

Основы теории педагогических заданий

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

В СТАТЬЕ ИЗЛАГАЮТСЯ ОСНОВЫ ТЕОРИИ ПЕДАГОГИЧЕСКИХ ЗАДАНИЙ, РАЗРАБАТЫВАЕМОЙ В ЛОГИКЕ ПРОЕКТА ОБЩЕЙ ТЕОРИИ ПЕДАГОГИЧЕСКИХ ИЗМЕРЕНИЙ[1]. ПРЕДСТАВЛЕНЫ ЭЛЕМЕНТЫ ИСТОРИИ ВОЗНИКНОВЕНИЯ ЗАДАНИЙ И ЛОГИКИ ИХ АНАЛИЗА, СФОРМУЛИРОВАНА СИСТЕМА ИСХОДНЫХ ОПРЕДЕЛЕНИЙ, ОПРЕДЕЛЕНЫ СУБЪЕКТИВНЫЕ, ИНТЕРСУБЪЕКТИВНЫЕ И ОБЪЕКТИВНЫЕ СВОЙСТВА ПЕДАГОГИЧЕСКИХ ЗАДАНИЙ. НА ОСНОВЕ МАТРИЦЫ ИСХОДНЫХ ТЕСТОВЫХ РЕЗУЛЬТАТОВ ВВЕДЕНЫ ФУНКЦИИ ВЕРОЯТНОСТИ ПРАВИЛЬНОГО ОТВЕТА НА ЗАДАНИЯ ТЕСТА, ПОНЯТИЯ И ПРИМЕРЫ ГРАФИЧЕСКОГО ОБРАЗА ЗАДАНИЯ ПО ЛЮБОЙ УЧЕБНОЙ ДИСЦИПЛИНЕ.

Ключевые слова: педагогическое задание, теория, качество и свойства заданий, измерение качества заданий.

Элементы истории педагогических заданий

В цикле работ по истории тестов отмечалось, что педагогические задания пришли к нам из глубины веков¹. Самый первый из сохранившихся источников — учебник XVII в. до н.э., содержащий 84 задания по математике. Он был написан на куске древнего папируса длиной более пяти метров². Сейчас этот исторический памятник древнего педагогического творчества хранится в Британском музее.

В конце XIX века н.э. в российской печати обсуждался вопрос о введении экзаменов. В циркуляре по Санкт-Петербургскому учебному округу «О производстве испытаний зрелости без послаблений», сообщалось, что во многих гимназиях испытания зрелости производятся крайне снисходительно, с весьма большими послаблениями. А это обстоятельство, в свою очередь, весьма неблагоприятно отзывалось на всём ходе учебного дела, приучая воспитанников по-

верхностно относиться к усвоению преподаваемых им наук, в надежде на ожидаемые послабления на испытаниях. В этом же циркуляре определён главный

принцип определения содержания контрольных заданий: «не обременяя требованиями мелких и второстепенных подробностей... необходимо удостовериться в знании существенно-важного в каждом предмете»³.

Первая попытка научного обоснования качества заданий была сделана в начале XX века. Это произошло во Франции. Авторы первого теста, созданного для измерения интеллектуальных способностей детей, А. Бине и Т. Симон провели эмпирическую проверку заданий, которые предполагалось включить в их тест. Для оценки пригодности заданий авторы использовали два основных критерия:

- 1) эмпирическую меру трудности каждого задания, которую они определяли по доле или по проценту правильных ответов в группах детей разного возраста.
- 2) информацию о степени совпадения результатов теста с мнением преподавателей.

Для достижения сопоставимости данных, полученных ими, с данными других исследователей и минимизации ошибок измерения, А. Бине и Т. Симон прилагали к тесту стандартную инструкцию по проведению тестирования. Интересен метод выявления дифференцирующей способности заданий, которым пользовались эти авторы. Результаты ответов испытуемых на каждое задание представлялись в виде точек

¹ См. сайт автора <http://testolog.mail.ru>

² Барр С. Россыпи головоломок / Пер. с англ. 3 изд. М.: Мир, 1987. С. 36.

³ Циркуляр по Санкт-Петербургскому учебному округу. 1901. № 5. С. 218.

на плоскости, где по оси абсцисс откладывались значения возраста, а по оси ординат — доля правильных ответов в каждой возрастной группе. Усредняя полученные точки и затем соединяя их линией, они получали графический образ (ломаную линию), на основании которого делали выводы об адекватности задания для того или иного возраста⁴. Позже этим же методом успешно воспользовалась М. Ричардсон⁵.

Исходные определения теории педагогических заданий

Научное определение понятия есть форма концентрированного выражения знаний. Знания концентрируются постепенно, по мере выявления системных свойств изучаемого предмета. Условием эффективности создаваемых понятий становится их адекватность сущности предмета и возникновение понятийной системы.

В этой работе продолжаются попытки создать систему понятий педагогической теории измерений⁶. Ключевое место в ней занимает авторская концепция трёх главных понятий: задания в тестовой форме, тестовое задание и педагогический тест⁷ как системы заданий возрастающей трудности. Общее для всех перечисленных понятий — понятие «педагогическое задание».

Ключевое место в педагогической теории измерений должна занять теория педагогических заданий.

Задание — это педагогическая форма, нацеленная на достижение учащимися запланированных результатов. Каждое задание создано для определённой цели, а потому, можно сказать иначе, имеет свою миссию и свои характеристики. Главная миссия заданий — побуждать к активной самообразовательной деятельности, овладению знаниями, умениями и навыками до требуемого уровня компетенции. Многократно мною утверждалось, что нельзя одну и ту же систему заданий использовать для разных целей и особенно для испытуемых разных уровней подготовленности, так как это повышает погрешность и снижает качество измерений⁸.

Задания могут создаваться для слабых или сильных учащихся, для интеллектуально одарённых или менее способных, для самостоятельной работы и компьютерной самопроверки уровня подготовленности, для аттестации или профессионального отбора. В зависимости от цели некоторые характеристики заданий могут меняться, в особенности по содержанию и по уровню трудности.

Задания выполняют синергетическую функцию объединения усилий педагога и обучаемого. Метафорически задание можно представить как мост между учеником и педагогом, открывающим возможности для их взаимодействия, объединяющим их усилия в учебном и воспитательном процессе. Если придерживаться и далее упомянутой метафоры, то качество и эффективность учебного процесса существенно зависят от качества моста. Чем лучше задания, тем (при прочих равных условиях) прочнее и весомее могут быть результаты взаимодействия ученика и педагога.

Лекционные и прочие формы педагогической деятельности представляют только одну, преподавательскую сторону педагогического процесса. Выполнение заданий учащимися (студентами) — вторая, не менее, а часто и более важная сторона педагогического процесса. Основные виды заданий — задача, вопрос, упражнение, творческое задание, задание в тестовой форме, тестовое задание, учебная проблема, курсовая и дипломная работа.

В наши дни, наряду с задачами и развивающими вопросами, заметную роль приобрели задания в тестовой форме. Обучение без заданий не эффективно, а нередко и вредно, потому что отучает от мышления и от самостоятельности в учебной работе. Отсутствие или недостаток опыта самостоятельного решения учебных задач переносится и на неумение выпускника школы или вуза успешно решать затем и жизненные задачи. В подлинной образовательной деятельности ведущая роль отводится не только педагогу, но и заданиям. А потому в педагогике не случайно говорят о задачном подходе к организации обучения. Как писал один из сторонников этого подхода Г.С. Костюк, понять новый учебный объект — значит решить какую-то, пусть маленькую, познавательную задачу⁹.

С этой точки зрения сложившаяся практика планирования, учёта и оценки педагогического труда по количеству проведённых

⁴ Подробнее см.: **Аванесов В.С.** Методологические и теоретические основы тестового педагогического контроля. Дисс... докт. пед. наук. С-Пб. гос. ун-т. 1994.

⁵ **Richardson M.W.** Notes on the Rationale of Item Analysis // *Psychometrika*, 1936, 1: 169–76.

⁶ См. ранее опубликованные работы на эту тему: **Аванесов В.С.** Основы педагогической теории измерений // *Педагогические измерения*. 2004. № 1. С. 15–21, 2004. № 1–2, 2005. № 1–4.

⁷ Там же.

⁸ Именно это сейчас происходит на Едином государственном экзамене.

⁹ **Костюк Г.С.** Избранные психологические труды. М. 1988; Цит. по кн.: **Балл Г.А.** Теория учебных задач. М.: Педагогика, 1990. С. 73.

лекций или уроков, без реального учёта количества и качества выполненных учащимися заданий, представляет собой анахронизм, мешающий улучшить образовательную деятельность.

Педагогическое задание можно также определить как средство развития, обучения и воспитания, способствующее развитию личности, повышению качества знаний, а также повышению эффективности педагогического труда. Развитие личности требует разработки системы интеллектуальных, предметно-дисциплинарных (языковых, математических, физических и т.п.), эстетических, трудовых и прочих заданий, отвечающих требованиям новых образовательных технологий. За единицу учебного времени такие задания дают наибольший прирост знаний, умений и навыков у наибольшего количества учащихся. Именно в вопросах производства и применения эффективных заданий российское образование стало сильно отставать от общемировых тенденций. Это одно из самых слабых мест российской педагогики и педагогической науки. Отсюда — актуальность проблемы измерения эффективности и качества педагогических заданий.

Измерение качества педагогических заданий означает процесс выявления их существенных свойств и отображение уровня проявления каждого свойства численными эквивалентами, адекватными количеству проявляемого свойства. Оценки обычно выражаются с помощью качественных и сравнительных понятий. Примеры качественных понятий — лёгкое или трудное задание, отлично подготовленный испытуемый, слабо знающий студент, отличный ответ и т.п. Специфическим условием измерений в общественных науках является формирование выборки испытуемых, обладающих, по предположению, интересующим свойством. Соответственно в матрицу результатов измерения включаются только те испытуемые, у кого такое свойство есть, хотя бы в самом минимальном количестве.

Эффективное задание позволяет учащимся за единицу учебного времени добиться большего эффекта, сравнительно с другими педагогическими заданиями: быстрее усвоить новые элементы знаний, правильнее их применять, улучшить воспроизведение знаний, повысить точность изложения.

Теория педагогических заданий призвана создать логически непротиворечивую систему научного знания, дающую целостный взгляд на существенные свойства заданий, независимо от конкретной учебной дисциплины и от уровня обучения. Следовательно, это может быть об-

щая прикладная педагогическая теория педагогической науки.

