовских центров тестирования также можно отнести к нетеоретическим формам начальной тестовой и тестоподобной деятельности.

Первый уровень преодолим при наличии научной литературы и поощрения открытости результатов, обязательной отчётности, понимания позитивной роли научной критики полученных результатов и отстранения бюрократии от непосредственного управления проблемами, принципиально требующими других, общественно-профессиональных форм управления. При государственном управлении сферой проверки знаний, отгороженной от общественной критики режимом засекречивания педагогических заданий и обобщённых статистических распределений, антипедагогические искривления, ложь, обман и провалы неизбежны.

Второй уровень функционирования педагогических измерений связан с практикой тестирования, опирающейся на отдельные элементы теории педагогических элементов. Это

уровень соединения теории с практикой, постепенного усиления практики теорией, развития методики педагогических измерений. Он характерен для использования в развивающихся педагогических учреждениях и для улучшающегося образования в целом. И неизбежен в силу того, что нельзя сразу перейти с низшего уровня на высший уровень. Этот уровень открывает возможности теоретизации, позволяет развивать личность и систему образования на основе развивающейся теории, что сказывается благотворно на соединении интересов личности, общества и государства.

Третий уровень включает активное развитие теории педагогических измерений и эффективную практическую деятельность по разработке научно обоснованных методов измерения (методологии). На этом уровне ценится теоретическое и методологическое знание, издаётся литература, выпускаются специализированные научные журналы, разрабатываются математико-статистические программы разработки тестов и шкалирования данных тестирования, ведутся научные дискуссии. Качественно проводимое тестирование также относится к практической стороне третьего уровня педагогических измерений.

Предмет третьего уровня — разработка теории педагогических измерений. Критикой тео-

Методология

58

Carlson S.B.
Creative classroom testing: Ten designs for assessment and instruction.
Princeton, N-J: ETS, 1985.

59

Frederiksen N. et. al. (Eds).
Diagnostic Monitoring of Skill and Knowledge Acquisition.... 510 pp.;
Frederiksen N. et. al. (Eds). Test Theory for a new generation of tests.
Lawrence Erlbaum Ass., Hillsday, N-J, 1993. 404 pp.

60

Freeman E.E.
The Redesign of Testing for the 21st century.
Princeton, N-J, 1986. 101 pp.

61

Mislevy, R.J.
Foundations of a new test theory. In: Frederiksen N. et. al. (Eds). Test Theory for a new generation of tests.
Lawrence Erlbaum Ass., Hillsday, N-J, 1993. 404 pp.

62

Osterlind S.J.
Constructing Test Item.
Boston etc. : Kluwer, 1989. 343 pp.

63

Wiersma W., Jurs S.G.
Educational Measurement and Testing. 2-nd. ed. Allyn and Bacon, Boston, 1990. 415 pp.

ПЕД
измерения

64

Ebel R.L.
Measuring Educational Achievement. Prentice-Hall, Inc. Englewood Cliffs, New Jersey, 1965. 481 pp.

65

<http://testolog.narod.ru>

66

Аванесов В.С.
Проблема психологических тестов. Вопросы психологии. 1978.

67

Аванесов В.С.
Тесты в социологическом исследовании. М.: Наука, 1982. 199 с.

68

Аванесов В.С.
Форма тестовых заданий. Уч. пособие. Москва. Центр тестирования, 2005. 152 с.

69

Аванесов В.С.
Композиция тестовых заданий. 1, 2 и 3 изд.: 3 изд. М.: Центр тестирования, 2002. 217 с.

70

Аванесов В.С.

рии и конструктивным преобразованием практики призвана заниматься методология педагогических измерений. Третий уровень противостоит псевдоизмерениям. Он предполагает открытую публикацию всех статистических результатов, представляющих интерес для тестологов, снятие всех незаконных запретов на публикацию реальных тестовых результатов, без указания данных об испытуемых. Особенно необходимы статистика параллельности заданий, меры общей и дифференцированной надёжности результатов, доказательства пригодности используемых заданий и полученных данных для различных целей. К сожалению, этой статистики тоже нет. Бюрократия своих привычек не меняет.

