

Теория

ПРОБЛЕМА ПЕДАГОГИЧЕСКОГО ИЗМЕРЕНИЯ ЛАТЕНТНЫХ КАЧЕСТВ

Вадим Аванесов

testolog@mail.ru

В статье рассматривается проблема измерений латентных качеств личности и формализуемых свойств тестовых заданий. В российской и иностранной литературе обычно пишут об измерении латентных переменных величин¹, но это не то же, что измерение латентных качеств. Величины возникают в процессе измерения.

Превращение интересующего латентного качества в измеряемую переменную величину требует проведения обширного цикла исследовательской работы, которая похожа на строительство моста, необходимого для соединения латентного качества с переменной величиной — а это самое трудное и интересное дело в подлинной культуре педагогических измерений.

Рассмотрены вопросы истории, теории, практики исследования латентных качеств личности и свойств тестовых заданий, даны новые определения ключевых понятий теории латентных качеств. С точки зрения данной теории подтверждён ранее сделанный вывод о непригодности т.н. КИМов ЕГЭ для проведения педагогических измерений.

Ключевые слова: латентное качество личности, латентное свойство заданий, педагогические измерения.

1

Например: *Bollen K.A.* Latent variables in psychology and the social sciences. Annual. Rev. Psychol. 2002. 53: 605–34; «Теория и практика измерения латентных переменных в образовании». // Материалы конференции. Славянск-на-Кубани, 24–25 июня 2005 г. Краснодарский край, 2005. 144 с.

Вопросы истории

Идея влияния невидимых факторов на видимые явления захватила сознание человечества в столь давние времена, что указание точного времени возникновения этой идеи становится сейчас делом невозможным. Известный зарубежный исследователь К.А. Боллен считает, что интерес к изучению такого влияния на поведение человека возник не позднее времени возникновения религиозного сознания². Такой интерес стал приобретать систематические формы научного исследования по мере появления первых университетов: вначале в арабских странах (Марокко, 859 г., Египет 988 г.), затем в Иране (1065 г.), и потом — в Европе (Болонья, 1088 г. и др.)³.

В период Возрождения скрытые качества рассматривались как некие силы, присущие всем вещам, причём каждая вещь представлялась как нечто двойственное: чувственно воспринимаемый предмет и внутреннее «скрытое качество»⁴. Всё, что не находило разумного объяснения, относилось за счёт действия неких скрытых сил и качеств, присущих каждой вещи.

В научно оформленном виде, приближенном к психологии и педагогике, идея измерения латентного качества личности впервые обнаруживается в трудах выдающегося учёного

Чарльза Спирмана. Обычно с его именем ассоциируется авторство классической (статистической) теории психологических измерений. В его работах⁵ можно обнаружить не только истоки упомянутой теории, но и начала развившейся позже другой теории — измерений латентных качеств (Latent Trait Theory).

Первая теория была построена вокруг известной в физике идеи недостижимости точных значений из-за возникновения неизбежных ошибок измерения. Соответственно, была выдвинута идея существования у каждого испытуемого наблюдаемого значения тестового балла, состоящего из двух ненаблюдаемых (латентных) при измерении значений, — истинного значения тестового балла и ошибочного компонента наблюдаемого (исходного) тестового балла.

Отношение дисперсий истинных компонентов измерения к общей дисперсии наблюдаемых данных по множеству испытуемых дали меру надёжности тестовых результатов — главного понятия классической (статистической) теории педагогических и психологических измерений.

Теория измерения латентных качеств (Latent Trait Theory) возникла в связи с попытками измерения интеллекта. Ч. Спирман выдвинул идею

2

Bollen K.A. op. cit.

3

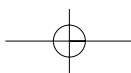
Манса R.
The Application of
Bologna Rules in Italian
Universities // В сб. трудов
Международной научно-практической конференции
«Образовательные измерения в
информационном обществе».
Киев, 26–29 мая 2010 г.

4

Суворов Л.Н.
Материалистическая диалектика.
М.: Мысль, 1980. 264 с.

5

Spearman Charles.
«General Intelligence»
objectively determined
and measured. *American
Journal of Psychology*
1904. 5: 201–293;
Spearman C. The proof
and measurement of association
between two things. *American
Journal of Psychology*, 1904. 15,
72–101.



общего (general) интеллектуального фактора, который влияет на многочисленные реальные проявления интеллекта, например, на различные задания соответствующего теста. На языке педагогических и психологических измерений понятие «фактор» часто рассматривается как эквивалентное понятию «латентное качество».

Спирман стал разработчиком метода факторного анализа, нацеленного на выявление общего интеллектуального фактора. Он исходил из плодотворной гипотезы: если вариацию по этому общему фактору статистически элиминировать из общей вариации результатов испытуемых, полученной после сложения баллов по всем заданиям теста, то ответы испытуемых одного уровня подготовленности на задания теста должны стать статистически независимыми.

Логика Спирмана была основана на идее метода, используемого при расчёте частных коэффициентов корреляции: определения меры связи результатов одного теста с другим тестом, без влияния результатов третьего, четвёртого и других тестов. Истинный уровень интеллектуального развития личности рассматривался как главная причина, вызывающая различия испытуемых по реально наблюдаемым тестовым баллам.

Постановка проблемы

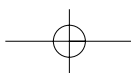
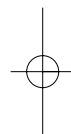
Любая сфера человеческой деятельности представляет собой совокупность деятельности лиц разного уровня понимания и подготовленности. Проблема педагогических измерений латентных качеств не является в этом смысле исключением.

В российской и зарубежной литературе встречаются тексты, в которых говорится об измерении латентных «переменных величин», без указания на педагогическую сущность измеряемого качества или свойства, без проведения их теоретического анализа. Такие тексты тяготеют к общему математико-статистическому стилю исследования педагогических проблем, а потому они оказываются мало приемлемыми. Нужны педагогический стиль, язык, предметы, объекты и методы исследования данной проблемы. Эти вопросы не исследованы с достаточной полнотой и ясностью. Данная статья — попытка восполнить сложившийся дефицит публикаций по указанной проблеме.

