

ИСТОРИЯ ПЕДАГОГИЧЕСКОЙ ТЕОРИИ ИЗМЕРЕНИЙ

Вадим Аванесов
testolog@mail.ru

Рассматриваются вопросы истории и теории педагогических измерений, уточняются цели, задачи, объект и предмет этой теории, исследована проблемная ситуация. Главным тормозом развития педагогических измерений в России назван т.н. Единый государственный экзамен (ЕГЭ). Исследованы причины краха его научного обеспечения. Много, если не всё, делалось в нём вопреки педагогической теории измерений. Закономерно, что результаты такого экзамена оказались некачественными.

Введение

Когда игнорируются проблемы измерения, результаты бывают далёкими от ожидаемых.

Maranell¹

Восемь лет прошло с того времени, когда в первом номере нового российского научно-методического журнала «Педагогические измерения» была напечатана статья «Основы педагогической теории измерений²». В статье были впервые выделены три главные части (предмета) этой теории: понятийный аппарат (язык) теории, содержание и форма тестовых заданий. Никакая другая теория этими ключевыми вопросами педагогических измерений не занимается. Отсюда вытекает актуальность и перспективность построения педагогической теории измерений.

Ключевое место в статье заняло истолкование смыслов определения трёх главных понятий педагогической теории — «задание в тестовой форме», «тестовое задание» и «педагогический тест». Тест кратко определялся как система заданий возрастающей трудности. Судя по

1
Maranell G.M.
Introduction. Pp. xi-xix.
In G.M. Maranell (Ed.)
Scaling: A source book
for behavioral scientists.
Chicago: Aldine. 1974.

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

3

Аванесов В.С. Проблема качества педагогических измерений // ПИ № 2 2004 г. Аванесов В.С. Проблема становления системы педагогических измерений. ПИ № 1, 2008 г. и мн. др. публикации журнала «Педагогические Измерения».

4

Аванесов В.С. Вопросы методологии педагогических измерений // Педагогические измерения. № 1. 2005. С. 3–27. <http://testolog.narod.ru/Theory34.html>

5

Аванесов В.С. Централизованное тестирование лучше Единого государственного экзамена // «Развитие тестовых технологий в России. Тезисы докладов Всероссийской научно-методической конференции / Под ред. Л.С. Гребнева. М.: Центр тестирования Министерства образования РФ. 2002. <http://testolog.narod.ru>.

6

Аванесов В.С. Единый государственный экзамен в фокусе научного исследования // Педагогические измерения. № 1. 2006. С. 3–31. <http://testolog.narod.ru/Education39.html>.

всему, эти смыслы не были восприняты. Поэтому вслед за первой статьёй по теории педагогических измерений публиковались новые статьи³. В частности, была предложена идея перехода от системы тестирования к системе педагогических измерений. Эта идея тоже не была замечена.

Различия между тестированием и педагогическими измерениями раскрываются в приводимых далее определениях. Идея перехода от тестирования к системе педагогических измерений возникла в результате сравнительного исследования вопросов методологии педагогических измерений⁴, практики централизованного тестирования⁵ и результатов Единого государственного экзамена (ЕГЭ)⁶.

Тесты имеют свою историю, которая предшествовала истории педагогических измерений. В литературе прошлого века чаще говорилось об истории тестов⁷, заметна реже — об истории педагогических измерений. Эти две истории отличаются в той мере, в какой понятие «тестирование» отличается от понятия «педагогические измерения». Первое — это преимущественно практика применения тестов, второе — научное обеспечение и научная сторона этой практики, но более высокого научного уровня, включая шкалирование и умелую интерпретацию тестовых результатов.

7

DuBois P.H. A History of psychological testing. Allyn and Bacon, 1970.

В данной статье сделана попытка рассмотреть связь двух историй и двух теорий — тестов и педагогических измерений.

О взаимосвязи истории и теории хорошо написал Н. Чернышевский. Он сделал это на примере искусства: «История искусства, — писал он, — служит основанием теории искусства, потом теория искусства помогает более совершенной, более полной обработке истории его; лучшая обработка истории послужит дальнейшему совершенствованию теории, и так далее, до бесконечности... Без истории предмета нет теории предмета; но и без теории предмета нет даже мысли о его истории, потому что нет понятия о предмете, его значении и границах»⁸.

Эта мысль является полезной и для понимания взаимосвязи интересующих нас двух историй и двух теорий — тестов и педагогических измерений.

Начало истории тестов: краткий обзор содержания опубликованных статей

*Без истории нет теории.
Гегель*

Вопросы истории тестов исследовались в цикле статей, напечатанных ранее в газете «Управление школой», а также в трудах В.М. Кадневского и других авторов.

8

http://wiki.kspu.karelia.ru/index.php/%D0%98%D1%81%D1%82%D0%BE%D1%80%D0%B8%D1%8F_%D0%BB%D0%B8%D1%82%D0%B5%D1%80%D0%B0%D1%82%D1%83%D1%80%D1%8B.

Первая статья называлась «Из глубины веков»⁹. В ней отмечалось, что всякая попытка определить точное время возникновения тестов напоминает стремление географов найти точное начало большой реки, вытекающей множеством ручейков из обширного болота. Примерно так же обстоит дело и с историей тестов. Для изучения истории тестов вначале потребовалось изучить предысторию — то самое болото, из которого и вытекает их подлинная история. Начало тестоподобных проверок учебных достижений прослеживалось от середины III тысячелетия до н.э.

Во второй статье¹⁰ родоначальником тестовых методов был назван Ф. Гальтон. Эта точка зрения была недавно поддержана и развита в статье В.М. Кадневского, В. Леммиш и Т. Ширшовой¹¹. В ней Ф. Гальтон был признан одним из основателей также и теории педагогических измерений. Он был тестологом, практиком и теоретиком, одновременно привнёсшим идеи измерения, статистики и математики в зарождавшуюся тогда науку об измерениях интеллектуальных и, надо особо отметить, двигательных способностей личности.

В третьей статье писалось о российском опыте преодоления недостатка выставляемых оценочных баллов, а также о создании первого теста во Франции. В СССР следствием постановления ВКП(б) «О порядке окончания в 1932/33 учебном году в начальной и средних школах» стала процентомания, от которой российское образование не может

оправиться и в наши дни. В постановлении отмечалось, что «проводимые испытания являются формой государственного и общественного контроля не только за работой учащихся, но и за качеством работы учителя и школы в целом».

Это утверждение содержало две ошибки.

Первая — отождествление государственного и общественного контроля. Госконтроль обычно вырождается в бюрократическое мероприятие, как это сейчас произошло с ЕГЭ. Общественный же контроль опирается на широкое участие родителей, педагогов и профессионалов-тестологов. Такого контроля в России не было никогда.

Вторая ошибка заключается в установлении жёсткой связи результатов контроля с качеством работы учителей и школы. Но не всё, особенно в наше время, зависит от школы и от учителя. Многое зависит также от условий обучения и места жительства, в которых работает школа, от качества работы дошкольных учреждений, от уровня и качества обеспечения учебной техникой, много другого, в том числе, от наличия цифровых образовательных ресурсов. Последних заметно не хватает, особенно на языках народов РФ.

Если условия образовательной деятельности неравные, то и экзаменационные требования к учащимся не могут быть едиными; требования должны быть дифференцированными в зависимости от содержания реально изученных курсов. Единого содержания уже почти нет, страна постепенно, отчасти вынужденно и стихийно

9

Аванесов В.С. Из глубины веков. Управление школой № 12, март, 1999 г. <http://testolog.narod.ru>

10

Аванесов В.С. Тесты: история и теория. Статья 2. Конец XIX века. Опубликовано в газете «Управление школой» № 14, апрель 1999 г. <http://testolog.narod.ru>

11

Кадневский, В., Леммиш В., Ширшова Т. Фрэнсис Гальтон: учёный-энциклопедист, один из первых создателей теории педагогических измерений // Педагогические измерения. № 1. 2012. С. 3–16.

12

Доклад Всемирного банка. № 18666-RU «Обновление образования в России (Региональный уровень)».

13

Тесты: история и теория. Статья 4: Тенденции XX века. Опубликовано в газете «Управление школой», № 18, май 1999 г. <http://testolog.narod.ru>

14

Аванесов В.С. Тесты в социологических исследованиях. М.: Наука, 1982. Digitized, Jul. 28, 2011, University of California. Length, 196 pp.

15

Аванесов В.С. Тесты: история и теория. Опубликовано в газете «Управление школой», № 18, май 1999 г. <http://testolog.narod.ru>

Статья 5. Середина XX века. Опубликовано в газете «Управление школой», № 18, май 1999 г. <http://testolog.narod.ru>

переходит на уровневое образование. Органы управления образованием и научно-педагогические учреждения на эту тенденцию реагируют с отставанием.

