

Теория

ЯЗЫК ПЕДАГОГИЧЕСКИХ ИЗМЕРЕНИЙ

Вадим Аванесов

testolog@mail.ru

В фокусе статьи — проблема становления русского языка педагогических измерений. Наибольший вклад в становление языка этой науки внесли статистическая, математическая и педагогическая теории. Статистическая и математическая теории развивались преимущественно в США, педагогическая — в США и в России.

Даны краткие определения основных понятий педагогической теории.

Ключевые слова: язык, педагогические измерения, тест, тестовое задание, задание в тестовой форме.

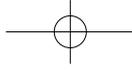
Проблемная ситуация

«...Словесный мир лишь в ничтожной мере даёт правдивую картину мира реального, а в общем и целом словесный мир искажает (фальсифицирует) мир реальный, причём — делает это не по невежеству, а по каким-то более серьёзным причинам¹.

Каждая наука стремится иметь свой собственный язык, позволяющий корректно характеризовать и исследовать интересные объ-

1

Зиновьев А.А.
Введение в логическую социологию. Социально-гуманитарные знания. М., 2000. № 4.



2

«...Организаторы ЕГЭ берут на себя часть ответственности за эту сумятицу, поскольку именно мы запустили в оборот слово «тест» применительно к ЕГЭ». *Болотов В.А.* ЕГЭ: промежуточные итоги. Вопросы образования. № 2. 2004. С. 155–167.

3

Аванесов В.С.
Из глубины веков.
<http://testolog.narod.ru>

4

Аванесов В.С.
Комплексный подход к противодействию коррупции в образовательной сфере // Государственная политика противодействия коррупции и теневой экономике в России. Материалы научной конференции. М.: Центр проблемного анализа и государственно-управленческого проектирования. М.: Научный эксперт, 2007. С. 717–723.
<http://testolog.narod.ru/Obrazov16.html>

5

Коррупция в сфере образования угрожает будущему молодёжи всего мира.
<http://www.epochtimes.ru/content/view/10999/2/>

екты. Педагогические измерения в этом процессе не являются исключением. Это сравнительно молодая наука, возникшая в самом начале XX века на стыке педагогики, психологии, математики, статистики, логики и философии.

Хотя становлению языка педагогических измерений уделялось внимание в каждом номере журнала ПИ, эту работу нельзя сейчас считать ни успешной, ни завершённой. Тем более что ситуация с формированием языка теории педагогических измерений в последнее время стала ухудшаться. Это происходит из-за того, что в России лексика педагогических измерений стала часто подменяться лексикой оценок, экзаменов и т.н. контрольно-измерительных материалов (КИМов) ЕГЭ. Вследствие чего стали распространяться несуразные выражения типа «тесты ЕГЭ», «АПИМЫ», «педагогические измерения посредством тестов КИМов ЕГЭ». Хотя наши читатели понимают, что всё перечисленное — это совсем не тесты, не тестовые задания, не задания в тестовой форме и не педагогические измерения, а только их видимость². Но как убедить в этом остальных, если пресса и руководители много лет пишут и говорят о «тестах» ЕГЭ?

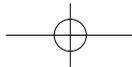
Относительно экзаменов тоже было много раз сказано и написано, что экзамены — нетехнологичная форма оценки знаний,

имеющая древнюю историю, превышающую четыре тысячи лет³. В наше время массовое педагогическое измерение посредством экзаменов невозможно и бессмысленно. Для получения качественных измерений целесообразно использовать тестовый процесс.

Использование КИМов ЕГЭ привело российское образование к тупику и к новому витку роста коррупции. В итоге Россия оказалась уже на 143 месте по уровню некоррупционности⁴. По данным ООН, в России ежегодная общая сумма взяток за поступление в вузы составляет 520 миллионов долларов⁵. Введение ЕГЭ породило новые формы коррупции, теперь уже вокруг него самого.

Корни неадекватного словоупотребления приходится искать в выступлениях руководителей образовательного ведомства страны⁶, а также в явном несоответствии названия и реальной продукции Федерального института педагогических измерений (ФИПИ). Этот институт должен был заниматься подлинными педагогическими измерениями, но со дня своего основания он фактически производил только КИМы ЕГЭ, не отвечающие элементарным требованиям качества и этического кодекса педагогических измерений⁷.

В 2009 году в Министерстве образования и науки отказались, наконец, от идеи оценить еди-



ным КИМом ЕГЭ выпускников школ и абитуриентов вузов⁸. Сделано это было своеобразно, посредством приказа №362, во круг которого возник спор, дошедший до Верховного Суда РФ. Этим приказом все КИМы для ЕГЭ 2009 г. разделили на две группы. Первую группу, с лёгкими заданиями, нацелили для аттестации выпускников школ. В неё вошли только КИМы по русскому языку и по математике. Вторую группу КИМов подготовили для приёмных экзаменов в вузы.

Это позволило избавиться, наконец, от главной причины производства некачественных КИМов ЕГЭ. Прежние годы делали единый КИМ для выпускников школ и абитуриентов всех вузов для ЕГЭ. Но это идея была ошибочная с точки зрения аксиом педагогических измерений, о чём автор этой статьи писал с самого начала затеи ЕГЭ⁹ и затем в течение всех восьми лет т.н. эксперимента по введению ЕГЭ в штатный режим. Наконец, поняли. Вследствие чего из аббревиатуры ЕГЭ теперь фактически выпала первая буква. Хотя номинально название «ЕГЭ» продолжает использоваться, в 2009 году от прежнего названия ЕГЭ в реальности остались теперь только две последние буквы.

Но в этом своём новом варианте государственный экзамен стал разрушать теперь систему аттестации выпускников школ, в

которой все выпускные экзамены сократились до двух. Кто сдаёт КИМ по русскому языку и математике, тот по воле Министерства образования и науки считается аттестованным и по всем остальным предметам за время обучения в школе.

Если это решение сохранится хотя бы ещё один год, то общее среднее образование по остальным школьным предметам станет разрушаться ещё быстрее, чем это было прежде. Потому что многие учащиеся изучают только то, что проверяется. Что не проверяется, изучается слабо, да и немногими, лишь наиболее сознательными¹⁰.

В нарушение этики¹¹ педагогического контроля критериальные значения баллов определяются теперь не до проведения аттестационного ЕГЭ, а после. Это должна делать ещё одна комиссия Рособнадзора. Именно там посмотрят на распределения результатов, и только затем решат, кого считать аттестованным, а кого — неаттестованным. Это существенный признак стагнации системы аттестации.

Случились ещё два события. 15 мая 2009 года, объединёнными усилиями Минобрнауки, Верховного Суда РФ, Минюста и Генпрокуратуры удалось отстоять ранее упомянутый приказ №362 Министерства образования и науки о проведении ЕГЭ¹² в 2009 г. от исков родителей тех детей, которые не хотели или не

Теория

6

См. также подборку противоречивых мнений руководителей образовательного ведомства по вопросу — есть ли в ЕГЭ тесты?// В статье: *Аванесов В.С.* Единый государственный экзамен в фокусе научного исследования. ПИ. №1. 2006. С. 22. <http://viperson.ru/wind.php?ID=535347&soch=1>

7

Code of fair testing practices in education.
Prepared by the Joint Committee on Testing Practices. 2004.
<http://www.apa.org/science/fairtestcode.html>

8

Аванесов В.С.
Удвоение ЕГЭ.
<http://viperson.ru/wind.php?ID=544972&soch=1>

9

Вадим Аванесов.
Единый государственный экзамен: Это вещь будет посильнее Фауста Гёте. Опубликовано в «Учительской газете». №49. 28 ноября 2000 г.

10

Вадим Аванесов.
Приказ № 362 — Ни шагу назад! 18 мая 2009 г.
<http://viperson.ru/wind.php?ID=561141&soch=1>

ПЕД
измерения

11

Code of fair testing practices in education. Prepared by the Joint Committee on Testing Practices. 2004. <http://www.apa.org/science/fairtestcode.html>

12

Вадим Аванесов. Приказ № 362 – Ни шагу назад! <http://viperson.ru/wind.php?ID=561141&soch=1>

13

Левада-центр опубликовал 19.05.2009 результаты опроса общественного мнения по поводу единого государственного экзамена (ЕГЭ). Соцопрос показал резко отрицательное отношение к ЕГЭ большинства респондентов. Тогда как недавно ВЦИОМ настаивал на том, что против ЕГЭ выступают исключительно двоечники, неврастеники и учителя-ретрограды. 36% россиян категорически против ЕГЭ. Оказалось, что только 16% россиян отдают предпочтение ЕГЭ по сравнению с обычными экзаменами. 36% отзываются о нём резко отрицательно, причём в основном это студенты и люди с высшим образованием. 27% опрошенных —

могли сдать аттестационный ЕГЭ. Кроме того, Левада-центр опубликовал новые данные опроса, в котором оказалось, что большинство населения страны вообще выступает против проведения ЕГЭ¹³.

Итак, проблемная ситуация такова: есть общемировой процесс развития теории, практики и, соответственно, языка педагогических измерений, но появился и специфический язык проведения КИМов ЕГЭ, получающий официальное распространение только в России. Эти два языка различны и, более того, интерферентны. Кроме этого, сейчас научный аппарат педагогических измерений несёт на себе весь груз прошлого неупорядоченного словоупотребления, который копился в течение века. Каждый автор приходил в педагогические измерения с языком той науки, которая была ему ближе.

Такая ситуация стала сильно затруднять научную коммуникацию и понимание сути педагогических измерений. Особенно мешает неразбериха и подмены тестов КИМами, экзаменами, оценением и т.п., и всё это вместо действительных педагогических измерений. Стало очевидным, что без чётких разграничений понятий — экзамены, тестирование и педагогические измерения и многих других — движение вперёд станет невозможно. Отсюда следует *актуальность* данной статьи.