Положение осложняется тем, что измерение латентных качеств по-разному воспринимается в России, странах Востока и Запада. В российской педагогике эта проблема пока не оценивается как заслуживающая внимания, о чём может свидетельствовать фактическое

Теория

43/0000



ПЕД	
	измерения

отсутствие каких-либо текстов по вопросам исследования латентных качеств в учебниках педагогики и во многих иных основных трудах педагогической направленности.

На уровне обыденного сознания делаются попытки оценивания латентных качеств, которые оканчиваются словесными или численными оценками, содержащими в себе существенные погрешности. Можно сказать, что оценивание возникает там, где нет возможности или умения измерить проявления интересующих свойств личности.

Нередко говорится об оценивании как деятельности, эквивалентной педагогическим измерениям. И на разработку таких систем оценивания тратятся большие деньги. Но эти расходы оказываются практически бесполезными. В России примером такого рода ошибочно задуманной и исполняемой государством системы является «ОСОКО». Достаточно заметить, что это — не система педагогических измерений латентных качеств, и что центральной опорой этой системы является неизменно Единый государственный экзамен, имеющий очень высокую погрешность измерения.

Фактическая отстранённость подавляющего большинства российских авторов от исследования вопросов теории ла-

тентных качеств, в сочетании с ошибочными министерскими опытами по ЕГЭ и ГИФО, привела к подмене реальной проблематики педагогических измерений своеобразными суррогатами, потребными, по всей видимости, только для проведения некачественных единых государственных экзаменов. Эта подмена ведёт российское образование в тупик.

Решение проблемы измерения латентных качеств личности позволяет существенно улучшить качество педагогических тестов. Несмотря на это существует много причин, из-за которых измерение латентных качеств не принимает массовый характер.

Одна из таких причин — психологическая. Зарубежным психологом Andrea diSessa было введено в научный оборот понятие «феноменологические примитивы». Под этим понимаются утверждения, основанные на домыслах⁶. Н.И. Браун и R.J. Mislevy приводят примеры таковых: «Тяжёлые предметы падают на землю быстрее лёгких», «Тест измеряет то, о чём говорится в его названии»⁷, «Интеллект — это то, что измеряется интеллектуальным тестом» и т.п.

Российская практика функционирования ЕГЭ даёт немало утверждений сходного толка, не получивших научного подтверждения. Например, что

6

Braun Henry I.,
Mislevy Robert J.
Intuitive test theory. CSE
Report 631, Los Angeles,
CA, May, 2004.

7

Там же.

«Единый государственный экзамен улучшает качество образования в стране», «Единый государственный экзамен эффективен⁸», «КИМы ЕГЭ — это тесты», что «ЕГЭ — это тоже тесты», а также утверждения: «ЕГЭ (КИМы) — это метод педагогических измерений», «КИМы — это методы педагогических измерений», «Федеральный институт педагогических измерений разрабатывает методы педагогических измерений», «КИМаи ЕГЭ можно одновременно и качественно измерить уровень подготовленности слабых выпускников школ и абитуриентов престижных вузов» и т.п.

Относительно последнего примитива можно определённо сказать, что одновременно измерить то и другое можно, но качественно это сделать посредством бланковых экзаменационных форм — нельзя. Отрицательный опыт проведения ЕГЭ вплоть до 2009 г., многократно подтверждал этот чисто теоретический тезис. Игнорирование теории мстит неразумной практике. Неслучайно, начиная с 2009 года, КИМы ЕГЭ фактически разделились: часть стала использоваться для аттестации выпускников школ, другая, большая часть — для приёма в вузы. Как и предсказывалось в одной из статей автора, ЕГЭ в прежнем едином, противоречащем теории виде так и не дожил до конца 2009 года⁹.

В литературе много раз говорилось о «тестах» ЕГЭ», без какого-либо упоминания о теории латентных качеств. Это определённо указывает на то, что т.н. контрольно-измерительные материалы (КИМы) ЕГЭ создавались и продолжают создаваться вне учёта положений данной теории. Между тем нацеленность на измерение латентного свойства личности является весомым признаком научности используемых методов, их отличия¹⁰ от псевдонаучных и ненаучных методов. Что должно подтверждаться всей совокупностью обязательно публикуемых эмпирических фактов. Но этого нет.

Очень важно видеть разницу между латентным качеством и переменной величиной. Причём различие это не столько лексическое, сколько сущностное. Не случайно с появлением Item Response Theory¹¹ сама теория педагогических измерений всё в большей мере стала превращаться в теорию измерения латентных качеств. Вследствие чего педагогические измерения, проводимые вне латентного контекста, начинают фактически принимать формы ненаучной деятельности. Этот факт позволяет лучше понять цель и смысл измерения латентных качеств личности, а также понять важность данного признака как критерия демаркации педагогических измерений от псевдоизмерений¹².

Теория

8

Аванесов В.С.
Спорное решение Кремлёвской комиссии по ЕГЭ. <http://obrazovanie.viperson.ru/wind.php?ID=618286&soch=1>

9

См. *Аванесов В.С.*
Доживёт ли Единый государственный экзамен до 2009 года? <http://testolog.narod.ru>

10

Аванесов В.С.
Проблема демаркации педагогических измерений Measurement // Педагогические измерения. 2009. № 3. С. 3–37.

11

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

Судя по отчётам Федерального института педагогических измерений, КИМы ЕГЭ являются методом педагогических измерений, и метрических недостатков в них нет. На самом же деле анализ имеющихся фактов указывает на прямо противоположную ситуацию¹³. Например, в отчётах нет даже упоминания о латентности оцениваемых институтом качеств личности. Об измерениях в современном понимании говорить в таких случаях нет никаких оснований¹⁴.

В последнее время появились элементы риторического признания в том, что «ЕГЭ далёк от совершенства», без конкретного указания — что и как надо переделывать, чтобы ЕГЭ наконец хотя бы на один шаг приблизился к мифическому «совершенству». Похоже, что 2010 году КИМы освободились негласно от претензий на обладание свойствами тестов, как это было все прошлые годы бюрократического «эксперимента».

Это означает, что, возможно, начался и процесс размежевания государственных экзаменов от государственного тестирования. Независимым и качественным государственным тестированием не было никогда, и быть таковым не может. Вместе с тем замена тестоподобных заданий задачами государственных экзаменов означает факти-

ческое прекращение попыток измерения латентных качеств личности и свойств заданий. Это может служить началом процесса явного разрыва ЕГЭ с подлинной теоретической основой педагогических измерений. В общем, началось движение вспять.