Определения исходных понятий теории педагогических измерений

Ранее педагогическое измерение было определено как *процесс* определения меры *интересующего свойства личности* испытуемых на *латентной интервальной шкале* посредством *качественного* теста, состоящего из *системы* заданий равномерно возрастающей трудности, позволяющего получать педагогически целесообразные ре-

зультаты, отвечающие критериям *надёжности, валидности, объективности и эффективности*. В этом определении курсивом выделены основные термины, позволяющие отграничить признаки педагогических измерений от прочих методов, научных, псевдонаучных и ненаучных⁷².

На языке философии Педагогические измерения можно рассматривать как специфическую форму научно-практической *педагогической* деятельности, имеющей своим предметом получение численных значений уровня образовательной подготовленности личности по интересующей учебной дисциплине или интересующего педагога свойству. Выделение курсивом слова «педагогической» имеет актуальный для России смысл, указывающий на принадлежность этой проблемы не образовательной бюрократии, как сейчас, а педагогической науке.

Любопытно схематическое представление самого простого, классического определения измерения, данного более чем полвека назад С.С. Стивенсом⁷³.

Здесь оно сопровождается примером создания шкалы отношения граждан к иммигрантам⁷⁴. Согласие респондента с содержанием каждого следующего вопроса указывает на более высокую меру его толерант-

ного (невраждебного) отношения к иммигрантам. Враждебное отношение и готовность использовать насилие по отношению к «иностранцам» измеряется шкалой национализма.

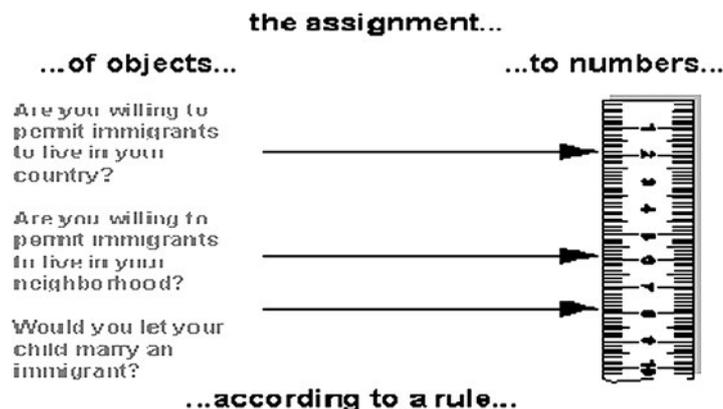


Рис. 4

Цель педагогического измерения — определить количество интересующего латентного свойства личности (меру интересующего признака), присущего данному испытуемому.

Результат педагогического измерения — латентная числовая величина, позволяющая установить числовое соотношение между испытуемыми по изучаемому свойству. Аналогично устанавливается числовое соотношение между используемыми в тесте заданиями, по изучаемым свойствам. Измеряемую величину образуют интересующее свойство и результаты теста, варьирующие, по уровню проявления, у множества испытуемых. Отсутствие вариации результатов и высокая

погрешность являются существенными признаками некачественности педагогического измерения.

Главный метод педагогических измерений — педагогический тест, представляющий собой систему вариативных заданий, равномерно возрастающей трудности, позволяющей качественно оценить структуру и эффективно измерить уровень подготовленности испытуемых по одной или нескольким учебным дисциплинам. Смысл словосочетания «система вариативных заданий» в том, что каждое задание теста имеет свои параллельные варианты.

В педагогических измерениях в обязательном порядке используются математико-статис-

Методология

71

Wright B.D.,
Linacre J.M.
The Differences Between
Scores and Measures.
Rasch Measurement
Transactions, 1989, 3:
3 p. 63. См. также раздел
«Raw scores are not mea-
sures». In: Measurement for
Social Science and
Education. A history of
social science measurement.

72

Аванесов В.С.
Проблема демаркации
педагогических измере-
ний. Педагогические из-
мерения. № 3. 2009.
С. 3–37. [http://
obrazovanie.viperson.ru/
wind.php?ID=425098](http://obrazovanie.viperson.ru/wind.php?ID=425098)

73

Scaling is the assignment
of objects to numbers
according to a rule.

74

Scaling. [http://www.
socialresearchmethods.net/
kb/scalgen.php](http://www.socialresearchmethods.net/kb/scalgen.php)

тические методы шкалирования получаемых результатов испытуемых и свойств заданий. В итоге получается шкала, имеющая нулевое начало, единицу измерения, достаточную протяженность. Это — свойства шкалы.

