

Почему КИМы ЕГЭ не являются педагогическими измерениями

*Вадим Сергеевич Аванесов,
профессор, доктор педагогических наук
testolog@mail.ru*

ПРОВЕДЕНИЕ ЕДИНОГО ГОСУДАРСТВЕННОГО ЭКЗАМЕНА (ЕГЭ) АКТУАЛИЗИРОВАЛО ПРОБЛЕМУ ДЕМАРКАЦИИ (РАЗГРАНИЧЕНИЯ) НАУЧНЫХ МЕТОДОВ ПЕДАГОГИЧЕСКОГО ИЗМЕРЕНИЯ ОТ МЕТОДОВ ПСЕВДОНАУЧНЫХ И НЕНАУЧНЫХ. В СФЕРЕ ПЕДАГОГИЧЕСКИХ ИЗМЕРЕНИЙ БЕССПОРНЫМ ПО КАЧЕСТВУ ИЗМЕРЕНИЯ И ОБЩЕПРИНЯТЫМ МЕТОДОМ В МИРЕ ЯВЛЯЕТСЯ ТЕСТ. ПО СРАВНЕНИЮ С НАУЧНО ОБОСНОВАННЫМИ ТЕСТАМИ, ВСЕ ДРУГИЕ МЕТОДЫ ОЦЕНКИ ЗНАНИЙ ОКАЗЫВАЮТСЯ БОЛЕЕ СУБЪЕКТИВНЫМИ, ПОДВЕРЖЕННЫМИ КОРРУПЦИИ, МЕНЕЕ ТОЧНЫМИ И ЭФФЕКТИВНЫМИ. НАУЧНОЙ ОСНОВОЙ ТЕСТИРОВАНИЯ ЯВЛЯЮТСЯ ПЕДАГОГИЧЕСКИЕ ИЗМЕРЕНИЯ.

Что такое демаркация?

В России сейчас применяются следующие основные формы и методы проверки уровня подготовленности: КИМы ЕГЭ, олимпиады, традиционные экзамены, текущее и итоговое оценивание. Эти формы, не являясь, в научном смысле, методами педагогических измерений, так или иначе, противопоставлены тестированию, которое при должном уровне работы, напротив, обладает свойствами педагогических измерений. Научной формой противостояния упомянутым формам псевдоизмерения может стать демаркация педагогических измерений от этих форм.

Демаркацией называется процесс разграничения научных теорий и методов от псевдонаучных и не научных. Необходимость в демаркации возникает там, где научные теории, понятия и методы не различаются или намеренно подменяются.

Псевдоизмерения можно определить как процесс производства и применения методов, позволяющих создать лишь внешнюю видимость педагогических измерений, которая чаще всего проявляется в ложных

названиях. Ярким примером псевдоизмерений являются КИМы ЕГЭ.

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

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

Успешное решение проблемы демаркации возможно при росте интереса к разработке языка педагогических измерений¹ и уточнения системы понятий, позволяющих отделить педагогические измерения от всех остальных форм педагогического оценивания, особенно субъективных и спорных.

Сущность педагогических измерений

Демаркация педагогических измерений предполагает чёткое выделение главного объекта от иных объектов. Педагогическое измерение представляет собой *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 и ненаучных.

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

Цель педагогических измерений — определить количество интересующего латентного свойства личности (меру интересующего признака), присущего данному испытуемому, определить значение интересующей латентной переменной у каждого испытуемого.

Уровни педагогических измерений — это наука и практика. Предмет науки — разработка теории педагогических измерений, предмет практики — разработка и применение тестов. Это и есть определение те-

стирования, которое нередко отличается от педагогических измерений пренебрежением к науке (теории). Между тем тестирование в отрыве от науки (педагогических измерений) приносит только вред. Критикой теории и конструктивным преобразованием практики призвана заниматься *методология педагогических измерений*.