С появлением ЕГЭ упомянутая идея постановления ВКП(б) 1923 года о контроле за работой учителя вновь начала ассоциироваться с качеством работы преподавателей. Не случайно появился документ Всемирного банка, призванный охладить пыл российских чиновников и разъяснить им смысл системы государственного контроля качества образования. «Задачей системы оценки качества образования (системы управления качеством), — написано там, — является не сбор информации в целях тотального контроля сверху, а мобилизация ресурсов школ для повышения качества. Поэтому создаваемые системы оценки должны... ориентироваться на самоконтроль школ и учителей как на важнейший ресурс повышения качества»¹².

Всякий метод государственного контроля кажется поначалу очень удобным власти, потому что он привычно работает исключительно на её стороне. Но при некачественном контроле такой метод имеет тенденцию оборачиваться своим эффектом в противоположную сторону. Особенно при отсутствии общественно-профессионального контроля. Это как раз и случилось с ЕГЭ. Как показывают результаты социологических опросов, ЕГЭ неуклонно теряет доверие со стороны основных групп населения страны.

В четвёртой статье¹³ отмечалось, что практика тестирования

в СССР характеризовалась серьёзными противоречиями; по мере роста числа тестов и тестовых исследований имели место попытки торможения тестовой работы и даже запрета на распространение тестов. В 1936 г. было принято Постановление Совета Народных Комиссаров под многозначительным извращением в системе Наркомпросов». Этот документ на долгое время затормозил применение тестов в СССР, мешал разработке и применению тестовых методов в сфере образования, профотбора и профориентации.

В США под теорией тестов понимались научные основы разработки тестов и тестовых заданий, в соответствии с двумя основными критериями качества тестовых результатов — надёжности и валидности. Эта теория начала развиваться более ста лет назад вместе с психологической теорией тестов. Тогда это была одна общая теория — психологических и педагогических тестов. За рубежом тесты применялись и в социологических исследованиях¹⁴.

В пятой статье говорилось о Постановлении Совета Народных Комиссаров и ЦК ВКП(б) от 3 сентября 1935 г.¹⁵, где утверждалось, что установленная Наркомпросами система оценки успеваемости не даёт представления о фактических знаниях учащихся и ведёт на практике к понижению уровня учёбы. Там же поручалось отделу ЦК ВКП(б) разработать, обязательно для всех школ СССР, нормы оценки успеваемости учащихся, с тем, чтобы один и тот же уровень знаний одинаково оцени-

вался во всех школах. Это партийное задание осталось невыполненным по сей день. Российский ЕГЭ наших дней эту проблему не только не решил, но и загнал её вглубь.

Для сравнения, в той же статье приводились результаты применения тестов в профессиональном отборе авиационных специалистов США во время Второй мировой войны. Приводилась такая статистика: в Первую мировую войну только 2% потерь авиации были связаны непосредственно с боевыми операциями; 8% было потеряно из-за дефектов материальной части, а больше всего — 90% потерь — было вызвано профессиональной непригодностью тех, кто пилотировал самолёты¹⁶. Эти факты убедили многих в необходимости проведения профессионального отбора кадров посредством применения тестов. Один из самых активных участников этого процесса J.P. Guilford посчитал, что на каждый доллар, затраченный на эту программу, американский бюджет сэкономят тысячу долларов¹⁷.

В шестой статье отмечалось, что запрещение педологии, а вместе с ней и тестов, стало одним из первых фактов идеологического давления на неугодные науки, научные направления и на отдельных учёных. В числе пострадавших от такого давления оказались авторы выпускавшегося в конце 20-х — начале 30-х годов сборника «Тесты: теория и практика». Это М.Я. Басов, М.С. Бернштейн, А.П. Болтунов, С.М. Василейский и другие. Как прозорливо отмечал в те годы П.П. Блонский, тесты — это больше, чем средство контроля; это средство рационализа-

ции школьного дела. Это мнение тоже не было услышано.

В седьмой статье¹⁸ тестирование в СССР периода пятидесятих-семидесятих годов двадцатого столетия было охарактеризовано одним словом — застой. Что связано с общей тенденцией деградации СССР. Причинами этой тенденции стали авторитаризм, национализм, ошибочная политика подбора управленческих кадров, в которой проявлял себя некий принцип, похожий на закон убывания плодородия почвы. В стране культивировался отрицательный подбор кадров, в соответствии с которым последующее поколение управленцев оказывалось хуже предыдущего. Похоже, что с этим сейчас дела обстоят не лучше.

В последние годы в России делались попытки называть тестированием процесс применения так называемых контрольно-измерительных материалов (КИМов) ЕГЭ¹⁹. Хотя уже было неоднократно показано, что ничего измерительного в этих материалах нет²⁰. Это лишь контрольные задания со скрываемой государственным органом мерой погрешностей результатов. Станным образом власть позволяет создателям КИМов ЕГЭ манипулировать словами и не показывать меру погрешности этих «материалов».

Педагогическое тестирование — это *практическая* деятельность по применению тестов в интересах личности и в целях улучшения качества образовательной деятельности. В данной статье педагогическое тестирование рассматривается как практическая сторона педагогических измере-

16

Методические основы профессионального отбора. Рига, 1968. 77 с.

17

Guilford J.P. Is Personnel Testing Worth the Money? Gen. Mgmt. Ser., 1950, № 176. P. 52–64.

18

Тесты: история и теория. Пятидесятые годы. «Управление школой», № 24, июнь, 1999 г. <http://testolog.narod.ru>

19

Ефремова Н.Ф. Тестовый контроль в образовании: Учебное пособие. М.: Университетская книга, Логос, 2007.

20

Аванесов В.С. Ошибочные цели — плачевные результаты // Педагогические измерения. № 4. 2010 г., С. 22–51.

21

Wright B.D. AND J.M. Linacre. Observations are Always Ordinal; Measurements, however, Must be Interval. Archives of Physical Medicine and Rehabilitation 70 (12) pp. 857–860, November 1989
<http://www.rasch.org/memo44.htm>

22

Educational & Psychological Measurement. Published by Sage Publications
<http://www.sagepublications.com>

23

Journal of Educational Measurement. Published by The National Council on Measurement in Education (NCME).

24

Smallwood M.L. A Historical Study of Examinations and Grading Systems in Early American Universities. Cambridge: Harvard Univ Press. Harvard Studies in Education v.24, 1935; Caldwell O.W., Curtis S.A. Then and Now in Education: 1845–1923. Harcourt, 1924. 400 p; McArthur D.L. Educational Assessment: A Brief History /McArthur D.L. (Ed). Alternative Approaches to the Assessment of Achievement. Kluwer Academic Publishers. Boston, 1987. 268 p.

ний. Как было отмечено, тестовые баллы — это результаты тестирования испытуемых, а не результаты педагогических измерений²¹. Поэтому их лучше называть «исходными баллами испытуемых».

Обычный порок тестирования — нарушение прав личности на объективную информацию об уровне собственной подготовленности. Но здесь дело не в тестах, а в субъективизме тех, кто применяет тесты. Нарушение прав происходит по отношению не только к учащимся, но и к педагогическим коллективам, если от имени государства используются некачественные тесты или, что ещё хуже, КИМы ЕГЭ, имеющие, как уже было неоднократно показано, высокие значения погрешностей измерений.

С появлением электронного обучения значимость этих двух теорий была осознана глубже, чем это было раньше.

Со времён Э. Торндайка в педагогическом сознании западных педагогов укоренялась мысль о том, что основная сфера применения тестовых форм это не столько педагогический контроль, сколько самоконтроль в процессе собственного учения. В России эта мысль получила, можно сказать, противоположное истолкование, что связано с иным типом государственного устройства и систем управления образованием, таким типом, где роль бюрократии преувеличена, а роль науки и учёных минимизирована.

Исторически сложилось так, что первыми (больше ста лет назад) появились труды по психологической теории измерений.

В 1936 году теоретики этого направления создали журнал высокого научного уровня под названием «Psychometrika». Сейчас он издаётся и распространяется по всему миру. Затем стало развиваться общее направление «Психологическое и педагогическое измерение²²». Язык и методы этого научного направления считались, видимо по инерции, общими, равно применимыми в психологии и педагогике.

Позже последовало издание журнала по педагогическим измерениям²³, где были сделаны попытки выделить собственный педагогический предмет исследования и создать собственные педагогические методы. Успеха, однако, эти попытки не имели из-за отсутствия общей методологии педагогических измерений.

Начало истории педагогической теории измерений

Теория педагогических измерений включает в себя методические правила проведения тестирования, и кроме того, три важных направления научной деятельности: формирование языка этой теории, теория формы и вопросы отбора содержания тестовых заданий. И это ещё не всё. Подлинное педагогическое измерение появляется после тестирования испытуемых, в процессе шкалирования их результатов, а также в процессе шкалирования заданий по их существенным свойствам.