Предмет этой теории составляет:

- система определений;
- объективные и субъективные свойства (характеристики) заданий;
- формы заданий, общие для всех учебных дисциплин;
- инвариантные принципы отбора содержания заданий, независимо от учебных дисциплин;
- вопросы измерения эффективности и качества самих заданий.

Краткая классификация заданий

В зависимости от цели применения все педагогические задания можно разделить на несколько классов.

Для обучения используются упражнения, задачи, вопросы, задания в тестовой форме.

Для контроля и самоконтроля знаний умений и навыков используются зачётные, экзаменационные и тестовые задания, задачи, вопросы, а также выпускные учебные работы.

Для создания гомогенного теста используется система тестовых заданий равномерно возрастающей трудности, имеющая общее содержание, приемлемую корреляционную и факторную структуру.

Аттестационные задания применяются для аттестации выпускников школ, средних профессиональных заведений и вузов. Трудность аттестационных заданий устанавливается на таком уровне, который точно соответствует квоте аттестуемых выпускников. В России сейчас положительно аттестуется свыше 99% выпускников школ. Следовательно, уровень трудности аттестационных заданий заведомо устанавливается чрезвычайно низким. А там, где он формально оказывается выше, систематически используются различные формы искажения результатов, ради достижения принятой квоты успешно окончивших школу. Расхожее среди чиновников название «аттестационный тест» имеет только внешнее отношение к тесту, где некоторые задания иногда представляются в тестовой форме. В таких «тестах» содержание всех заданий преимущественно лёгкое, а потому тестов там, по определению, нет.

Для приёма в вузы нужно использовать множество тестов, различающихся по содержанию и по уровню трудности, в зависимости от требований самого вуза. Также неоднократно мною отмечалось, что единый подход к комплектова-

нию студенческого контингента множества различных вузов вреден и опасен своими отрицательными последствиями. Проблема профессионального отбора и классификации абитуриентов по направлениям подготовки одним методом не решается: ЕГЭ и КИМы эту проблему не только не решают, но существенно ухудшают условия нормальной работы вузов.

Задания для оценки профессиональной компетентности называются *сертификационными*. Сертификационные тесты (если они действительно будут созданы) предполагаются различающимися по трудности. Испытуемый сертифицируется в зависимости от содержания и от уровня трудности успешно выполняемых заданий.

Принципы теории педагогических заданий

Взаимосвязанные принципы контроля знаний и педагогических измерений уже излагались в работах автора¹⁰. Напомним, что это были принципы объективности, справедливости и гласности, научности и эффективности, систематичности и всесторонности, связи контроля с образованием, обучением и воспитанием. Опираясь на эти общие принципы, можно попытаться сформулировать принципы частной теории педагогических заданий.

Принцип обновления заданий. Хотя некоторым педагогическим заданиям уже тысячи лет, массовое применение тестовых методов и новых образовательных технологий требует специально организованного процесса создания, творческого обновления, улучшения и постоянной замены заданий. По сути, нужно ставить вопрос о создании своеобразной культуры *непрерывного изменения* и замены заданий в тестовых системах, открытия новых самокупаемых методических центров, способных качественно производить и эффективно применять тестовые задания в практике. Это предмет специальной научной методологии, теории, методики и практики педагогических измерений.

Учитывая важную роль принципов в деятельности педагогов, это предложение можно сформулировать как принцип *обновления заданий*, применение которого позволит получить измерения повышенной точности и преодолеть пропасть, возникшую между практикой тестирования и наукой о педагогических измерениях.

Принцип соответствия уровня трудности заданий уровню подготовленности испытуемых. Задание называется соответственным

(адекватным) для испытуемого, если испытуемый может его решить с какой-либо вероятностью успеха (больше нуля). Наиболее адекватны для адаптивного тестирования задания с примерно 50% вероятностью удачного решения. Абсолютно не решаемое задание не адекватно уровню подготовленности испытуемых. В процессе обработки данных тестовые матрицы часто приводят к пригодному для композиции теста виду. Это означает, что из матрицы удаляют все задания, не соответствующие испытуемым по уровню их подготовленности, а потому не подходящие для создания теста.

Отсюда следует важный принцип педагогического измерения: уровни трудности заданий теста должны соответствовать уровням подготовленности испытуемых. Этот принцип обладает симметрией: уровни подготовленности испытуемых также должны соответствовать уровням трудности заданий теста. Вот почему его можно кратко назвать *принципом соответствия*. Нарушение этого принципа повышает ошибки измерения. Чем больше несоответствие, тем ниже точность педагогических измерений. Принцип соответствия удачно дополняет систему принципов, сформулированную ранее¹¹.

Не соответствующими бывают не только некоторые задания, но и, как это ни покажется странным, некоторые испытуемые. Это те, чья подготовка выходит за пределы измеряемого уровня подготовленности. Везде можно найти таких «обучающихся», которые, фактически не учась, только делают вид, что учатся¹². Вследствие этого они оказываются не подготовленными к тестированию. По хорошему тесту, состоящему

¹⁰ **Аванесов В.С. и др.** Основы педагогики и психологии высшей школы / Под ред. А.В. Петровского. М.: МГУ, 1986; **Аванесов В.С.** Основы научной организации педагогического контроля в высшей школе. М.: МИСиС, 1989; **Аванесов В.С.** Принципы педагогического контроля. <http://testolog.narod.ru/Theory44.html>

¹¹ **Аванесов В.С.** Научные проблемы тестового контроля знаний. М.: Иссл. центр, 1994. С. 72–73; **Аванесов В.С.** Проблема качества педагогических измерений // Педагогические измерения. 2004. № 2. С. 9–11.

¹² Каждый 10-й школьник г. Череповца — прогульщик или двоечник и заканчивает своё весьма среднее образование без аттестата. Некоторые подростки прогуливают занятия, начиная с 1-го класса. Причём: это дети не только из трудных семей: многие из них имеют вполне благополучных и состоятельных родителей. По мнению Татьяны Скидановой, за последние 10 лет в обществе сформировалось мнение, что образование не настолько важно, что можно и без него найти себе место в жизни: на рынке, торговой базе, челночить, т.е. найти вид деятельности, не требующий образования. И сейчас идёт не только материальное расслоение общества, но и духовное: есть люди, желающие получить престижное образование и сделать карьеру, и есть те, кто считает, что всё это бесполезно и не нужно. <http://www.5ballov.ru/news/newsline/index.shtml?2006/06/05/52867>

из заданий с несколькими правильными ответами, они получают нули¹³, а это и есть показатель их несоответствия содержанию, и, возможно, и цели теста. В процессе тестирования выясняется, что тест создан не для них¹⁴.

Не соответствующими для данного теста оказываются также и те испытуемые, кто на все задания отвечает правильно, но в отведённое для всех время мог бы решить, возможно, и другие задания, если бы таковые были в тесте¹⁵. Для них нужен другой тест, повышенного уровня трудности.

Принцип технологичности педагогических заданий. В российском образовании используются задания преимущественно в нетестовой форме. Большинство из них — вопросы, задачи и упражнения, которые в тестовой культуре не используются. Нередко вопросы и задачи называются тестами. Но ответы на них иногда бывают столь многословными, что для выявления истинности требуются большие затраты интеллектуальной энергии, в то время, как технологичная методика тестирования предполагает чёткую и быструю дифференцируемость правильного ответа от неправильного. В этом смысле традиционные вопросы и ответы не технологичны; их не рекомендуется включать в тест. То же относится и к задачам, имеющим громоздкие формулировки.

Технологичность заданий определяется как вариант композиции, которая позволяет вести

процесс тестирования с помощью технических средств, и делать это точно, быстро, экономично и объективно. Задания становятся технологичными, если их содержание правильно и быстро понимается испытуемыми, и если форма заданий способствует процессу компьютеризации тестирования.

Главные препятствия для достижения технологичности тестового задания — это обилие слов, потребность в устных разъяснениях, расплывчатость и неопределённость формы, плохая формулировка содержания, отсутствие необходимой техники и программ. Использование автоматизированного обучения и контроля на основе достижений новой педагогики, педагогических измерений, психологии, кибернетики и компьютерной техники, взятых в разумном соотношении, образует основу именно того, что сейчас называют педагогической технологией.

В этой технологии самым узким местом оказалось неумение делать задания, без чего нельзя сейчас ни объективно проверить знания, ни создать современную автоматизированную контрольно-обучающую программу, ни наладить такую форму организации учебного процесса, как дистанционное обучение. В идеальном случае учебная программа, каждый её модуль сопровождаются заданиями в тестовой форме. Однако путь к достижению этого идеала лежит через трудности создания качественных тестовых заданий. Технологичность заданий обеспечивается правильностью формы заданий и корректностью их содержания. С точки зрения содержания тестовое задание оценивается по конкретности и абстрактности, глубине, обобщённости и полноте содержания.

Задания становятся технологичными, если их содержание точно и быстро понимается испытуемыми, и если форма заданий способствует процессу компьютеризации тестирования. Именно использование автоматизированного обучения и контроля на основе достижений новой педагогики, педагогических измерений, психологии, кибернетики и компьютерной техники, взятое в разумном соотношении, образует основу того, что сейчас называют педагогической технологией.

Принцип соизмеримости шкал для оценки испытуемых и заданий. Для проведения педагогических измерений очень полезной оказалась идея L.L. Thurstone, а вслед за ним и G. Rasch, логарифмического преобразования исходных шкал трудности заданий и уровня подготовленности испытуемых. В итоге получается одна общая логарифмическая шкала, в которой

¹³ В ряде областей РФ по КИМах — ЕГЭ часть выпускников получают нули, хотя они могли хотя бы что-то угадать. Раньше я думал, что это следствие неудачного сочетания факторов незнания и невезения при угадывании. На самом же деле это результат нежелания этой не лучшей части выпускников подвергаться Единой Государственной Экзекуции, в условиях гарантированной тройки, выставляемой им на основе Положения о проведении ЕГЭ, утверждённого Минюстом РФ. Я называю таких реальных двоечников «троечниками в законе», число которых тоже растёт, однако и этот результат Гособнадзором незаконно засекречен. См. **Аванесов В.С.** Таким ЕГЭ двоечники очень довольны. http://www.gazeta.ru/education/2006/04/05_e_580108.shtml

¹⁴ Если тестируются студенты, я называю их «не студенты». Доброжелательные оппоненты отмечают, что у них в кармане студенческий билет, а поэтому они всё-таки студенты. «Формально — да, а по результатам тестирования — нет».