Система педагогического тестирования — организация сбора и обработки данных посредством применения тестов на множестве испытуемых. Она может оказаться научно обоснованной, не обоснованной и даже антинаучной. Научной основой системы тестирования является теория педагогических измерений. Предметом практики педагогических измерений являются создание теста, проведение тестирования, шкалирование тестовых результатов, проведение рейтинга и мониторинга (в части разработки показателей текущих и итоговых учебных достижений). И, кроме того, построение системы качественных показателей образовательной деятельности².

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

Качественный педагогический тест образует система тестовых заданий равномерно возрастающей трудности, позволяющая получать объективные, надёжные и валидные результаты измерения уровня и структуры подготовленности испытуемых.

Единицей педагогического измерения является логит уровня подготовленности испытуемого и логит уровня трудности задания. Исходное значение логита подготовленности испытуемого вычисляется как натуральный логарифм отношения долей правильных ответов испытуемого к доле неправильных ответов. Далее в процес-

¹ Аванесов В.С. Язык педагогических измерений. Педагогические измерения. № 2. 2009 г.

² Аванесов В.С. Основы теории педагогических измерений // Педагогические измерения. № 1. 2004 г. С. 15-21.

се педагогического измерения логиты шкалируются. Среднее арифметическое шкалы принимается равным нулю, стандартное отклонение — равным одному логиту.

Идея возрастающей трудности тестовых заданий была впервые выражена в работе А.Бине и Т.Симона³ в начале XX века. Затем она проходила во многих, если не во всех работах автора данной статьи по тестовой проблематике, начиная с 70-х годов XX века. Наиболее чёткое, аргументированное и системное выражение этой же идеи мы находим в классической работе Best Test Design⁴. Основу любого теста составляют содержательная педагогическая теория (концепция) подготовленности испытуемых по учебной дисциплине и математико-статистические характеристики как каждого задания в отдельности, так и теста в целом. Содержательная теория позволяет обосновать качество измерения на языке учебной дисциплины, математико-статистические характеристики — качество каждого задания и теста в целом формальными и вычислительными методами.

Только тест даёт возможность получить результаты, соответствующие международным стандартам качества педагогической информации⁵, эмпирически проверяемые по критериям надёжности, валидности, объективности и эффективности. Ни один другой метод в этом отношении сравниться с тестом не может. Столь строгие требования к обоснованию качества методов педагогических измерений позволяют обезопасить личность, общество и государство от некачественных оценок, повысить ответственность разработчиков тестов за результаты своей деятельности.

Необходимость разграничения педагогических измерений от экзаменов

Экзамен — это традиционная нетехнологичная форма педагогического контроля уровня подготовленности, в процессе которой делаются попытки оценить у испытуе-

мого, в основном, знания, умения и навыки. Экзамены применяются в России, в программах международного исследования PISA, на Тайване и в других странах Юго-Восточной Азии. Они обычно содержат вопросы и задачи. Задания в тестовой форме там иногда бывают, но тестов нет.

Экзамены имеют две основные формы — устную и письменную. Преимущества устного экзамена — возможность оценить речевые способности учащихся, провести углублённую проверку знаний по избранным темам, в небольших учебных группах. Но для массового процесса измерения уровня подготовленности учащихся и студентов устные экзамены не годятся. Результаты на них заметно зависят от речевых способностей испытуемых и от личности оценивающего педагога. Письменные экзамены нетехнологичны, имеют повышенную трудоёмкость, оценки по ним не менее субъективны, хотя выглядят более аргументированными.

Недостатком экзамена является также несравнимость оценок, полученных учащимися разных школ по одной и той же дисциплине, и отсюда — принципиальная невозможность корректно установить — где процесс обучения поставлен лучше? Ответ на этот вопрос можно дать только в случае использования стандартной тестовой программы с применением технических средств контроля. В наше время эффективное измерение подготовленности большого числа испытуемых посредством экзаменов экономически неэффективно.

