

# Основные направления развития педагогических измерений

*Вадим Сергеевич Аванесов,  
доктор педагогических наук*

• педагогические измерения • демаркация • форма • содержание • язык • критерии качества результатов •

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)<sup>1</sup>. В распоряжении Правительства РФ № 910-з объяснялось, что ЕГЭ вводится с целью обеспечения

объективности и унификации итоговой аттестации и вступительных испытаний в системе профессионального образования<sup>2</sup>.

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 обо 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 достиже-

<sup>1</sup> Постановление Правительства РФ от 26 июля 2000 г. № 1072-п. ОвД, № 18, 2000 г., с. 29.

<sup>2</sup> <http://www.r-komitet.ru/obraz/EGE.htm>

ния 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.

С научно-педагогической точки зрения идея ЕГЭ оказалась уязвимой<sup>3</sup>. Преобладание внешней заданности в целях, содержании и технологии образования приводит к ослаблению внутренней мотивации учеников..., к гиперболизации формальных ценностей образования — получение отметки, сдача экзаменов, поступление в вуз, его окончание<sup>4</sup>.

Ещё более идея ЕГЭ оказалась несостоятельной с точки зрения теории педагогичес-

ких 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 взглянуть на один пример контрольного «материала», не имеющего ничего общего с педагогическими измерениями<sup>5</sup>. Такой плохой старт стал одной из причин неудовлетворительного финиша. Некачественные контрольные «материалы» ЕГЭ дискредитировали благие замыслы Правительства РФ.

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 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 — 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 польза<sup>6</sup>.

<sup>3</sup> Хуторской А.В. ЕГЭ — пример научно необоснованного нововведения. Фрагмент из книги: «Педагогическая инноватика»: Учеб. пос. 2-е изд. М.: Изд. центр «Академия», 2010. С. 97–101. <http://testolog.narod.ru/Other20.html>. См. также Педагогические измерения. 2010. № 4. С. 17–21.

<sup>4</sup> Хуторский А.В. Дидактическая эвристика. Теория и технология креативного обучения. М.: Изд-во МГУ, 2003. 416 с.

<sup>5</sup> 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.)

<sup>6</sup> DeVellis R.F. Scale Development: Theory and Application. Applied Social Research Methods Series, Vol. 26. Sage Publications, 1991.

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 учащимся учиться для экзамена, а не для науки»<sup>7</sup>. Государственный экзамен довёл отмеченную наклонность до абсурда.

Странным образом верховная власть дала исполнителям ЕГЭ право игнорировать всеобщую, по существу, критику в их адрес, предоставила бюджетные средства на проведение асоциального бюрократического контроля<sup>8</sup> и на производство некачественной<sup>9</sup> продукции. И сверх того дала право не отчитываться должным образом о своей работе.

### ЕГЭ провалился?

Важным объективным свидетельством провала ЕГЭ стал уже упоминавшийся выше документированный отчёт Общественной палаты РФ<sup>10</sup>. 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.

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 государственного экзамена и используемых там материалов. На критику дефектов разработчики этих материалов не реагировали, как будто они жили в другом государстве. В таком асоциальном варианте ЕГЭ был, конечно же, обречён на провал<sup>11</sup>.

Реакция на этот прогноз была неожиданной. Вместо привычного молчания в Министерстве образования и науки РФ, без какого-либо признания ошибочности ЕГЭ, тихо и быстро фактически 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 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<sup>12</sup>. 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 пресинг создавал в социуме ложное представление, что Единый государственный экзамен жив и вечно будет жить. Случившаяся в Министерстве образования и науки под-

<sup>7</sup> Пирогов Н.И. Избр. пед. соч. М.: 1952. С. 219–220.