Исторически сложилось так, что теории педагогических измерений предшествовала разработка вопросов теории оценивания²⁴.

Оценивание можно определить как экспертный процесс выставления каких-либо баллов по заранее предписанным правилам. Автор этой статьи полагает научный статус оценивания ниже статуса педагогического измерения.

История возникновения общей теории измерений в общественных науках обстоятельно исследована Б. Райтом²⁵. Он был уверен, что создаваемая им теория пригодна для всех общественных наук, включая педагогику. Так оно и случилось, и для того времени это был правильный ход мысли. Но рано или поздно, в науках возникают тенденции расширения и разделения, выделения собственного предмета. А потому естественным образом возникла проблема отделения педагогической теории измерений от общей теории измерений в психологии, социологии и в других общественных науках.

Теория педагогических измерений, как научная сторона процесса тестирования, стала развиваться с небольшим опозданием от теории тестов. Начало теории педагогических измерений положил Эдвард Торндайк^{26, 27}. В своих трудах он затрагивал широкий круг философских, психологических и педагогических вопросов измерений, а также профессионального отбора.

Новаторство Э.Л. Торндайка проявилось в том, что он ранее других понял необходимость развития не только теории обучения, но также и двух других теорий — учения (самостоятельного изучения) и теории педагогических измерений. С появлением элек-

тронного обучения значимость этих двух теорий была осознана глубже, чем это было раньше.

Со времён Э. Торндайка в педагогическом сознании западных педагогов укоренялась мысль о том, что основная сфера применения тестовых форм — это не столько педагогический контроль, сколько самоконтроль в процессе собственного учения. В России эта мысль получила, можно сказать, противоположное истолкование, что связано с иным типом государственного устройства и систем управления образованием, таким типом, где роль бюрократии преувеличена, а роль науки и учёных минимизирована.

Исторически сложилось так, что первыми, больше ста лет назад, появились труды по психологической теории измерений. В 1936 году теоретики этого направления создали журнал высокого научного уровня, под названием «Psychometrika». Сейчас он издаётся и распространяется по всему миру. Затем стало развиваться общее направление «Психологическое и педагогическое измерение»²⁸. Язык и методы этого научного направления считались, видимо, по инерции, общими, равно применимыми в психологии и педагогике.

Позже последовало издание журнала по педагогическим измерениям²⁹, где были сделаны попытки выделить собственный педагогический предмет исследования и создать собственные педагогические методы. Успеха, однако, эти попытки не имели из-за отсутствия общей методологии педагогических измерений.

25

Wright B.D. A History of Social Science Measurement. University of Chicago, MESA Psychometric Laboratory; Wright B.D. Fundamental measurement in social science and education. / Research Memorandum No. 33a, MESA Psychometric Laboratory, University of Chicago, March 30, 1983.

26

Thorndike E.L. Theory of Mental and Social Measurements. Science Press. 1904.

27

Thorndike E.L. Educational Measurements of Fifty Years Ago / J. of Educational Psychology, 1913, 6: 551–552.

28

Educational & Psychological Measurement. Published by Sage Publications <http://www.sage-publications.com>

29

Journal of Educational Measurement. Published by The National Council on Measurement in Education (NCME).

Три теории измерений

Разработка педагогической теории измерений проходила на основе имеющихся теорий. Сейчас наиболее известны три теории измерения:

1. Статистическая или, иначе, классическая теория тестов (classical test theory, СТТ). Так она называлась в начале XX века.

Сейчас слова «классическая теория» можно рассматривать как метафору. В подтверждение такого мнения можно привести два аргумента.

Во-первых, эта теория имеет точное, а не нынешнее метафорическое название. На английском языке она называется true-score model of measurement. Поэтому её лучше называть не теорией, а моделью измерений.

Во-вторых, эта модель — перенос идей теории физических измерений на измерения в психологии и педагогике самого начала XX века. Отсюда упрощённая, мало реалистичная для наших дней аксиоматика. Её точнее называть моделью истинного и ошибочного компонентов измерения. Разумеется, эта модель уже обросла тысячами исследований подлинно теоретического характера. Поэтому название «теория» реально существует, оно приемлемо, за исключением случаев сравнения этой модели с другими моделями педагогического измерения.

2. Вторую теорию можно очень условно перевести на русский язык как математическая теория измерений (Item Response Theory, IRT)³⁰.

3. Третья теория — больше чем теория. Это одновременно ещё методология и метрическая система. Потому она получила название метрической теоретической системы Г. Раша³¹ (Rasch Measurement, RM). Когда эту модель измерения называют теорией, то и другие модели можно называть теориями, как это случилось, например, с IRT. Хотя другие позже появившиеся модели измерения так и остались моделями. Да и более точное название IRT — это Latent Trait Model of Measurement.

Если за этими двумя названиями моделей измерения сохранять статус теорий, то отмеченное лексическое несоответствие лучше преодолевать созданием теории педагогических измерений, что и было сделано в статье журнала ПИ № 1, 2004 г.

Все перечисленные — относительно общие теории, применимые не только в педагогике, но и в психологии, социологии, в медицинских и иных исследованиях. Все теории неоднократно рассматривались в журнале «Педагогические измерения (ПИ)». Но общие теории не давали и не могли дать адекватные ответы на специфические педагогические проблемы. Нужна была специфическая педагогическая теория.

Неоднократно приходилось читать о делении теорий измерения на современные и несовременные. С подобным подходом трудно согласиться, хотя бы потому, что устаревших (ненужных) теорий в педагогических измерениях нет. Каждая теория и сопутствующие ей методы дают свою

30

Аванесов В.С. Item Response Theory: Основные понятия и положения. Статья первая // Педагогические измерения. № 2. 2007. С. 3–28; Аванесов В.С. Истоки и основные понятия математической теории педагогических измерений (Item Response Theory). Статья вторая // Педагогические измерения. № 3. 2007. С. 3–36.

31

Аванесов В.С. Применение тестовых форм в Rasch Measurement // Педагогические измерения. № 4. 2005. С. 3–20.

уникальную информацию о качестве заданий.

Например, в RM используется проверка заданий на соответствие требованиям модели Раша. Но получаемая при этом статистика оценки адекватности задания (Fit statistics) заметно уступает по точности, значениям коэффициентов корреляции, используемых в статистической (классической) модели измерения для обоснования пригодности каждого задания при включении его в тест. Не случайно во всех профессионально сделанных вычислительных пакетах значения коэффициентов корреляции приводятся. Сторонники RM иногда не принимают во внимание эту статистику, что надо признать как недостаток.

Другой пример. В IRT основной предмет исследования — формальные свойства каждого тестового задания. Эти свойства определяются по одной из математических моделей, применяемых к эмпирическим данным результатов ответов множества испытуемых на данное задание. Но ничего педагогического в этой математической модели нет, хотя в педагогических применениях математической модели это даёт хорошие результаты. В подлинном тестовом процессе все теории оказываются нужными и важными.

Поскольку перечисленные теории одинаково применимы к психологическим, социологическим, медицинским и другим измерениям, их можно отнести к общим междисциплинарным метрическим теориям. Следовательно, педагогическую теорию измерений

можно тогда отнести к специальной педагогической теории.

В педагогической теории измерений выделены три предмета исследования: педагогические понятия теории, принципы отбора педагогического (предметного) содержания заданий и логические принципы формулирования педагогических тестовых заданий.

Существенные признаки педагогических измерений — это педагогически обоснованный отбор содержания теста, точный подбор формы заданий для того или иного содержания, система основных педагогических понятий теории³². Удивительно, что измерения педагогических латентных свойств, в том числе относящихся к педагогической науке, до сего дня некоторые зарубежные авторы считают задачами психометрики, с чем согласиться невозможно.

Проблемная ситуация

К началу XXI века многие страны подошли с разным уровнем развития педагогической науки и образовательных технологий. Особенно заметными стали различия в вопросах изменения форм учебной деятельности, научной организации самостоятельной индивидуальной работы учащихся и студентов, в вопросах развития и применения цифровых образовательных ресурсов, новой учебной техники, новых методов обучения, контроля и самоконтроля.

Хотя к концу XX века научно-технический потенциал России был ещё высок, для многих спе-

33

МГУ теряет позиции.
http://www.stoletie.ru/lenta/mgu_terajet_pozicii_303.htm.
 11.09.2012.