¹⁵ Такие испытуемые оказываются, образно говоря, «недомеренными». Но это дефект не испытуемых, а разработчиков. С каждым годом число таких «недомеренных» в КИМах-ЕГЭ растёт, что является ещё одним засекреченным показателем неизбежной «инфляции» ЕГЭ. При общем ухудшении качества образования в стране ежегодный прирост «баллов ЕГЭ-КИМов» по некоторым территориям превышает сотни, а иногда и тысячи процентов. В то же время, действительный потенциал самых талантливых выпускников школ недомерен. Это фактические метрические дефекты «КИМов-ЕГЭ».

стало удобно сравнивать испытуемых и задания, что открывает дорогу к адаптивному обучению и адаптивному тестовому контролю. Результаты такого логарифмирования представлены в таблице 1 этой статьи.

Принцип содержательной и логической правильности заданий. В педагогической практике заметное распространение получают различные формы заданий и задач, нарушающие данный принцип. В основном это задачи с неправильными условиями.

Пример из работы Я.И. Груденова. В треугольнике АЕК угол $A = 62^\circ$, $E = 75^\circ$, $K = 53^\circ$. Вычислить внешние углы треугольника¹⁶. Между тем, в евклидовой геометрии не может существовать треугольник, сумма углов которого не равна 180° . Есть задачи, принципиально не решаемые, например: «Вычислить сторону прямоугольника, если его площадь равна 435 м^2 ». Узнать длину стороны по площади невозможно. Некоторые задания даже имеют своего рода идеологическое обоснование¹⁷.

Логическое обоснование качества педагогических заданий

Логика определяется как наука о формах правильного мышления. Форма мысли — это то, что остаётся после отвлечения от содержания мысли. Одна из важнейших задач логики — критика мышления и его результатов. Основная задача логики — вскрывать алогизмы, учить давать точные определения, отделять правильно сформулированные утверждения от неправильно сформулированных. В тестовом процессе логика помогает правильно формулировать задания, находить в них формальные ошибки, делать задания понятными для испытуемых.

В теории и методике педагогических измерений основными объектами логической рефлексии являются система определений этой теории, логические принципы разработки заданий и ответов к заданиям.

Логической основой формы задания с выбором одного правильного ответа из нескольких предлагаемых является закон исключённого третьего (впервые сформулированный Аристотелем). Выбор правильного ответа даёт истинное суждение, а выбор неправильного — ложное суждение. Третьего не дано. Из этого закона следует методическое правило: в каждом задании с выбором одного ответа правильный ответ должен быть, что придаёт однозначность замыслу самого задания и не допускает противоречивых толкований у испытуемых.

Огромное количество примеров алогичных заданий дают производители так называемых «КИМов для ЕГЭ». И это «дело» поставлено на поток. Посмотрим алогичный пример, вместе с принятой там пунктуацией.

Обвести кружком номер правильного ответа:

1. ВОССТАНИЕ СПАРТАКА БЫЛО

- а) в 73 г. до н.э.
- б) в 73 г. н.э.
- в) в 1973 г.
- г) пока ещё не было¹⁸.

Ранее мною был сформулирован принцип логической однородности подбора дистракторов в задании. Пример нарушения общей логики задания и этого принципа можно увидеть в задании:

2. ПРИЕХАВ В СТОЛИЦУ

- 1) всех поражает шум и суета;
- 2) у вас не будет времени сходить в музей;
- 3) идите сразу на Красную площадь;
- 4) у меня было прекрасное настроение¹⁹.

Задание в тестовой форме полезно рассматривать как высказывание. Высказывание — грамматически правильное предложение, взятое вместе с выражаемым им смыслом. Смысл, выражаемый грамматически правильным предложением, называется суждением. Тестовое задание формулируется из точных терминов и никогда не содержит метафоры, лишние слова и лишние знаки. Пример метафорично сформулированного задания в открытой форме:

3. ОТЦОМ РУССКОЙ ФИЗИОЛОГИИ СЧИТАЕТСЯ _____.

В логике выделяется четыре основных свойства правильного мышления: *определённость, правильность, непротиворечивость, обоснованность*. Применительно к тестовой теории и практике эти общие свойства правильного мышления приобретают функции специфических регулятивов тестовой деятельности, вследствие чего они приобретают значение принципов. Рассмотрим их подробнее.

Определённость содержания теста. Оп-

¹⁶ Психологическое обоснование целесообразности широкого использования задач, не имеющих решений. // Новые исследования в педагогических науках. 1964. Вып. 2.

¹⁷ Чуканцов С.М. Для подготовки к практической, трудовой деятельности: «важно, чтобы ещё на школьной скамье ученик получил правильное представление о том, что не всякая задача и не при любых условиях может быть решена. О задачах на реализованные ситуации с ложными данными // Математика в школе. 1977. № 2.

¹⁸ Российская газета. 2001, 23 марта.

¹⁹ Педагогическая диагностика. 2004. № 2. С. 88.

ределённость содержания теста образует предмет педагогического измерения. В случае *гомогенного* теста возникает вопрос об уверенности в том, что все задания теста проверяют знания именно по определённой учебной дисциплине, а не по какой-то другой. Например, в физических расчётах используется немало математических знаний и потому в систему физического знания обычно включается та математика, которая используется при решении физических задач. Неудача в математических расчётах порождает неудачу при ответах на задания физического теста. Отрицательный балл ставится, соответственно, за незнание физики, хотя испытуемый допустил ошибки математического толка.

Если в тест включено много таких заданий, которые для правильного решения требуют не столько физических знаний, сколько умений выполнять усложнённые расчёты, то это может быть примером неточно определённого содержания теста по физике. Чем меньше пересечения знаний одной учебной дисциплины со знаниями другой, тем определённее выражается в тесте содержание каждой учебной дисциплины. Во всяком тестовом задании заранее определяется, что однозначно считается ответом на задание, с какой степенью полноты должен быть правильный ответ.

В гетерогенном тесте это достигается при явном выделении заданий одной учебной дисциплины в отдельную шкалу. При этом нередко встречаются задания, хорошо работающие не только на одну, но и на две, три и даже на большее число шкал.

Задания по математике для обычной средней школы перенасыщены интеллектуальным компонентом. Этот компонент затрудняет освоение математических знаний и операций, что составляет суть этой учебной дисциплины, и отвращает от математики миллионы детей, недостаточно подготовленных для решения задач с интеллектуальным компонентом.

Включение большого числа заданий, имеющих повышенное интеллектуальное содержание, в учебники (якобы из благих побуждений: дети ведь должны учиться мыслить — кто с этой демагогией может спорить!) на самом деле уже много лет фактически убивает математическое образование в стране.

Между тем интеллектуальное развитие детей — творческая задача сама по себе. Она не решается механическим включением трудных заданий в учебники или примитивным принуждением. Интеллектуальный компонент нужно вклю-

чать в задания только для тех, кто может и хочет их решать. Отсюда становится понятной важная проблема культурной организации измеряемого уровня образования, которую не следует путать с ошибочным, чиновно внедряемым профильным образованием учащихся средней общеобразовательной школы.

Логическая правильность. Немалую роль в формировании теста как системы играет требование логической правильности заданий, включённых в тест. Логическая правильность в формулировании тестовых заданий достигается при соблюдении условия соразмерности объёма определяющего понятия объёму определяемого. Известные в литературе примеры (2):

4. ЧЕТЫРЕХУГОЛЬНИК, У КОТОРОГО ВСЕ СТОРОНЫ РАВНЫ, НАЗЫВАЕТСЯ _____.

5. ПРЯМОУГОЛЬНИК, У КОТОРОГО ВСЕ СТОРОНЫ РАВНЫ, НАЗЫВАЕТСЯ _____.

Если на четвёртое задание даётся ответ «квадрат», допускается несоразмерность: объём определяющего понятия — «четырёхугольник» — больше объёма определяемого понятия «квадрат».

Ещё пример задания с ромбом:

Обвести кружком номера всех правильных ответов:

6. РОМБ — ЭТО

- 1) квадрат
- 2) четырёхугольник
- 3) прямоугольник
- 4) параллелограмм
- 5) параллелограмм с равными сторонами

Непротиворечивость содержания заданий.

Непротиворечивость содержания заданий требует, чтобы относительно одной и той же мысли не возникали суждения, одновременно утверждающие и отрицающие её. Недопустимо существование двух исключаящих ответов на одно и то же задание теста.

Если испытуемому даётся инструкция: «Обведите кружком номер правильного ответа», а затем в одном из ответов утверждается, что правильного ответа нет, либо все ответы правильные, то это пример алогичного мышления.

В заданиях с выбором одного или нескольких ответов иногда встречаются ответы, вообще не связанные с содержанием задания. Такие ответы довольно легко распознаются испытуемыми как ошибочные, и потому весь тест оказывается неэффективным. Для повышения эффек-

тивности дистракторов²⁰ вместе с заданием проходят апробацию на типичной выборке испытуемых. И если обнаружатся такие ответы к заданиям, которые испытуемые вообще не выбирают, то они удаляются, как не выполняющие свою функцию — дистрактора, призванного отвлечь внимание незнающих испытуемых от правильного ответа. Неэффективные дистракторы вредны для теста; так как они снижают точность измерений.

Обоснованность. Обоснованность содержания тестовых заданий означает наличие у них оснований истинности. Обоснованность связана с аргументами, которые могут быть приведены в пользу той или другой формулировки заданий теста.

При отсутствии доказательных аргументов в пользу правильности сформулированного задания оно в тест не включается, ни под каким предлогом. То же происходит, если при экспертном обсуждении возникает хотя бы один контраргумент или допускается условие, при котором это утверждение может оказаться двусмысленным или ложным.

Идея обоснованности содержания теста тесно переплетается с принципом *содержательной правильности тестовых заданий*. В тест включается только то содержание учебной дисциплины, которое является объективно истинным и поддается некоторой рациональной аргументации. Соответственно, спорные точки зрения, вполне приемлемые в науке, не рекомендуются включать в содержание тестовых заданий.

Неточно или двусмысленно сформулированные задания, порождающие несколько правильных или условно правильных ответов, тоже не включаются в тест. Отсюда возникает необходимость вводить дополнительные условия истинности, что удлиняет само задание и усложняет его семантику.