<sup>8</sup> «Скажите, пожалуйста, что за беспредел у нас творится? Почему такое ощущение, что для детей введен 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 науч. ред. Л.Н. Духаниной. М.: Изд. дом Гос. унив. высшей школы экономики, 2010. <http://www.oprf.ru/files/jkbhfhbd84729.pdf>

<sup>9</sup> Провал ЕГЭ: большинство оценок оказались фальшивыми. [http://gory.shulakov.ru/rossiya/news\\_2009-11-12-13-00-08-168.html](http://gory.shulakov.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>

<sup>10</sup> Уроки проведения ЕГЭ — 2010. Аналитический доклад. По материалам обращения граждан на горячую линию Общественной палаты Российской Федерации / Под науч. ред. Л.Н. Духаниной. М.: Изд. дом Гос. унив. высшей школы экономики, 2010. <http://www.oprf.ru/files/jkbhfhbd84729.pdf>

<sup>11</sup> Аванесов В.С. Доживёт ли Единый государственный экзамен до 2009 года? [http://www.socpolitika.ru/rus/social\\_policy\\_research/analytics/document8205.shtml](http://www.socpolitika.ru/rus/social_policy_research/analytics/document8205.shtml)

<sup>12</sup> Деменчёнок О.Г. Погрешность баллов Единого государственного экзамена // Педагогические измерения. 2011. № 4.

мена большинства ранее *единых* экзаменов *неединиными* 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 Федеральный институт 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 реагировали<sup>13</sup>.

ФИПИ погрузился в бесплодный «эксперимент» на целых восемь лет, без опубликованной научной программы. 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 возможности 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 существовать<sup>14</sup>. Оно стало частью бюрократического режима контроля.

Настоящие научные статьи и научные монографии по педагогическим измерениям в ФИПИ не готовились и не издавались. Производимым там изделиям быстро нашлось размытое и ложное название: *контрольно-измерительные материалы (КИМы) ЕГЭ*. Размытое, потому что под название «материалы» можно подвести даже канцелярский чих, попавший на служебную бумагу. Ложное, потому что эти КИМы не являются ни тестами, ни обычными экзаменами<sup>15</sup>, ни педагогическими измерениями<sup>16</sup>.

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 дожило до прогнозированного автором этой статьи срока в своём прежнем, подчеркнём, *едином* виде<sup>17</sup>. Разве это не признак провала прежнего ошибочного замысла проведения *Единого* государственного экзамена?

<sup>13</sup> Аванесов В.С. Единый государственный экзамен, или «Эта вещь будет посильнее Фауста Гёте» // Учит. газета. 2000. № 49. <http://www.ug.ru/00.49/t25.htm>.

<sup>14</sup> Михалёва Т.Г. Двадцать лет централизованному тестированию в России. Что дальше? // Педагогические измерения. 2011. № 1. С. 88–105.

<sup>15</sup> Аванесов В.С. Ни тест, ни экзамен // Университет и школа. 2002. № 1–2. С. 26–27.

<sup>16</sup> Обоснование этого вывода смотрите в работе автора: Являются ли КИМы ЕГЭ методом педагогических измерений? // Педагогические измерения. 2009. № 1. С. 3–26. Вторая редакция — 25 мая 2009 г. и др. работы <http://viperson.ru/wind.php?ID=425098>

<sup>17</sup> Аванесов В.С. Доживёт ли Единый государственный экзамен до 2009 года? [http://www.socpolitika.ru/rus/social\\_policy\\_research/analytics/document8205.shtml](http://www.socpolitika.ru/rus/social_policy_research/analytics/document8205.shtml)

Коррупция не уменьшилась, а увеличилась, теперь уже вокруг ЕГЭ. И это уже не мнение, а профессионально установленный точный факт<sup>18</sup>. А значит, это одновременно и факт провала одной из главных целей введения ЕГЭ — снижения 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 рисунке 1 два всплеска,