Результаты олимпиад не являются педагогическими измерениями

Олимпиада — это форма развития творческого мышления молодёжи и форма проверки уровня подготовленности в одинаковых для всех участников условиях. Внешне различие между тестированием и олимпиадой несколько напоминает отличие между быстрой и обычной игрой в шахматы. В России сложилась традиция трёх-четырёхчасового проведения экзаменов и олимпиад, в странах Запада чаще используется примерно сорокаминутное тестирование. Если придерживаться аналогии с шахмата-

³ Binet A., Simon T.H. The Development of Intelligence in Young Children. Vineland, N-J: The Training School, 1916.

⁴ Wight B.D., Stone M. H. Best Test Design. MESA PRESS. Chicago. 1979.

⁵ Аванесов В.С. Вопросы методологии педагогических измерений // Педагогические измерения. № 1. 2005 г. С. 3–27.

ми, то на результат олимпиад и тестирования влияет не только фактор мастерства, но и фактор лабильности мышления, специальной тренировки по формированию алгоритмов быстрого выполнения заданий.

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

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

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

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

В России сотни лет существует традиция неспешного выполнения заданий и задач, а потом проверки и перепроверки своих ответов. В этом смысле проведение олимпиад больше соответствует российским традициям, а проведение быстрого тестирования — западным и восточным традициям. Поскольку процент детей, не готовых быстро отвечать на задания теста, представляется автору достаточно высоким, таким испытуемым лучше подходят олимпиады. Отсюда и общественные приоритеты в пользу олимпиад и других традиционных форм неспешного выполнения заданий и субъективированного оценивания.

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

Оценка участникам олимпиады выставляется на основе числа решённых задач и качества проведённого обоснования решений. Каждая задача обычно ассоциируется с неким числом баллов, даваемых за её правильное решение. Чем труднее задача, тем больше баллов может получить участник олимпиады. Положительная сторона — это равенство условий и документированность результатов участников. Каждый собственноручно пишет свои решения и обоснование решений. Недостаток олимпиад — нетехнологичность форм их проведения, возможности субъективных воздействий, нередкие утечки содержания олимпиадных заданий, субъективные методы оценивания результатов. Отсюда частые недовольства итогами олимпиад со стороны участников и наставников.

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

одно выделенное и хорошо определённое латентное свойство личности. Именно здесь видится главное отличие олимпиад от педагогических измерений.

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

Почему КИМы ЕГЭ невозможно считать тестами

Как следует из публикации Федерального института педагогических измерений (ФИПИ)⁶, там уже в самом начале эпопеи с КИМаи ЕГЭ возникли противоречия по вопросу соотношения КИМов ЕГЭ и тестов. Одни сотрудники и авторы, вслед за руководителями Министерства образования и Рособнадзора, называли КИМы ЕГЭ тестами. Другие проявляли осторожность, считая, что «говорить о ЕГЭ, как тесте, основываясь на результатах первого года эксперимента, конечно, рано» (с.34). Похоже, что тоже лелеяли надежду на чудесное превращение КИМов в тесты.

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

понятия оказалось слишком узкое, для научного анализа абсолютно непригодное. Этот вывод вытекает из логического закона обратного отношения между объёмом и

содержанием понятий. А потому в логическом смысле КИМы обречены оставаться неконкретным названием не определённых наукой объектов. К тому же, если на секунду допустить, что слово «материалы» является частью научного понятия «контрольно-измерительные материалы», то объём такого понятия оказался бы слишком широк для того, чтобы оно было в состоянии выразить что-то существенное или конкретное.

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

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

И действительно, все варианты к каждому заданию, по каждому предмету должны иметь общее содержание, примерное равенство средних арифметических показателей, а также показателей вариации по каждому заданию. Только тогда обеспечивается справедливость и объективность выставляемых оценок⁷. В КИМах ЕГЭ задания не параллельны по трудности, что делает их принципиально несправедливыми при оценке испытуемых.