Модель измерения определяется как структурное построение, позволяющее соединить латентную переменную величину с наблюдаемыми значениями этой величины⁷⁵.

Педагогической шкалой называется упорядоченная совокупность значений испытуемых на латентной переменной величине, служащей основой для измерения данной величины. Идеальный образ шкалы педагогического измерения — линейка с нулевой отметкой посередине, с равными интервалами значений, с положительными значениями справа и отрицательными значениями — слева.

Поскольку отрицательные значения баллов испытуемых критически воспринимаются в образовательной среде, часто используются методы трансформации данных, такие, чтобы баллы всех испытуемых были представлены на шкале положительных значений. Наиболее часто применяемая трансформационная шкала исходных баллов испытуемых — так называемая Т-шкала.

Шкалирование определяет-ся как та часть процесса измерения, в котором исходные значе-

ния тестовых баллов испытуемых и меры трудности заданий переводятся в стандартные значения с общей единицей измерения, с удобными для интерпретации значениями средней арифметической и вариации данных. В Rasch Measurement результатом шкалирования является расположение объектов на шкале логитов.

Другие важные определения теории педагогических измерений сформулированы в работах, представленных в журнале «Педагогические измерения» и на сайте автора⁷⁶.

Необходимость методологии педагогических измерений

Ранее методология педагогических измерений была определена как учение об основных положениях, формах, методах, принципах научного исследования и организации эффективной практики в разработке показателей качества подготовленности учащихся и показателей педагогической деятельности⁷⁷.

Одна из важных задач методологии — правильное определение целей развития педагогических измерений. Недостаточно внимание к вопросам методологического анализа целей ЕГЭ и контрольных материалов привело к печальным результатам⁷⁸.

75

Bollen K.A.
Structural Equations with
Latent Variables. N-Y,
Wiley & Sons, 1989.
514 pp.

76

<http://testolog.narod.ru>

77

Вопросы методологии педагогических измерений. Педагогические измерения. № 1. 2005.
<http://testolog.narod.ru>

78

Аванесов В.С.
Ошибочные цели — плачевные результаты. Педагогические измерения. № 4. 2010. Вторая, расширенная версия этой статьи представлена по адресу:
<http://obrazovanie.viperson.ru/wind.php?ID=425098>

Второй важный вопрос методологии — это критика и развитие теории. В российской и мировой литературе нет достаточной ясности относительно теоретической основы педагогических измерений. На Западе больше развиты статистические и математические теории педагогических измерений, в то время как собственно педагогическую теорию измерений там пока не удалось создать. В первую очередь это касается трёх главных вопросов теории: вопросов композиции формы и содержания педагогических тестов, а также вопросов состояния понятийного аппарата педагогических измерений.

Раньше в качестве такой основы ошибочно рассматривались статистические теории⁷⁹. Затем — математическая теория Item Response Theory (IRT). В России IRT иногда провозглашалась как «единственная» и «современная» теоретическая основа педагогических измерений. Но это ошибочно. Все существующие сейчас теории применимы не только к педагогическим, но и к психологическим и социологическим измерениям. Хотя эти теории широко применяются для проведения педагогических измерений, в них нет собственно педагогического содержания, как нет методов и принципов решения содержательных педагогических проблем.

К тому же это теории формальные, не имеющие, по существу, педагогического содержания. В то время как именно содержание теории указывает на её предметную отнесённость. А это означает, что эти теории в значительной мере являются общими, или общенаучными, не нагруженными содержанием собственно педагогики⁸⁰.

Вопрос о необходимости создания собственно педагогической теории измерений был поставлен в первом номере российского научно-методического журнала «Педагогические измерения»⁸¹. В той статье отмечалось, что основное содержание теории педагогических измерений могли бы составить собственный язык (тезаурус) теории, принципы формулирования заданий, а также содержание теста, тестовых заданий и заданий в тестовой форме.

Внимание читателей обращалось на такие ключевые вопросы теории педагогических измерений, как система понятий теории, вопросы формы и содержания тестовых заданий, возможности обучающего потенциала заданий в тестовой форме, подходов к педагогической интерпретации результатов измерения — содержательно ориентированной или критериально ориентированной или нормативно ориентированной.