Неистинность содержания тестовых заданий отличается от некорректности их формулировки. Неистинность, как отмечалось выше, определяется соответствующим ответом, в то время как некорректно сформулированное задание может продуцировать ответы как правильные, так и неправильные, а то и вызывать недоумение. Некорректность формулировки обычно выясняется в процессе обсуждения содержания заданий с опытными педагогами-экспертами.

Матрица для анализа качества педагогических заданий

Для выявления интерсубъективных и объективных свойств заданий используются матрицы тестовых результатов²¹.

Матрицей называется математическая форма упорядочения и хранения информации, имеющей некоторую общность содержания. Применительно к тестовой технологии матрица является формой организации, сохранения, представления и обработки данных. Тестовые результаты мыслятся в форме матриц. По строкам матриц всегда располагаются фамилии испытуемых или их номера, по столбцам — номера заданий. Вопрос: а почему нельзя иначе, разве в математике не всё равно, что писать по строкам, а что — по столбцам? В математике всё равно, а в тестовой литературе матрицы принято располагать именно так²². Такого рода конвенция удобна для коммуникации при интерпретации исходных, промежуточных и конечных результатов.

Исходная тестовая матрица обычно имеет вид *прямоугольной* матрицы. Из соображений достоверности статистических выводов, число (N) испытуемых (строк) всегда должно быть больше числа (m) заданий (столбцов); чем больше отношение $\frac{N}{m}$, тем выводы о характеристиках испытуемых более обоснованы. Часто ставится вопрос о минимально допустимом соотношении этих двух чисел для обоснования качества тестовых заданий. Хорошо, если число строк в десять раз превышает количество заданий. Минимально допустимое соотношение числа испытуемых к числу заданий — не менее пяти.

Элементы тестовых матриц — это числа, отражающие оценки каждого испытуемого, полученные ими по каждому заданию. Если обозначить символом i номер испытуемого, а j — номер задания, то общий элемент матрицы x_{ij} трансформируется в уникальное, получая номер строки и номер столбца. Этим номерам ставятся в соответствие получаемые баллы. Например, если первый по счёту испытуемый ответил неправильно по восьмому заданию

²⁰ Это неправильные ответы в заданиях с выбором одного или нескольких правильных ответов, отвлекающие внимание незнающих испытуемых от правильных ответов.

²¹ О матрицах см., например: Педагогические измерения: Тезаурус // Педагогические измерения. 2005. № 1. С. 28–32; **Аванесов В.С.** Введение в статистические и математические методы педагогических измерений // Педагогические измерения. 2005. № 4. С. 106–115.

²² В книге «Matrix Algebra for Social Scientists» (1963) классик психометрики Р. Horst называл такое расположение natural order of a matrix.

теста, то элемент с именем x_{i8} принимается равным нулю. Это записывается так: $x_{i8} = 0$.

G. Rasch рассматривал все значения тестовых матриц, как итог противоборства каждого испытуемого i , с предлагаемыми ему заданиями (j). Исход каждого противоборства оценивается баллом x_{ij} . Значение балла зависит от соотношения уровня подготовленности тестируемого и уровня трудности задания, при условии заранее принятого правила (конвенции): что считать «победой» испытуемого или задания. В упрощённом подходе допускаются только два исхода. Первый — правильное решение, что означает победу испытуемого; он получает один балл на пересечении номера своей строки и номера задания. Вторым исходом — ошибочное выполнение задания. Тогда в том же пересечении записывается нуль баллов. Каждая строка представляет результаты одного испытуемого, каждый столбец — результаты ответов множества испытуемых по одному только заданию данного номера.

Матрицы исходных тестовых результатов обычно готовятся в электронных таблицах «Excel»²³. Для работ вузовского или школьного уровня

«Excel» удобен тем, что это часть стандартного компьютерного офиса, а потому практически доступен. Кроме того, данные этих таблиц принимают все, или почти все профессионально сделанные пакеты многомерной статистики²⁴, что сильно облегчает работу над созданием качественного теста.

Исходя из дидактических соображений здесь используется ранее приводившийся пример небольшой учебной матрицы тестовых результатов, в которой всего 13 испытуемых и 10 заданий. Что делает саму матрицу и операции с её элементами обозримыми на одной странице. Поскольку

это дидактический пример, вопрос о достоверности выборочных статистик в таких случаях не ставится.

Таблица 1 представляет собой расширенный вариант ранее использованного примера²⁵. Для того, чтобы понять смысл последующей части работы, здесь придётся воспроизвести небольшую часть понятийного и формального аппарата. Напомним, что первый столбец таблицы представляет номера испытуемых, с 1 по 13. Общее число испытуемых в данном примере равно 13 ($N = 13$). Вектор-столбцы $X_1, X_2, X_3, \dots, X_{10}$ представляют ответы испытуемых по десяти заданиям.

Пример матрицы тестовых результатов приводится в таблице 1.

Сложение по строке даёт исходные тестовые баллы испытуемых (Y_i), подлежащие далее шкалированию. Суммирование баллов испытуемых $\sum Y_i$ даёт число 65. Сложение по столбцам характеризуют задания. Чем больше правильных ответов на задание, тем легче оно оказывается для данной группы испытуемых. Внизу таблицы:

R_j — число правильных ответов, полученных в заданиях;

W_j — указывает на число неправильных ответов испытуемых в каждом задании.

p_j — доля правильных ответов по заданию j ; определяется по формуле $p_j = \frac{R_j}{N}$.

В результате деления получается нормированный, по числу испытуемых, статистический показатель — доля правильных ответов (p_j) на задание j . Значения p_j приводятся в третьей строке нижней части таблицы.

Деление W_j на N даёт долю неправильных ответов q_j ; $q_j = W_j / N$.

Все значения долей q_j представлены в четвёртой строке нижней части таблицы.

Сумма по строке p_i и значение среднего арифметического балла совпадают, или иначе, $\sum p_i = M$. Потому что p_j является не только долей правильных ответов, но и принимается в качестве средней арифметической для каждого задания. Данное равенство отражает сумму всех элементов матрицы тестовых заданий, но только в случае, если для получения Y_i используются одинаковые весовые коэффициенты (c_j) значимости заданий. Например, когда в тесте все c_j равны, например, единице²⁶. Все упомянутые расчёты делаются с целью проверить: какие задания в тестовой форме могут стать *тестовыми заданиями*, а какие — не могут стать таковыми.

²³ В центрах тестирования используются другие средства, типа SQL Server и т.п.

²⁴ Это компьютерные программы, содержащие методы множественного дисперсионного, факторного, регрессионного и многих других статистических методов, активно применяемых в профессиональной работе по созданию качественных педагогических тестов. Все расчёты прилагаются к тесту.

²⁵ **Аванесов В.С.** Основы научной организации педагогического контроля в высшей школе. М.: МИСиС, 1989; **Аванесов В.С.** Проблема качества педагогических измерений // Педагогические измерения. 2004. № 2. С. 15.; **Аванесов В.С.** Применение тестовых форм в Rasch Measurement // Педагогические измерения. 2005. № 4. С. 17.

²⁶ Для математически ориентированных читателей вектор Y_i можно представить как линейную функцию от переменных X : $Y_i = c_1 X_{i1} + c_2 X_{i2} + c_3 X_{i3} + \dots + c_k X_{ik}$. Если в качестве вероятностного аналога эмпирической средней арифметической рассматривать математическое ожидание, то очевидно равенство $E(Y_i) = c_1 E(X_{i1}) + c_2 E(X_{i2}) + c_3 E(X_{i3}) + \dots + c_k E(X_{ik})$. При c_j равных единице, Y_i становится элементарной суммой баллов.

Таблица 1

№	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	Y_i	p_i	q_i	p_i/q_i	$\ln p_i/q_i$
1.	1	1	1	0	1	1	1	1	1	1	9	0,90	0,10	9	2,20
2.	1	1	0	1	1	1	1	1	1	0	8	0,80	0,20	4	1,39
3.	1	1	1	1	0	1	1	0	1	0	7	0,70	0,30	2,33	0,85
4.	1	1	1	1	0	1	0	1	0	0	6	0,60	0,40	1,50	0,40
5.	1	1	1	1	1	1	0	0	0	0	6	0,60	0,40	1,50	0,40
6.	1	1	1	1	0	0	1	0	0	0	5	0,50	0,50	1,00	0
7.	1	1	0	1	1	0	1	0	0	0	5	0,50	0,50	1,00	0
8.	1	1	1	1	1	0	0	0	0	0	5	0,50	0,50	1,00	0
9.	1	0	1	0	1	1	0	0	0	0	4	0,40	0,60	0,66	-0,42
10.	0	1	1	0	0	0	0	1	0	1	4	0,40	0,60	0,66	-0,42
11.	1	1	1	0	0	0	0	0	0	0	3	0,30	0,70	0,43	-0,84
12.	1	1	0	0	0	0	0	0	0	0	2	0,20	0,80	0,25	-1,39
13.	1	0	0	0	0	0	0	0	0	0	1	0,10	0,90	0,11	-2,21
R_j	12	11	9	7	6	6	5	4	3	2	65				
W_j	1	2	4	6	7	7	8	9	10	11					
p_i	0,923	0,846	0,692	0,538	0,462	0,462	0,385	0,308	0,231	0,154	5				
q_i	0,077	0,154	0,308	0,462	0,538	0,538	0,615	0,692	0,769	0,846					
$p_i q_i$	0,071	0,130	0,213	0,248	0,248	0,248	0,236	0,213	0,178	0,130					
q_i/p_i	0,083	0,182	0,445	0,859	1,164	1,164	1,597	2,246	3,329	5,493					
$\ln q_i/p_i$	-2,489	-1,704	-0,810	-0,152	0,152	0,152	0,468	0,809	1,202	1,703					

До начала сложения из матрицы удаляются неподходящие задания и испытуемые, после чего производится два упорядочения — по вектору — столбцу Y и по вектору-строке R . Одно касается испытуемых. В первой строке представляются баллы самого успешного испытуемого, во второй — менее успешного и т.д., по нисходящей сумме баллов. Последнюю строку матрицы занимает наименее подготовленный испытуемый. Другое упорядочение проводится для заданий. На первом месте ставится самое лёгкое задание (по нему есть наибольшее число правильных ответов), на втором — менее лёгкое, и так далее, до последнего, самого трудного задания. Такое построение строк и столбцов удобно.