Сложившаяся проблемная ситуация указывает на недостаточность проделанной работы. Одновременно это может служить и свидетельством трудности самой проблемы данной статьи — становления подлинного языка педагогических измерений.

Методы оценивания педагогических измерений

Надо заметить, что оценивание и педагогические измерения — разные методы. В России подготовлен очень дорогостоящий проект т.н. оценивания всех обучающихся страны (ОСОКО). Там проявляет себя и своеобразный язык КИМов ЕГЭ. В противовес этому проекту автор статьи предлагает создать в стране ничего не стоящую бюджету систему педагогических измерений¹⁴.

Оценивание представляет собой общее название всех известных методов оценки знаний посредством словесных эквивалентов (отлично, хорошо и т.д.) и (или) соответствующими числовыми оценками. Примерный аналог этого понятия в английском языке — «evaluation», в то время как измерения имеет другой английский эквивалент — «measurement».

Идеального метода для использования в системах оценивания и измерений уровня подготовленности выпускников школ

и абитуриентов вузов различной профессиональной направленности нет. Автор этой статьи неоднократно отмечал, что в массовом процессе аттестация выпускников школ — это преимущественно вопрос оценки достигнутого учащимися образовательного минимума, в то время как вопрос приёма в вузы — это совсем другая проблема, профессионального отбора, требующая тестов с высокой различающей способностью, на отдельных интервалах континуума подготовленности. Качество отбора зависит от квоты приёма, от качества тестов и избранной модели.

Соответственно, с точки зрения теории педагогических измерений и теории профессионального отбора качественная и социально ответственная работа по приёму в вузы требует принципиально иного подхода. Одним набором заданий эти две разные задачи не решаются. Это мнение автора данной статьи все годы «эксперимента» с КИМами упорно не замечалось, не обсуждалось и, соответственно, не принималось к практическому использованию.

На сегодняшний день сложились три основных метода, используемых в национальных и международных системах оценивания и педагогического измерения уровня подготовленности учащихся и студентов.

Первый, самый распространённый — это метод *тестирова-*

ния. Он применяется в США, странах Европы, в Казахстане и странах Юго-Восточной Азии. Актуальность тестового метода объясняется некоторыми его преимуществами перед другими методами. Кратко напомним о пяти основных преимуществах:

1. Высокая *научная обоснованность* настоящего теста.
2. *Технологичность* тестовых методов.
3. Объективно фиксируемая и открыто проверяемая *точность измерений*.
4. *Одинаковость правил измерения и адекватной интерпретации* тестовых результатов.
5. *Сочетаемость тестовой технологии* с другими современными образовательными технологиями.

Но тестовый метод имеет и отрицательные последствия — в смысле развития речи учащихся и студентов, отрицательного влияния на содержание учебного процесса. Он также имеет социально-психологические издержки, особенно в случаях, если тестовая оценка становится решающей для формирования жизненной и профессиональной траектории развития молодежи¹⁵. Поэтому применение тестов должно обязательно сопровождаться применением и других методов, компенсирующих недостатки тестов.

Второй метод — традиционные экзамены, где иногда используются задачи повышенной

Теория

индифферентны как к одному, так и к другому способу проверки знаний.

Напомним, что 4 мая (накануне начала досрочной сдачи ЕГЭ) ВЦИОМ «пропиарил» госэкзамен. Социологи ВЦИОМа сообщили, что против него выступают исключительно двоечники, неврастеники и учителя-ретрограды. В то время как «за» ЕГЭ проголосовала прогрессивная часть педагогического сообщества и отличники с хорошистами, уверяли социологи ВЦИОМа.

<http://www.dailyonline.ru/m/4627/>

Вадим Аванесов.

Проблема становления системы педагогических измерений.

<http://viperson.ru/wind.php?ID=425098>,
<http://testolog.narod.ru>

15

Avanesov V.S.

Consequences of the EGE in RUSSIA. KEDI Journal of Educational Policy. Special Issue: pp. 89–100. Guest Editor. Hoi Suen. V. 3, № 1, 2006. <http://testolog.narod.ru>

ПЕД
измерения

16

Лить Ю-Сю.
Образовательная реформа на Тайване: От единого вступительного экзамена к многовариантным формам приёма в вузы // Педагогические Измерения. №1. 2007. С. 71–76.

17

Баккер С.
Экспертное заключение// В сб.: «Оценка образовательных достижений в рамках национальных экзаменов». Материалы и тезисы докладов Межд. конф. 13–15 декабря 2004г. Стр. 258. М.: Изд-во «Уникум-центр», 2005. 279 с.

18

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

19

Аванесов В.С.
Единый государственный экзамен в фокусе научного исследования. ПИ, № 1. 2006. С. 22.
<http://viperson.ru/wind.php?ID=535347&soch=1>

трудности. Они применяются в большей или меньшей степени в программах международного исследования PISA, в России, на Тайване¹⁶ и других стран Юго-Восточной Азии. Преимущество этого метода — положительное влияние на развитие речи учащихся, возможности индивидуализации контроля, углублённой проверки знаний по избранным темам. Лучше обеспечивается также связь между содержанием обучения и содержанием экзамена. Но для массового процесса измерения уровня подготовленности в генеральной совокупности испытуемых эти методы не годятся.

Третий метод — т.н. контрольно-измерительные материалы (КИМы). Применяются исключительно в рамках проведения российского ЕГЭ. Качество этих материалов является главной тайной Министерства образования и науки РФ. Однако по имеющимся данным, их качество ниже допустимого уровня. Возможно, что именно этим объясняется отсутствие внятных публикаций о качестве КИМов ЕГЭ. Тайной являются и реальные цели, поставленные перед КИМами ЕГЭ. Они оказались непонятными даже известному приглашённому международному эксперту¹⁷ по вопросам тестирования.

Подлинные цели ЕГЭ Б.Виноградов усматривает в соединении с ваучерами ГИФО. Это

создание механизмов перехода государственных вузов в образовательные бизнес-центры. Уже в 2001 году были выбраны экспериментальные регионы, определена шкала стоимости ГИФО в зависимости от оценок по ЕГЭ¹⁸. Но затем что-то сорвалось. Можно предположить, что этот план ваучеризации образования рухнул из-за некачественных КИМов ЕГЭ. Об этом автор писал в уже упомянутой выше статье «Единый государственный экзамен в фокусе научного исследования»¹⁹.

Использование КИМов ЕГЭ вместо тестирования и методов педагогических измерений есть прямой путь в образовательный тупик и к новому витку роста коррупции в сфере образования. Качество образования ухудшается зримо и во множестве регионов. По данным ООН, в России ежегодная общая сумма взяток за поступление в вузы составляет 520 миллионов долларов²⁰, по уровню некоррупционности Россия уже оказалась на 143 месте²¹. Введение ЕГЭ породило новые витки коррупции, теперь уже вокруг него самого.

Постановка проблемы

.....Вещи, качество, события вообще воспринимаются так, как они называются. Сепир Э. Избранные труды по языкознанию и культурологии. М.: Прогресс, 1993. С. 227.

Для научного понимания феномена педагогических измерений надо, говоря словами А.А. Зинovieва, осуществить логическую обработку языка. Причём эта обработка должна охватить не отдельно взятые понятия, а весь их комплекс. Результатом её должна явиться не сумма разрозненных фрагментов, а целостная теория (система, концепция), построенная в соответствии с правилами логики и методологии науки.

Ситуация с языком науки становится сложной в период, когда её статус не определён должным образом. Тогда в этой науке начинается употребление языков нескольких наук. В педагогических измерениях раннего периода вначале использовались язык статистики и математики. О том, что мы имеем дело именно с математико-статистическими теориями, было чётко сказано в классической работе²² конца шестидесятых годов прошлого века.

Принято полагать, что начало научному этапу становления педагогических измерений дал Чарльз Спирман²³. Хотя он занимал должность профессора психологии, язык и методы этого автора были преимущественно статистическими. Не случайно и созданная им теория позже стала называться статистической теорией педагогических и психологических измерений. Позже многие авторы называли её также

классической теорией. В историю науки эта теория так и вошла с двумя именами. Один из смыслов второго названия теории заключался в том, что статистика — это и есть настоящий (классический) язык психолого-педагогических измерений.

Однако последующее развитие науки показало, что смысл названия «классическая теория» оказался гипертрофированным. Спустя полвека, в середине XX века появилась другая, Item Response Theory, не менее классическая теория, и более продвинутая в научном отношении. А потому сейчас самое подходящее время называть теорию, начало которой положили Ч. Спирман, Г. Галиксен²⁴ и тысячи других авторов именно статистической теорией педагогических измерений.

До построения целостной и развитой теории педагогических измерений дело может пойти не скоро, особенно если принять во внимание синтетический характер знаний этой области. В силу ряда причин в России долгое время не было условий для становления языка педагогических измерений. Нужных условий нет, правда, и сейчас. Это одна из причин, почему ошибочная лексика продолжает порождать ошибочную практику.

Например, при обсуждении поправок в ст. 15 Закона РФ «Об образовании», в части, касающейся введения ЕГЭ, надо бы-

Теория

20

Коррупция в сфере образования угрожает будущему молодежи всего мира.

<http://www.epochtimes.ru/content/view/10999/2/>

21

Аванесов В.С.

Комплексный подход к противодействию коррупции в образовательной сфере // Государственная политика противодействия коррупции и теневой экономике в России. Материалы научной конференции. М.: Центр проблемного анализа и государственного управления проектного управления. М.: Научный эксперт, 2007.

С. 717–723.

<http://testolog.narod.ru/Obrazov16.html>

22

Lord F.M. Lord F.M., Novick M.

Statistical Theories of Mental Test Scores. Addison-Wesley Publ. Co. 1968, Reading, Mass. 560 pp.

23

Spearman Ch.

The proof and measurement of association between two things.

American J. of Psychology. 15, 72–101.