34

Дмитрий Ливанов. РИА
 Новости. http://ria.ru/edu_news/20120621/678237385.html

35

Игорь Реморенко
 (@Remorenko_Igor).
 Posted Monday 14th May
 2012 from Twitlonger.
 Что на самом деле пред-
 лагает Ямбург?: <http://www.novayagazeta.ru/society/52529.html>

36

Уроки проведения ЕГЭ-
 2010. Аналитический
 доклад по материалам
 обращения граждан на го-
 рячую линию Обществен-
 ной палаты Российской
 Федерации / Под науч-
 ной редакцией Л.Н. Духа-
 нниной. Издательский дом
 Государственного уни-
 верситета Высшей школы
 экономики. Москва, 2010.
<http://www.oprf.ru/files/jkbhfhbd84729.pdf>

циалистов стала очевидной отри-
 цательная тенденция ослабления
 научно-технического и образова-
 тельного потенциала государст-
 ва — вначале советского, а затем
 и российского. Низок междуна-
 родный рейтинг российских ву-
 зов³³. Всё меньше иностранных
 студентов стали приезжать на
 учёбу в российские вузы. А это до-
 вольно объективный счётный по-
 казатель снижения привлекатель-
 ности российского образования.

«Мы полностью утратили меж-
 дународную конкурентоспособ-
 ность в этих сферах. Мы её просто
 потеряли. В СССР была хорошая
 наука и хорошее высшее образова-
 ние, но их уже нет. То, что мы име-
 ем сегодня, не соответствует да-
 же минимальным требованиям». «На повестке дня — радикальные
 изменения»³⁴. Эти слова можно
 истолковать как официальное
 признание краха образовательной
 политики, проводившейся в пост-
 советской России.

Одна из причин краха лежит
 в недостаточном понимании сути
 проблем педагогических изме-
 рений. Вот свежий пример. «Не
 ясно, — заявил переназначенный
 недавно на новый срок замести-
 тель министра образования и на-
 уки РФ Игорь Реморенко, — поче-
 му так рьяно критикуется ЕГЭ³⁵.
 Автору этой статьи и очень мно-
 гим другим неоднократно приходи-
 лось писать о некачественности
 этого экзамена, о его отрицатель-
 ном влиянии на систему образо-
 вания. Ассоциальные аспекты ЕГЭ
 зримо выявлены в докладе Об-
 щественной палаты³⁶. Разве этого
 мало для появления желания по-
 нять недостатки ЕГЭ?

Все соображения сходятся
 к тому, что некачественное управ-
 ление сферой образования и про-
 ведение ЕГЭ явились главными
 причинами неприятия педагоги-
 ческой теории измерений, равно
 как и общего краха образователь-
 ной политики последних лет.
 А случившийся по воле Мини-
 стерства образования и науки пе-
 реход от некачественного центра-
 лизованного тестирования к про-
 ведению в России ещё более нека-
 чественного Единого государст-
 венного экзамена можно сравнить
 с отступлением из прошлого века
 в позапрошлый век.

В стране был сформирован
 ошибочный политический заказ
 на создание именно *Единого* го-
 сударственного экзамена и вы-
 делены большие деньги, которые
 направлялись в бывшее Мини-
 стерство образования для орга-
 низации работ. Никакого общест-
 венного контроля или проведения
 независимой экспертизы внутри
 страны не предполагалось.

Не случайно, что за прошед-
 шее время педагогическая теория
 измерений не получила в России
 официального признания. Её не
 принимали и не отвергали, не ру-
 гали и не хвалили. Те, кто зани-
 мались Единым государственным
 экзаменом (ЕГЭ), её просто не
 замечали, как не замечали и дру-
 гие исследования, сопутствующие
 этой теории. Похоже, что была
 уверенность в возможности обой-
 тись без всего этого. Как показало
 прошедшее время, случилось
 прямо противоположное: теория
 вполне обошлась без таких специ-
 алистов и продолжает развиваться,
 а вот ЕГЭ, оставленный сей-

час без должного теоретического обеспечения, проявляет всё больше признаков стагнации. Теория, в отличие от практики, может существовать независимо от мнения отдельных лиц или даже мнения целого министерства, как это случилось в России.

Судя по возникшей ситуации, причину незаинтересованного отношения к педагогической теории измерений надо искать не в самой теории, а в ЕГЭ, который был создан бывшим Министерством образования РФ вопреки положениям данной теории. В результате такого рода бюрократической деятельности произошла обычная подмена качественного тестового метода измерений некачественным ЕГЭ.

Вторую причину отторжения теории можно усмотреть в проводимой в последние годы образовательной политике, слепо отмечающей всё, что не укладывается в рамки принятых в ней догматов. Например, когда в Японии заметили, что национальные экзамены вызвали серьёзное падение уровня образования, там их отменили³⁷. Государственные экзамены, в отличие от национальных, наносят ещё больший ущерб. И действительно, в России государственные экзамены оказались разрушительными для дела образования, но здесь нет никакой возможности их отменить. Власть не слышит мнения учёных и педагогов³⁸.

Вместе с тем, возникшая проблемная ситуация в некотором смысле обострила рефлексию относительно сути ЕГЭ и педагогических измерений, что начало способствовать поиску аргумен-

тов в пользу создания в России других, подлинных методов педагогических измерений. Вероятно, это и есть один из немногих плюсов ЕГЭ. Остальное — минусы.

ЕГЭ — главный тормоз развития педагогических измерений

С началом проведения в России Единого государственного экзамена возникла потребность как-то обосновать применение своеобразных, придуманных там т.н. контрольно-измерительных материалов (КИМов) ЕГЭ.

ЕГЭ пытались сделать похожим на педагогические измерения, но это было только в самом начале. Тогда ещё делались некие попытки поиска подходящей для него научной основы, однако единства или хотя бы приемлемых взглядов в стане разработчиков ЕГЭ так и не было.

Поначалу разработчики не смогли определиться с формами — это будет экзамен или тест? Работники бывшего Центра тестирования настаивали на знакомой им тестовой форме технологичного единого государственного тестирования, в то время как педагоги-предметники, впервые привлечённые к разработке заданий для ЕГЭ, склонялись, наоборот, к тому, что им было известно и понятно — к нетехнологичной экзаменационной форме вопросов.

Поскольку успех всего дела на первом этапе в большей мере зависел от предметников, чиновники Минобробразования приняли их сторону, хотя это был тупиковый

37

Валянский С.И. Правда о «реформе образования». В сб.: Что стоит за реформой образования? Российская академия наук. Физический институт им. П.Н. Лебедева. Материалы научно-общественного семинара ФИАН. Методологический семинар ФИАН. Доклады. Выпуск 5. <http://www.netda.ru/fian/fian5a.htm>

38

Эдуард Днепров. Единый государственный экзамен: замыслы и итоги. // Педагогические измерения. № 3. 2011; *Александр Привалов.* О ЕГЭ как об идеальном тупике // Педагогические измерения. № 3. 2011; *Хуторской А.В.* ЕГЭ — пример научно необоснованного нововведения. Фрагмент из книги: *Хуторской А.В.* Педагогическая инноватика: Учебное пособие. 2 изд. М.: Изд. центр «Академия», 2010. С. 97–101. Опубликовано также в ж. «Педагогические измерения». № 4. 2010. С. 17–21. А также критические работы мн. др. авторов.

выбор. Таким образом, на самом раннем этапе возникновения ЕГЭ победили сторонники не тестовой, а экзаменационной формы. Этому способствовало и общее отрицательное отношение к тестам, которое сложилось в России давно. И такое отношение из-за ЕГЭ снова начало укрепляться в России. В подтверждение правильности этого мнения приведём выдержки из решения Всероссийской конференции математиков:

«Многие недавние решения и проекты руководителей образования вызывают у нас серьёзное беспокойство. Прежде всего, это переход на 12-летнее обучение, замена конкурсного экзамена единым тестом. Мы опасаемся, что они не улучшат образование, но, наоборот, приведут к снижению его уровня». За проведение единого экзамена по математике выступило 22 участника конференции, против — 88. Всего участвовали в голосовании 131 участник конференции³⁹.

Крах научного обеспечения ЕГЭ

На первых порах странный гибрид экзаменов и тестоподобных форм в ЕГЭ дал повод руководителям бывшего Министерства образования утверждать об оригинальности избранного ими пути. После многократной критики автором этой статьи утверждения такого рода там затихли, а информация о качестве таких материалов стала засекреченной. Она и теперь является главной тайной Министерства образования и науки. Однако по изредка

появляющимся фрагментарным данным, качество контрольных «материалов» там остаётся ниже всякой критики. В них нет ничего метрического.

Необученные вопросам теории педагогических измерений, и в частности, не понимающие логику тестовых форм, методисты-предметники создали затем такие задания, которые сразу же пробили огромную брешь в качестве Единого государственного экзамена. Часть заданий была алогичной⁴⁰, часть ответов в заданиях легко угадывалась. В итоге госэкзаменуемые с низкой подготовкой незаслуженно получали сравнительно высокие баллы.

Вторая и третья части этого госэкзамена, в силу нетехнологичности используемых форм, потребовали гигантских затрат времени, ручного труда и бюджетных средств, породили субъективизм в оценках и множество ошибок, но Министерство образования утвердило всё-таки такой вариант единого экзамена, надеясь со временем разобраться и в этом вопросе.