Далее нам понадобятся значения последнего столбца таблицы 1. $\ln(p_i/q_i)$. — значения натурального логарифма отношения доли правильно решённых заданий p_i , испытуемым под номером i , к доле неправильно решённых заданий q_i . Это очень интересная мера, позволяющая провести логарифмическое шкалирование исход-

ных тестовых баллов испытуемых. Средняя арифметическая в логарифмической шкале равна нулю, стандартное отклонение равно единице.

Значения логарифмированных тестовых баллов полезно обозначить, вслед за классиком американской тестологии F.M. Lord, как θ , где θ означает переменную величину, называемую им по-английски ability. Это стало общим названием измеряемой латентной величины. На русском языке для педагогических измерений правильнее дать другое название — уровень подготовленности испытуемых. Конкретные значения испытуемых дополняются подстрочными символами θ_{i0} , где подстрочный индекс i обозначает номер испытуемого, а ноль обозначает, что это исходное (нулевое) значение латентного тестового балла испытуемого, подлежащего дальнейшему уточнению. Таким образом, $\theta_{i0} = \ln(p_i/q_i)$. Все эти значения для каждого испытуемого, напомним, приводятся в последнем столбце таблицы 1.

Симметрично, в Rasch Measurement вводится логарифмическая мера трудности заданий β . Исходные значения логарифма значения уровня трудности каждого задания $\beta_{j0} = \ln(q_j/p_j)$ читатель найдёт в последней строке таблицы 1.

Свойства и показатели качества педагогических заданий

Качество педагогических заданий формируется составом и структурой, т.е. системой его существенных свойств. Состав задания образуют его элементы: номер задания, инструкция по выполнению задания, текст задания (слова, термины, предложения), формульный аппарат, рисунки, графики, чертежи, схемы, правильные ответы и неправильные (дистракторы).

Качество заданий представляется системой показателей различного уровня сложности. Элементарный показатель назовём индикатором. В педагогическом тесте в качестве индикатора выступает тестовое задание. Система индикаторов образует показатель²⁷. В педагогических измерениях одним из показателей является тест.

Главные показатели качества заданий — меры трудности, коэффициенты вариации, корреляции. Представляют интерес и множество производных показателей — крутизна характеристической кривой задания, значения факторных нагрузок, информационная функция задания и другие.

В литературе иногда пишут о «надёжности» и «валидности» заданий, но эти две традиционные категории теории педагогических измерений правильнее применять к характеристике теста в целом, а не к отдельным его индикаторам (заданиям).

Структуру задания образуют взаимосвязь и расположение перечисленных выше элементов задания.

Свойства заданий можно разделить на три класса: субъективные, интерсубъективные и объективные.

Привлекательность. Как и образование в целом, хорошее педагогическое задание отвечает познавательным потребностям обучающихся. Для того чтобы задание стало интересным, оно

должно быть абсолютно понятным, соответствовать уровню интеллектуального развития и индивидуальному уровню подго-

товленности по учебному предмету. Ввиду естественных индивидуальных различий в восприятии и понимании информации, легко прийти к важному для практики выводу: нет заданий, одинаково подходящих для всех. В идеально организованном учебном процессе для каждого учащегося, в каждый текущий момент учебного процесса, надо искать своё задание, развивающее личность. А в массовом учебном процессе это можно сделать только в системе адаптивного компьютеризованного обучения и контроля, в котором применяются тестовые технологии. Альтернативы этому подходу нет. Отсюда его актуальность

Трудоёмкость задания. Главное отличие задачи от тестовых заданий состоит во времени решения, в количестве вычислений и в количестве умственных действий. Увлечение только задачами или только тестовыми заданиями формирует соответствующий тип мышления. На самом деле нужны различные формы обучения и контроля знаний. В качественно организованном учебном процессе применение учебных форм координирует педагог. *Насильственное внедрение тестовых форм в качестве единственных или главных в обучении или контроле наносит вред, особенно для становления мыслительной, речевой и письменной культуры личности.*

Одна из причин образовательного кризиса заключается в отсутствии учёта средней трудоёмкости заданий для учащихся и студентов. Каждый преподаватель требует выполнения своих заданий и не может принять во внимание, что и другие преподаватели тоже дают задания. Обычное отсутствие реальной координации по количеству и трудоёмкости ежедневно предлагаемых заданий приводит к учебной перегрузке, и как следствие, к регулярному отказу от выполнения заданий по причине нехватки времени.

К числу показателей трудоёмкости заданий можно отнести число выполняемых умственных операций и время выполнения. В практике тестирования стремятся подбирать такие задания, чтобы цепочка умственных действий была не длинной, число вычислений было небольшим, а затраты времени — оказались минимальными.

Трудность заданий. Трудность задания определяется как статистическая мера его решаемости испытуемыми. Чем больше решаемость, тем легче задание. Доля правильных ответов — самая популярная мера для определения трудности задания. Трудность заданий может опре-

²⁷ Аванесов В.С. К вопросу о социальных показателях. В сб.: «Современные концепции уровня, качества и образа жизни». М.: Институт соц. исследований АН СССР, 1978. С. 128–133.

Таблица 2

деляться умозрительно, на основе предполагаемого числа и характера умственных операций, необходимых для успешного выполнения заданий. Поскольку этим чаще занимаются психологи, то в этой науке часто используется другое понятие — сложность задания.

Знание объективированной меры трудности — обязательное требование к тестовым заданиям. Эту мысль можно усилить, связав с ранее изложенным материалом: если неизвестна эмпирическая мера трудности задания, то это задание — не тестовое. В лучшем случае оно будет заданием в тестовой форме, в худшем — ни тем, ни другим. Мера трудности задания определяется после эмпирической апробации заданий, составления матриц и подсчёта долей неправильных ответов q_j в каждом задании (j). В качестве показателя трудности в классической теории тестов²⁸ долго использовалась статистика p_j .

G. Rasch в числе первых обратил внимание на смысловую ошибочность показателя p_j : увеличение значения p_j указывает не на возрастание трудности, а, наоборот, на возрастание лёгкости, если можно применить такое слово. Поэтому в последние годы с показателем трудности заданий стали ассоциировать противоположную статистику — долю неправильных ответов (q_j). Если тест — система параллельных заданий равномерно возрастающей трудности, то в нём нет места заданиям с неизвестной мерой трудности.

В западной статистической теории тестов многие годы рассматривались только эмпирические показатели трудности. В новых вариантах психологических и педагогических теорий тестов больше внимание стало уделяться характеру умственной деятельности учащихся в процессе выполнения тестовых заданий различных форм²⁹.

Таким образом, требование известной трудности оказывается важнейшим системообразующим признаком тестового задания. Если тест — это система параллельных заданий равномерно возрастающей трудности, то в нём нет места заданиям с неизвестной или неадекватной мерой трудности. В настоящем тесте, сделанном по новым технологиям, нет места и заданиям одинаковой трудности.

Для удобства в разработке теста и в интерпретации результатов тестирования F. Baker разделил задания на пять примерных уровней трудности³⁰ (табл. 2):

ГРАДАЦИИ ТРУДНОСТИ	МЕРА
Очень трудные задания	$\beta_j > 2,6$
Трудные задания	$1,5 < \beta_j < 2,59$
Задания среднего уровня трудности	$-1,49 < \beta_j < 1,49$
Лёгкие задания	$-2,59 < \beta_j < -1,5$
Очень лёгкие задания	$\beta_j < -2,6$

На основе этой уровневой классификации, задания таблицы 1 можно оценить следующим образом. Первое задание очень лёгкое, второе и третье — лёгкие, четвёртое, пятое, шестое, седьмое и восьмое — задания среднего уровня трудности. Девятое и десятое — трудные задания. Очень трудных заданий в таблице 1 нет.

Нередко встречаются задания, ответы на которые оцениваются сплошными нулями (предельно трудные) или единицами (предельно лёгкие) задания. Те и другие образуют отдельную группу так называемых «нетестовых» заданий.

Нетестовые задания — это те, которые имеют нетестовую форму. Но нетестовые задания могут иметь тестовую форму и не быть тестовыми по существу, то есть не отвечать содержательным и другим требованиям к тестовым заданиям. Форма — условие необходимое, но недостаточное. Нетестовые задания не дифференцируют подготовленных испытуемых от неподготовленных. В своих крайних проявлениях это такие задания, на которые либо все испытуемые отвечают правильно, либо все отвечают неправильно³¹. Все такие задания удаляются из матрицы как не соответствующие уровню подготовленности испытуемого контингента. По ответам на слишком лёгкие или слишком трудные задания все испытуемые выглядят одинаково — удачливыми или неудачливыми испытуемыми. Иначе говоря, нетестовые задания в силу ошибочной композиции — неправильно подобранного содержания или ошибок в форме, а возможно, того и другого, не дифференцируют испытуемых по уровню подготовленности.

В этой связи полезно немного перефразировать и слегка уточ-

²⁸ Gulliksen H. Theory of Mental Tests. N.Y. Wiley. 1950 и др.

²⁹ Tatsuoка K.K. Item construction and psychometric models appropriate for constructed response. Princeton, N-J, 1993; Frederiksen, N., Mislevy R.J., Bejar I. J. (Eds). Test theory for a new generations of tests. Lawrence Erlbaum Ass. Publ. 1993, Hillsdale, N-J, и др.

³⁰ Baker F.B. The Basics of Item Response Theory. 2 ed. Hieneman, Portsmouth, New Hampshire, 2001. p. 7.

³¹ Примеры таких заданий и матриц, с необходимыми расчётами, были приведены в моей работе «Основы научной организации педагогического контроля в высшей школе. М.: МИСиС, 1989.

нить первую часть определения педагогического теста: «тест — это система фасетных заданий равномерно возрастающей трудности; таких заданий, которые обладают свойством дифференцировать³² испытуемых по уровню их подготовленности». В тесте нет места заданиям, не обладающим отмеченными свойствами.

Если слабо подготовленные испытуемые отвечают на задание так же правильно, как и хорошо подготовленные испытуемые, то это может означать высокую угадываемость правильных ответов или наличие каких-то других дефектов — задания или организации тестирования. В любом случае итогом становится *низкое дифференцирующее свойство задания*³³. Такие задания являются плохим индикатором и не подходят для создания теста. Эти задания — не тестовые.