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

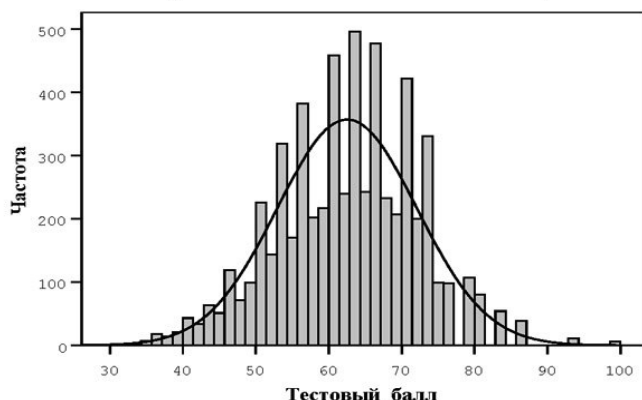


Рис. 1

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.

Другой факт искусственно сжатого (справа) трансформированного распределения результатов по русскому языку, но теперь уже во всероссийском масштабе представлен на рис. 2<sup>19</sup>. Этот так называемый «тест»

<sup>18</sup> В МВД отметили двойной рост коррупции после введения ЕГЭ. <http://www.gzt.ru/topnews/accidents/-mbpch-prizyvaet-chinovnikov-sfery-obrazovaniya-/307628.html?subscribenam=content&category=-mbpch-prizyvaet-chinovnikov-sfery-obrazovaniya-&date=270510>

<sup>19</sup> Русский язык. 2.2.1. Характеристика контрольных измерительных материалов по русскому языку 2010 г. Отчёт ФИПИ. 2010.

Русский язык 2010 г.

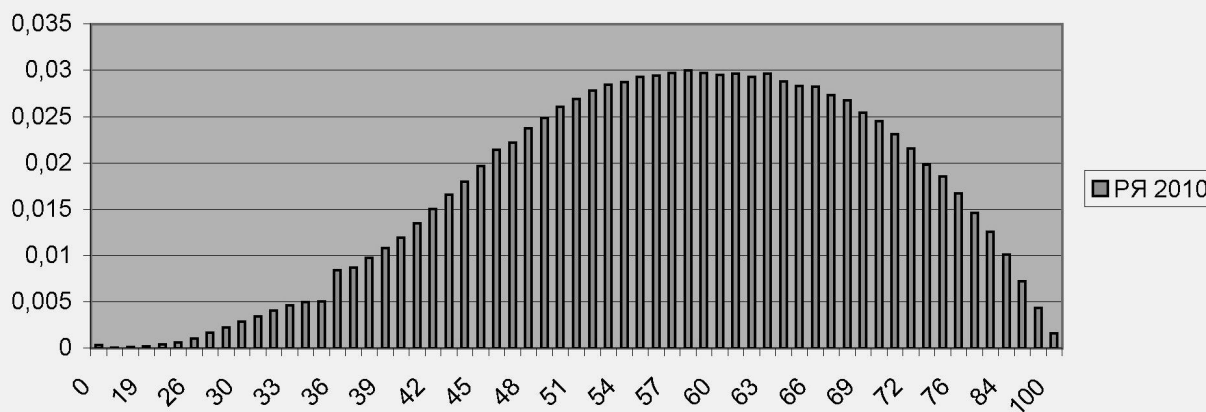


Рис. 2. Распределение участников экзамена по полученным тестовым баллам в 2010 г.

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.

Справа, на рис. 2, распределение сжато так, что коррупционные 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.

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 Рособнадзором 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)<sup>20</sup>). 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.

<sup>20</sup> 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](http://www.fipi.ru/binaries/1085/1_razdel_11.pdf)