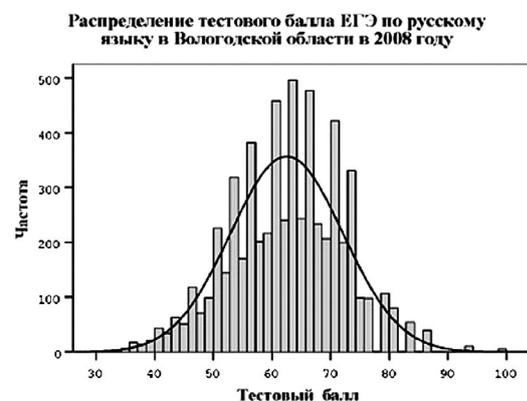
В России КИМы ЕГЭ часто называли и продолжают называть тестами. Но в таких

⁶ «Единый Государственный экзамен. Научные основы, методология и практика организации эксперимента» / Под ред. В.А. Болотова. М.: Логос, 2002. 208 с.).

⁷ **Аванесов В.С.** Проблема объективности педагогических измерений // Педагогические Измерения. № 3. 2008 г.

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

Пример такой гистограммы с очень странным распределением так называемых тестовых баллов ЕГЭ можно увидеть в Вологодском отчёте. Этот отчёт был получен благодаря содействию Общественной комиссии Совета Федерации.



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

Напротив, тест, как научный метод, интересен тем, что при его разработке обязательно нужно использовать методы обоснования качества получаемых результатов. Все остальные методы этой методологической

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

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

Отличие педагогических измерений от педагогического оценивания

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

В зарубежной педагогической литературе широко используется понятие «evaluation», что можно перевести как «оценивание», имея в виду не только конечный результат, но и процесс формирования оценки. Соответственно выделяются два основных вида оценивания — формирующее и итоговое⁸. Цель первого — оказывать, посредством оценивания, формирующее влияние на текущий процесс обучения, в смысле его улучшения, за счёт установления обратной связи от учащегося и студента к преподавателю. Цель второго — получить итоговые результаты обучения.

Английское educational evaluation можно рассматривать как одну из начальных форм численного оценивания, в которой, однако, нет результатов измерений. На Западе понятие «evaluation» нередко включает в себя все традиционные педагогические и экспертные оценки⁹. Там в это поня-

⁸ Bloom B.S. a .o. Handbook on Formative and Summative Evaluation of Student Learning. N-Y., McGraw-Hill, 1971. 923 pp.

⁹ Аванесов В.С. Вопросы методологии педагогических измерений. Педагогические Измерения. № 1. 2005. С. 3-27.

тие раньше иногда включали и измерение. Но в последние тридцать лет от этого отказались по причине необходимости дифференцировать оценивание от более наукоёмкого процесса педагогического измерения. В России делают наоборот.

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

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

Можно сказать и так: измерение — самый высокий уровень численного оценивания, где к оценкам добавляется ещё и мера. Все остальные формы оценивания, включая учительские оценки, это применение числа, но не меры. Рассматривать их как равноценные понятия — значит низводить измерение до уровня оценивания. Итак, существенным критерием демаркации педагогических измерений от педагогического оценивания является признак наличия меры, точнее, интервальной шкалы.

Сейчас уже можно говорить о системе оценивания, которая пока не внедрена, но уже широко обсуждается на различных совещаниях. Она называется общероссийской системой оценки качества образования (ОСОКО) и мыслится её создателями как не имеющая аналогов в мире. Фактически система ОСОКО призвана установить не только полный контроль уровня подготовленности всех учащихся и студентов, но и

тотальный контроль за работой всех образовательных учреждений России,

во всех краях, областях, и в автономных республиках, что противоречит принципам подлинного федерализма, а соответствует, скорее, принципам унитарного государства.