Измерение интересующих латентных качеств личности кардинально зависит от качества теста и от точности измерения основных свойств тестовых заданий. Игнорирование этой зависимости ведёт к имитации педагогических измерений, примером которых, как было показано ранее, являются т.н. «КИМы ЕГЭ».

Этической нормой для тестологов является прежде всего выявление свойств заданий и обоснование качества используемого метода измерения. Только после этого возникает моральное право обсуждения результатов измерения латентных свойств личности и принятия решений относительно приёма в вуз и т.п.

Отсутствие теоретической определённости латентных качеств отрезает всякую надежду на их качественное измерение. Можно утверждать, что в этом смысле всякое упоминание термина «педагогические измерения» вне контекста измерения именно латентных качеств не имеет научного смысла.

12

Аванесов В.С.
Проблема демаркации педагогических измерений // Педагогические измерения. № 3, 2009. С. 3–37.

13

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

14

Там же

Процесс научно-педагогического измерения латентных качеств личности требует не только теоретизации, но и аргументации. Важно определить ведущие понятия, проверить логическую правильность названия измеряемого качества, определить предмет измерения, а также системы индикаторов, понятийных и эмпирических, указывающих на наличие или отсутствие интересующего качества¹⁵. Требуется аксиоматика и математические формализмы, выбор подходящей модели и стандартизация условий измерения. И, наконец, полученные результаты подлежат аргументированной интерпретации с точки зрения их точности и соответствия ясно сформулированной цели.

Все без исключения теории педагогических измерений направлены на оценку истинных компонентов измерений латентных качеств, явно не наблюдаемых. И все теории исходят из аксиомы устойчивого существования истинных компонентов измерения. Этот процесс наблюдался не только в педагогических, но и в психологических, социологических, экономических и медицинских измерениях. В силу приведённых аргументов название Latent Trait Theory надо отнести к теории более высокого, междисциплинарного уровня, включающей в себя все остальные теории педа-

гогических, психологических и социологических измерений.

Соответственно, в наше время сложилось несколько методов измерения латентных качеств личности. Среди них — математическая теория педагогических измерений (Item Response Theory¹⁶), метрическая система Rasch Measurement¹⁷ и другие. Эти теории имеют существенные различия.

Mary E. Lunz и B.D. Wright подчёркивают важность получения осмысленных результатов измерений с позиций именно теории измерения латентных качеств. Они считают, что результаты должны позволять делать широкий индуктивный вывод, выходящий за пределы конкретного набора заданий, испытуемых, экспертов¹⁸. По сути, эти авторы ставят вопросы генерализации получаемых результатов измерений, о необходимости иметь качественные методы, по которым принимаются решения о приёме в вузы или при профотборе. Они справедливо утверждают, что наиболее эффективно эти цели достижимы при использовании моделей измерения, основанных на модели измерения латентных качеств¹⁹.

Исследование латентных качеств — предмет научного процесса педагогических измерений. Практические работники и управленцы далеки от идей измерения латентных качеств,

Теория

15

Об этом процессе подробнее см., например, на стр. 87–105 книги: *Аванесов В.С.* Тесты в социологическом исследовании. М.: Наука, 1982. 199 с.

16

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

17

Аванесов В.С. Метрическая система Rasch Measurement // Педагогические измерения. 2010. № 2, С. 3–36.

18

Lunz Mary E., Benjamin D. Wright. Latent Trait Models for Performance Examinations. Ch.6. in the book: Applications of Latent Trait and Latent Class Models in the Social Sciences. Edited by Jurgen Rost and Rolf Langeheine. «The purpose of a performance

ПЕД
измерения

examination is to infer candidate abilities that go beyond the particular sample of tasks, items, and judges encountered».

19
Whether the goal is to make reproducible pass/fail decisions or to position candidates according to demonstrated ability, the performance examination must measure candidate ability consistently. This is most efficiently accomplished by using a latent trait model... Там же. Р. 80.

20
Аванесов В.С.
Проблема демаркации педагогических измерений // Педагогические измерения. 2009. № 3, С. 3–37.

21
Ковешникова Ю.Н.
Улучшение качества контрольных работ посредством расчёта статистических характеристик заданий // Педагогические измерения. 2010. № 3.

но они тем не менее пытаются руководить этим процессом. Результаты такого руководства могут быть только отрицательными. Вероятно, поэтому они занимаются в России не педагогическими измерениями, а государственными экзаменами — главной до настоящего времени формой государственной имитации педагогических измерений²⁰.

Важно отметить, что латентные качества исследуются не одной, а всеми существующими научными теориями педагогических измерений. Такая позиция позволяет выработать интегрированную и сбалансированную позицию по отношению к положительным сторонам и реальным возможностям каждой теории. Периодически появляющиеся новые теории обычно не отвергают, а дополняют одна другую; они различаются не по признаку мнимой «современности», а по этапам их применения, по лексике, моделям и методам измерения, а также по прогнозируемости индивидуальных результатов и релевантности поставленным целям.

Измерение уровня подготовленности личности включает преимущественно знания и способности. В отрыве от таких качеств суждения об измерении латентных переменных величин становятся неосмысленными. Да сами величины есть результат концептуализации и

следующей за этим остальной части процесса измерения. Таким образом, распространённое словосочетание «измерение латентных величин» полезно заменить на «измерение латентных качеств личности и формализуемых свойств заданий».

В последние годы к качествам личности стали часто относиться и компетенции. Это спорно с метрической точки зрения, поскольку сама идея проверки компетентности является прагматической, основана на логике проверки соответствия интересующего свойства личности заданному стандарту минимуму. Для этого не обязательно создавать методы педагогического измерения. Хватает и традиционных контрольных работ²¹.

При оценивании компетентности измерений фактически нет. Достаточно установить соответствие оценки достигнутого уровня заранее установленной границе. При массовой оценке компетентности фактически исчезает необходимость рассматривать интересующее свойство личности как континуальное и измеримое на значительном интервале. Достаточно получить некоторые значения, позволяющие решить, достигнут или не достигнут граничный уровень. Иначе процесс оценивания становится слишком дорогим, как в нынешнем ЕГЭ.