В качестве уступки оппонентам-сторонникам технологичного тестирования было решено использовать тестовую форму заданий, более или менее апробированную к тому времени в бывшем Центре тестирования. Тестовую форму заданий создатели ЕГЭ пытались использовать для самой лёгкой по содержанию первой части Единого государственного экзамена.

Небольшое отрезвление от отождествления тестов и КИМов ЕГЭ последовало также после смены руководства Рособназдо-

39

Всерос. конф. «Математика и общество. Математическое образование на рубеже веков». Обращение Всероссийской конференции «Математика и общество. Математическое образование на рубеже веков» \\ Математическое просвещение. Третья серия, 5 вып., п.3, 2001 г. Дубна. «Ратмино». 18–22 сентября 2000 г.

40

Вот пример: 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.).

ра. Было заявлено, что тестирование и ЕГЭ не следует рассматривать как синонимы⁴¹. После чего процесс размежевания тестов от КИМов пошёл с заметным ускорением.

С годами КИМы всё больше скатывались к традиционному экзамену. В итоге ЕГЭ оказался вовсе вне какой-либо метрической теории и вне внятно артикулированных наук — как педагогики, так и педагогических измерений. Были потуги создать что-то вроде КИМологии, но и они провалились. Отсутствие научной поддержки сыграло для ЕГЭ губительную роль. В научных публикациях неоднократно приводились факты некачественности результатов этого экзамена, отягощённые к тому же многочисленными фактами асоциальности самой процедуры проведения ЕГЭ⁴².

Для научного обеспечения проведения ЕГЭ при бывшем Министерстве образования РФ было создано специальное научное заведение — Федеральный институт педагогических измерений (ФИПИ). Как следует из редких научных публикаций этого института, одни сотрудники и авторы, вслед за руководителями Министерства образования и Рособнадзора, называли упомянутые КИМы ЕГЭ тестами. Другие считали, что говорить о ЕГЭ, как о тесте, рано (с. 34). Третьи полагали, что «контрольные материалы для Единого экзамена не сводятся к тестам в узком смысле слова»⁴³.

В общем, там воцарилась неразбериха, итогом которой стала некая псевдонаучная деятельность, которая больше всего под-

ходила под название ещё одной псевдонауки — «ЕГЭведение». Кредо этой псевдонауки можно усмотреть в такой цитате: «КИМы — это стандартизованная экзаменационная работа, создаваемая в соответствии с требованиями теории педагогических измерений, позволяющая с достаточной объективностью и надёжностью провести независимую государственную аттестацию выпускников образовательных учреждений и отбор абитуриентов вузов»⁴⁴. Казалось бы, необходимость теории, как основы ЕГЭ, признаётся, но только на словах. Реальные результаты оказались диаметрально противоположными.

Получалось, что ЕГЭ позволит единым набором КИМов провести аттестацию и приём в вузы якобы в соответствии с требованиями этой самой теории. В то время как требования не только педагогической, но и любой другой теории измерений в общественных науках были, наоборот, в том, что для повышения точности измерений уровень трудности заданий должен соответствовать уровню подготовленности испытуемых. Чем больше соответствие, тем выше точность педагогических измерений.

В ЕГЭ это требование нарушалось. Научные аргументы сторонники ЕГЭведения слушать не хотели, ибо тогда сразу обрушилась бы вся ошибочная конструкция, которую они выстроили. Если бы эту конструкцию не подпирали чиновники профильного министерства. Эти боялись возможного гнева верховной власти за выпуск бракованного изделия. А потому

41*Глебова Л.Н.*

Никакого ЕГЭ в 9 классе нет. http://www.gazeta.ru/education/2009/04/09_n_2972208.shtml

42

Уроки проведения ЕГЭ-2010. Аналитический доклад по материалам обращения граждан на «горячую» линию Общественной палаты Российской Федерации / Под научной редакцией Л.Н. Духаниной. Издательский дом Государственного университета Высшей школы экономики. Москва, 2010. <http://www.oprf.ru/files/jkbhfbhd84729.pdf>

43

«Единый государственный экзамен. Научные основы, методология и практика организации эксперимента» (с. 14); Под ред. В.А. Болотова. М.: Логос, 2002. 208 с.

44

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

скрывали всё, что только можно скрыть.

Под название «КИМы ЕГЭ» можно подвести любой «материал» — задачи, вопросы, упражнения и т.п. — всё это действительно материалы, но не для создания методов педагогических измерений. Создание КИМов можно охарактеризовать как деятельность, лишь *похожую* на измерение, но судя по имеющимся фактам, не соответствующую научному процессу педагогических измерений. В то время как создание теста — это разработка научно обоснованного метода педагогического измерения.

Основные ошибки были сделаны на первом этапе.

Вот выдержки из статьи, предназначенной для научного обеспечения ЕГЭ: «Согласно общему решению, выработанному рабочей группой и коллективами авторов, в 2001 г. постарались пойти по простейшему и наиболее «прозрачному» для пользователей пути, совместив основания для структуризации и выделив три части контрольных измерительных материалов. В первую, наиболее лёгкую часть попали задания с выбором ответов. Во вторую были включены более трудные задания, требующие от испытуемых дополнения в виде краткого, заранее регламентированного автором задания ответа. В третью, наиболее трудную часть вошли задания со свободно конструируемым ответом».

Мнения разделились при оценке трудности контрольных измерительных материалов, которую, как и следовало ожидать, вузовские педагоги считают не-

достаточной, а школьные — избыточной и предлагают облегчить аттестационную часть теста. Различаются точки зрения по поводу оптимального числа частей КИМ. Многие рекомендуют включить в них дополнительную особо трудную часть для отбора абитуриентов по профилирующим дисциплинам в престижные вузы, причём перед выполнением этой части заданий предлагают делать перерыв⁴⁵...».

Три противоречивых основания, разделяющие задания одного метода не по содержанию и уровню трудности, а по признаку формы, принятые *общим решением* всех без исключения разработчиков в основу создания одного ЕГЭ — достаточное условие для провала этого государственного экзамена. Если бы не было мощной поддержки чиновников.

Подтверждающие фактические результаты о полной метрической бездарности созданного ЕГЭ появились к моменту выхода в свет второго номера журнала «Педагогические измерения». Данные оказались ошеломляющими. Тем не менее, сенсационная статья не могла быть опубликована в журнале по причине несогласия авторов с редакционными правками. Тем не менее, ввиду важности этой статьи для истории ЕГЭ и педагогических измерений, она сохранилась в редакционном портфеле журнала.

Вот выдержки из той неопубликованной статьи, в которой были визуализированы результаты влияния трёх единогласно избранных оснований (факторов) при создании ЕГЭ⁴⁶:

45

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

46

Имена могут быть названы, если авторы пожелают.

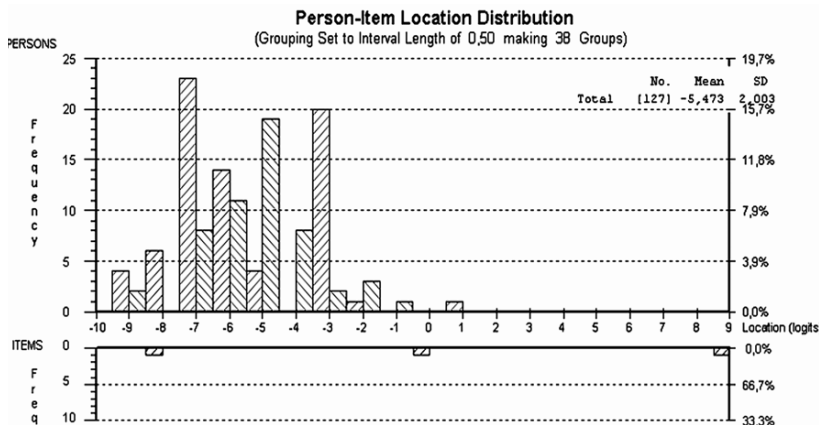


Рис. 1. Соответствие⁴⁷ между уровнем знаний учащихся и трудностью частей А, Б и С по математике

«На рис. 1 представлены две гистограммы. Первая (вверху) свидетельствует о распределении уровня знаний учащихся, вторая гистограмма (внизу) — о среднем уровне трудности заданий трёх частей КИМа — А, Б и С.

Вот что писали авторы той статьи. Цитата: «На этом рисунке по оси абсцисс откладываются значения латентной переменной (уровень знаний математики и уровень трудности частей теста), по оси ординат — число регионов (вверху) и число частей теста⁴⁸ (внизу). С левой стороны оси ординат указаны абсолютные значения регионов/частей теста, а с правой — соответствующие относительные величины (проценты).