В процессе обработки данных тестовые матрицы часто приводят к пригодному для композиции теста виду. Это означает, что из матрицы удаляют всё не подходящее для этого задания.

Одномерность задания — четвёртое требование к свойствам тестового задания. Содержательная сторона одномерности может быть описана словами «предметная чистота содержания задания». Под этим понимается отсутствие пересечения содержания задания одной учебной дисциплины с содержанием другой, что определяется экспертно. Отсутствие пересечения порождает признак

гомогенности теста. Это означает, что тест измеряет именно то знание, которое задумано измерять, и ничего сверх того. Это, казалось бы, естественное требование для гомогенного (одномерного) теста, очень часто нарушается из-за стремления увязать проверку знаний по отдельной дисциплине со стремлением проверить что-то ещё. Неоднократное нарушение этого требования приводит к ухудшению качества измерения, к потере так называемого свойства одномерности теста или

шкалы, если держать в уме ещё одно очень короткое определение: тест — это шкала. Одномерность определяется также статистическими методами.

Вариация тестовых баллов. Вариация баллов — пятое свойство и, одновременно, обязательное требование к тестовым заданиям. Мера *вариации (дисперсии) тестовых баллов* — это показатель. Значения дисперсий, рассчитанные по этой формуле для каждого задания, представлены в пятой строке нижней части примерной таблицы 1 тестовых результатов. Чем выше дисперсия, тем выше дифференцирующая способность задания. Для данных, представленных в дихотомической шкале один или ноль максимум дисперсии наблюдается при $p = q = 0,5$. В таблице 1 представлены значения $p_j q_j$ — дисперсии тестовых баллов по каждому заданию для случаев использования оценок 1 и 0». Из этой строки видно, что для заданий, оцениваемых единицей и нулём, максимальное значение дисперсии получается при p_j и q_j , равных $0,5$ ³⁴.

В отличие от R_j и W_j , значения долей p_j и q_j можно отнести к статистическим показателям, или кратко, к статистикам, что открывает возможность проверки достоверности этих мер, полученных в разных выборках. Кроме того, определяются значения так называемых доверительных интервалов, в пределах которых могут находиться значения этих долей в генеральной совокупности испытуемых. В статистике естественным образом принимается $p_j + q_j = 1$.

Так как цель разработки теста — это измерение уровня подготовленности испытуемых, то качественно это можно сделать только с помощью системы заданий равномерно возрастающей трудности. И если все, без исключения, испытуемые отвечают на задание одинаково правильно, то это означает, что задание не дифференцирует знающих испытуемых от незнающих. Нет, соответственно, и никакой вариации в значениях ответов. По данному заданию в матрице будут стоять одинаковые оценки. Такое задание из матрицы удаляется из-за отсутствия вариации баллов.

Нет вариации и в очень трудном задании, где нет ни одного правильного ответа; в матрице стоят, соответственно, одни нули. Вариация по нему тоже равна нулю, что означает необходимость и его удаления из проектируемого теста. Оно не тестовое.

Часто ставится вопрос: а как проявят себя удаляемые задания в других выборках испытуемых? Ответ зависит от подбора групп, а точнее

³² На Западе принята лексика «дискриминантная способность задания». То же пишут и на русском языке, или что хуже, пишут «дискриминативная способность», хотя слово «дискриминация» в обоих языках имеет другой, политический, и далеко не позитивный смысл. Поэтому вместо всего этого лучше говорить и писать о дифференцирующем свойстве задания и теста.

³³ Что проявляется в значениях коэффициента корреляции числового вектора ответов испытуемых на задание с числовым вектором исходных тестовых баллов испытуемых. Производным от такого значения корреляции является параметр крутизны графического образа задания, материал о котором излагается далее.

³⁴ Если q_j заменить равным значением $(1 - p_j)$ и в произведении $p_j(1 - p_j)$ раскрыть скобки, то получим $p_j - p_j^2$. Производная данной функции по p_j равна $1 - 2p_j$. Приравняв $1 - 2p_j$ нулю, получим доказательство максимума дисперсии $p_j q_j$ при $p_j = 0,5$.

от статистического плана формирования выборочных совокупностей. Ключ к правильному ответу на этот вопрос надо искать в смысле понятия «target group»; это множество испытуемых, для которых предназначен разрабатываемый тест. Соответственно, если задания проектируемого теста ведут себя неодинаково в разных группах, то это может быть указанием на ошибки в формировании выборок испытуемых, различий в преподавании. На языке статистики это означает, что испытуемые целевой и экспериментальных групп должны принадлежать одной генеральной совокупности.

Дифференцирующим свойством задания называется его способность различать испытуемых по уровню подготовленности, на данном уровне измеряемого континуума подготовленности θ . Чем выше дифференцирующее свойство задания, тем лучше деление испытуемых на подготовленных и не подготовленных, тем выше и качество теста. Это свойство легко понять из сравнения результатов, например, применения пятибалльной и одиннадцатибалльной шкал. Вторая шкала лучше различает (дифференцирует) испытуемых, чем первая. Ещё лучше это делает, например, тридцатибалльная шкала теста. Она делит всех испытуемых на тридцать групп (классов). Эту логику можно распространить на стобалльную шкалу, часто применяемую в массовом тестировании.

Значение коэффициента корреляции задания с тестом (r_{xy}) — второй показатель дифференцирующей способности заданий. Это распространённое в практике название представляет собой сокращённый вариант более правильного понятия — корреляция оценок, полученных испытуемыми по заданию под номером x_i , с суммой баллов тех же испытуемых. Корреляция является стандартной мерой дифференцирующей способности задания. Чем выше r_{xy} , тем лучше оно дифференцирует испытуемых по уровню подготовленности. Задание в тестовой форме нельзя называть тестовым, если баллы по этому заданию не коррелируют (не связаны) с суммой исходных тестовых баллов³⁵.

Для расчёта r_{jy} формируется два вектор-столбца, один из которых — задание (X_j), другой — критерий (Y). Между значениями этих двух векторов и устанавливается мера связи, если такая существует. Мера связи определяется посредством расчёта коэффициента корреляции r_{jy} , где символом r обозначается так называемый классический коэффициент корреляции Пирсона, j представляет номер коррелируемого

задания, а символ Y — числовой вектор-столбец тестовых баллов испытуемых.

Формулы для расчёта коэффициентов корреляции и примеры такого расчёта уже приводились мною ранее³⁶. Проверим, например, меру связи ответов испытуемых по заданию № 7 с суммой баллов испытуемых по всему тесту. Для этого строится вспомогательная таблица 3, в которой использованы соответствующие данные таблицы 1.

В колонке X_7 приводятся значения баллов, полученных испытуемыми в седьмом задании. Сумма этих баллов равна 5.

Во второй колонке представлены тестовые баллы (Y_i); в таблице представлено без индекса i , что позволяет не перегружать формулы; $\sum Y_i = 65$.

В третьей колонке даются произведения баллов каждого испытуемого по седьмому заданию (X_7) и по сумме баллов (Y); $\sum X_7 Y = 34$. Это сумма попарных произведений X и Y .

В четвёртой и пятой колонках — квадраты значений X_7 и Y ; соответственно, $\sum X_7^2 = 5$ и $\sum Y^2 = 387$.

При расчёте используются, последовательно, четыре формулы:

1) Вначале находится сумма квадратов отклонений баллов испытуемых от среднего арифметического балла в интересующем задании (SS_7). Это делается по формуле:

$$SS_7 = \sum X_7^2 - \frac{(\sum X_7)^2}{N} = 5 - \frac{5^2}{13} = 3,077. \quad (1)$$

Расчёт коэффициента корреляции. Таблица 3

2) Затем находится сумма квадратов отклонений тестовых баллов испытуемых от среднего арифметического балла по всему тесту (SS_Y). Это делается также по формуле:

$$SS_Y = \sum Y^2 - \frac{(\sum Y)^2}{N} = 387 - \frac{4225}{13} = 62. \quad (2)$$

3) Находится так называемая скорректированная на средние значения сумма попарных произведений X и Y , по формуле:

³⁵ Некоторые авторы пишут о так называемых «сырых» баллах, используя метафорический перевод слов с английского «raw scores», что может натолкнуть на мысль о возможности существования и «не сырых», например, «пропаренных или поджаренных» баллов.

³⁶ **Аванесов В.С.** Основы научной организации педагогического контроля в высшей школе. М. МИСиС, 1989. С. 96–100.

Таблица 3

Испытуемые	X ₇	Y	X ₇ Y	X ²	Y ²
1.	1	9	9	1	81
2.	1	8	8	1	64
3.	1	7	7	1	49
4.	0	6	0	0	36
5.	0	6	0	0	36
6.	1	5	5	1	25
7.	1	5	5	1	25
8.	0	5	0	0	25
9.	0	4	0	0	16
10.	0	4	0	0	16
11.	0	3	0	0	9
12.	0	2	0	0	4
13.	0	1	0	0	1
Σ :	5	65	34	5	387

$$SP_{XY} = \sum XY - \frac{\sum X \sum Y}{N} = 34 - \frac{5 \cdot 65}{13} = 9. \quad (3)$$

В последней формуле $\sum XY$ представляет собой сумму произведений баллов каждого испытуемого по седьмому заданию и по Y_7 , тестовому баллу испытуемых. Вторая часть формулы $\frac{\sum X \sum Y}{N}$ представляет собой коррекцию на средние значения X и Y .

4) Рассчитывается коэффициент корреляции по формуле:

$$r_{XY} = \frac{SP_{XY}}{\sqrt{SS_X \cdot SS_Y}}. \quad (4)$$

Подставляя в эту формулу результаты проведенных расчётов, получаем

$$r_{XY} = \frac{9}{\sqrt{3,007 \cdot 62_Y}} = 0,652.$$

³⁷ Ряд материалов интересующийся читатель может найти в работе автора «Научные основы тестового контроля знаний». М.: Иссл. центр, 1994; **Аванесов В.С.** Методологические и теоретические основы тестового педагогического контроля. Дисс... докт. пед. наук. С-Пб. Госуниверситет, 1994.

³⁸ Point-bi-serial. См. например, **Гласс Дж., Стэнли Дж.** Статистические методы в педагогике и психологии / Пер. с англ. М. Прогресс, 1976. С. 150.