Можно отметить и философский аспект проблемы изме-

рения латентных качеств личности. Как отмечается в литературе, в наше время типичными чертами мышления и поведения европейского человека являются солипсизм, прагматизм, агностицизм. Он признает реальным лишь то, что осязает, что может доказать, установить, чем может овладеть и т.п. Для такого стиля мышления характерно ограничение феноменами (в кантовском смысле), тем, что «является». Такой стиль мышления не способен постичь целое, проникнуть в сущность бытия, увидеть перспективу²².

Человек такого типа мышления не способен осмыслить идеи существования латентных качеств, а потому готов отвергнуть любые попытки измерения таких качеств. Для него не имеет смысла заниматься тем, что реально не видно или не ощутимо.

Положение дел на Западе даёт иную картину. С первых дней возникновения идей измерения интеллекта, знаний, мотивов и мировоззренческих ценностей сознания там была принята концепция латентности этих качеств, что означает невозможность их прямого измерения, а равно и нежелательность субъективного оценивания. Проблема, следовательно, заключалась в том, что нужно было разрабатывать качественные методы, которые на основе эмпирических индикаторов

могли бы свидетельствовать о количестве интересующего латентного качества у каждого испытуемого. В России проблема такого рода ранее не формулировались.

Подмена теории измерений латентных качеств личности т.н. «измерением латентных величин» приводит к дефициту метрической и педагогической составляющих подлинных педагогических измерений, когда требуется использование двух языков:

- первый язык, педагогики, нужен на уровне формулирования гипотез, концептуализации, содержания интересующего качества, формы и заданий проектируемого теста. Понятие «латентная величина» здесь не уместно в принципе. Оно слишком общее, абстрактное, непедagogическое;
- второй, формальный язык, логично использовать лишь на этапе математико-статистической обработки и интерпретации данных, когда возникает необходимость использования языка педагогических измерений, математики и статистики.

Итак, решение проблемы измерения латентных качеств личности сопряжено с необходимостью научного определения понятий, выделения основных направлений исследовательской деятельности и формирования двух (по меньшей мере) языков исследования

*Бессонов Б.Н.,
Ващекин Н.П.,
Тихонов М.Ю.,
Урсул А.Д.*
Образование – прорыв
в XXI век. Монография.
М. Изд-во МГУК, 1998.
242 с.

этих качеств, а также демаркации педагогических измерений от элементарной оценочной деятельности и от суррогатов измерений типа КИМов ЕГЭ.

Основные направления решения проблемы измерения латентных качеств личности и свойств заданий

Из множества направлений изучения латентных качеств личности в данной статье выделяются четыре.

Первое направление связано с разработкой понятийного аппарата. Сюда входят определения латентного качества и всех основных понятий теории измерения латентных качеств.

Второе направление определяет содержание педагогической теории измерения латентных качеств личности и свойств заданий. В мировой науке это, пожалуй, наименее исследованная часть проблемы.

Третье направление отражает точку зрения Ч. Спирмана, считавшего, что если вариация результатов по нескольким тестам зависит от одного общего латентного фактора, то статистическое устранение влияния такого фактора должно привести к идее статистической независимости результатов этих тестов.

Факторный анализ позволяет уменьшить размерность

векторного пространства, редуцировать вариацию результатов испытуемых по множеству заданий к вариации результатов на одной, наиболее важной латентной переменной величине, которая получает название «фактор». Фактор как осмысленную латентную переменную величину образуют только те задания или тесты, которые коррелируют между собой.

В рамках этого направления формулируются аксиомы педагогических измерений. Особое значение имеют две основные аксиомы — локальной независимости результатов испытуемых, одного уровня подготовленности, на одно задание, в зависимости от ответов на любое другое задание.

Четвёртое направление рассмотрения проблемы латентных качеств относится к анализу форм связи между уровнем развития интересующего латентного качества и вероятностью правильного ответа на задания теста. Это направление стало главной темой диссертации Ф. Лорда²³. Он нашёл, что применительно к каждому отдельному заданию эта связь имеет вид т.н. логистической функции²⁴, представленной на рис. 1, из учебного пособия Ф. Бейкера²⁵. Вероятность правильного ответа испытуемых на задание теста есть непрерывная возрастающая, нелинейная гладкая функция.

23

Lord F.M.
A theory of test scores.
New York: Psychometric
Monograph 7.
Psychometric Society.
1952.

24

Lord F.M.
The relation of test score
to the trait underlying the
test. Educational and
Psychological
Measurement. 1953. 13,
517–548.

25

Baker F.B.
The basics of Item
Response Theory. Second
edition. Published by the
ERIC Clearinghouse on
Assessment and
Evaluation. University of
Wisconsin. ERIC
Clearinghouse on
Assessment and
Evaluation. 2001. 176 pp.

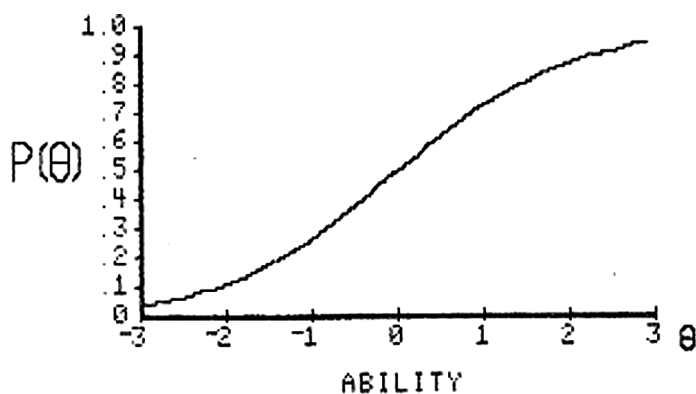


Рис. 1.

Чем выше уровень подготовленности, тем большей должна быть вероятность правильного ответа.

В зарубежной психометрической литературе сложилась традиция называть прямую линию (ось абсцисс) континуумом, на котором теоретически можно представить уровни способности любого, даже мифического испытуемого. Поскольку на Западе подлинный предмет измерений — не столько уровень знаний, сколько уровень способности к овладению знаниями, то ось абсцисс там традиционно называется ability.