Методология

79

Lawley D.N.
On Problems Connected with Item Selection and Test Construction // Proceedings of the Royal Society of Edinburgh. Section A Mathematical and Physical Sciences. 43 v. LXI, part 111, p. 273–287, 1943; Gulliksen H. Theory of Mental Tests. N-Y. Wiley. 1950. 486 p.; Cronbach L.J., Gleser G.C., Nanda H., Rajaratnam N. The dependability of behavioral measurements: Theory of generalizability for scores and profiles. N-Y: John Wiley and Sons. 1972.

80

Аванесов В.С.
Три статьи. Item Response Theory: Основные понятия и положения. Статья первая. Педагогические измерения. № 2. 2007; Истоки и основные понятия математической теории педагогических измерений (Item response Theory). Статья вторая; Понятия и методы математической теории педагогических измерений (Item Response Theory). Статья третья. Педагогические измерения. № 4. 2009.

81

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

ПЕД
измерения

82

Аванесов В.С.
Проблема демаркации педагогических измерений. Педагогические измерения. № 3. 2009. С. 3–37. <http://obrazovanie.viperson.ru/wind.php?ID=425098>

83

Там же.

84

*Haladyna T.M.,
Downing S.M.,
Rodriguez M.C.*
A Review of Multiple-Choice Item-Writing Guidelines for Classroom Assessment. Applied measurement in education, 15(3), 309–334.

85

Whipple G.M.
Manual of Mental and Physical Tests. Warwick & York, Baltimore, Md., 1910.

86

Odell C.W.
Traditional examinations and new-type tests. N-Y., The Century Co., 1928.

87

*Hawkes H.F.,
Lindquist E.F., Mann C.R.*
The construction and use of achievement examinations. Boston: Houghton Mifflin, 1936.

Основные направления развития педагогических измерений

Первое и самое актуальное направление развития — это полное, по возможности, отграничение (*демаркация*) педагогических измерений от контрольных материалов государственных экзаменов и от иных псевдометрических форм и методов, распространяемых сейчас в России под видом средств педагогических измерений. Необходимость демаркации вызвана расширяющейся практикой смешения научных форм и содержания педагогических измерений с ненаучным и антинаучными элементами проведения в России ЕГЭ.

Демаркации педагогических измерений было посвящено специальное исследование автора, в котором было выделено 13 критериев, позволяющих отделить педагогические измерения от прочих методов⁸². Основные критерии, сформулированные в том исследовании, — это организация в стране процесса педагогических измерений; развитие культуры научно обоснования тестов; обязательность предварительной эмпирической апробации всех заданий, включаемых в методы измерения, открытость результатов измерений для научного анализа и критики, организа-

ция теоретических исследований, технологичность всех этапов, разработка стандартов метрической деятельности, пакетов для математико-статистической обработки данных, технических и этических стандартов на разработку и применение тестов и др.⁸³

Второе направление развития педагогических измерений — это совершенствование форм тестовых и нетестовых педагогических заданий. Сложность такой работы вытекает из противоречия, реально существующего между теоретическим и практическим мышлением. Большинству чиновников и практиков тестового процесса форма заданий считается знакомой и вполне понятной, а потому они не видят здесь никаких проблем. Их логика проста: в любом деле в первую очередь важна не форма, а важно содержание. Однако в педагогических измерениях форма играет очень важную роль, как способ связи и упорядочения элементов содержания в общей композиции тестовых заданий. Нарушение требований формы резко ухудшает понимаемость содержания заданий, а следовательно и качество измерения.

Теоретическое мышление находит отношение практиков к форме ограниченным и вредным, тормозящим развитие тестовой культуры. От качества формы зависит понимание со-

держания заданий, технологичность тестового процесса, точность и эффективность измерений. Нарушение требований формы является также одним из самых распространённых источников погрешностей в контрольных материалах, где нарушения формы и логики содержания заданий — явление частое, происходящее год от года. Пока нет мастерского владения формами тестовых заданий, не будет и качественных тестов.