Критерии демаркации педагогических измерений

1. Наличие в стране процесса педагогических измерений. Среди возможных критериев отличия педагогических измерений от псевдоизмерений главным представляется наличие в стране самого процесса измерений. Без него надежды на появление качественных тестов тщетны. Отмеченный процесс включает абстрагирование интересующего свойства, выбор математической модели измерения, разработку заданий в тестовой форме, проверку и публикацию их тестовых свойств, поиск наилучших методов композиции тестовых заданий в систему заданий возрастающей трудности. При этом каждое задание должно иметь множество параллельных заданий со своими статистическими характеристиками, график зависимости вероятности правильных ответов испытуемых в зависимости от уровня их подготовленности, статистику соответствия характеристик задания требованиям модели измерения. Далее идут требования к форме и содержанию заданий, а также требования к тесту в целом¹⁰.

2. Научная обоснованность теста. Это обеспечивается обязательной публикацией концепции теста, применением содержательных и метрических теорий, на основе которых готовится тест. Повышению научной обоснованности теста способствует определённость предмета педагогического измерения. Наиболее распространённым предметом педагогических измерений является уровень подготовленности испытуемых. Предметом могут быть и другие свойства личности, и каждое требует концептуализации, операционализации, структурирования, оснащения индикаторами, эмпирическим основанием обоснованности выделенного для педагогического измерения свойства личности. Таких работ в стране практически нет. И это тоже критерий оценки состояния дел с педагогическим измерением в стране.

¹⁰ Аванесов В.С. Композиция тестовых заданий. М.: Центр тестирования, 2003. 237 с.

3. *Обязательность эмпирической апробации всех заданий, включаемых в тест.* Это требование является обязательным не только для педагогических измерений, но и для любого метода оценки личности. Именно в процессе апробации появляется возможность отделить тестовые задания от заданий в тестовой форме и элементарных вопросов. Делается это посредством проведения статистического анализа характеристик каждого задания. Эмпирическая апробация позволяет провести тщательный подбор заданий теста, соответствующих конкретной цели педагогического измерения. Подбор заданий выполняется на основе специальных компьютерных программ.

4. *Открытость результатов педагогических измерений.* Наличие качественного отчёта является существенным признаком демаркации тестов от псевдотестов. Вся система признаков педагогических измерений должна излагаться в отчёте по разработке каждого теста. Это особенно важно в отношении упомянутых ранее критериев эффективности и качества тестовых результатов — надёжности, валидности и объективности. Отчёт по разработке теста подлежит обязательной публикации. Без подробного отчёта по разработке, а также пособия по применению теста, как системы заданий равномерно возрастающей трудности, не бывает. Отсутствие требуемых отчётов является важным критерием разграничения педагогических тестов от псевдотестов и нетестов.

Этические нормы тестологической службы США, Великобритании, Австралии и многих других стран Запада и Востока запрещают применять тесты, если нет опубликованной информации об их качестве. В ещё большей мере этот запрет относится к иным методам. Для тестового метода обязательна публикация подлинных статистических данных по каждому заданию и по тесту в целом, необходимость построения графических образов каждого задания теста, обязательность проведения т.н. дистракторного анализа, построение графиков информационных функций каждого задания и теста в целом, проверка адекватности данных избранной модели педагогических измерений. Список обязательных требований, относящихся к мировым стандартам разработки и применения настоящих тестов, довольно обширен¹¹.

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

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

Качество дистрактора обычно оценивается несколькими методами. Чаще всего используется подсчёт долей (процентов) испытуемых, которые выбирают тот или иной дистрактор при ответе на задание теста. Чем выше процент испытуемых, выбравших данный неправильный ответ, тем выше его привлекательность для незнающих учащихся и студентов.

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

6. *Наличие теории или теорий, позволяющих обосновать качество методов педагогических из-*

¹¹ 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>

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

7. Технологичность теста. В тесте недопустимы нетехнологичные и субъективные оценочные формы, которые используются сейчас во второй и третьей частях КИМов ЕГЭ. В наше время для массового контроля уровня подготовленности испытуемых традиционные экзамены уже не применяются ввиду их субъективности и нетехнологичности. Используются только тестовые технологии. Яркий пример эффективного применения тестовых технологий даёт всемирно известная служба Educational Testing Service¹².