Ось абсцисс предполагается континуальной, т.е. непрерывно протяжённой от минус бесконечности до плюс бесконечности. Поскольку примерно 99 процентов результатов всех испытуемых располагаются на шкале в пределах от -3 логита до +3 логита, где логит — стандар-

тизованная единица измерения, то ось абсцисс этого графика нередко представляется на протяжении данного отрезка. Хотя фактические пределы вариации результатов испытуемых в мировой практике педагогических и психологических измерений встречаются в пределах ± 5 логитов.

Все четыре упомянутых выше направления исследований представлены в кратком варианте в настоящей статье.

О дефектах КИМов ЕГЭ

После десяти лет бюрократических экспериментов с созданием КИМов ЕГЭ недавно было признано наконец что ЕГЭ «далёк от совершенства». Однако ЕГЭ далёк не от «совершенства», а от элементарной пригодности для используемых целей. Только в КИМах ЕГЭ

ПЕД
измерения

26

Анализ результатов государственной итоговой аттестации выпускников 2009 года общеобразовательных учреждений Томской области. Информационно-аналитический отчет и методические рекомендации // Под общей ред. П.И. Горлова. Томск. Дельтаплан. 2009. 340 с.

Аванесов В.С.
Педагогическое измерение латентных качеств // Педагогическая диагностика. 2003. № 4. С. 69–78.

28

Unmeasured variables, factors, unobserved variables, constructs, or true scores are just a few of the terms that researchers use to refer to variables in the model that are not present in the data set.

Bollen K.A.
Latent variables in psychology and the social sciences. Annual. Rev. Psychol. 2002. 53: 605–34, p. 607.

встречаются невероятные разрывы в расположении заданий на континууме трудности. То и другое является важным метрическим показателем дефективности разрабатываемых в этом институте методов педагогических измерений.

Именно на это указывают реальные данные третьего, после Вологодского и Кемеровского, обнаруженного правдивого отчёта по проведению ЕГЭ (математика) в 2009 году²⁶.

Таблица 6 на стр. 28 этого отчёта прямо указывает на несоответствие использованного набора заданий критерию равномерности возрастания трудности, на существенные разрывы в уровнях трудности использованных заданий, подтверждающие русскую поговорку: где пусто, где густо. Отличников, получающих сто баллов, оценивают лишь одним заданием, испытуемых среднего уровня подготовленности — всего лишь двумя заданиями. Зато много лёгких заданий. Так легко и просто создаётся видимость приемлемых баллов и успехов образовательной деятельности.

Появление такого рода данных, дискредитирующих российские государственные экзамены, может оказаться главной причиной продолжающегося незаконного засекречивания распределения реальных статистических результатов по всем

остальным субъектам РФ и в отчётах самого ФИПИ. В сложившихся условиях для изменения ситуации сокрытия абсолютно несекретных статистических распределений по КИМам ЕГЭ достаточно было бы указания В.В. Путина. Но до него научная критическая информация, видимо, не доходит.

Другой недопустимый приём, используемый в ЕГЭ — минимально допустимые критерийные значения аттестации испытуемых определяются не до начала экзамена, а после его проведения. Так не делается нигде, потому что это нарушает этику проведения педагогического контроля.

Определение латентного качества

Латентными называются положительные и отрицательные качества личности, которые явно не видимы, а потому не поддаются прямому, непосредственному измерению²⁷. В зарубежной литературе используются и такие названия, как ненаблюдаемые переменные величины, факторы, конструкты или истинные значения тестовых результатов, используемые для обозначения теоретических показателей, влияние которых изучается методами математико-статистического моделирования²⁸.

В педагогических измерениях наблюдаемо не само интересующее качество личности, а признаки его проявления, такие, как умение правильно решить задачу, дать точное название, правильно ответить на задание теста. Такие проявления называются эмпирическими индикаторами²⁹.

Поэтому нередко различия между баллами испытуемых по латентному качеству рассматривают как причину, вызывающую отличия между испытуемыми по реально наблюдаемым тестовым баллам. По внешним количественным проявлениям исследователь оценивает значения латентного качества у каждого испытуемого. Все наблюдаемые ментальные проявления личности считаются индикаторами латентного качества. Но это не совсем верное предположение, потому что на результаты влияет не только латентное качество, но и качество метода измерения, время тестирования и условия проведения теста.

Педагогическая латентная величина, в отличие от некоторых физических величин, существует не априори. Она возникает в процессе измерения, вначале как конструкт интересующего качества, затем как шкала. Идея и методы измерения латентных качеств реализуются в тесной зависимости от эмпирических результатов.

Латентное качество — это, во-первых, концепция, а не переменная величина. Процесс выделения и именованя латентного качества называется концептуализацией, и это подтверждает важность начальной организации теоретической части процесса измерений. J. Loevinger проводит такое интересное различие между исследуемым качеством личности и психологическим образом (представлением, конструктом) этого качества: качества присущи людям, конструкты существуют в сознании исследователей и в научных психологических журналах³⁰.

Во-вторых, наблюдаемы не сами интересующее качества личности или свойства заданий, а признаки их проявления, такие, как умение правильно решить задачу, дать точное название, правильно ответить на задание теста, мера трудности заданий. Такие проявления называются эмпирическими индикаторами. Таким образом, латентное качество представляется (репрезентируется) системой эмпирических индикаторов, ответы на которые могут указывать на наличие признаков интересующего качества у каждого испытуемого. Полезно добавить, что идея латентности не опровергается подавляющим большинством исследователей³¹.

В-третьих, для обоснования существования латентного

Теория

29

29
Аванесов В.С.
Вопросы методологии педагогических измерений // Педагогические измерения. 2005. № 1.

30

«Traits exist in people; constructs (here usually about traits) exist in the minds and magazines of psychologists.»
Loevinger J. 1957.

Objective tests as instruments of psychological theory. Psychol. Rep. 3: 1–18.

31

Исключение представляет известный зарубежный психолог бихевиористского направления *B.F. Skinner*, 1976. About Behaviorism. New York: Vintage Books.

качества необходимы правила соотнесения эмпирических фактов с концептуальными положениями. Нужны также правила интерпретации эмпирических результатов. Это делается в факторном анализе — в исторически первом методе выявления латентных качеств на основе множества эмпирических показателей.

В-четвёртых, латентные переменные величины обычно моделируют интересующее качество с помощью математических моделей IRT. R.B. Cattell полагал, что латентное качество является источником (source trait) для внешне наблюдаемого свойства (surface trait)³².