ПЕД	
	измерения

ло искать подходящее определение ЕГЭ и КИМов. В итоге там написали: «Единый государственный экзамен представляет собой форму независимой оценки уровня учебных достижений обучающихся с использованием заданий стандартизированной формы (контрольные измерительные материалы), выполнение которых позволяет установить уровень учебных достижений обучающихся по освоению федерального компонента государственного образовательного стандарта среднего (полного) общего образования». Получилось совсем не подходящее, слишком короткое, общее и не точное определение.

Опираясь на реалии, понятие «КИМы ЕГЭ» можно определить иначе: это подготовленные государственными органами задания с нарушениями требований логики и тестовой формы, для получения независимых от общества и системы образования оценок сомнительного качества, с засекреченными результатами. В результате некомпетентной работы законодателей Россия погрузилась в пучину министерского языка, принятого в КИМах ЕГЭ.

Ещё один пример некачественных определений даёт нам зарубежная практика тестирования. Из США к нам пришла традиция называть «тестами» любые задания с выбором одного правильного ответа, а то и просто

любые наборы вопросов. В российских образовательных учреждениях эта традиция имеет, похоже, ещё большую распространённость, хотя уже многие знают, что тест — это не отдельное задание, а система заданий возрастающей трудности, представленных в одной из четырёх тестовых форм. Знают, но ничего поделать не могут. Слишком сильна привычка называть тестами любой набор задач, вопросов, или упражнений, подготовленный для контроля знаний студентов или школьников.

Обыденный смысл слова test охватывает любой метод проверки, как научный, так и не научный. Научное мышление рассматривает тест не как элементарный перевод общего смысла иностранного слова, а как научно обоснованный метод, как результат, полученный данным методом, а также как следствие умелой интерпретации результатов. Разъяснения автора данной статьи о том, что тест содержит не вопросы или задачи, а кратковременные и технологичные задания, сформулированные как предложения в повествовательной форме, при ответе на которые получают высказывания, истинные или ложные, в массовом и чиновном сознании не принимаются во внимание заметным образом.

Проблему осложняет положение самой теории педагогических измерений. Все известные в

этой теории методы обоснования качества тестовых результатов неизменно оказывались статистическими и математическими. Так сложилось исторически, таким был идеал научности тестовых результатов. Однако попытки построения собственно педагогической теории педагогических измерений, других идеалов научности не прекращались ни в прошлом, ни в наше время. Известны попытки создания педагогической теории педагогических измерений в США²⁵.

Для решения проблемы становления приемлемого языка педагогических измерений требуется анализ причин неудовлетворительного состояния этого языка, проведение терминологической работы и создание пер-вых языковых систем.

В России проблемой создания педагогической теории педагогических измерений занимался автор этой статьи. Идея необходимости перехода от тестирования к системе педагогических измерений возникла как результат сравнительного исследования вопросов методологии педагогических измерений и тестирования²⁶, научного анализа практики централизованного тестирования²⁷, введения российского Единого государственного экзамена (ЕГЭ)²⁸ и личного опыта взаимодействия с практиками-работниками центров тестирования. Последние чаще вынуждены выполнять указания

сверху, нежели рекомендации учёных.

В наше время существенный признак педагогических измерений — это педагогически обоснованный отбор содержания теста, точный подбор формы заданий, для того или иного содержания, система основных педагогических понятий этой теории²⁹, получение данных об испытуемых и заданиях на интервальной шкале натуральных логарифмов. Последнее достигается использованием соответствующих алгоритмов трансформации исходных результатов практического тестирования посредством компьютерных программ типа RUMM 2020, Winsteps и др.

Многолетнее отставание педагогических измерений от потребностей практики тестирования и фактическое отсутствие учёных в штате большинства центров тестирования на пространстве СНГ превращает практику тестирования в зону постоянного риска получения ненадёжных и невалидных — практически непригодных для обозначенных целей — результатов. Этот вывод легко подтверждается независимой экспертизой работы любого центра тестирования. Вывести тестирование из зоны перманентного риска получения ошибочных оценок личности может только стабильная опора на научные педагогические измерения.

Теория

25

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

26

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

27

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>

28

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

ПЕД
измерения

29

Аванесов В.С.
Научные проблемы тестового контроля знаний. М.: Иссл. Центр, 1994. 135 с.

30

Здесь лучше было бы написать слово «единого» языка, но после введения в России единого государственного экзамена (ЕГЭ) это слово приобрело негативный смысл редукции процесса педагогических измерений к элементарному подсчёту числа правильных ответов на вопросы, или иначе, к т.н. контрольно-измерительным материалам (КИМам) ЕГЭ. Они, конечно же, надо повторить, не являются педагогическими измерениями.

Аргументация для такого вывода представлена в статье: *Аванесов В.С.* Являются ли КИМы ЕГЭ педагогическими измерениями? ПИ. №1. 2009. С. 3–26. Вторая редакция этой статьи, с таблицами и гистограммами, представлена по адресу <http://viperson.ru/wind.php?ID=563869&soch=1>

38

Службы ETS в США и Европе являются исключением, подтверждающим общее состояние тестирования, существующего вне необходимого им научного обеспечения. В странах бывшего СССР все центры тестирования созданы на небольших бюджетах образовательных ведомств, а потому они обречены, можно сказать, на производство некачественных оценок, если они не получают статус независимых центров педагогических измерений, имеющих научную опору среди профессорско-преподавательского состава вузов.

Причины отсутствия общепринятого языка педагогических измерений

Можно выделить четыре причины отсутствия требуемого языка педагогических измерений.

Первая причина — это недостаточная исследованность вопроса сущности самой науки о педагогических измерениях. Для начала можно отметить, что эта наука синтетична. Синтетичность здесь понимается в смысле необходимости синтеза научных достижений и лексики ряда наук. В педагогических измерениях сейчас фактически используется несколько языков — педагогики, статистики, математики и, кроме того, в заметной мере, языки психологии, психометрики, педагогической технологии,

информатики и других наук. Свойство синтетичности объективно мешает формированию общего³⁰ языка педагогических измерений.

Педагогические измерения имеют одновременно свойства как фундаментальной, так и прикладной науки. Например, физические измерения основаны на фундаментальных понятиях пространства, времени, массы, на физических законах и свойствах физических тел, на соотношениях изучаемых параметров, а также на свойствах чисел и аксиомах собственно физической теории измерений. Педагогические измерения тоже опираются на педагогические концепции, понятия, свойства, индикаторы интересующего свойства, на свойства чисел и аксиомы.

Это несомненные признаки фундаментальности, присущие любому измерению. Но одновременно педагогические измерения призваны разрабатывать методы измерения, используемые в педагогической практике. А это можно отнести к приложениям педагогической науки. Хотя сама российская педагогика данное приложение не рассматривала своей частью. В тридцатые годы XX века в СССР педагогические и психологические измерения назывались буржуазной тестологией.

Вторая причина отсутствия общепринятого языка педагогических измерений — это задержка с развитием собственно педа-

2' 2009

гогической теории измерений. Целый век использовалась статистическая теория, а последние полвека — математическая теория (Item Response Theory, IRT). Они и представляли язык педагогических измерений. Работки автора по педагогической теории педагогических измерений во многом не принимались во внимание. Это явление скорее типичное, чем нетипичное, очень знакомое в истории российской науки и практики.

Третья причина отсутствия общего языка — чисто российская. Именно в России возник и крепнет, пока что, официально принятый язык т.н. контрольных измерительных материалов (КИМов) единого государственного экзамена (ЕГЭ), измерительный статус которых научно не определён и не подтверждён каким-либо удовлетворительным образом. Напротив, появились первые данные, свидетельствующие, что КИМы ЕГЭ вообще не являются методом педагогических измерений³¹.

Четвёртая причина отсутствия общего и устойчивого языка педагогических измерений заключается в том, что с течением времени начинают различать явления, которые ранее не различались³².

С научно-исторической точки зрения, ни КИМы ЕГЭ, ни исходные результаты даже настоящего тестирования не могут рассматриваться как педагогические измерения.

Они оказываются всего лишь первым этапом сбора данных для последующего логарифмического шкалирования для получения трансформированных баллов на интервальной шкале. Можно сказать так. Тестирование — это только начало педагогического измерения, или иначе, это незаконченное педагогическое измерение. В новейшей литературе только шкалированные результаты тестирования признаются полноценными педагогическими измерениями.

Первая дискуссия по проблеме языка

«..Коль скоро недочёт в понятиях случится,

Их можно словом заменить.

Словами диспуты ведутся,

Из слов системы создаются;

Словам должны вы доверять:

В словах нельзя ни йоты изменять».

Гёте. Фауст

Как следует из публикации Федерального института педагогических измерений (ФИПИ) по вопросам ЕГЭ³³, в самом начале там возникли противоречия по вопросу соотношения лексики КИМов ЕГЭ и тестов. Одни сотрудники и авторы, вслед за руководителями министерства образования и Рособнадзора, называли КИМы ЕГЭ тестами. Другие считали, что «говорить о

Теория

31

31

Аванесов В.С.

Являются ли КИМы ЕГЭ методом педагогических измерений? ПИ. №1. 2009.

32

Зиновьев А.А.

На пути к сверхобществу. М.: ЗАО Изд-во «Центрополиграф», 2000. 638 с.

33

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

ПЕД
измерения

34

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

35

Какая-то подделка под науку ... присвоила себе название знания и ставила возврату последнего почти неодолимые преграды. Руссо Ж.-Ж. Сочинения /Пер. с фр. Н.И. Кареев и др. Сост. и ред. Т.Г. Тетенькина. Калининград: Янтарный сказ, 2001. 416 с.

36

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

37

Аванесов В.С.
Из глубины веков. <http://testolog.narod.ru>

ЕГЭ как тесте, основываясь на результатах первого года эксперимента, конечно, рано» (с. 34). Вероятно, надеялись на лучшее будущее. Третьи полагали, что «контрольные материалы для единого экзамена не сводятся к тестам в узком смысле слова» (с. 14).