Чем выше значения r_{xy} , тем больше вероятность у задания быть включённым в тест. Если взять $r^2 \times 100\%$, то получим значение так называемого коэффициента детерминации, выраженного в удобной для интерпретации

процентной мере связи задания с суммой баллов. Для взятого примера коэффициент детерминации у седьмого задания равен $(0,652)^2 \times 100\% = 42,5\%$, что можно интерпретировать так: 42,5% вариации исходных тестовых баллов испытуемых связано с вариацией баллов по одному только седьмому заданию, что указывает на достаточно хороший вклад седьмого задания в общую дисперсию тестовых баллов испытуемых.

При значениях $0,2 < r_{xy} < 0,5$ во внимание начинают приниматься и другие характеристики: такие как мера корреляции задания с другими заданиями, факторная чистота задания и прочие; их рассмотрение здесь потребовало бы другого стиля изложения³⁷. В качестве нижней границы включения заданий в тест обычно рассматриваются значения $r_{xy} = 0,300$, и самой нижней, в исключительных случаях, $r_{xy} = 0,200$. Нулевая корреляция свидетельствует об отсутствии у задания системных свойств, присущих тестовому заданию. Такие задания, равно как и задания с отрицательными значениями r_{xy} , устраняются из тестовых материалов как не выдержавшие эмпирической проверки.

Теоретически предпочтительнее рассчитывать другие варианты коэффициента корреляции Пирсона. Первый из них называется point-biserial — одним из бисериальных коэффициентов корреляции, для случая, когда одна переменная представлена дихотомической оценкой³⁸.

$$r_{pb} = \frac{M_1 - M_0}{s_Y} \sqrt{\frac{(n_0)(n_1)}{n(n-1)}}, \quad (5)$$

где r_{pb} означает названный коэффициент;

M_1 — среднее арифметическое по всему тесту для испытуемых, получивших по данному заданию один балл;

M_0 — среднее арифметическое по всему тесту для испытуемых, получивших по данному заданию ноль баллов;

n_1 — число испытуемых, получивших в задании один балл, n_0 — число испытуемых, получивших в задании ноль баллов.

При использовании этой формулы из таблицы 2 опираются на такие данные: один балл по седьмому заданию получили первый, второй, третий, шестой и седьмой испытуемые. Сложение полученных ими баллов по Y даёт $9 + 8 + 7 + 5 + 5 = 34$; среднее арифметическое $M_1 = 34/5 = 6,800$. Ноль баллов по этому же заданию получили 4, 5, 8, 9, 10, 11, 12, и 13 испытуемые. Сложение полученных ими баллов по Y даёт $6 + 6 +$

$5 + 4 + 4 + 3 + 2 + 1 = 31$; среднее арифметическое $M_0 = 31/8 = 3,875$, при $n_1 = 5$, $n_0 = 8$; $n = 13$.

Подстановка полученных данных в формулу 11.9 даёт

$$r_{pb} = \frac{6,800 - 3,875}{2,273} \sqrt{\frac{(5)(8)}{13(13-1)}} = 0,651.$$

Сравнение $r_{pb} = 0,651$ и $r_{xy} = 0,652$ подтверждает сходство полученных значений и практическую достаточность использования любой одной из этих формул. Расчёт другого, так называемого (bi-serial) бисериального коэффициента корреляции задания с критерием проводится обычно в целях теоретического исследования. На малых выборках он даёт завышенные значения, что объясняется обычными отклонениями исходных тестовых баллов от нормального распределения³⁹.

Параметр крутизны характеристической кривой задания теста — это третий показатель дифференцирующей способности заданий. Кривая — это функция вида

$$P_j = \{x_{ij} = 1 | \beta_j a_j\} = \frac{\exp a_j(\theta - \beta_j)}{1 + \exp a_j(\theta - \beta_j)}, \quad (6)$$

рассматриваемой ниже. Он обозначается символом a_j . При условии, что уровень трудности заданий соответствует уровню подготовленности испытуемых, и при наличии достаточно репрезентативной выборки, оценки параметров a_j проводятся по формуле, приведенной в работе F.M. Lord

$$a_j = \frac{r}{\sqrt{1-r^2}}, \quad (7)$$

где r представляет ранее посчитанный коэффициент корреляции ответов на задание j теста с общей суммой баллов теста.

По уровню дифференцирующей способности F. Baker делит тестовые задания на семь уровней⁴⁰ (Табл. 4):

Дифференцирующая способность	Значения параметра крутизны функции, a_j
1) практически отсутствует	0
2) очень низкая	0,01–0,34
3) низкая	0,35–0,64
4) средняя	0,65–1,34
5) высокая	1,35–1,69
6) очень высокая	>1,70
7) отличная (совершенная, perfect)	с неопределяемого большого числа до + бесконечности

Встречающиеся в практике задания с отрицательными значениями a_j не относятся к тестовым заданиям, а потому подлежат удалению. Пример такого задания представлен в разделе «Графические образы заданий».

Требования к тестовым заданиям

Ранее для заданий в тестовой форме были сформулированы такие требования:

- краткость;
- технологичность;
- правильность формы;
- логическая форма высказывания;
- одинаковость правил оценки ответов;
- наличие определённого места для ответов;
- одинаковость инструкции для всех испытуемых;
- правильность расположения элементов задания;
- адекватность инструкции форме и содержанию задания⁴¹.

Для того, чтобы задания в тестовой форме могли бы далее превратиться в тестовые задания, к ним добавляются требования иметь рассмотренные выше свойства:

- трудности;
- вариации тестовых баллов;
- положительной корреляции баллов по заданию с баллами по всему тесту⁴²;
- одномерности. Последнее свойство выражается содержательно и формально.

Формально одномерность — главное свойство заданий гомогенного теста. Это означает, что все задания теста измеряют одно и то же интересующее свойство. Формальным выражением идеи одномерности

Таблица 4

³⁹ Формулу расчёта бисериального коэффициента интересующиеся читатели могут найти в работе: **Гласс Дж., Стэнли Дж.** Статистические методы в педагогике и психологии / Пер. с англ. М. Прогресс, 1976. С. 150.

⁴⁰ **Baker F.B.** The Basics of Item Response Theory. 2 ed. Hieneman, Portsmouth, New Hampshire, 2001. p. 7. В оригинале этой классической работы автор пишет о «дискриминантной», в смысле способности задания различать испытуемых по уровню их подготовленности. Но на русском языке слово «дискриминация» имеет совсем другой смысл. Поэтому, в отличие от других текстов, здесь использовано понятие «дифференцирующая способность задания».

⁴¹ **Аванесов В.С.** Основы педагогической теории измерений // Педагогические измерения. 2004. № 1. С. 17.

⁴² Там же.

измеряемого свойства является требование так называемой локальной независимости ответов испытуемых на задания теста. Логика такова: если на вероятность правильного ответа влияют только различия в уровне подготовленности испытуемых, то для испытуемых одинакового уровня подготовленности вероятность правильного ответа на одно задание не должна зависеть от вероятности правильного ответа на любое другое задание;

— адекватность уровня трудности задания уровню подготовленности испытуемого. В этом случае показателем эффективности становятся значения так называемой информационной функции задания⁴³.

Кроме того, педагогические задания должны:

- формулироваться из точных терминов, не содержать метафор, лишних слов и знаков.
- иметь педагогически корректное содержание;
- быть логически точными и непротиворечивыми;
- быть изложенными на понятном для всех учащихся языке изученной дисциплины (предмета);
- быть обоснованными теоретически и эмпирически.
- иметь алгоритм правильного и, по возможности, быстрого решения.

Было бы хорошо, если бы все педагогические задания в той или иной мере могли отвечать изложенным выше требованиям. Но этого нет и не будет до тех пор, пока культура педагогических измерений не укоренится в педагогическом обществе.

Теоретическое обоснование качества педагогических заданий имеет два источника. Первый — в предметной области, где формируется система научной аргументации в пользу избранного содержания задания. Второй источник обоснования

качества задания находится в теории педагогических измерений, в которой формулируются общие требования к тестовым заданиям⁴⁴.

Эмпирическое обоснование качества тестовых заданий проводится на выборочной совокупности испытуемых, с применением статистических и математических методов обработки дан-

ных. Для этой цели специально разработаны статистические пакеты SPSS, Winsteps, RUMM-2020 и множество других.

Объективные свойства заданий

Объективные свойства педагогических заданий — это содержание, форма и технологичность.

Содержание заданий ранее определялось как специально отобранная и признанная профессиональным сообществом система элементов объективного опыта человечества, применение и усвоение которой необходимо для успешной профессиональной деятельности индивида в избранной им сфере и в процессе жизнедеятельности. Содержание образования задаётся учебным планом и соответствующими ему программами учебных дисциплин; оно отражается в учебниках, учебных пособиях и в других средствах обучения⁴⁵. В хороших образовательных системах существуют Национальные (общественно-профессиональные) стандарты уровня и структуры подготовленности учащихся и студентов, достижение которых контролируется общественными профессиональными организациями и независимыми Центрами тестирования. Для оценки подготовленности используются тысячи тестовых заданий различного уровня и направления, в зависимости от профиля вуза.

В России утвердилась идеология государственных стандартов подготовки, которые исходят из идей минимальной достаточности, с последующим контролем знаний со стороны государственных органов управления образованием и подчинённых им центров тестирования.

С точки зрения содержания тестовое задание оценивается по конкретности и абстрактности, глубине, обобщённости, полноте знаний предмета.

Форма педагогического задания определяется как такое расположение элементов, которое наилучшим образом позволяет выполнить предназначенную функцию контроля или обучения⁴⁶. Это ключевая проблема теории педагогических измерений. От правильного выбора тестовых форм зависит точность выражения содержания теста и качество тестовых оценок. Образовавшийся в практике застой с применением только формы с выбором одного правильного ответа из 3–4 дистракторов подменяет педагогический принцип контроля принципом угадывания. Что неизбежно приведёт к росту справедливой критики и к дискредитации тестового метода. Для предотвращения нежелательных последст-

⁴³ Birnbaum, A. (1968). Test scores, sufficient statistics, and the information structures of tests. In F.M. Lord & M.R. Novick, *Statistical theories of mental test scores*. Reading, Mass.: Addison-Wesley.

⁴⁴ См. например, Аванесов В.С. *Композиция тестовых заданий*. М.: Центр тестирования. 2002.