32

Cattell R.B.
Personality. N-Y,
McGraw-Hill, 1950.

33

Sydenham P.H. (Ed).
Handbook of
Measurement Science.
J. Wiley & Sons.
Chichester, 1982. v.1.

34

Аванесов В.С.
Метрическая система
измерения Rasch
Measurement // Педаго-
гические измерения.
2010. № 2.

35

Аванесов В.С.
Понятие и методы мате-
матической теории педа-
гогических измерений
(Item Response Theory):
статья третья. / Педаго-
гические измерения.
2009. №4. С. 5.

Латентное качество измеряемо

Педагогические измерения не имеют права на существование вне системы научно обоснованных понятий теории латентных качеств.

Измерение. Р.Н. Sydenham определяет измерение как эмпирический процесс объективного количественного представления свойств интересующих объектов для их научной характеристики³³. В этом определении утверждение о сугубо эмпирическом процессе измерения является очень спорным, если не сказать неприемлемым, хотя бы потому, что всякий эф-

фективный метод опирается на теорию.

Можно сказать, что процесс педагогического измерения столь же эмпиричен, сколь и теоретичен. Необходимы концептуализация, определение понятия интересующего качества, проведение теоретического анализа предмета измерения, операционализация основного понятия, что связано с представлением в виде системы эмпирических индикаторов (тестовых заданий). Далее анализируется содержание метода измерения, теоретическое обоснование валидности используемых эмпирических индикаторов измеряемого качества.

Понятие «Педагогические измерения» ранее было сформулировано³⁴ как процесс определения меры интересующего латентного качества личности испытуемого на *интервальной шкале* посредством теста, состоящего из *системы* заданий равномерно возрастающей трудности, позволяющего получать педагогически целесообразные результаты, отвечающие критериям *надёжности, валидности, объективности и эффективности*. В этом довольно длинном и схоластическом определении курсивом выделены основные термины, позволяющие отграничить педагогические измерения от прочих методов — научных, псевдонаучных и ненаучных³⁵.

Педагогическое измерение можно также определить как процесс отображения уровней проявления интересующего латентного свойства испытуемых в числовую шкалу. Желательно, чтобы процесс отображения давал значения тестовых баллов в интервальной шкале. Это означает шкалу с равными интервалами, у которой есть нулевое значение, единица измерения, правила измерения и правила интерпретации результатов. В педагогических измерениях требованиям интервальной шкалы отвечают тестовые баллы, получаемые по модели G. Rasch³⁶.

Педагогические измерения проводятся на основе различных моделей.

Модель измерения в литературе определяется как структурное построение, позволяющее соединить латентные переменные с наблюдаемыми переменными³⁷. Модель измерения является статистической моделью, связывающей наблюдаемые индикаторы (обычно задания теста) с одним или несколькими латентными качествами личности³⁸.

Модель измерения, основанная только на содержательном анализе заданий и на элементарных оценках, можно назвать педагогическим *оцениванием*. Более широким аналогом этого понятия в западной литературе является понятие «педа-

гогическое оценивание» (educational evaluation), которое можно рассматривать как одну из начальных форм педагогического измерения, в котором нет средства измерения (теста), но уже есть подобие числовой шкалы, на которой могут градуироваться, обычно с долей субъективности, результаты испытуемых. Различия между оцениваем и педагогическим измерением — предмет специальной статьи.

Величина. Латентная величина возникает не произвольно или как данное естественным образом, извне, а как результат определённого оперирования индикаторами интересующего качества. Латентная величина возникает вначале как идея, затем как концепция, выражаемая содержанием теста, и только затем как результат измерения.

Настоящие педагогические измерения — это измерения не «латентных величин», а латентных качеств личности или свойств интересующих педагогике объектов. Сама теория педагогических измерений есть не что иное, как теория измерения именно этих качеств и свойств.

Такое понимание величины имеет место не только в гуманитарных науках, но и в технике, и даже в математике. Например, известная из школьного курса математическая величина «дискриминант квадратного уравнения» возникает как результат

Rasch G.
Probabilistic Models for Some Intelligence and Attainment Tests. With a Foreword and Afteward by B.D. Wright. The Univ. of Chicago Press. Chicago & London, 1980. 199 p.

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

A latent variable model is a statistical model that relates a set of variables (so-called manifest variables) to a set of latent variables. См. Latent variable model. Wikipedia, the free encyclopedia.

ПЕД	
	измерения

целенаправленного поиска показателей числа корней квадратного уравнения, а затем и вычисления корней по определённой формуле, если корни есть. У квадратных уравнений дискриминант сам по себе явно не выражен, он принимает различающиеся значения в разных уравнениях, что даёт основания называть эту величину переменной, зависимой от значений коэффициентов и свободного члена уравнений.

Каждое интересное свойство личности можно назвать величиной, если наблюдаемые объекты по этому свойству могут принимать различные значения, желательно на интервальной шкале. Сущность, подтверждаемая подходящей концепцией и термином, а также наблюдаемая дисперсия значений объектов являются самыми важными признаками переменной величины.

Переменной величиной называется такая, которая в условиях изучаемого процесса меняет своё значение. Переменная величина y называется функцией переменной x , если каждому допустимому значению аргумента x соответствует определённое значение y . В процессе измерения переменной величины объекты часто принимают различающиеся значения.

Постоянной величиной называется такая, которая при изучении какого-либо процес-

са не меняет своего значения. Например, сумма внутренних углов треугольника, отношение длины окружности к диаметру.

Операциональные определения понятий обеспечивают правильный переход от теоретического уровня исследования к эмпирическому уровню. Основными элементами перехода выступают понятийные индикаторы различного уровня общности. Так, основному понятию ставится в логическое соответствие некоторое небольшое (обычно не более пяти-шести) число частных понятий, становящихся понятийными индикаторами первой ступени.

Затем этим индикаторам ставятся в соответствие ещё более частные понятия, которые становятся индикаторами второй и далее ступени, вплоть до эмпирических индикаторов, соответствие которых основному понятию проверяется эмпирически. Применительно к педагогическим измерениям в качестве эмпирических индикаторов обычно выступают задания, подобранные с целью проверить знания по какой-либо учебной дисциплине.