Из рис. 1 следует, что уровень трудности части А теста ниже, а уровень трудности части Б теста — выше среднего уровня знания математики по регионам. Уровень трудности части С имеет запредельное значение и значительно превышает уровень знания мате-

матики в регионах. С точки зрения точности измерения латентных переменных, средняя трудность теста не должна отличаться от среднего уровня знаний учащихся более чем на 1 логит⁴⁹. В данном случае уровень трудности теста на 5,47 логита превышает средний уровень знания математики учащихся в регионах. Это означает, что тест⁵⁰ неэффективен с точки зрения точности измерения — точность измерения уровня знаний по математике у этого значительно ниже точности того теста, который бы соответствовал уровню знаний учащихся по математике, при прочих равных условиях».

Разработчики ЕГЭ в таких случаях говорят — это не тест, а КИМ ЕГЭ, а подходить к КИМу с тестовыми требованиями несправедливо. Но и в этой логике есть большой порок. Вопрос ведь стоял о создании метода педагогического измерения именно Федеральным институтом педагогических измерений. По нормальной логике, обнаруженный дефект их изделия просто немислим. Это факт полной метрической непригодности созданного метода. При этом критика некачественности этого метода Министерством образования и науки РФ игнорируется. Тем самым оно не оставляет никаких возможностей культурного воздействия на ошибочную практику.

47

Это как раз случай полного несоответствия из-за зияющих разрывов по мере трудности между частями одного и того же метода. Тесты с такими разрывами не позволяют даже студентам, поскольку нарушается известное по текстам L.L.Thurstone требование континуальности измерения. Авторы статьи не согласились тогда с такой интерпретацией их результатов, поскольку занимали руководящие должности в вузе. Вследствие чего статья с важными результатами пошла, к сожалению, не в набор, а в архив редакции журнала.

48

Авторы этой статьи были сторонниками тестов. Отсюда и соответствующая лексика. Напомним, что тестов в ЕГЭ никогда не было и нет сейчас.

49

Правильнее должно быть 0,5 логита. (Ред.)

50

По мере выявления кардинальных различий между КИМами ЕГЭ и тестами сторонников тестов постепенно отодвигали от разработки КИМов ЕГЭ, в то время как влияние сторонников традиционной экзаменационной формы усиливалось. Это обстоятельство сыграло для государственного экзамена отрицательную роль. В научном отношении он оказался без какой-либо основы вообще.

Хотя тогда использовалась и наукообразная лексика типа «тест», «шкалирование», т.н. «калибровка заданий» и т.п. С годами исчезла и эта лексика. КИМы ЕГЭ всё больше скатывались к традиционному экзамену. Таким дорогостоящим образом исправлялась допущенная в самом начале ошибка совмещения экзамена и тестовой формы, а также совмещения трёх сильно различающихся уровней оценивания в одном ЕГЭ. Это стало торжеством абсурдного с метрической точки зрения мышления!

После 2005 года научные публикации по метрическим вопросам ЕГЭ, по существу, прекратились, и это стало началом кризиса. Федеральный институт педагогических измерений начал реально превращаться в методический центр по созданию некачественных КИМов ЕГЭ.

Новая история теории педагогических измерений

Автор этой статьи жил в период возрождения в СССР тестов и теории педагогических измерений, был активным участником этих событий.

Можно было бы сказать, что педагогическая теория измерений разрабатывалась в ответ на запрос образовательной педагогической практики. Но в свете её фактической не востребоваемости такое утверждение было бы скорее неправдой, чем правдой. Тем не менее, началом целенаправленных исследований по созданию измерений в России можно считать решение бывшего Комитета по высшей школе о создании Учебного центра при Исследовательском центре бывшего Госкомитета СССР по высшей школе.

Конечно, некоторые авторы могут возразить, что и раньше были публикации, тяготеющие больше к вопросам теории педагогических измерений, чем к вопросам практики тестирования. Но там всё-таки было больше тяготения к теории тестов. Здесь в качестве отправного пункта принят факт создания специализированной кафедры педагогических измерений в упомянутом Учебном центре. Насколько известно, в истории СССР, а затем и России создание специализированной кафедры в научном учреждении случилось впервые.

Подходящих учебных материалов на русском языке тогда не было, а то, что имелось в печати, для учебного курса не годилось. Такие материалы пришлось создавать в процессе работы кафедры. Началась целенаправленная работа по исследованию возможностей тестовых форм для улучшения качества учебного процесса в вузах. Первым заметным итогом той работы стало учебное посо-

бие для профессорско-преподавательского состава вузов⁵¹. Именно в том пособии были сформулированы и раскрыты вопросы истории, теории, методики и практики педагогического контроля.

В те годы писать о тестировании и педагогических измерениях не было принято. У многих ещё хранился в памяти упоминавшийся выше запрет на применение тестов. Отсюда и вынужденная редакционная замена непривычных понятий «тестирование» и «измерения» на приемлемые тогда понятия «проверка» и «контроль».

Система педагогических измерений включает в себя всю научную составляющую разработки и применения тестов для получения качественных тестовых результатов и последующего шкалирования, проведение рейтинга, мониторинга и создания качественных показателей образовательной деятельности. Основной предмет педагогической теории измерений — разработка качественных тестов для измерения уровня и структуры *подготовленности* учащихся и студентов. Такие тесты используются почти во всех современных образовательных технологиях.

Исходные понятия теории педагогических измерений

В тот период было начато решение несколько проблем педагогических измерений⁵². Первой из них стала проблема разработки исходных понятий теории⁵³. Понятия педагогической теории измерений целенаправленно начали формулироваться в период напи-

сания автором этой статьи диссертации на соискание учёной степени доктора педагогических наук. Для того чтобы избежать обычной путаницы в толковании сути рассматриваемых научных положений, необходимо было с самого начала дифференцировать два главных понятия: педагогическое тестирование и педагогические измерения.

Педагогическое измерение — это преимущественно научная деятельность, результатом которой является методология и теория организации практического тестового процесса, такого, который позволяет в итоге получить объективированные результаты на интервальной, а возможно, и на пропорциональной⁵⁴ шкале.

Педагогический тест — это научно обоснованный и технологичный метод исследования результатов учебной деятельности посредством системы заданий возрастающей трудности, позволяющей качественно оценить структуру и измерить уровень подготовленности испытуемых.

В учебном процессе преимущественно применяются не тесты, а *задания в тестовой форме*, если последние сделаны правильно, с учётом формальных, содержательных, логических и семантических требований. Но правильная разработка заданий в тестовой форме встречается редко. То, что сейчас иногда называют «тестами ЕГЭ», не является ни тестами, ни педагогическими измерениями⁵⁵.

Н.Ф. Ефремова считает, что тесты являются наиболее эффективной и объективной формой

51

Аванесов В.С. Основы научной организации педагогического контроля в высшей школе: Пособие для слушателей учебного центра Государственного центра СССР. М.: МИСиС, 1989.

52

Аванесов В.С. Методологические и теоретические основы тестового контроля». Дисс... докт. пед. наук. СПб госуниверситет, 1994. 339 с.

53

Аванесов В.С. Определение исходных понятий. <http://testolog.narod.ru/Theory46.html>; Аванесов В.С. Понятийный аппарат теории педагогических измерений// <http://testolog.narod.ru/Theory26.html>

54

Каргин Ю.Н. Идея педагогических измерений в шкале отношений // Педагогические измерения. № 2. 2012.

55

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

оценивания знаний⁵⁶. Она относит тест не к формам педагогического измерения, а формам оценивания, таким, например, как экзамен, зачёты, КИМы ЕГЭ и т.п. С этим трудно согласиться. Отнесение разных объектов к одному классу может способствовать лишь риску потери уникальных свойств теста, в отличие от экзаменов и пр.

Тесты обычно являются средством итогового, по содержанию, внешнего контроля. В России методом внешнего, относительно школ, контроля стал Единый государственный экзамен. Но, по сути, это не тест и не экзамен, в обычном понимании смысла этих педагогических форм⁵⁷.

Измерения можно также определить как процесс перехода от изучения интересующего свойства к созданию переменной величины, являющейся операциональным выражением данного свойства. Педагогические измерения имеют дело с латентными педагогическими свойствами личности, формируемыми в процессе образовательной деятельности и воспитания.

В теории педагогических измерений утверждается: качество измерения тем выше, чем лучше уровень подготовленности испытуемого соответствует уровню трудности заданий. Объективность педагогических измерений возникает как следствие интеграции методов обоснования надёжности и валидности тестовых результатов с методами шкалирования.

Результат измерения интересующего свойства выражается точкой на прямой линии, пред-

ставляющей меру интересующего свойства. Отсюда и возникло одно из самых коротких определений педагогического измерения⁵⁸.

Цели и задачи педагогических измерений

Вторая проблема, которая решалась в последние годы — это формулирование целей и задач.