Особое внимание привлекло тогда такое вот определение: «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»³⁴.

Откуда видно, что КИМы ЕГЭ считаются не тестами, а *экзаменационной работой*, которую якобы можно создать в гармонии с требованиями педагогических измерений. В этом определении критерии качества тестовых результатов оказались бесосновательно и произвольно перенесёнными на оценку качества экзаменационных работ. Как будто такое возможно. Это была надежда на светлое будущее КИМов. В отчётах ФИПИ 2007 и 2008 гг. скороговорка на половину страницы и непонятно откуда взявшиеся цифры на этот сюжет есть. При отсутствии надлежащего обоснования эти цифры воспринимаются похожими на вымысел.

К такому невдохновляющему выводу подводят гистограммы распределения исходных результатов по КИМах ЕГЭ. Все доступные гистограммы оказались искусственно сжатыми по краям исходной шкалы, чтобы не были видны зияющие «успехи», а точнее, дефекты данного этапа сбора результатов КИМов ЕГЭ, повсеместно представляемые как результаты разработки «тестов ЕГЭ». Впрочем, терминологические подмены такого рода — явление исторически не новое³⁵. И только совсем недавно со стороны нового руководства Рособнадзора последовало отрезвляющее пояснение, что тестирование и ЕГЭ не следует рассматривать как синонимы³⁶.

Основные дотеоретические понятия

Вопрос о дотеоретических понятиях имеет свою историю. Это понятия, возникшие в естественном языке много лет назад, когда ещё не было заметных следов теоретической работы.

Экзамены сложились в период, превышающий четыре тысячи лет³⁷. Особенно важно проводить экзамены там, где нужно проверить качество речи учащихся. В наше время формированию речи учащихся в большинстве образовательных учреждений уделяют заметно

меньше внимания, а потому это качество стало сильно ухудшаться. С точки зрения массового процесса педагогических измерений экзамены стали формой нетехнологичной, субъективной и экономически неэффективной.

Задание. Под этим именем можно представить многие педагогические формы, используемые для упражнений учащихся, проверки и самопроверки уровня их подготовленности. В результате теоретического анализа было определено, что *задание* — это педагогическая форма, нацеленная на проверку усвоения учащимся изучавшихся тем и разделов курса. Каждое задание имеет как бы свою миссию, создаётся для определённой цели и имеет свои характеристики. Главная миссия заданий — побуждение учащихся к активной самообразовательной деятельности, к развитию учащихся и овладению необходимыми знаниями, умениями и навыками до требуемого уровня компетенции³⁸.

Педагогические задания выполняют как обучающие, так и контролирующие функции. Обучающие задания применяют учащиеся для активизации собственного учения, усвоения учебного материала, саморазвития, а также применяют педагоги для обучения учащихся. Всё это свидетельствует об обучающем потенциале заданий. Невнимание к обучающим возмож-

ностям заданий в тестовой форме стало одной из причин отставания российского образования от положения дел с этим в других странах³⁹.

Контроль. Это слово можно рассматривать синонимом другого слова — проверка, которую можно определить как педагогическую деятельность, позволяющую определить уровень учебных достижений. Контроль может быть субъективным или объективным.

Это понятие автор данной статьи использовал исключительно в момент, когда не было возможностей рационально, продуктивно писать и говорить о тестах⁴⁰. Соответственно, в название диссертации автора на соискание учёной степени доктора педагогических наук пришлось включить слово «контроль» вместо «педагогических измерений» или «тестирования». Но тогда это был вынужденный шаг. Сейчас стало намного легче писать о педагогических измерениях.

Педагогический контроль. Данное словосочетание содержит в себе элементы дотеоретического и теоретического знания, поскольку позволяет отчленивать традиционные формы контроля от современных технологических форм. Сейчас основным предметом педагогического контроля рассматривается преимущественно подготовленность, включающая в себя знания, умения, навыки и компетенции.

Теория

Теория

38

Аванесов В.С.
Основы педагогической теории измерений // Педагогические измерения. №1. 2004. С. 15–21. и другие, в №№ 1–2, 2004г. и в №№ 1–4 2005.

39

Аванесов В.С.
Основы теории педагогических заданий. ПИ, №№ 2, 3, 2006.

40

Аванесов В.С.
Методологические и теоретические основы тестового педагогического контроля. Дис. д-ра пед. наук. С-Пб, 1994. 339.

Попытка перейти от донаучных форм контроля к научным была предпринята автором этой статьи в пособии для слушателей учебного центра⁴¹. В этом пособии было отмечено, что в педагогических измерениях пробивает дорогу традиция, давно утвердившаяся в психологическом тестировании. Это формулирование тестовых заданий не в виде традиционных задач, вопросов, указаний, требований, а в виде утверждений в повествовательной форме предложения, которые затем превращаются в высказывания, истинные или ложные.

Оценивание. 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 итоговое⁴². Цель первого — оказывать, посредством оценивания, формирующее влияние на текущий процесс обучения, в смысле его улучшения, за счёт установления обратной связи от студента к преподавателю. Цель второго — получить итоговые результаты обучения.

Оценки нередко путаются с отметками. Последние являются численными представителями оценочных суждений. Они вошли в жизнь в советский период,

директивным путём — Постановлением Совета Народных Комиссаров РСФСР от 10 января 1944 года. Постановление было принято, как писалось, «в целях более чёткой и точной оценки успеваемости» и замены «словесной системы» оценки успеваемости — отлично, хорошо, посредственно, плохо и очень плохо — цифровой пятибалльной системой отметок 5, 4, 3, 2, 1. Этим постановлением предписывалось ввести цифровую систему отметок со следующего дня — 11 января 1944 г. — от момента публикации⁴³.

Три источника и три составные части теории педагогических измерений

Язык педагогических измерений формируют три основные теории — это статистическая, математическая и собственно педагогическая теории педагогических измерений. С публикации проекта педагогической теории педагогических измерений берёт начало линия на развитие нашего журнала⁴⁴. Вряд ли можно утверждать, что какая-то одна теория лучше двух других. Ни одна из этих теорий не отменяет и не заменяет другие. Все они нужны, применяются в меру их потенциальных возможностей.

Статистическая (классическая) теория позволяет дать оценку заданиям и тесту в целом,

41

Аванесов В.С.
Основы научной организации педагогического контроля в высшей школе. М.: МИСиС, 1989. 167 с.

42

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

43

Сборник постановлений РСФСР. М., 1944. С. 43.

44

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

провести предварительную оценку надёжности и валидности тестовых результатов. Затем появилась математическая теория педагогических и психологических измерений (Item Response Theory)⁴⁵. Последнюю в России и других странах нередко называют современной, но такое название ряд авторов считают неоправданным, и с этим надо согласиться. Математическая теория педагогических измерений нацелена на углублённый анализ характеристик каждого отдельного задания, формирования их в оптимальную систему измерения, с требуемыми характеристиками качества результатов.

Педагогическая теория требуется для оценки качества содержания, формы и педагогического понятийного аппарата педагогических измерений.

Краткий анализ каждой из теорий даётся ниже. Но прежде обратимся к языку практики.

Язык практики педагогических измерений

Основным предметом практики педагогических измерений являются создание теста, проведение тестирования и шкалирования тестовых результатов.

Тестирование — это практическая деятельность, практика применения тестов — педагогических, психологических и др.

Педагогическое тестирование определяется как преимущественно *практика* применения тестов для оценки уровня и структуры подготовленности.

Система педагогического тестирования — организация сбора и обработки данных посредством применения тестов на множестве испытуемых. Эта система может оказаться научно обоснованной, не обоснованной и даже антинаучной. Научной основой системы тестирования должна стать теория педагогических измерений.

Исходные результаты тестирования уже не рассматриваются как педагогические измерения; это лишь первый этап сбора данных для последующего логарифмического шкалирования. Можно сказать и так: тестирование — это незаконченное педагогическое измерение. В новейшей литературе только шкалированные результаты тестирования признаются полноценными педагогическими измерениями.

Нынешняя практика тестирования даёт немало фактов для разных выводов — как против тестов, так и в их пользу. Голоса «против» звучат иногда громче голосов «за». Важно понять простую истину: сами по себе тесты — полезный метод, но их применение в практике может оказаться как полезным, так и вредным. Многое зависит от цели, качества, соблюдения принципа добровольности, возможности

Теория

45

45

Аванесов В.С.
Там же. См. две статьи автора по IRT в ПИ № 2 и № 3, 2007 и др.

46

<http://epm.sagepub.com>. Позже этот журнал стал публиковать работы и по математической теории измерений, применимой также с равным успехом к педагогическим и к психологическим измерениям.

47

<http://testolog.narod.ru>

48

Глас Дж., Стэли Дж. Статистические методы в педагогике и психологии: Пер. с англ. Л.И. Хайрусовой. Общ. ред. Ю.П. Адлера. М.: Прогресс, 1976. 495 с.; *Наследов А.Д.* Математические методы психологического исследования. Анализ и интерпретация данных. М.: Изд-во «Речь», 2007. 392 с.

49

Ким В.А. Коррекция баллов на угадывание. ПИ, №4, 2006. С. 47; *Деменчёнок О.Г.* Влияние угадывания на значение тестового балла: корректировать или устранять? ПИ. №1. 2007. С. 56; *Ким В.С.* Коррекция исходных баллов испытуемых. ПИ. №3. 2007. С. 37. и др.

потенциального или реального ущерба личности.

Язык статистической (классической) теории педагогических измерений

Как уже отмечалось, первым языком теории педагогических измерений стал язык статистики, созданный Чарльзом Спирманом. Это язык оказался единственным для теории как педагогических, так и психологических измерений. В начале XX века не делали заметных различий между этими двумя направлениями применения статистической теории. Соответственно, такую объединительную линию, ввиду общности статистического языка, много лет проводил и продолжает это делать научный журнал *Educational and Psychological Measurement*⁴⁶.