На Западе развитие формы протекало в рамках прикладного направления педагогики, называемого Item Writing⁸⁴. Первое описание формы заданий с выбором одного правильного ответа можно найти в трудах G.M. Whipple⁸⁵. В США форму тестовых заданий исследовали C.W. Odell⁸⁶, H.F. Hawkes, E.F. Lindquist, C.R. Mann⁸⁷, W.S. Monroe, J.C. DeVoss, F.J. Kelly⁸⁸, H.D. Risland⁸⁹, G.M. Ruch⁹⁰, M.W. Richardson, J.T. Russel, J.M. Stalnaker, L.L. Thurstone⁹¹, J. Follman, B. Hall, R. Wiley, J. Hartmanh⁹² и мн. др. Немного позже, в 20–40-х годах, стали делаться попытки механизировать, а затем и автоматизировать формы контроля знаний.

В те годы экспериментальным сравнением тестовых форм в СССР успешно занимался Г.С. Костюк⁹³. «Форма теста, — писал он в 1928 году, является одной из тех проблем тестиро-

вания, научное изучение и решение которой чрезвычайно важно и неотложно»⁹⁴. Этот призыв был проигнорирован в то время, это был период утверждения авторитаризма. В наше время на важность формы тестовых заданий недавно обратил внимание А.Г. Войтов⁹⁵.

Опираясь на приведённые труды, автор этой статьи наметил свой подход, суть которого — изменение логической основы для заданий в тестовой форме. Если обычной основой заданий являются вопросы и ответы, то в нашей практике разработки заданий в тестовой форме используются утверждения (суждения), которые при выборе правильных ответов превращаются в истинные высказывания, а при выборе неправильных ответов — в ложные утверждения.

Число ответов желательно иметь больше, чем один. Тогда знания можно проверить шире, глубже, полнее, с меньшей вероятностью угадать один правильный ответ. А это повышает валидность и надёжность педагогических измерений⁹⁶.

Третье направление — улучшение содержания тестовых заданий. Содержание теста и его вариантов является приемлемым, если по итогам ответов на задания любого варианта одного и того же теста достигается хорошая сравнимость с другими вариантами. Иначе говоря,

Методология

88

Monroe W.S., DeVoss J.C., Kelly F.J.
Educational tests and measurements. Boston: Houghton Mifflin, 1924.

89

Risland H.D.
Constructing tests and grading in elementary and high school subjects. N-Y, Prentice-Hall, 1937.

90

Ruch G.M.
The objective or new-type examination. Glenville, Ill: Scott Foresman, 1924.

91

Richardson M.W., Russel J.T., Stalnaker J.M., Thurstone L.L.
Manual of examination methods. University of Chicago, Board of Examinations, 1933.

92

Follman J., B. Hall, R. Wiley, J. Hartmanh.
Relationship Between Objective Test Formats. Educ. Rev., 1947: 26; 150–151.

93

Костюк Г.С.
О зависимости результатов тестирования от формы теста // Тесты: теория и практика. М.: Моск. психологическое объединение, 1928. 205 с.

| | |
|-----|-----------|
| ПЕД | |
| | измерения |

94

Костюк Г.С.
Избранные психологические труды. М., 1988. С. 73.

95

Войтов А.Г.
Форма и формы тестовых заданий // Педагогические измерения. № 2. 2005. С. 25–34.

96

Композиция тестовых заданий. М. Издания 1998, 2002 гг.: Форма тестовых заданий. М.: Центр тестирования, 2005. 155 с.

97

Аванесов В.С.
Проблема качества педагогических измерений. Педагогические измерения. № 2. 2004. С. 3–31.

98

Аванесов В.С.
Тестирование как основа формирования честной и объективной оценки учебных достижений // Материалы 2-й ежегодной конф. «Новые подходы в преподавании бизнеса и экономики. С. 241–245. USAID, Amer. Councils for International Education. Бишкек, 29–30 апреля 2002.

испытуемому должно быть всё равно, на какой вариант теста он отвечает. В контрольных материалах ЕГЭ результат сильно зависит от меры трудности заданий, доставшихся испытуемому. В то время как в хорошем тесте все варианты каждого задания сопоставимы между собой по уровню трудности, а также по содержанию учебных тем, проверяемых у каждого испытуемого.