Сейчас цели тестирования и педагогических измерений лучше разделять. Цель тестирования — получение тестовых результатов уровня подготовленности испытуемых, цель же педагогических измерений в основном — получение шкалированных значений уровня подготовленности испытуемых и уровня трудности заданий.

Цели педагогических измерений можно разделить на метрические, педагогические, психологические и социально-политические.

Пример метрической цели — построить шкалу уровня подготовленности испытуемых или шкалу уровня трудности заданий. В рамках метрической цели нередко используется сравнительно узкое, фокусированное определение цели педагогического измерения — как позиционирование испытуемого на прямой линии, представляющей латентную переменную величину⁵⁹.

Для позиционирования испытуемых на шкале их надо тестировать посредством заданий, определяющих переменную величину, а затем посредством шкалирования убедиться в том, что полученные тестовые результаты помогают определить место испы-

56

Ефремова Н.Ф. Тестовый контроль в образовании. Учебное пособие. М.: Университетская книга, Логос. Москва, 2007.

57

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

58

A measure is a location on a line. Measurement is the process of constructing lines and locating individuals on lines. In: Wright, D. N. and M. H. Stone. Best Test Design. Chicago, 1979.

59

Wright B.D., Stone M.H. Best Test Design. P. 3.

туемого на латентной переменной величине.

Пример педагогической цели — провести классификацию испытуемых по уровню их подготовленности, для комплектования уровневых классов учащихся школы.

Психологи используют тестовые результаты в целях выявления психологических затруднений в учебном процессе.

И наконец, пример социально-политической цели — применение тестов для обеспечения равного доступа учащихся к качественному образованию. Для достижения последней цели вместо педагогических измерений в России стали использоваться некачественные КИМы ЕГЭ.

Основную задачу педагогических измерений Ю.Н. Каргин видит в поиске оптимальных оценок показателей подготовленности испытуемых и трудности заданий по наблюдаемым результатам тестирования⁶⁰.

С момента возникновения психометрики считается, что латентная переменная величина образует так называемый континуум измеряемого свойства⁶¹. Это означает непрерывность значений интересующего свойства на шкале, отсутствие каких-либо ограничений на принимаемые значения слева и справа от среднего арифметического результата.

«Контрольно-измерительные материалы» обернулись метафорой, не содержащей в себе ничего метрического. В реальности был только госконтроль посредством преимущественно нетестовых заданий. Противоречия между теорией педагогических измерений

и практикой разработки КИМов ЕГЭ скрыть не удавалось.

Образовательно-политический фактор

*В российской системе образования одновременно сосуществуют три процесса, три состояния: реформа, застой и откат назад*⁶².

Действительно, в России уже много лет мирно уживаются три названных процесса, три состояния: реформа, застой и откат назад или, иначе, деградация образования. Это вызвано образовательной политикой, действующими силами и традициями. Развитие образования востребовано родителями и общественностью, обеспечивается педагогами-новаторами, снизу, разрушительные же реформы навязываются бюрократией.

А потому в России в какой-то день всё может превратиться в политический фактор, даже ЕГЭ и теория педагогических измерений. Высокие значения погрешностей получаемых оценок ЕГЭ оказались неприемлемыми для экзаменуемых и их родителей. На дефекты ЕГЭ в Правительстве РФ не обращали должного внимания⁶³. Вопрос качества ЕГЭ власть не считает важным. А потому этот вопрос с момента возникновения некачественного ЕГЭ стал постепенно превращаться в фактор ошибочной образовательной политики. Наверное поэтому перед началом ежегодного государственного экзамена высшие руководители страны обещают его улучшить.

60

Каргин Ю.Н. Педагогические измерения в шкале отношений // Педагогические измерения. № 2. 2012. С. 14.

61

Guilford J.P. Psychometric Methods. McGraw-Hill, 1954.

62

Основные положения концепции очередного этапа реформирования системы образования // Народное образование. № 8. 1997. С. 5.

63

Впрочем, обращают внимание, но реагируют весьма специфично.

64

Юлия Котова. Президенту доложили об успехе ЕГЭ. <http://www.monich1.ucoz.ru/publ/1-1-0-78FirefoxHTML%5CShell%5COpen%5CCommand>

65

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

66

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

67

Дмитрий Ливанов. РИА Новости. http://ria.ru/edu_news/20120621/678237385.html

68

Там же.

69

Петренко В. Учебники в России всё хуже и хуже. http://www.gazeta.ru/education/2006/09/20_a_837211.shtml

В.В. Путин признал, что существует *очень много проблем* с Единым государственным экзаменом, *которые надо как-то решать*⁶⁴. Но вслед за этой сентенцией ничего путного не последовало. Д.А. Медведев неоднократно заявлял о необходимости разобраться в сути проблем, возникающих с проведением ЕГЭ⁶⁵, и о корректировке⁶⁶ этого экзамена. Была даже создана Кремлёвская комиссия по ЕГЭ, которая тоже не смогла в нём что-то улучшить.

Никто так и не объяснил гражданам — что конкретно нужно изменить в этом ошибочно скроенном экзамене, у которого по сей день нет ни автора, ни проекта, ни теоретического обоснования, ни научно проверенных результатов, но на который вхолостую тратится много государственных денег — по некоторой информации, порядка 33 миллиона рублей в год.

Похоже, что В.В. Путин и Д.А. Медведев теперь решили дистанцироваться от этой политической проигрышной темы, поскольку процент населения, поддерживающий идею проведения ЕГЭ, с каждым годом снижается.

В 2012 году очередные обещания улучшить ЕГЭ было поручено дать главе администрации Президента РФ Сергею Иванову. В его речи прозвучали элементы критики. Ещё критичнее выступил новый министр образования и науки, который в довольно жёстком тоне заявил, что «российские научные и образовательные организации за последние 20 лет оказались полностью неконкурентоспособными. А потому нуждаются в радикальной реформе⁶⁷».

Это был ответ на критику образовательной системы со стороны Г. Грефа. Последний, в частности, заметил, что российские вузы «принимают дубов, а выпускают липу»⁶⁸. Министр мог бы ответить на это, что «дубы» и «липы» появились в огромных количествах после того, как в Центре стратегических исследований, возглавлявшимся ранее Г. Грефом, возникла сомнительная идея проведения в России ЕГЭ. Но он так не ответил.

Ввязавшись в проведение ЕГЭ без должной теоретической проработки проблемы, бывшее Министерство образования допустило стратегическую ошибку. Были кейсовые обсуждения, оппонентов не слышали, да и слушать не хотели. В общем, всё было так же, как и сейчас. Всё это — последствия своеобразного стиля работы. А потому ошибки в управлении образованием продолжают множиться.

О необходимости изменений немало говорилось и бывшими министрами. Например, в 2006 году А. Фурсенко признал, что, например, качество содержания российского образования совершенно не отвечает современным требованиям, и что 80% учебников допускаться к обучению не должно. После чего он допускал эти учебники в школу и продолжал управлять сферой образования ещё шесть лет, ничего не сделав для исправления ситуации. Итог таков — учебники в России становятся всё хуже и хуже⁶⁹, образование же продолжает деградировать.

Не было ясности в ключевых целях и задачах проводимых реформ, в методах и формах образо-

вательной деятельности, в критериях достижения объективно сформулированных целей. В общем, во всех российских образовательных реформах последних 20-ти лет не было заметной опоры на существующие теории и методологии. Это были бюрократические реформы.

С годами ЕГЭ всё больше удалялся от требований, предъявляемых к педагогическим тестам, постепенно превращаясь, таким образом, в отчасти технологизированный массовый государственный экзамен. Это и есть его настоящее название. Используемые в нём т.н. контрольно-измерительные материалы не имеют признаков пригодности их для проведения педагогических измерений. И это уже доказанные факты, признавать которые министерство не спешит.

Для бедного образовательного бюджета страны затраты на общественно неодобряемые «механизмы» самого управления оказались чрезмерными. Достаточно привести пример очень странных, трудно объяснимых затрат миллиарда рублей на спорные, если не сказать ошибочные, государственные образовательные стандарты. Можно также упомянуть издание дорогостоящих, но малополезных школьных учебников, не говоря уже о напрасных затратах на проведение ЕГЭ.

Очень спорны цели, задачи и методы создания Общероссийской системы оценки качества образования, как и уже произведённые на неё расходы. В итоге многие начинания чиновников подорвали доверие педагогической общественности к органам

управления образованием. Эти органы могут качественно работать только в связке с общественно-профессиональными объединениями. Такова принципиальная сущность эффективного управления образованием и наукой.

Самой популярной в последние годы формой отторжения критики стало её замалчивание. Яркое подтверждение тому — нулевая реакция Министерства образования и науки на критический отчёт Общественной палаты по результатам проведения ЕГЭ⁷⁰. Поневоле возникает вопрос: зачем нужна Общественная палата, если власть не обращает внимания на её публикации и не принимает меры к исправлению возникшей ситуации? Отсутствует реакция и на критические статьи очень многих авторов.