Самые распространённые статистические характеристики — доля правильных и неправильных ответов, коэффициент корреляции между ответами испытуемых по заданию и исходными тестовыми баллами испытуемых, показатели средней тенденции (средняя арифметическая, мода и медиана), показатели вариации исходных тестовых результатов. При разработке теста вычисляются пять показателей вариации — размах данных, сумма квадратов отклонений от средней арифметической, дис-

персия (варианса), стандартное отклонение, коэффициент вариации. Кроме того, для исходных тестовых баллов считается коэффициент асимметрии и мера эксцесса.

После вычисления и анализа данных по этим значениям вычисляют коэффициент надёжности тестовых результатов. Все перечисленные статистики удобно вычислять стандартными статистическими пакетами «Статистика», SPSS и им подобными. Формулы для вычисления статистических характеристик читатель найдёт в статьях нашего журнала, на сайте автора⁴⁷, в учебниках по статистике⁴⁸. Ряд статистических вопросов классической теории педагогических измерений связан с вероятным угадыванием правильных ответов.

Вероятность угадывания правильных ответов

С первых дней своего существования статистическая теория педагогических измерений начала уделять внимание феномену возможности угадывания правильных ответов, при применении заданий с выбором одного правильного ответа. Математико-статистическая сторона этого явления уже была рассмотрена в работах авторов журнала⁴⁹. В КИМах ЕГЭ этот вопрос обходят стороной. Хотя угадывание

является там особо существенным источником погрешностей измерения — тем большим, чем меньше число предлагаемых ответов в каждом задании с выбором одного правильного ответа.

При выборе из четырёх ответов могут быть угаданы правильные ответы примерно в четвертой части от общего числа заданий. В итоге испытуемые получают баллы, которые они не заслужили. Это и есть одна из самых распространённых форм искажения результатов за счёт формы заданий с выбором одного правильного ответа.

С целью преодоления феномена угадывания в ЕГЭ было принято решение использовать задания открытой формы, где угадывания практически нет. Правильные ответы по таким заданиям там оцениваются в два раза выше; в предположении, что задания открытой формы труднее для испытуемых, чем задания с выбором одного правильного ответа.

Но задания открытой формы для проведения массовых исследований не годятся из-за нетехнологичности. Ведь не случайно настоящие тестовые службы в массовых обследованиях используют только задания с выбором, используя далее автоматизированную считку результатов с бланков посредством оптических сканеров очень высокой производительности, от 5 до 50 тысяч бланков в час. Хороший при-

мер успешного применения такой техники даёт тестовая служба Республики Казахстан. Там результаты тестирования по всей республике становятся известными испытуемым в тот же день.

Трудность заданий зависит, главным образом, не от формы, а от содержания заданий! Не случайно даже в ЕГЭ ряд заданий с выбором одного правильного ответа оказывается труднее ряда заданий открытой формы. На этот очевидный результат ошибочного волевого «шкалирования» оценок за выполнения заданий разной формы в КИМах ЕГЭ не обращается внимания. Здесь скрываются и другие источники ошибок измерения, вызванные *неверным толкованием* тестовых форм и их возможностей, субъективным приписыванием весовых коэффициентов заданиям открытой формы, ошибками сложения баллов, не обладающих свойством, допускающим такое сложение, и прочими несуразностями в работе с КИМами ЕГЭ.

Применение языка статистики для анализа КИМов ЕГЭ

Центральной моделью статистической теории педагогических измерений обычно является т.н. нормальное распределение результатов. Существует много причин, по которым методы для массового оценивания

Теория

2009

ПЕД
измерения

ориентируются именно на это распределение. Отклонение от модели вызывает обоснованные подозрения в некачественности данных.

Феномен угадывания заметно отклоняет результаты от нормального распределения, что даёт основания усомниться в качестве результатов. Тот же эффект наблюдается при списывании. Этот вывод подтверждается доступными данными по КИМаМ ЕГЭ. Посмотрим пример распределения исходных баллов по восьми КИМаМ⁵⁰, представленных в табл. 1. Кроме угадывания, на отклонение от нормальности распределения влияет также преимущественное пред-

почтение или иначе, ориентирование КИМов ЕГЭ на аттестацию выпускников или на приём в вузы.

На диапазоне до 20 баллов фактически не работают КИМы по физике, химии, биологии, географии и истории России. Единственным исключением из общего ряда является КИМ по математике. Русский язык занимает промежуточное положение. Из этой же таблицы легко видеть, что КИМ по математике ориентирован противоположным образом, преимущественно на приём в вузы. В нём задания оказались труднее, а потому здесь появилось много испытуемых с низкими оценками в левой части

Таблица 1. Распределение участников ЕГЭ в Вологодской области по диапазону тестовых баллов в 2007 году

№ п/п	Общеобразовательный предмет	Диапазоны тестовых баллов							
		0–10		11–20		21–30		31–40	
		чел.	%	чел.	%	чел.	%	чел.	%
1	Русский язык	1	0,0	16	0,2	242	2,7	869	9,8
2	Математика	18	0,2	247	2,8	814	9,3	1037	11,9
3	Физика	0	0,0	9	0,5	75	4,5	242	14,6
4	Химия	0	0,0	1	0,3	24	7,3	50	15,2
5	Биология	1	0,1	6	0,4	100	6,4	290	18,5
6	История России	1	0,1	1	0,1	32	4,1	110	14,1
7	Обществознание	0	0,0	3	0,1	78	2,1	433	11,6
8	Английский язык	0	0,0	5	1,8	12	4,2	18	6,4
9	Немецкий язык	0	0,0	0	0,0	2	6,7	3	10,0
10	География	0	0,0	0	0,0	1	2,4	10	23,8

50

Статистические материалы результатов ЕГЭ на этапе государственной (итоговой) аттестации в Вологодской области в 2008 году
<http://ege.edu35.ru/static/static.htm>

Теория

Продолжение табл. 1

№ п/п	Общеобразовательный предмет	Диапазоны тестовых баллов											
		41–50		51–60		61–70		71–80		81–90		91–100	
		чел.	%	чел.	%	чел.	%	чел.	%	чел.	%	чел.	%
1	Русский язык	2133	24,0	2974	33,5	1771	19,9	681	7,7	159	1,8	39	0,4
2	Математика	2212	25,3	1907	21,8	1637	18,7	718	8,2	146	1,7	7	0,1
3	Физика	501	30,1	421	25,3	286	17,2	102	6,1	22	1,3	4	0,2
4	Химия	67	20,4	82	24,9	63	19,1	26	7,9	14	4,3	2	0,6
5	Биология	418	26,7	364	23,2	225	14,4	96	6,1	50	3,2	17	1,1
6	История России	152	19,5	177	22,7	182	23,3	81	10,4	31	4,0	14	1,8
7	Обществознание	1121	30,1	1431	38,4	532	14,3	122	3,3	6	0,2	0	0,0
8	Английский язык	23	8,1	32	11,3	44	15,5	63	22,3	61	21,6	25	8,8
9	Немецкий язык	4	13,3	1	3,3	5	16,7	5	16,7	8	26,7	2	6,7
10	География	9	21,4	12	28,6	6	14,3	2	4,8	1	2,4	1	2,4

шкалы исходных данных. В КИ-Ме по математике лучше дифференцирует испытуемых левая часть шкалы,

Сходная картина наблюдается и в табл. 2, где представлены данные 2008 г. по Вологодской области. Это означает, что концепция КИМов в тот период не менялась.

Предположение о неработоспособности левой части шкалы подтверждается и табл. 3, где 37–39 баллов считаются неудовлетворительным уровнем.

О том же свидетельствует и средний балл результатов (табл. 3). Неработоспособность левой части шкалы может быть

следствием двух, по меньшей мере, причин: это либо результат заметного облегчения заданий, либо допускались нарушения процесса проведения ЕГЭ.

О несправедливости оценок испытуемых, получаемых по ЕГЭ, свидетельствуют данные отчёта по Кемеровской области за 2007 г. В табл. 4 (в оригинале табл. 1) представлены проценты выполняемости каждого варианта КИМа ЕГЭ по физике⁵¹.

Анализ таблицы позволяет утверждать, что в этом КИМе нет ничего похожего на тест как систему заданий возрастающей трудности. Здесь недопустимо большая вариация заданий по

51

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

47

ПЕД
измерения

трудности, между вариантами одного и того же номера КИМа. Из-за чего ЕГЭ превратился не в средство педагогического измерения, а в лотерею. Уже по одной этой причине единый экзамен, как государственный, не имеет права на применение, поскольку этим некачественным экзаменом нарушается право каждого испытуемого на объективную и справедливую оценку уровня его подготовленности.

В нарушение законов РФ весь основной массив результа-

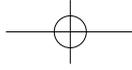
тов за годы применения КИМов ЕГЭ оказался засекреченным для независимых экспертов и общественности. Фактор тайны консервировал их некачественность все восемь лет своеобразного «эксперимента», что и сделало их окончательно непригодными для работы в «штатном режиме» в 2009 г. В итоге к 2009 году пришлось срочно менять содержание КИМов, в связи с их раздвоением для аттестации и для приёма в вузы, предусмотренным спорным приказом № 362.

52

Статистические материалы результатов ЕГЭ на этапе государственной (итоговой) аттестации в Вологодской области в 2008 году.