8. Наличие стандартов. Каждое задание настоящего педагогического теста имеет содержательную и формальную аргументацию, необходимость которой специально оговорена в зарубежных стандартах на разработку тестов¹³. Отсутствие подобных стандартов на русском языке является объективным препятствием для создания тестов с проверяемым качеством измерения. Полезно напомнить, что на разработку КИМов ЕГЭ стандарты не существуют. Без разработки общественно профессиональных стандартов настоящую культуру педагогических измерений поднять будет трудно.

Стандартами предусмотрена обязательная проверка результатов теста по критериям надёжности и валидности. Это идеи точности результатов испытуемых и адекватности метода измерения поставленной цели. Одной цели. Могут ли результаты теста

быть валидными относительно другой цели? Могут, но только отчасти, и в меньшей обычно мере. Не бывает теста, который одинаково пригоден для разных целей.

9. Сочетаемость тестовой технологии с другими современными образовательными технологиями. Вторая и третья формы КИМов не технологичны, а вопросы образовательных технологий, похоже, в России отодвинуты в сторону. Так образовательная политика ведёт страну к технологическому отставанию от стран Запада и Востока, где технологии уже внедрены в образовательную практику и показывают хорошие результаты.

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

10. Наличие программно-вычислительного обеспечения. Для разработки настоящих тестов используются специальные компьютерные программы типа RUMM 2020 и WINSTEPS, а также ряда других, которые позволяют трансформировать счётные данные в измерения, получать графические образы каждого задания и создать систему заданий с требуемыми свойствами теста. Если графический образ удовлетворяет требованиям настоящего теста, то задание с таким графиком включается в тест, если не удовлетворяет, то в тесте такому заданию места нет. Без таких программ качественных педагогических измерений не бывает. Таким образом, ещё один критерий демаркации педагогических измерений от псевдоизмерений — это использование специальных компьютерных программ, которые позволяют получать настоящие педагогические измерения.

11. Наличие статистических характеристик каждого задания. Компьютерные программы позволяют получить статистические характеристики каждого задания теста. А это является важным критерием разграничения тестов от псевдотестов и не тестов.

12. Время проверки. Есть и другой простой метод демаркации тестов от псевдотестовых материалов. Если время проверки уровня подготовленности превышает при-

¹² Educational Testing Service. Princeton, N-J. <http://www.ets.org/portal/site/ets/menuitem>

¹³ Am. Educ. Research Ass., The Am. Psychological Ass., National Council on Measurement in Education. Standards for Educational and Psychological Testing. 1999. V.1-3.

мерно 40 минут, то этот признак свидетельствует о потере одного из самых существенных свойств теста — кратковременности процесса контроля знаний.

14. Наличие системы решающих правил и адекватность интерпретации результатов. После проведения государственного экзамена выпускников школ в Рособрнадзоре определили допустимый процент тех, кого можно лишить аттестата зрелости. 28863 выпускникам 2009 года — а это примерно 3% — на основании некачественных оценок по КИМах ЕГЭ не выдали аттестаты зрелости. Так не делается ни в одной стране мира. Из этических соображений критерии оценки и решающие правила в массовых процессах всегда определяются не после проверки, а до неё. Кстати, по министерским критериям гений русской литературы Александр Сергеевич Пушкин, не знавший математики, живи он в 2009 году, был бы однозначно лишён ат-

тестата зрелости. Страна потеряла бы ещё одного гения.

* * *

Все перечисленные критерии и научные аргументы однозначно подкрепляют главный вывод автора — КИМы ЕГЭ не являются методом педагогических измерений. Но тогда напрашиваются два вопроса, адресованных верховной власти.

Первый: кто и почему разрешает Федеральному институту педагогических измерений Министерства образования и науки РФ производить некачественные КИМы ЕГЭ вместо подлинных методов педагогических измерений?

И второй вопрос: как долго в России будет продолжаться подмена подлинных методов педагогических измерений псевдоизмерениями? □