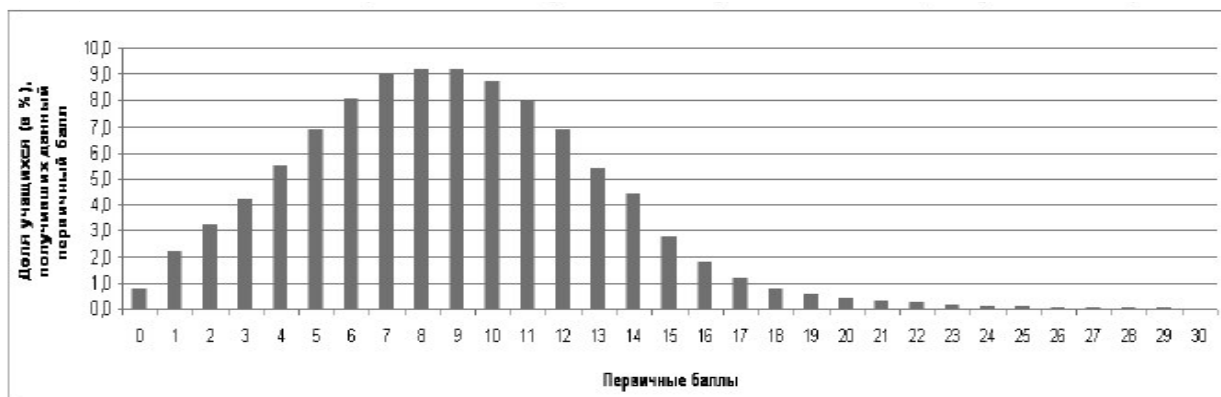


Рис. 3. Распределение исходных результатов по ЕГЭ (математика – 2010)

В классической (статистической) теории педагогических измерений уже давно сложилась традиция считать результаты тестов с асимметричным распределением баллов как относительно невалидные по двум 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 факт<sup>21</sup>, представленный в таблице.

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 недопустимо 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?

<sup>21</sup> Источник: Государственное учреждение «Областной центр мониторинга качества образования». Единый государственный экзамен 2007. Физика. Сборник аналитических материалов. Кемерово 2007.

Таблица

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

#### Задания группы «А, В»

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

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

### Спорное решение Государственной думы

Развитию педагогических измерений в России мешают контрольные 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 в сторону разрушения образования в стране.

Госдума приняла законопроект, в котором написано, что контрольные измерительные материалы (КИМы) ЕГЭ признаются конфиденциальной информацией<sup>22</sup>. Это было сделано, опять-таки, казалось бы, из благих побуждений — сохранить упомянутые материалы от несанкционированного до-

ступа 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 деньги<sup>23</sup>.

По замыслу депутатов Госдумы, придание материалам юридического статуса конфиденциальности позволит удержать ЕГЭ от дальнейшего интенсивного саморазруше-

<sup>22</sup> 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](http://www.proforientator.ru/index.php?option=com_content&view=article&id=462:2010-09-10-20-29-49&catid=2:5)

<sup>23</sup> Римский В.Л. ЕГЭ способствует росту коррупции в системе образования // Педагогические измерения. 2011. № 2.



ния. 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 системы<sup>24</sup>.

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, отнесённую к государственной тайне, и на конфиденциальную информацию. Отсюда можно вывести, что КИМы ЕГЭ законом «Об информации...» не могут быть отнесены ни к государственному 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 также неизвестно.

И уже совсем не может быть юридического термина «контрольные измерительные материалы», которым сегодня оперирует послушное правительству большинство Госдумы. Инициаторам конфиденциальности и каждому голосовавшему за предложение правительства полезно задать один-един-

ственный вопрос: где они видели измерительные 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 смысл<sup>25</sup>, относительно которых режим конфиденциальности и правового режима если и возможен, то в гротескном или абсурдном виде.

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 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, имеющих в базе, то это может означать только одно — такой испытуемый имеет высокий уровень подготовленности. И ему нечего бояться и ловчить. Он готов к тестированию. Легко видеть, что открытость заданий в таком вари-

<sup>24</sup> **Аванесов В.С.** Основы научной организации педагогического контроля в высшей школе: Учеб. пос. для слушателей Учебного центра. М.: МИСиС, 1987. 107с.