С точки зрения теории педагогических измерений *не бывает* тестов в бланковом формате, одинаково точно измеряющих испытуемых разного уровня подготовленности. Это аксиома, которую никому ещё не удалось опровергнуть. Практика проведения некачественного ЕГЭ лишь подтверждает высокую погрешность получаемых оценок. От такого метода больше вреда, чем пользы. Такая погрешность оценок унизительна для страны. Дети имеют право на объективное измерение уровня их подготовленности. К сожалению, ЕГЭ это право нарушает. Кодекс профессиональной чести тестологов на Западе не разрешает использовать методы измерения с неизвестными показателями точности. В России это разрешает делать государство.

Сложившееся отношение к педагогическим измерениям абсурдно. Альбер Камю был прав, утверждая, что чувство абсурдности поджидает нас на каждом шагу. Мы уже много лет живём в антиобразовательном абсурде. И не видно пока просвета.

Основные результаты развития педагогической теории измерений

Как уже отмечалось выше, в работах последних лет был сформулирован понятийный аппарат педагогической теории, даны определения основным понятиям⁷¹. В качестве таковых выделены: *задание в тестовой форме, тестовые задания, педагогический тест, испытуемые, тестовый процесс и содержание теста*.

Предметом специального анализа стали также два других важных понятия педагогической теории измерений. Первое понятие — *форма тестовых заданий*. Это способ связи, упорядочения элементов задания. Тысячи примеров заданий четырёх основных форм по многим учебным дисциплинам читатель найдёт в книгах автора^{72,73}.

Показано, что логическое преимущество задания в тестовой форме, по сравнению с привычными вопросами экзаменов и госэкзамена, заключается в возможности превращения заданий в тестовой форме после ответа испытуемых в форму высказываний, истинных или ложных, что создаёт надёжную основу для безошибочного дифференцирования испытуемых по уровню их подготовленности. Задания в тестовой форме выгод-

но отличается от обычных вопросов свойствами краткости, лучшей понимаемости смысла заданий, быстроты ответа учащихся, технологичности и эффективности.

В современном тестировании часто применяется дифференцирование испытуемых по их ответам на каждое задание, не только на два, но и на большее число уровней. Это означает, что оценки даются за ответы правильные, неправильные, а также за частично правильные, с разными уровнями градации правильности. Такое оценивание ответов испытуемых называется по-английски Partial Credit Model. Оно имеет преимущества перед традиционным дихотомическим оцениванием посредством единицы и нуля. Главное преимущество — повышение надёжности тестовых результатов.

В качестве главной *проблемы* теории педагогических измерений названа разработка показателей качества и эффективности образовательного процесса.

Предмет педагогических измерений — уровень подготовленности испытуемых, который представляет собой единство знаний, умений, навыков, представлений. К понятию «подготовленность» можно отнести и часто используемое сейчас понятие «компетентность», представляющую собой прагматическую, или утилитарную, проекцию общего содержания образования.

При педагогических измерениях одно и то же свойство личности проявляется у разных испытуемых в разной степени. Обычно большему проявлению интересующего свойства ставится в соот-

71

Аванесов В.С. Определение исходных понятий. 3 редакция, 7 апреля 2007. <http://testolog.narod.ru/Theory46.html>; Аванесов В.С. Понятийный аппарат теории педагогических измерений // <http://testolog.narod.ru/Theory26.html>

72

Аванесов В.С. Форма тестовых заданий // Труды Иссл. центра. М., 1991. 33 с. Аванесов В.С. Форма тестовых заданий. М.: Центр тестирования, 2005.

73

аванесов В.С. Композиция тестовых заданий. М.; Аванесов В.С. Педагогическое измерение латентных качеств // Педагогическая диагностика, №4, 2003г.;

ветствие и большее число. Реже поступают наоборот. Это бывает в случаях измерения общественно неодобряемых психологических свойств личности.

Кроме упомянутых выше результатов, в рамках развития педагогической теории измерений исследованы два традиционных критерия качества⁷⁴ — надёжности и валидности тестовых результатов, принятых в статистической теории измерений, и два относительно малоизвестных критерия — объективности⁷⁵ и эффективности⁷⁶ тестовых результатов, используемых преимущественно в РМ и в адаптивном тестировании.

Сформулированы также принципы отбора содержания тестовых заданий⁷⁷, концепция пяти этапов тестового педагогического процесса⁷⁸, дана обновлённая классификация видов проверяемых знаний школьников и студентов⁷⁹. Самое большое внимание в рамках разработки педагогической теории измерений было уделено вопросам формы⁸⁰ и принципам композиции тестовых заданий⁸¹.

Три главных свойства заданий подлежат обязательному шкалированию и доскональному теоретическому и эмпирическому педагогическому исследованию — это уровень понимаемости смысла содержания заданий, мера их

трудности и дифференцирующей способности. В России система такого рода метрической работы по изучению главных свойств педагогических заданий пока ещё не сложилась.

Не были обойдены вниманием и вопросы становления Национальной системы педагогических измерений⁸². Эту проблему Правительство РФ пыталось решить по-своему, и безуспешно, посредством ОСОКО (Общероссийской системы оценки качества образования), в центре которой стоял тот же ЕГЭ. Новые варианты ОСОКО не лучше прежней системы.

Определено пять основных направлений развития теории и практики педагогических измерений: *демаркация* (отграничение) педагогических измерений от контрольных материалов государственных экзаменов и от иных псевдометрических форм и методов, распространяемых сейчас в России под видом средств педагогических измерений; разработка новых форм тестовых заданий; существенное обновление содержания тестов и тестовых заданий; дальнейшая разработка критериев качества и эффективности педагогических измерений; улучшение языка педагогических измерений⁸³.

81

Аванесов В.С. Теоретические основы разработки заданий в тестовой форме: Уч. пособие для профессорско-преподавательского состава высшей школы. М.: МГТА, 1995. 95 с. *Аванесов В.С.* Композиция тестовых заданий. М.: Центр тестирования, 2003. 217 с.

82

Аванесов В.С. Проблема становления Национальной системы педагогических измерений. Харьков, 13–14 декабря, 2007г. <http://testolog.narod.ru/Education19.html>

83

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

74

Аванесов В.С. Проблема качества педагогических измерений // Педагогические измерения. № 2. 2004.

75

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

76

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

77

Аванесов В.С. От заданий в тестовой форме — к тестовым заданиям // Педагогические измерения. № 4. 2010. С. 105–110.

78

Аванесов В.С. Пять этапов педагогических измерений. <http://testolog.narod.ru/Theory58.html>

79

Аванесов В.С. Знания как предмет педагогического измерения // Педагогические измерения. № 3. 2005.

80

Аванесов В.С. Форма тестовых заданий. М.: Центр тестирования, 2006. 137 с.

Выводы и предложения

1. История тестов и практический современный опыт стран с развивающимся образованием позволяют сделать вывод о необходимости скорейшей трансформации государственной формы контроля знаний учащихся и студентов в общественно-профессиональные формы.

2. История педагогических измерений подтверждает важную роль науки в организации систем контроля качества образования. Как показывает опыт внедрения российского ЕГЭ, без развития педагогических измерений существующие формы государственного экзамена вырождаются в примитивные бюрократические и асоциальные формы контроля.

3. Практика игнорирования педагогической теории измерений уже нанесла большой вред образовательной деятельности в России, и может нанести ещё больший вред, если образовательная политика не начнёт меняться в лучшую сторону.

4. Актуальны пять основных направлений развития теории и практики педагогических измерений:

→ Демаркация (отграничение) педагогических измерений от контрольных материалов государственных экзаменов и от иных псевдометрических форм и методов.

→ Разработка новых форм тестовых заданий.

→ Существенное обновление содержания тестов и тестовых заданий.

→ Централизованная разработка показателей качества и эффективности педагогических измерений и образовательной деятельности.

→ Улучшение языка педагогических измерений.

5. Для каждого уровня обучения и для каждого учебного предмета предлагается создать тысячу тщательно отобранных по содержанию и технологичных по форме заданий возрастающей трудности, доведённых до каждого педагога и учащегося. Эта тысяча способна радикально изменить качество массового образования к лучшему. Потому что появляется принципиально новая и эффективная единица учёта качества собственной учебной работы школьников и студентов.

6. Наличие двух и более тысяч таких заданий, умелое их применение для самостоятельной работы каждого учащегося позволят совершить рывок в сфере образования. Потому что такие технологичные задания начинают выполнять главную функцию развития учащихся и студентов. Всё остальное лишь способствует этой функции.

Это предложение можно рассматривать как одну из современных электронных и метрических версий известного в прежней педагогике задачного подхода в обучении. Разработка систем заданий возрастающего уровня трудности по всем изучаемым темам — проблема государственной важности.