Таблица 2. (в оригинале 8). Распределение участников ЕГЭ в Вологодской области по диапазону тестовых баллов в 2008 году⁵²

№ п/п	Общеобразовательный предмет	Диапазоны тестовых баллов											
		0–10		11–20		21–30		31–40		41–50		51–60	
		чел.	%	чел.	%	чел.	%	чел.	%	чел.	%	чел.	%
1	Русский язык	0	0,0	0	0,0	9	0,1	169	2,3	839	11,5	2205	30,4
2	Математика	86	1,2	650	8,9	831	11,4	1518	20,9	2213	30,4	1236	17,0
3	Физика	0	0,0	1	0,1	8	0,8	109	10,8	256	25,4	340	33,8
4	Химия	0	0,0	0	0,0	5	1,7	28	9,4	33	11,1	67	22,5
5	Биология	0	0,0	0	0,0	3	0,3	112	10,8	256	24,6	273	26,2
6	География	0	0,0	0	0,0	1	1,7	9	15,5	13	22,4	21	36,2
7	История России	0	0,0	0	0,0	24	3,5	114	16,4	146	21,0	156	22,4
8	Обществознание	0	0,0	0	0,0	0	0,0	107	4,0	442	16,3	778	28,7
9	Английский язык	1	0,4	2	0,8	9	3,6	19	7,6	21	8,4	18	7,2
10	Немецкий язык	0	0,0	1	3,2	1	3,2	2	6,5	5	16,1	10	32,3
11	Французский язык	0	0,0	0	0,0	0	0,0	0	0,0	1	50,0	0	0,0
12	Информатика и ИКТ	0	0,0	0	0,0	1	1,4	3	4,2	10	13,9	14	19,4



Продолжение табл. 2

№ п/п	Общеобразова- тельный предмет	Диапазоны тестовых баллов							
		61–70		71–80		81–90		91–100	
		чел.	%	чел.	%	чел.	%	чел.	%
1	Русский язык	2693	37,1	1152	15,9	181	2,5	17	0,2
2	Математика	543	7,5	175	2,4	21	0,3	3	0,04
3	Физика	215	21,4	58	5,8	17	1,7	2	0,2
4	Химия	67	22,5	66	22,1	28	9,4	4	1,3
5	Биология	239	23,0	136	13,1	22	2,1	0	0,0
6	География	8	13,8	5	8,6	1	1,7	0	0,0
7	История России	141	20,3	101	14,5	12	1,7	1	0,1
8	Общество- знание	1069	39,5	306	11,3	5	0,2	0	0,0
9	Английский язык	50	20,0	58	23,2	57	22,8	15	6,0
10	Немецкий язык	5	16,1	2	6,5	5	16,1	0	0,0
11	Французский язык	1	50,0	0	0,0	0	0,0	0	0,0
12	Информатика и ИКТ	30	41,7	11	15,3	1	1,4	2	2,8

Примерно такие же выводы позволяет сделать анализ гистограмм. Статистически достоверен коэффициент асимметрии результатов КИМа ЕГЭ по физике 2007 г.⁵³ на рис. 1. Этот КИМ был нацелен на приём в вузы, или на учащихся с более высокой подготовкой, из-за чего он некачественно оценивал знания испытуемых 2 волны ЕГЭ. Но одновременно он имеет зримый дефект несостоятельности шкалы исходных баллов при оценке знаний учащихся с отличной подготовкой по физике.

Если бы шкала не была сжата, то этот дефект стал бы более зримым.

Например, на искусственно сжатой по краям гистограмме результатов по русскому языку осмысленность шкалы исходных результатов видна, начиная только с уровня 17 баллов (рис. 2)⁵⁴.

На некоторых гистограммах границы угадывания находятся ещё выше.

В КИМах по математике, напротив, непрогнозируемым образом ведут себя баллы на правой части шкалы, что видно

Теория

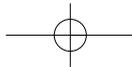
15/08/09

53

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

54

Единый государственный экзамен. Сборник аналитических материалов. Русский язык, 2007. Кемерово.



ПЕД
измерения

Таблица 3. (в оригинале 10) соответствия тестовых баллов ЕГЭ выпускников общеобразовательных учреждений Вологодской области отметкам по пятибалльной шкале в 2008 году⁵⁵

Общеобразовательный предмет	Количество выпускников	Пятибалльная шкала				Средний балл
		«2»	«3»	«4»	«5»	
Русский язык	6005	0–39	40–57	58–71	72–100	62,5
		1,1	28,7	53,2	17,1	
		64	1721	3195	1025	
Математика	6338	0–5	6–12	13–18	19–30	44,2
		9,1	38,0	42,3	10,6	
		575	2411	2679	673	
Физика	905	0–37	38–52	53–67	68–100	55,8
		2,8	36,5	48,3	12,5	
		25	330	437	113	
Химия	261	0–35	36–55	56–72	73–100	63,7
		1,9	26,8	39,8	31,4	
		5	70	104	82	
Биология	827	0–34	35–54	55–71	72–100	58,1
		1,1	38,2	45,8	14,9	
		9	316	379	123	
География	43	0–34	35–50	51–66	67–100	56,0
		2,3	25,6	53,5	18,6	
		1	11	23	8	
История России	542	0–32	33–49	50–67	68–100	57,7
		2,0	26,9	46,5	24,5	
		11	146	252	133	
Обществознание	2332	0–38	39–54	55–66	67–100	60,5
		1,33	23,8	45,7	29,2	
		0	556	1065	681	
Английский язык	222	0–30	31–58	59–83	84–100	68,8
		3,2	20,3	54,5	22,1	
		7	45	121	49	
Немецкий язык	27	0–30	31–58	59–83	84–100	57,3
		7,4	44,4	40,7	7,4	
		2	12	11	2	
Французский язык	2	0–30	31–58	59–83	84–100	55,5
		0,0	50,0	50,0	0,0	
		0	1	1	0	
Информатика и ИКТ	69	0–38	39–56	57–72	73–100	62,4
		4,3	26,1	53,6	15,9	
		3	18	37	11	

**Таблица 4. Решаемость заданий ЕГЭ по физике в 2007 году
по вариантам (1 волна)
Задания группы «А, В»**

Теория

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

Продолжение табл. 4

Номер варианта	A8, %	A9, %	A10, %	A11, %	A12, %	A13, %	A14, %	A15, %
1	23,19	31,88	34,78	60,87	50,72	56,52	34,78	44,93
2	24,64	37,68	39,13	59,42	59,42	63,77	50,72	17,39
3	42,67	36,00	37,33	45,33	56,00	69,33	48,00	14,67
4	36,11	16,67	34,72	41,67	63,89	79,17	47,22	19,44
5	45,83	11,11	31,94	48,61	37,50	58,33	70,83	40,28
6	32,43	54,05	56,76	18,92	29,73	64,86	41,89	18,92
7	62,50	12,50	43,06	50,00	38,89	63,89	29,17	27,78
8	33,33	36,23	56,52	52,17	63,77	62,32	33,33	15,94
9	45,71	35,71	57,14	50,00	62,86	70,00	52,86	57,14
10	43,66	47,89	46,48	69,01	61,97	66,20	32,39	23,94
11	63,24	22,06	51,47	29,41	42,65	83,82	63,24	29,41
12	51,52	13,64	65,15	48,48	78,79	68,18	42,42	19,70

ПЕД
измерения

Окончание табл. 4

13	50,00	57,81	43,75	57,81	26,56	62,50	34,38	32,81
14	36,36	25,76	53,03	53,03	42,42	71,21	56,06	43,94
15	62,71	15,25	55,93	57,63	28,81	86,44	55,93	37,29
Среднее	43,34	30,41	46,91	49,23	49,81	68,24	46,14	29,34

55

Таблица составлена на основе распоряжений Федеральной службы по надзору в сфере образования и науки (Рособрнадзор) «Об установлении шкалы перевода баллов единого государственного экзамена в пятибалльную систему оценивания, используемую для выставления отметок в аттестат о среднем (полном) общем образовании в 2008 году»: № 1190-08 от 05.06.2008, № 1228-08 от 10.06.2008, № 1102-08 от 27.05.2008, № 1062-08 от 23.05.2008, № 1136-08 от 30.05.2008, № 1035-08 от 30.05.2008, № 1101-08 от 27.05.2008, № 1100-08 от 27.05.2008, № 1026-08 от 20.05.2008, № 1027-08 от 20.05.2008.

52

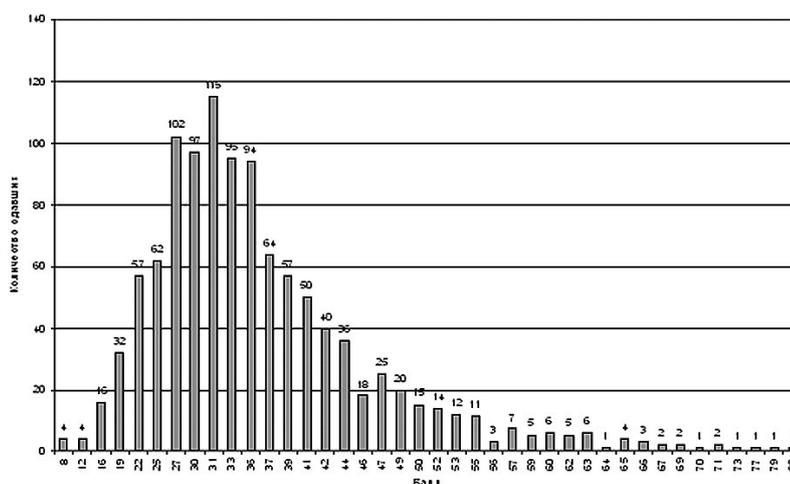


Рис. 1. Результаты по физике в ЕГЭ по баллам в Кемеровской обл. (2 волна)