<sup>25</sup> **Аванесов В.С.** Мифология ЕГЭ // Первое сентября. 2001. № 50. <http://ps.1september.ru/2001/50/2-1.htm>

анте 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 самообразования. 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, за рубежом 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 время. И потом ещё надо будет вписать правильно полученные извне ответы в специальный бланк. Гораздо эффективнее в таких случаях тактика использования отведённого времени на быстрое решение тех заданий, которые испытуемый может решить самостоятельно.

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.

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.

### 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 веков<sup>26</sup>. 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 корни<sup>27</sup>.

Позже, во второй половине XX века, возникла культура педагогических измерений, включающая в себя, как тогда писали, «теорию тестов», культуру формулирования заданий<sup>28</sup>, 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

<sup>26</sup> Аванесов В.С. Возникновение и развитие тестового контроля. Гл. 2 дисс. на соиск уч. степени докт. пед. наук «Методологические и теоретические основы тестового педагогического контроля». СПб.: СПбГУ, 1994. 339 с.

<sup>27</sup> Аванесов В.С. Тесты: история и теория. Из глубины веков // Управление школой. 1999. № 12, март. С. 14.

<sup>28</sup> 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.

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 возраста<sup>29</sup> и 2) информацию о степени совпадения результатов теста с мнением преподавателей. А. Бине и Т. Симону принадлежит также важная идея расположения заданий теста в зависимости от возрастания меры их трудности<sup>30</sup>. 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, Дальтон-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 запрещены<sup>31</sup>. 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 последую-

щее 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 нормированием, 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

<sup>29</sup> 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.

<sup>30</sup> «...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.

<sup>31</sup> **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.: Речь, 2006. 440 с. Общая психодиагностика. <http://medbookaide.ru/books/fold1002/book2002/p1.php>

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 перевода, 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. Adkins<sup>32</sup>, N.E. Glonlund<sup>33</sup>, C.M. Lindvall & A.J. Nitko<sup>34</sup>, W.A. Mehrens & Lehmann<sup>35</sup>, A.J. Nitko<sup>36</sup>, G. Sax<sup>37</sup>, A.G. Wesman<sup>38</sup> и мн. др.

В девяностые годы XX века в педагогических измерениях США наметился явный успех, связанный с появлением работ R.E. Bennet & W. Ward<sup>39</sup>, S.B. Carlson<sup>40</sup>, N. Frederiksen<sup>41</sup>, E.E. Freeman<sup>42</sup>, R.J. Mislevy<sup>43</sup>, S.J. Osterlind<sup>44</sup>, W. Wiersma & S.G. Jurs<sup>45</sup> и мн. др. Ближе к созданию теории педагогических измерений подошёл в США R. Ebel<sup>46</sup>.

В СССР и России проблемой развития педагогических 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 работы по ключевым вопросам развития педагогических измерений. Почти все они представлены на сайте автора<sup>47</sup>. Ранее этого периода были опубликованы труды по психологическим<sup>48</sup> и социологическим<sup>49</sup> измерениям.

В качестве основы теории педагогических измерений была разработана система понятий педагогических измерений, принципы разработки формы<sup>50</sup> и содержания тестовых заданий<sup>51</sup>, концепция пяти этапов тестового педагогического процесса. В этой концепции бюрократическому режиму проведения контроля посредством контрольных материалов места нет.

<sup>32</sup> Adkins Dorothy C. Test construction: Development and Interpretation of Achievement Tests. 2 ed. Columbus, Ohio, Ch.E. Merrill Publ. Co..., 1974. 164 pp.

<sup>33</sup> Gronlund N.E. Measurement and Evaluation in Teaching. The MacMillan Co. N-Y., 1965.

<sup>34</sup> Lindvall C.M., Nitko A.J. Measuring pupil achievement and aptitude. 2nd ed. Harcourt Brace Jovanovich, Inc., 1975.

<sup>35</sup> Mehrens W.A., Lehmann I.J. Measurement and Evaluation in Education and Psychology. N-Y., Halt, Rinehart and Winstin Inc., 1978.