Шкалирование определяет-ся как та часть процесса измерения, в которой исходные значения тестовых баллов испытуемых и меры трудности заданий переводятся в стандартные меры с общей средней арифмети-

ческой, общей единицей вариации данных и общей единицей измерения. В Rasch Measurement и IRT результатом шкалирования является расположение объектов на шкале логитов.

Процесс измерения латентных качеств охватывает испытуемых, разработчиков тестов и тестовых заданий, а также тех, кто применяет тесты и интерпретирует результаты. Для большинства педагогов самым интересующим свойством (или качеством) личности является уровень подготовленности учащихся (студентов) по одной или нескольким учебным дисциплинам. Большинство известных зарубежных педагогических тестов — SAT, GRE, LSAT — созданы по модели одномерного шкалирования результатов³⁹.

Содержание педагогической теории измерения латентных качеств личности и свойств заданий

Ранее были определены три главные составляющие педагогической теории измерений. Это понятийный аппарат, содержание теста и форма тестовых заданий⁴⁰. Отдельно были рассмотрены вопросы организации процесса педагогических

измерений, а также практического применения тестов (test administration)⁴¹. Исходя из идеи эквивалентности теории педагогических измерений и теории измерения латентных качеств личности, компоненты этих теорий следует принимать одинаковыми.

В данной статье дополнительно исследуются вопросы выбора объектов педагогических измерений, подходы к измерению свойств заданий, продолжена разработка понятийного аппарата, дана классификация общих и частных теорий. Рассмотрим эти вопросы по порядку.

Помимо упомянутой ранее аксиомы локальной независимости при исследовании латентных качеств личности важное значение имеет ещё одно положение IRT: это свойство монотонности графика каждого задания теста. Монотонное возрастание предполагает обязательным такое свойство функции (графика), которое ставит в соответствие уровень подготовленности испытуемых с вероятностью правильного ответа на задание; чем выше подготовленность, тем большей должна быть вероятность правильного ответа.

Одномерность — это свойство шкалы, образуемой тестом. Если тест измеряет одно качество, его называют одномерным (гомогенным). Если тест измеряет несколько качеств — многомерным (гетерогенным).

Latent Trait Analysis and Item Response Theory (IRT) Models. Wikipedia <http://ourworld.com-puserve.com/home-pages/jsuebersax/lta.htm>

Аванесов В.С. Методологические и теоретические основы тестового контроля. Дис... докт. пед. наук. СПб, Госуниверситет, 1994. 339 с.

Там же.

42

Но не «и» выпускников школ, как это много лет пытались сделать чиновники Минобрнауки в Едином государственном экзамене. Именно в этом был настоящий смысл т.н. «единого» экзамена. Этот смысл был бесславно похоронен в предсказанный срок. См. *Аванесов В.С.* Доживёт ли Единый государственный экзамен до 2009 года? <http://testolog.narod.ru>. Теперь в аббревиатуре ЕГЭ фактически остались две последние буквы.

43

Borsboom Denny, Mellenbergh, Gideon J., van Heerden, Jaap. The Theoretical Status of Latent Variables. University of Amsterdam. *Psychological Review?* 2003. Vol. 110, № 2, 203–219.

44

Аванесов В.С. Композиция тестовых заданий. Учебная книга. 3 изд., доп. М., Центр тестирования, 2002. 240 с.

Объекты педагогических измерений

Можно выделить несколько обобщённых объектов педагогических измерений.

Первый такой объект — это латентные качества личности.

Примерами являются «уровень подготовленности студентов или⁴² выпускников школ», «знание учебной дисциплины», «способность понимать» и многое другое. Примеры латентных переменных в психологии — «интеллектуальное развитие», «Я-концепция», «Тревожность»; в социологии — «Бесправие», «Национальная предубеждённость».

Понятие «качество» иногда отождествляется с понятием «свойство». Например, в психологии часто утверждается о качествах *или* свойствах личности.

D. Borsboom, G.J. Mellenbergh и J. van Heerden приводят пример высказывания: «Если бы Альберт Эйнштейн не имел выдающийся интеллект, то он не смог бы сформулировать свой знаменитый закон $e = mc^2$. Эти авторы поставили вопрос о смысле данного высказывания и о зависимости между значением на латентной переменной величины «уровень интеллекта» и эмпирически существующими фактами проявления интеллекта Альберта

Эйнштейна в форме данного закона.

Ответы на это высказывание позволяют соотнести наблюдаемые ответы испытуемых относительно закона Эйнштейна ($e = mc^2$) с ненаблюдаемым непосредственно высоким уровнем развития интеллекта выдающегося учёного. При этом уровень подготовленности по ненаблюдаемому качеству личности играет решающую роль в результатах испытуемых по данному заданию⁴³. Связь между латентным качеством и наблюдаемыми результатами описывается графиком нелинейной, т.н. логистической функции, называемым по-английски item characteristic curve.

Примерами второго обобщённого педагогического объекта измерения являются, например, такие формальные свойства заданий, как уровень трудности отдельных заданий и теста в целом, а также уровень дифференцирующей способности заданий. Количественными показателями этой способности являются меры вариации баллов испытуемых, корреляции с критерием и др. Подробнее об этом можно прочитать в ранее опубликованных работах автора и других исследователей⁴⁴.

Существуют три принципиально различающихся подхода к изучению свойств педагогических заданий.

Первый подход — содержательный. Он проявляется при выборе содержания заданий проектируемого теста, при аргументации уместности использования того или иного фрагмента учебных знаний, при содержательной валидации результатов тестирования. Важно подчеркнуть, что для разработки теста этот содержательный подход является необходимым, но недостаточным. Важно понимать, что это — не единственно необходимый подход. Обязательно нужны и другие.

Второй подход — эмпирический и статистический. Он позволяет проверить каждое задание эмпирически, на выборке подходящих испытуемых (target group), оценить уровень трудности и дифференцирующей способности заданий посредством применения статистических показателей.

При втором подходе существенную роль играют статистические оценки погрешности измерений, которые разработчики тестов обязаны обнаруживать. Эти обязательства вытекают из этического кодекса тестологов США и многих стран мира. Но в России такого кодекса нет. Нет, соответственно, и тестов, тестовой технологии, нет и должного развития педагогических измерений. Но зато есть КИМы ЕГЭ и перманентная закрытость главных показателей их некачественности. Такого ро-

да закрытость общественно важной статистики уже давно вошла в противоречие с Законом РФ.