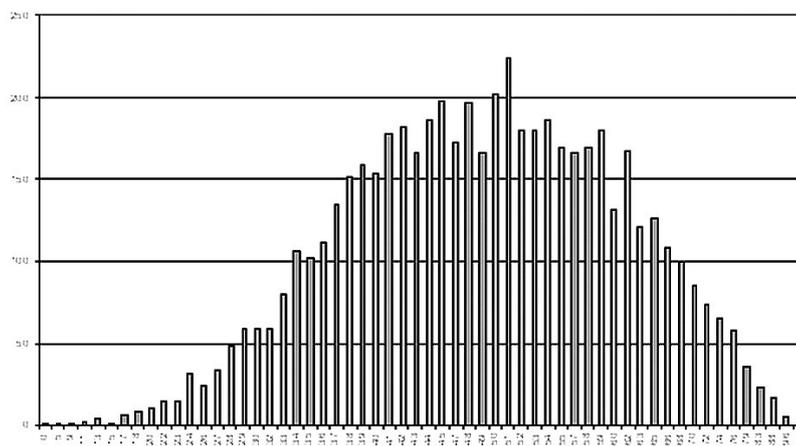


Рис. 2

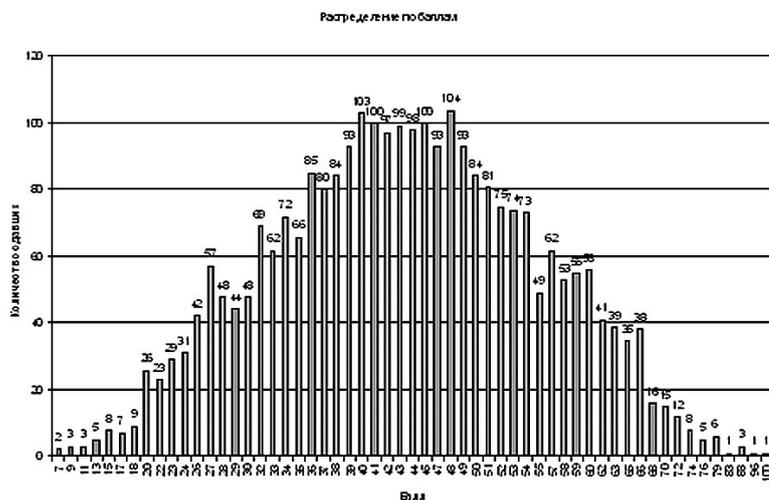


Рис. 3. Распределение результатов ЕГЭ — 2007 год по баллам в Кемеровской области (2 волна)

после уровня 76 баллов. Именно после этого уровня видна несостоятельность шкалы исходных баллов. Можно предположить, что это свойство КИМа ЕГЭ во многом порождается субъективными оценками комиссий, создаваемых для оценки ответов испытуемых по части «С» КИМа ЕГЭ. По данным гистограммы легко видеть высокую потенциальную погрешность оценивания у наиболее подготовленных испытуемых. Что согласуется с ранее опубликованными данными⁵⁶, в которых стандартная ошибка измерения превышала допустимые значения.

Фактически нацелен на приём в вузы оказался КИМ по физике, представленный на рис. 3.

Здесь оказалась недостаточной различающая способность для слабо подготовленных аттестуемых, но вполне впечатляющая различающая способность КИМа для средне подготовленных абитуриентов вуза.

Основные критерии оценки результатов педагогических измерений

К настоящему времени исследованы четыре основных критерия качества и эффективности педагогических измерений.

Надёжность. Это был первый критерий, предложенный новатором статистической теории педагогических и психологических измерений Ч. Спирманом. Идея надёжности берёт начало

Теория

15/08/09

56

Хлебников В.А.
Краткий анализ технологии и результатов Единого государственного экзамена. ПИ. № 4. 2008. С. 25–40.

в идее точности измерения. Зная погрешность, легко корректировать меру точности результатов. Существует множество коэффициентов надёжности тестовых результатов. В упомянутых ранее стандартных пакетах есть ключевая опция — расчёт коэффициента надёжности измерения альфа.

Язык и методы обоснования надёжности на русском языке представлены в ряде работ автора⁵⁷.

Валидность (от англ. слова valid — действительный, пригодный, действенный) — один из основных критериев качества педагогических измерений. Обсуждение вопроса валидности предполагает поиск ответов на три, по меньшей мере, вопроса.

Первый — валидность чего? Раньше считалось — теста. В последние годы валидность стала соотноситься с тестовыми результатами. Валидность результатов — это характеристика их возможности измерять именно интересующее свойство, и никакое другое.

Второй вопрос — от чего зависит валидность тестовых результатов. От цели разработки и применения теста, от содержания и формы тестовых заданий, от качества теста, условий проведения измерений и от интерпретации результатов. Валидность результатов зависит также от подбора контингента испытуемых, от их физического и психического состояния, и от других условий.

Третий вопрос — о критериях оценки валидности результатов педагогических измерений. Используются формальный (правильность формы), содержательный, концептуальный и статистический критерии.

В отечественной и западной тестовой литературе много пишется не только о валидности тестов, но и, ошибочно, о валидности отдельных заданий. Научные разработки последних лет позволяют подойти по-новому и к этому вопросу, а именно: вместо валидности оперировать другим понятием — эффективностью заданий. Естественным следствием такого подхода является обогащение определения теста: в этом случае тест становится не просто системой заданий, а система эффективных заданий. Понятно, что система неэффективных заданий порождает неэффективный тест, а в некотором пределе — совсем не тест⁵⁸.

Объективность педагогических измерений — это такое отражение интересующего свойства личности на числовой шкале, которое адекватно действительному распределению испытуемых по латентному свойству. Можно сказать, что объективность является самым важным и вместе с тем наименее исследованным критерием качества педагогических измерений. Объективность может возникнуть как результат применения системы методов измерения⁵⁹.

57

Аванесов В.С.
Проблема качества педагогических измерений. ПИ. №2. 2004. С. 3–27.

58

Подробно о валидности на русском языке см. Аванесов В.С. Проблема качества педагогических измерений. ПИ. № 2. 2004. С. 3–27.

59

Objective Measurement.
<http://www.meaningful-measurement.com/Objective%20Measurement.pdf>

Идея объективности результатов педагогических измерений была осознана вместе с возникновением первых тестов. По мнению J. Mac Keen Cattell, тест — это средство для получения объективных оценок интересующего свойства личности. Для организации тестирования он считал необходимым создание условий, приближённых к эксперименту. Требуются одинаковость инструкций, равное время на выполнение заданий каждому участнику тестового процесса, добровольность участия испытуемых в тестировании, статистическая обработка данных, ограничение времени тестирования — не более одного часа⁶⁰. Подробнее о критерии объективности — в статье автора⁶¹.

Эффективность педагогических измерений — сравнительно новый критерий оценки теста и тестовых заданий. Проблема эффективности педагогических тестов является частью общей проблемы эффективности форм и методов педагогической деятельности. Естественно поставить вопрос — почему тестирование относится к эффективной форме организации контроля знаний, а сам тест считается эффективным и объективным методом диагностики уровня и структуры знаний? Краткий ответ на этот вопрос заключается в том, что настоящий тест экономичен, потому что не требует тех больших затрат живого труда преподава-

телей, которые сейчас имеют место⁶².

Язык математической теории педагогических измерений Item Response Theory (IRT)

В России название IRT переводили такими словами, как «теория латентных черт», «теория характеристических кривых заданий», «теория моделирования и параметризации педагогических тестов», «современная» теория тестов и т.д. Столь заметные различия в переводах одного только названия IRT уже само по себе являются свидетельством неблагополучия в понимании её сути. Не лучшим образом обстояло дело с переводом на русский язык исходных понятий и положений IRT. Проблема с русским названием IRT заключалась в том, что английское название этой теории не точное, не полное, отчасти устаревшее и метафоричное. А потому прямо не переводимое, в принципе.

IRT определяется как математическая теория педагогических измерений (МТПИ), основным предметом которой является анализ формальных, математико-статистических свойств отдельных заданий и теста в целом⁶³.

В своё время модели этой теории, как и большинство авторов на Западе, модели IRT назывались современными⁶⁴. В последовавших затем у нас

Теория

Ирина

60

Cattell J. McKeen. Mental Tests and Measurements. Mind, 1890, v. 15, p. 373–380.

61

Аванесов В.С. Проблема объективности педагогических измерений. ПИ. № 3. 2008.

62

Подробнее об этом критерии см.: Вадим Аванесов. Проблема эффективности педагогических измерений. ПИ. № 4. 2008. С. 3–24.

63

Аванесов В.С. Item Response Theory: основные понятия и положения. ПИ. № 2. 2008.

64

Аванесов В.С. Основы научной организации педагогического контроля в высшей школе. М.: Иссл. Центр, 1989. 167 с.

65

Точного перевода названий этих теорий на русский язык пока нет.

66

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

67

Аванесов В.С.
Item Response Theory: основные понятия и положения.
ПИ.№№ 2–3, 2008.

68

Например, название классического труда того времени Gulliksen, H. «The Theory of Mental Test Scores». N-Y, Wiley, 1950.

69

В России автор этого доклада ранее использовал понятие «тестология», теперь применяется только «педагогические измерения».

70

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

публикациях других российских авторов современные модели превратились в «современную теорию», что сразу же высветило два существенных недостатка такого толкования. Выходило, что пользователи теории IRT- современные авторы, в то время как опирающиеся на другие теории — не современные. Между тем два сравнительно новых варианта классической теории измерения — Random Sampling Theory, а также Theory of Generalizability⁶⁵ существенно не менее современны, чем IRT.

В главном фокусе изучения IRT находятся, строго говоря, отдельные задания, изучаемые посредством математических моделей, хотя и разработке теста уделено большое внимание. Модель измерения определяется как структурное построение, позволяющее соединить так называемые латентные переменные с одним или с большим числом эмпирически наблюдаемых переменных⁶⁶.

Основной язык математической теории педагогических измерений рассмотрен автором в двух статьях⁶⁷. Важно отметить, что вопросами определения педагогических понятий теории измерений, содержания педагогических тестов и формы тестовых заданий IRT (МТПИ) не занимается. Этим занимается педагогическая теория измерений (ПТИ).