<sup>36</sup> Nitko A.J. Educational Tests and Measurement: An Introduction. Harcourt Brace Jovanovich, Inc., N-Y., 1983. -674pp.

<sup>37</sup> Sax G. Principles of Educational Measurement and Evaluation. Belmont, Ca: Wardsworth Publ. Co., Inc., 1980.

<sup>38</sup> Wesman A.G. Writing the Test Item // In: Thorndike R.L. (Ed.) Educational Measurement . 2nd ed. Am. Council on Education.1971. 768 pp.

<sup>39</sup> Bennet R.E., Ward W. (Eds.) Construction vs. Choice in Cognitive Measurement. Hillsdale, N-J. 1993.

<sup>40</sup> Carlson S.B. Creative classroom testing: Ten designs for assessment and instruction. Prinston, N-J: ETS., 1985.

<sup>41</sup> Frederiksen N. et. al. (Eds). Diagnostic Monitoring of Skill and Knowledge Aquisition... 510 pp.; Frederiksen N. et.al. (Eds). Test Theory for a new generation of tests. Lawrence Erlbaum Ass., Hillsday, N-J, 1993. 404pp.

<sup>42</sup> Freeman E.E. The Redesign of Testing for the 21st century. Prinston, N-J., 1986. 101 pp.

<sup>43</sup> 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.

<sup>44</sup> Osterlind S.J. Constructing Test Item. Boston etc.: Kluwer,1989. 343 pp.

<sup>45</sup> Wiersma W., Jurs S.G. Educational Measurement and Testing. 2-nd. ed. Allyn and Bacon, Boston, 1990. 415 pp.

<sup>46</sup> Ebel R.L. Measuring Educational Achievement. Prentice-Hall, Inc. Englewood Cliffs, New Jersey, 1965. 481 pp.

<sup>47</sup> <http://testolog.narod.ru>

<sup>48</sup> Аванесов В.С. Проблема психологических тестов // Вопросы психологии. 1978.

<sup>49</sup> Аванесов В.С. Тесты в социологическом исследовании. М.: Наука, 1982. 199 с.

<sup>50</sup> Аванесов В.С. Форма тестовых заданий: Учеб. пос. М.: Центр тестирования, 2005. 152с.

<sup>51</sup> Аванесов В.С. Композиция тестовых заданий. 1, 2 и 3 изд.; 3 изд. М.: Центр тестирования, 2002. 217с.

### Уровневое функционирование 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 измерений<sup>52</sup>. 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 негодных 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 научного после себя это 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 центрах 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 требующих 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 методологии педагогических измерений. Он характерен для использования в развивающихся педагогических учреждениях и для улучшающегося образования в целом. И неизбежен — 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.

<sup>52</sup> 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 measures». In: Measurement for Social Science and Education. A history of social science measurement. <http://www.rasch.org/memo62.htm>, а также: All measures are numbers. But not all numbers are measures. In: Wright B.D. (1997). Fundamental measurement for outcome evaluation. Physical medicine and rehabilitation: State of the Art Reviews. 11(2): 261–288. [www.rasch.org/memo66.htm](http://www.rasch.org/memo66.htm)

Предмет третьего уровня — разработка теории педагогических измерений. Критикой теории 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 необходима 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.

### Определение 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 результаты, 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 ненаучных<sup>53</sup>.

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. Стивенсом<sup>54</sup>.

Здесь оно сопровождается примером создания шкалы отношения граждан к иммигрантам<sup>55</sup>. Согласие респондента с содержанием каждого следующего вопроса указывает на более высокую меру его толерантного (невраждебного) отношения к иммигрантам. Враждебное отношение и готовность использовать насилие по отношению к «инородцам» измеряются шкалой национализма.

*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» означает, что каждое задание теста имеет свои параллельные варианты.