Настоящие педагогические измерения должны иметь опубликованные оценки погрешности измерений, общие и дифференцированные. Отсутствие упомянутых оценок есть верный признак недостаточного качества полученных результатов.

Третий подход представляет собой комбинацию содержательного и теоретико-вероятностного подходов. Выдвигается идея латентного качества, подбирается его точное название, создаётся система заданий в тестовой форме, проверяются их тестовые свойства. После чего выбирается математическая модель, связывающая уровень подготовленности испытуемых по интересующему латентному качеству личности с вероятностью правильного ответа.

В предыдущих статьях уже приводились три основные модели педагогических измерений, а потому приводить их здесь нет необходимости. Как пишут Wim J. van der Linden и Ronald K. Hambleton, главная особенность математической теории педагогических измерений (IRT), используемой при данном подходе, — это определение математической функции, которая ставит вероятность правильного ответа

ПЕД
измерения

испытуемого в соответствие от уровня его подготовленности⁴⁵.

Полезно выделить ещё одну группу специфических объектов, вполне подпадающих под название метаобъектов педагогических измерений. Это критерии эффективности и качества, которыми руководствуются разработчики тестов. Они названы здесь метаобъектами по причине их вторичности. Необходимость сверяться с критериями качества и эффективности возникает в процессе исследования свойств заданий и качеств испытуемых, т.е. после начала процесса измерения латентных качеств личности и формальных свойств заданий. Отсутствие аргументированной информации по этим критериям — ещё один верный признак отсутствия подлинных измерений⁴⁶.

Два предположения теории педагогических измерений латентных качеств

В рассматриваемой теории важное место занимают два положения (assumptions). Первое называется *аксиомой локальной независимости результатов*, которую можно сформулировать так: для испытуемых одного уровня подготовленности вероятность правильного ответа на одно задание не должна зависеть от вероятности правильно-

го ответа по любому другому заданию теста. Важно добавить, что эта аксиома применима не ко всем заданиям вообще, а только к заданиям теста.

Другое важное положение, подлежащее обязательной проверке, — *одномерность* (unidimensionality) тестовых результатов, в соответствии с которой все задания теста создаются с прицелом на измерение только одного интересующего качества личности. После выделения латентного фактора все остаточные корреляции между заданиями теста заметно снижаются. Чем больше такое снижение, тем приемлемее становится гипотеза одномерности (гомогенности) теста.

Если вероятность правильного ответа на одно задание зависит от вероятности правильного ответа на другое задание, то надо признать, что такие задания обладают т.н. цепным эффектом. Теория разработки таких заданий, очень полезных для обучения, но непригодных для качественного тестирования, а также практические примеры цепных заданий представлены в работах автора этой статьи⁴⁷ и ряда авторов журнала «Педагогические измерения». Цепные задания обладают высоким обучающим потенциалом, однако в России этот потенциал продолжает пока оставаться официально незамеченным и неиспользуемым.

45

Wim J. van der Linden,
Ronald K.
Hambleton.
Handbook of modern item
response Theory. 1997
Springer-Verlag, New
York, Inc. p. v.

46

Производство т.н.
КИМов ЕГЭ и имею-
щаяся по ним информа-
ция однозначно указы-
вают на отсутствие в них
признаков педагогичес-
ких измерений.

47

Аванесов В.С.
Форма тестовых зада-
ний. 2 изд. М. Центр тес-
тирования. 2005. 155 стр.

Алгоритм измерения латентных качеств личности

1. Формулируются цели и задачи.
2. Определяется и обсуждается предмет измерения. Обычно это интересующее качество личности. Точное название качества даёт впоследствии название переменной величине.
3. Исследуется содержание преподаваемого курса, что служит содержательной основой для разработки заданий в тестовой форме.
4. Пишутся задания в тестовой форме.
5. Проводится апробация заданий в тестовой форме в целевой группе испытуемых для тестовых заданий, пригодных для создания теста.
6. Включённая в тест система заданий рассматривается как система эмпирических индикаторов измеряемого качества. Именно такая система тестовых заданий позволяет совершить переход от словесного описания интересующего качества личности к созданию переменной величины. У каждого задания проверяются системные статистические свойства, показатели средней тенденции и показатели вариации, определяются параметры графика, шкалирование заданий по уровню их трудности.
7. Аналогично определяются параметры графиков зависимо-

сти вероятности правильного ответов испытуемого от уровня трудности заданий. Цель такой работы — позиционирование испытуемых на шкале уровня подготовленности, порождаемой тестом.

8. Проверяются профили ответов испытуемых на задания теста, определяется мера правильности каждого профиля.

9. Проводится факторный анализ результатов тестирования для выявления одномерной шкалы тестовых результатов и для выбраковки заданий, не образующих латентную переменную величину.

Для гомогенного теста существует неформальное правило, разделяемое практически всеми разработчиками тестов. Это правило можно истолковать так: при разработке теста все задания должны измерять одно интересующее качество личности, но задания должны различаться между собой по содержанию и по формальным свойствам⁴⁸. Главные различия — по элементам общего содержания одной и той же учебной дисциплины и по уровню трудности.

Г. Смирнова⁴⁹ выделяет такие этапы:

- удаление экстремальных заданий;
- проведение дистракторного анализа;
- проверка способности тестовых заданий образовать тест как

Теория

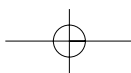
48/49

48

На английском языке это правило выражается так: In test construction, the guiding principle is «all items must be about the same thing, but then be as different as possible».

49

Смирнова Г.И. Применение программы RUMM–2020 для разработки педагогического теста // Педагогические измерения. 2010. № 3.



ПЕД	
	измерения

систему заданий возрастающей трудности (у неё это называется проверкой совместимости заданий теста);

- построение графических образов заданий теста и их интерпретация;
- проверка равномерности возрастания уровня трудности заданий теста;

- проверка соответствия уровня трудности теста уровню подготовленности испытуемых.

Если делаются попытки измерения одним тестом двух и более качеств личности, то тест должен быть *гетерогенным*. Технология разработки таких тестов существенно усложняется, а потому эти вопросы здесь не затрагивались.