Язык педагогической теории измерений (ПТИ): основные понятия

Примерно до середины 50-х годов XX века наука о разработке тестов называлась преимущественно «Теорией тестов»⁶⁸. И лишь с начала XXI века в качестве научной основы педагогических измерений повсеместно утвердился термин «Теория педагогических измерений»⁶⁹. Так сложилось исторически, таким был идеал научности тестовых результатов. Однако попытки построения собственно педагогической теории педагогических измерений, других идеалов научности не прекращались ни в прошлом, ни в наше время. Известны попытки создания педагогической теории педагогических измерений в США⁷⁰.

В России проблемой создания педагогической теории педагогических измерений занимался автор данной работы. Идея необходимости перехода от тестирования к системе педагогических измерений возникла как результат сравнительного исследования вопросов методологии педагогических измерений и тестирования⁷¹, научного анализа практики централизованного тестирования⁷², негативного опыта введения российского Единого государственного экзамена (ЕГЭ)⁷³ и личного опыта взаимодействия с практи-

ками–работниками центров тестирования.

В качестве педагогической основы теории педагогических измерений могут служить:

Система основных понятий педагогических измерений⁷⁴. В данную систему включены понятия:

Задание в тестовой форме

Тестовые задания

Педагогический тест

Испытуемые

Тестовый процесс

Содержание теста.

Все эти понятия определены в работах автора.

Ещё два понятия ПТИ стали предметом специального анализа в двух книгах автора этой статьи.

Форма тестовых заданий определялась как способ связи, упорядочения элементов задания. Содержание теста существует, сохраняется и передаётся в одной из четырёх основных форм заданий. Вне тестовых форм ни тест, ни его содержание не существуют. По форме все известные в теории и практике тестовые задания можно разделить на четыре основные группы. Вот их точные названия: задания с выбором одного или нескольких правильных ответов, задания открытой формы, задания на установление соответствия и задания на установление правильной последовательности. Тысячи примеров заданий различных форм, по многим, если не всем учебным дисциплинам читатель

найдёт в книге автора⁷⁵. У книги по форме было немало подражателей, но чего-то нового с точки зрения теории у них не было замечено, кроме искажений авторской концепции формы.

*Композиция тестовых заданий*⁷⁶ — это название другой книги, посвящённой вопросам творческого соединения формы и содержания. Композицию не случайно считают уделом сосредоточения многих проблем, связанных с замыслом и исполнением, содержанием и формой. Хорошо сделанное задание — это всегда лучшее, на текущий момент, содержание и наилучшая форма; то и другое — результат, появляющийся вследствие осознания важности их взаимосвязи. Разработка теста начинается с композиции заданий в тестовой форме.

Кроме перечисленных основных понятий, основу ПТИ могут составить:

- концепция пяти этапов тестового педагогического процесса⁷⁷;
- формулировки целей педагогических измерений⁷⁸;
- результаты исследования содержания⁷⁹, форм⁸⁰ и принципов композиции тестовых заданий⁸¹;
- математические модели⁸² и аксиоматика теории педагогических измерений⁸³.

Ключевое место в первом номере журнала ПИ 2004 г. заняла авторская концепция трёх главных понятий — задания в тестовой

Теория

71

Аванесов В.С.

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

72

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>

73

Аванесов В.С.

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

74

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>

ПЕД
измерения

75

Аванесов В.С.
Форма тестовых заданий. М.: Центр тестирования, 2005.

76

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

77

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

78

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

79

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

80

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

81

Аванесов В.С.
Теоретические основы разработки заданий в тестовой форме: Уч. пособие для профессорско-преподавательского состава высшей школы. М.: МГТА, 1995. 95 с.

вой форме, тестового задания и педагогического теста⁸⁴ как системы заданий возрастающей трудности.

В этих работах были даны определения следующих основных понятий теории педагогических измерений:

- *Латентные качества личности.* Латентными называются интересующие исследователей положительные и отрицательные качества личности, не поддающиеся непосредственному измерению. Примерами являются «подготовленность студентов», «знание учебной дисциплины», «способность понимать» «интеллектуальное развитие» и многое другое. Попытки измерения подобных качеств⁸⁵ на уровне обыденного сознания оканчиваются словесными или численными оценками, содержащими в себе немалые погрешности. Методом измерения латентных качеств (свойств) личности является тест.

- *Педагогическое задание в тестовой форме.* Это технологичное средство интеллектуального развития, образования и обучения, способствующее активизации учения, повышению качества знаний, а также повышению эффективности педагогического труда.

Логическое преимущество задания в тестовой форме заключается в возможности его превращения, после ответа студента,

в форму истинного или ложного высказывания.

Задания в тестовой форме выгодно отличается свойствами эффективности, краткости, лучшей понимаемости смысла заданий, быстротой ответа учащихся и определения меры трудности каждого задания, технологичности.

Научно-методическая ориентация педагогов на широкое использование обучающего потенциала заданий в тестовой форме в системе e-Learning даёт начало новой образовательной технологии. В этой технологии задания начинают выполнять функцию не только самоконтроля, но и организации образовательной и самообразовательной деятельности, делая это самым технологичным и эффективным образом. С точки зрения теории педагогических измерений⁸⁶ главным средством обучения в e-learning могут и должны стать не тесты, а задания в тестовой форме.

Взаимосвязь содержания и формы. При рассмотрении этого общего диалектического принципа применительно к тестам невольно возникает ассоциация с искусством. Настоящее искусство представляет собой, по словам Гегеля, законченное внутри себя соединение содержания с вполне соответствующей ему формой⁸⁷. Точно так же и настоящий педагогический тест можно охарактеризовать как резуль-

тат взаимовлияния содержания заданий с наиболее подходящей формой. При этом, в отличие от широко распространённого философского истолкования активности содержания и пассивности формы, тестовую форму надо рассматривать вместе с содержанием как активную сторону взаимодействия.

Тестовое задание определяется как составная единица теста, отвечающая содержательно-педагогическим требованиям к заданиям в тестовой форме и, кроме того, статистическим требованиям: известной трудности, достаточной вариации тестовых баллов испытуемых по заданию, положительной корреляцией ответов по заданию с исходными тестовыми баллами испытуемых.

Не все задания в тестовой форме могут стать тестовыми заданиями. Это заметно различающиеся понятия. Задания имеют шанс стать тестовыми только после эмпирической проверки меры их трудности и других статистических свойств, на типичных группах испытуемых. Корреляция ($r > 0,3$) указывает на способность заданий различать хорошо подготовленных учащихся от тех, кто подготовлен плохо.

Чем больше значение коэффициента корреляции, тем выше *дифференцирующая способность задания*. В последние годы вместо понятия «дифференцирующая

способность задания» автор этой статьи чаще применяет понятие «различающая способность задания». Это связано с тем, что слово «дифференцирующая» в зарубежной теории педагогических и психологических измерений используется для нескольких целей.

Для того чтобы задание можно было включить в тест, его проверяют по множеству и других формальных математико-статистических требований. Это обычно выполняется с помощью статистических пакетов RUMM-2020 и WINSTEPS, которые созданы специально для углублённого анализа формальных свойств тестовых заданий, а также для проведения дуального (conjoint) шкалирования уровней трудности заданий и уровней подготовленности испытуемых.

Именно этот смысл, а не «калибровка», как нередко пишут на русском языке, следует из английского словосочетания *test calibration*. Свойства тестовых заданий были рассмотрены в статьях и работах автора⁸⁸. Задания в тестовой форме имеют шанс стать тестовыми только после эмпирической проверки их статистических свойств, на типичных группах испытуемых.

• *Педагогический тест* — система заданий равномерно возрастающей трудности, позволяющая оценить структуру и качест-

Теория

Аванесов В.С. Композиция тестовых заданий. М.: Центр тестирования. 2003. 217 стр.

82

Аванесов В.С. Математические модели педагогического измерения. Научное издание. М.: Иссл. центр проблем качества подготовки специалистов, 1994. 26 с.

83

Аванесов В.С. Педагогическое измерение латентных качеств // Педагогическая диагностика. №4. 2003.

84

Там же.

85

Понятия «качество», «свойство», «признак» удобно рассматривать как обобщённый аналог английского понятия *trait*.

86

Серия статей в журнале «Педагогические измерения», а также на сайте <http://testolog.narod.ru>

87

Гегель Г. Энциклопедия философских наук // Соч., Т. 1: Логика, 1929. 368 с.

ПЕД
измерения

88

Подробнее см.
Аванесов В.С. Композиция тестовых заданий.
М.: Центр тестирования, 2002. С. 163.

89

Ronald K. Hambleton H.
Swaminathan H.
Jane Rogers.
Fundamentals of Item Response Theory. p. 6,
Sage publ., 1991.

90

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

91

Аванесов В.С.
Форма тестовых заданий. М.: Центр тестирования, 2005. 156 с.

венно измерить уровень подготовленности испытуемых. Тесты применяются для объективизации итогового контроля результатов обучения.

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.

Авторы очень известной на Западе книги *Fundamentals of Item Response Theory*⁸⁹ пишут, что очень трудно или даже невозможно, создать параллельные варианты теста, для проверки надёжности тестовых результатов. И действительно, все варианты к каждому заданию, по каждому предмету, должны иметь общее содержание. Они должны иметь хотя бы примерное равенство средних арифметических показателей, а также показателей вариации по каждо-

му заданию. Только тогда обеспечивается справедливость и объективность выставляемых оценок⁹⁰ испытуемых.

Создание параллельных вариантов заданий заметно облегчается при овладении методикой композиции фасетных заданий⁹¹. В рамках одной только математической теории создать тест с качественными фасетными заданиями практически невозможно. Поэтому авторов упомянутой зарубежной книги по математической теории измерений понять можно.

Остаётся напомнить, что *facet* — это форма записи возможных параллельных вариантов задания, что является обязательным требованием при разработке теста, имеющего на каждое задание набор параллельных заданий. Все элементы одного фасета априорно считаются одинаково трудными по содержанию, что требует последующего эмпирического подтверждения. Каждому испытуемому компьютерная программа подбирает один вариант из фасета.