В педагогических измерениях в обязательном порядке используются математико-статистические методы шкалирования получаемых результатов испытуемых и свойств заданий. В итоге получается шкала, имею-

<sup>53</sup> Аванесов В.С. Проблема демаркации педагогических измерений // Педагогические измерения. 2009. № 3. С. 3–37. <http://obrazovanie.viperson.ru/wind.php?ID=425098>

<sup>54</sup> Scaling is the assignment of objects to numbers according to a rule.

<sup>55</sup> Scaling. <http://www.socialresearchmethods.net/kb/scalgen.php>

щая нулевое начало, единицу измерения, достаточную протяжённость. Это — свойства шкалы.

*Модель измерения* определяется как структурное построение, позволяющее соединить латентную переменную величину с наблюдаемыми значениями этой величины<sup>56</sup>.

*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 применяется трансформационная шкала исходных баллов испытуемых — так называемая Т-шкала.

*Шкалирование* определяется как та часть процесса измерения, в котором исходные значения 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 автора<sup>57</sup>.

## 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 деятельности<sup>58</sup>.

Одна из важных задач методологии — правильное определение целей развития педагогических измерений. Недостаточное внимание к вопросам методологического анализа целей ЕГЭ и контрольных материалов привело к печальным результатам<sup>59</sup>.

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 теории<sup>60</sup>. 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 от-

<sup>56</sup> Bollen K.A. Structural Equations with Latent Variables. N-Y, Wiley & Sons, 1989. 514 pp.

<sup>57</sup> <http://testolog.narod.ru>

<sup>58</sup> Вопросы методологии педагогических измерений // Педагогические измерения. 2005. № 1. <http://testolog.narod.ru>.

<sup>59</sup> Аванесов В.С. Ошибочные цели — плачевные результаты // Педагогические измерения. 2010. № 4. Вторая, расширенная версия этой статьи представлена по адресу: <http://obrazovanie.viperson.ru/wind.php?ID=425098>

<sup>60</sup> 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.

несённость. А это означает, что эти теории в значительной мере являются общими, или общенаучными, не нагруженными содержанием собственно педагогики<sup>61</sup>.

Вопрос о необходимости создания собственно педагогической теории измерений был поставлен в первом номере российско-го научно-методического журнала «Педагогические измерения»<sup>62</sup>. 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, критериально ориентированной или нормативно-ориентированной.

### 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 методов<sup>63</sup>. Основные критерии, сформулированные в том исследовании — это организация в стране процесса педагогических измерений; развитие культуры научного обоснования тестов; обязательность предварительной эмпирической апробации всех заданий, включаемых в методы измерения, открытость результатов измерений для научного анализа и критики, организация 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.<sup>64</sup>.

*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, 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 понимание 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 содержания 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.

<sup>61</sup> Аванесов В.С. Три статьи. Item Response Theory: Основные понятия и положения. Статья первая // Педагогические измерения. 2007. № 2; 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. 2009. № 4.

<sup>62</sup> Аванесов В.С. Основы педагогической теории измерений // Педагогические измерения. 2004. № 1. С. 15–21.

<sup>63</sup> Аванесов В.С. Проблема демаркации педагогических измерений // Педагогические измерения. 2009. № 3. С. 3–37. <http://obrazovanie.viperson.ru/wind.php?ID=425098>

<sup>64</sup> Там же.



На Западе развитие формы протекало в рамках прикладного направления педагогики, называемого Item Writing<sup>65</sup>. Первое описание формы заданий с выбором одного правильного ответа можно найти в трудах G.M. Whipple<sup>66</sup>. В США форму тестовых заданий исследовали C.W. Odell<sup>67</sup>, H.F. Hawkes, E.F. Lindquist & C.R. Mann<sup>68</sup>, W.S. Monroe, J.C. DeVoss & F.J. Kelly<sup>69</sup>, H.D. Risland<sup>70</sup>, G.M. Ruch<sup>71</sup>, M.W. Richardson, J.T. Russel, J.M. Stalnaker & L.L. Thurstone<sup>72</sup>, J. Follman, B. Hall, R.Wiley. & J.Hartmanh<sup>73</sup>. 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. Костюк<sup>74</sup>. «Форма теста, — писал он в 1928 году, является одной из тех проблем тестирования, научное изучение и разрешение которой чрезвычайно важно и неотложно»<sup>75</sup>. 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. Войтов<sup>76</sup>.