Четвёртое направление — разработка критериев качества и эффективности педагогических измерений

Все годы развития педагогических измерений в теории рассматривались два основных критерия качества — надёжность и валидность. Эти два критерия и по сей день считаются основными. В российской практике всё ещё продолжают говорить и писать слова «надёжность и валидность тестов» вместо более правильного выражения «надёжность и (или) валидность результатов тестирования».

Оба эти критерия качества педагогических измерений, их основные составляющие — понятия, методы и формулы — были рассмотрены во втором номере нашего журнала⁹⁷.

Помимо этих двух традиционных критериев качества результатов, в последние годы большое внимание привлёк к себе критерий объективности

тестовых результатов⁹⁸. По сути, это был самый первый критерий качества тестовых результатов, с которых начинали свою работу классики западной психометрики. Ввиду широты самого понятия «объективности» и недостижимости её полного достижения, внимание большинства исследователей переключилось на критерии надёжности и валидности. Однако с момента возникновения методологии Rasch Measurement внимание многих авторов опять привлёк критерий объективности.

Основные требования, позволяющие достигнуть объективированные результаты измерения, были сформулированы в работах последователей Г. Раша и в нашей статье. В сжатом варианте таких требований всего пять:

1. Параметры математических моделей для измерения свойств заданий и испытуемых не должны быть взаимно зависимы. Это главное научное достижение G. Rasch. Вся технология RM вытекает из свойства независимости параметров испытуемых от параметров заданий, и наоборот.

2. Метод измерения должен быть сравнительно лёгким, компьютеризованным, полностью, по возможности, технологичным. Это требование позволяет привлечь к проведению измерений большое число школьных педагогов и вузовских преподавателей.

3. Для начинающих исследователей считается наиболее приемлемой одномерная модель измеряемого свойства. Продвинутое авторы, имеющие подходящее математическое и статистическое образование, тяготеют к многомерным моделям измерения, которые чаще оказываются интереснее одномерных моделей.

4. Получаемые значения в процессе педагогических измерений должны отвечать требованиям линейности шкалы, что обеспечивает сравнительно удобные варианты математических аксиом и операций, а также интерпретации результатов.

5. Монотонность отображения измеряемого свойства испытуемых или заданий в числовую шкалу. Смысл этого требования прост: испытуемые, имеющие более высокий уровень подготовленности, должны получать и более высокий балл в RM⁹⁹. Измерения по теории Rasch отвечают всем этим требованиям.

И, наконец, четвёртым, по счёту, но не по важности, стал критерий эффективности тестов и тестовых заданий. Если относительно первых трёх критериев речь шла о качестве результатов тестирования, то четвёртый критерий используется для принятия решения об эффективности каждого отдельного задания или теста в целом¹⁰⁰.

В процессе тестирования приходится принимать во внимание затраты времени и средств на измерение интересующего свойства испытуемых, искать возможности применения кратковременных тестов вместо длительных методов педагогического оценивания или четырёхчасового некачественного государственного экзамена.

Фактор времени — один из главных признаков различия между тестом и государственным экзаменом. Качественно разработанный тест всегда короче по времени, точнее, адекватнее для совокупности испытуемых, технологичнее, имеет меньшую погрешность измерения, объективнее, справедливее и экономнее, чем любой экзамен.

Эффективность рассматривается как комплексный критерий оценки полезности тестов¹⁰¹, тестовых заданий и тестовых результатов, по широкому кругу экономических, социальных, психологических. Эффективность тестов зависит от целей, методологии, теории, методики и от текущей практики педагогических измерений.

Пятое направление развития педагогических направлений — это развитие языка этой прикладной науки. Основные вопросы развития этого направления представлены в статье автора¹⁰².

Методология

99

Аванесов В.С.
Проблема эффективности педагогических измерений. ПИ. № 4. 2008. (Переработано).

100

Аванесов В.С.
Методологические и теоретические основы тестового педагогического контроля. Дис. на соиск. уч. степени д-ра пед. наук. Санкт-Петербург, 1994. 339 с.

101

Аванесов В.С.
Проблема эффективности педагогических измерений. Педагогические измерения. № 4. 2008 (Переработано).

102

Аванесов В.С.
Язык педагогических измерений. Педагогические измерения. № 2. 2009. С. 29–60.