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 измерений<sup>77</sup>.

*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, 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 правильных выражений «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 журнала<sup>78</sup>.

<sup>65</sup> 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.

<sup>66</sup> Whipple G.M. Manual of Mental and Physical Tests. Warwick & York, Baltimore, Md., 1910.

<sup>67</sup> Odell C.W. Traditional examinations and newtype tests. N-Y., The Century Co., 1928.

<sup>68</sup> Hawkes H.F., Lindquist E.F., Mann C.R. The construction and use of achievement examinations. Boston: Houghton Mifflin, 1936.

<sup>69</sup> Monroe W.S., DeVoss J.C., Kelly F.J. Educational tests and measurements. Boston: Houghton Mifflin, 1924.

<sup>70</sup> Risland H.D. Constructing tests and grading in elementary and high school subjects. N-Y, Prentice-Hall, 1937.

<sup>71</sup> Ruch G.M. The objective or new-type examination. Glenville, Ill: Scott Foresman, 1924.

<sup>72</sup> Richardson M.W., Russel, J.T., Stalnaker J.M., Thurstone L.L. Manual of examination methods. University of Chicago, Board of Examinations, 1933.

<sup>73</sup> Follman J., Hall B., Wiley R., Hartmanh J. Relationship Between Objective Test Formats. Educ. Rev., 1947: 26: 150–151.

<sup>74</sup> Костюк Г.С. О зависимости результатов тестирования от формы теста // Тесты: теория и практика. М.: Моск. тестологическое объединение, 1928. 205с.

<sup>75</sup> Костюк Г.С. Избранные психологические труды. М. 1988. с. 73.

<sup>76</sup> Войтов А.Г. Форма и формы тестовых заданий // Педагогические измерения. 2005. № 2. С. 25–34.

<sup>77</sup> Композиция тестовых заданий. Издания 1998, 2002 гг.: Форма тестовых заданий. М.: Центр тестирования 2005 г. 155 с.

<sup>78</sup> Аванесов В.С. Проблема качества педагогических измерений // Педагогические измерения. 2004. № 2. С. 3–31.

Помимо этих двух традиционных критериев качества результатов, в последние годы большое внимание привлёк к себе критерий объективности тестовых результатов<sup>79</sup>. 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 достичь объективированных результатов 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. Продвинутые авторы, имеющие подходящее математическое и статистическое образование, тяготеют к многомерным моделям измерения, которые чаще оказываются интереснее одномерных моделей.

<sup>79</sup> **Аванесов В.С.** Тестирование как основа формирования честной и объективной оценки учебных достижений // Матер. 2-ой ежегодной конф. «Новые подходы в преподавании бизнеса и экономики. С. 241–245. USAID, Amer. Councils for International Education. Бишкек, 29–30 апреля 2002 г.

<sup>80</sup> **Аванесов В.С.** Проблема эффективности педагогических измерений // Педагогические измерения. 2008. № 4. (Переработано).

<sup>81</sup> **Аванесов В.С.** Методологические и теоретические основы тестового педагогического контроля. Дис. ... докт. пед. наук. СПб., 1994. 339 с.

<sup>82</sup> **Аванесов В.С.** Проблема эффективности педагогических измерений. Педагогические измерения. 2008. № 4. (Переработано).

<sup>83</sup> **Аванесов В.С.** Язык педагогических измерений // Педагогические измерения. 2009. № 2. С. 29–60.

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 RM<sup>80</sup>. 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 целом<sup>81</sup>.

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 тестов<sup>82</sup>, 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 автора<sup>83</sup>. □