⁴⁵ Гендина Н.И., Колкова Н.И. *Нормативно-методическое обеспечение учебного процесса в вузе. Стандарты высшего образования*. Кемерово, 1998.

⁴⁶ Аванесов В.С. *Применение тестовых форм в Rasch Measurement // Педагогические измерения*. 2005. № 4.

вий предлагается быстрее перейти к применению в практике заданий с выбором нескольких правильных ответов, сдвоенных заданий, тематических и текстовых заданий. Они позволяют проверять различные виды знаний, делать самоконтроль интереснее и качественнее⁴⁷.

Технологичность является одним из современных требований к тестовым заданиям. *Технологичность* заданий определяется как вариант композиции, которая позволяет вести процесс тестирования с помощью технических средств, и делать это точно, быстро, экономично и объективно. Задания становятся технологичными, если их содержание правильно и быстро понимается испытуемыми и если форма заданий способствует процессу компьютеризации тестирования.

Главные препятствия для достижения технологичности задания — неопределённость формы, плохая формулировка содержания, отсутствие необходимой техники и программ. Именно использование автоматизированного обучения и контроля на основе достижений новой педагогики, педагогических измерений, психологии, кибернетики и компьютерной техники, взятое в разумном соотношении, образует основу того, что сейчас называют педагогической технологией.

В этой технологии самым узким местом оказалось неумение делать задания, без чего нельзя сейчас ни объективно проверить знания, ни создать современную автоматизированную контрольно-обучающую программу, ни наладить такую форму организации учебного процесса, как дистанционное обучение. В идеальном случае учебная программа, каждый её модуль сопровождаются заданиями в тестовой форме. Однако путь к достижению этого идеала лежит через трудности создания качественных тестовых заданий. Технологичность заданий обеспечивается правильностью формы заданий и корректностью содержания заданий. С точки зрения содержания, тестовое задание оценивается по конкретности и абстрактности, глубине, обобщённости, полноте содержания.

Вероятностные функции педагогических заданий

Педагогу традиционной формации бывает нелегко объяснить, что у каждого педагогического задания может быть свой графический образ. Для выявления свойств заданий и для построения графических образов применяются различные математические модели-функции. Извест-

ны три основные модели педагогических измерений. В качестве одной из них используется так называемая логистическая функция, известная, по мнению F. Baker, в биологии с 1844 года. Это функция вида

$$Y = \frac{e^x}{1 + e^x}. \quad (8)$$

В 1958 г. у G. Rasch возникла идея выразить вероятность правильного ответа испытуемого i на задание j посредством такой функции. Она приняла такой вид, который на Западе называют логистической функцией:

$$P_j(\theta) = \{x_{ij} = 1 | \beta_j\} = \frac{\exp(\theta - \beta_j)}{1 + \exp(\theta - \beta_j)}, \quad (9)$$

где $P_j(\theta)$ это вероятность правильного ответа на задание под номером j , вместо x стоит разность $\theta - \beta_j$, $x_{ij} = 1$, если ответ испытуемого i на j -ое задание правильный; θ — уровень подготовленности (знаний), латентная переменная. Поэтому индекс i при символе θ опущен. β_j — уровень трудности j -го задания теста, измеряемой на латентном континууме трудности заданий. \exp — константа e , иррациональное число, равное, округлённо, 2,71828.

Функцию (1) можно записать и другими способами, например, в строку:

$$P_j = \{x_{ij} = 1 | \beta_j\} = \exp(\theta - \beta_j) / (1 + \exp(\theta - \beta_j)), \quad (10)$$

или $P_j(\theta) = \{x_{ij} = 1 | \beta_j\} = \frac{1}{1 + e^{-L}}$.

или, совсем короче $P_j(\theta) = \frac{1}{1 + e^{-L}}$.

А также $P_j(\theta) = \frac{1}{1 + \exp(-L)}$,

где L представляет разность параметров испытуемого i и задания j , $(\theta_i - \beta_j)$. Если берётся конкретное задание под номером j , то разность записывается с индексом соответствующего номера β_j . Если берётся конкретный испытуе-

⁴⁷ **Аванесов В.С.** Форма тестовых заданий. М.: Центр тестирования, 2005; **Аванесов В.С.** Проблема формы тестовых заданий. В сб. Тестовые формы контроля по русскому языку как иностранному (РКИ). Стр. 20–30. II Всероссийская научно-практическая конференция. Доклады и сообщения. Центр международного образования МГУ им. М.В. Ломоносова., 12–13 апр. 2005 / Под ред. Н.П. Андрушиной, О.А. Усковой. М., 2005.

мый i с присущим ему уровнем подготовленности, то разность пишется $(\theta_i - \beta)$.

Понятно, что вместо неизвестного значения x в формуле (1) G. Rasch подставил разность получаемых из таблицы 1 исходных значений двух параметров θ и β . Напомним, что первый параметр — $\theta_i = \ln p_i/q_i$ — означает уровень подготовленности испытуемого в шкале натуральных логарифмов. Вторым — $\beta_j = \ln q_j/p$ — означает уровень трудности задания под номером j , представленный в той же шкале.

Логарифмические оценки уровня знаний и уровня трудности заданий дали возможность сравнить их, что оказало огромное влияние на развитие зарубежной педагогической теории и практики. Впервые появилась возможность непосредственно сопоставить испытуемого и задание, начать процесс адаптивного обучения.

Чем выше крутизна функции, тем уже интервал, на котором это задание работает. Таким образом, возникла мысль об улучшении модели G. Rasch за счёт введения в выражение (10) второго, после параметра β_j , параметра a_j . Параметр a_j даёт информацию о задании с точки зрения оценки его дифференцирующей способности, на заданном интервале континуума измерения. Геометрически значение параметра a_j выражается крутизной характеристической кривой, аналитически — значением производной функции в точке перегиба. После введения в выражение параметра a_j получается двухпараметрическая модель педагогического измерения.

$$P_j = \{x_{ij} = 1 | \beta_j, a_j\} = \exp a_j(\theta - \beta_j) / (1 + \exp a_j(\theta - \beta_j)) \quad (11)$$

Если ещё раз на оси абсцисс отложить значения логитов уровня подготовленности, а по оси ординат — значения вероятности правильного ответа на задание j , то характеристические кривые одинакового уровня трудности, но с изменяющимися значениями параметра a_j приобретают графические образы, представленные на рис. 1⁴⁸.

При a_j , принимаемых равными единице во всех заданиях, получается система кривых, отличающихся только уровнем трудности заданий, что выражается сдвигом по оси (см. рис. 2).

Чем задание труднее, тем оно располагается правее. Значение проекции точки перегиба функции на ось θ в точности равно параметру трудности задания.

Графический пример системы заданий, построенных на основе модели G. Rasch, представлен по публикации В.С. Кима в нашем журнале⁴⁹.

По оси абсцисс рис. 2 отложен латентный уровень подготовленности испытуемых (θ), по оси ординат — вероятность правильного ответа на задание (j), с уровнем трудности β_j . Добавление третьего параметра c_j , ассоциируемого со значением вероятности угадывания правильного ответа в заданиях с выбором, даёт трёхпараметрическую модель:

$$P_j = \{x_{ij} = 1 | \beta_j, a_j, c_j\} = c_j + (1 - c_j) \exp a_j(\theta - \beta_j) / (1 + \exp a_j(\theta - \beta_j)). \quad (12)$$

Алгоритм вычисления вероятности правильного ответа на задания

Как было уже показано, исходные значения параметров θ_j и β_j берутся из таблиц, создаваемых по аналогии с примерной таблицей 1. Далее обе шкалы θ и β приводятся к общей средней арифметической и к общему стандартному отклонению, а затем методом максимального правдоподобия либо другими вычислительными методами уточняются значения интересующих параметров⁵⁰.

В основу математической теории педагогических измерений положено понятие функции — вероятности правильного ответа испытуемого в зависимости от уровня его подготовленности от уровня трудности заданий.

Посмотрим здесь сравнительно простой случай, в предположении, что исходные значения θ_i и β_j уже известны из таблицы 1, и надо найти первую оценку вероятности правильного ответа испытуемых на задание одного фиксированного уровня трудности. Для этого последовательно берутся значения переменной величины θ_i , с некоторым шагом, и для каждого вычисляют значение искомой вероятности. Это позволит посмотреть — как меняется вероятность правильного ответа на одно и то же задание, с уровнем трудности $b = 1.0$ для испытуемых, имеющих различный уровень подготовленности. Для этого достаточно провести небольшой вычислительный эксперимент, в котором надо последовательно (с шагом +1) брать разные уровни подготовленности, и полученные данные свести в таблицу 1.

⁴⁸ Аванесов В.С. Композиция тестовых заданий. М.: Центр тестирования, 2003. С. 186.

⁴⁹ Ким В.С. Анализ результатов тестирования в процессе Rasch Measurement // Педагогические измерения. 2005. № 4. С. 42.

⁵⁰ По вычислительным методам параметров в IRT на небольшом и доступном примере планируется подготовить отдельную статью.

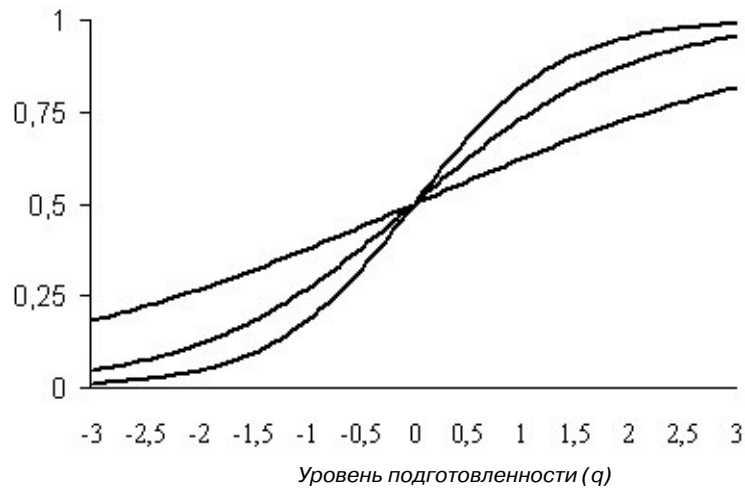


Рис. 1. Вероятностные функции (графические образы) заданий одинакового уровня трудности, отличающиеся значениями параметра дифференцирующей способности a_j : $a_1 = 1,5$; $a_2 = 1,0$; $a_3 = 0,5$.

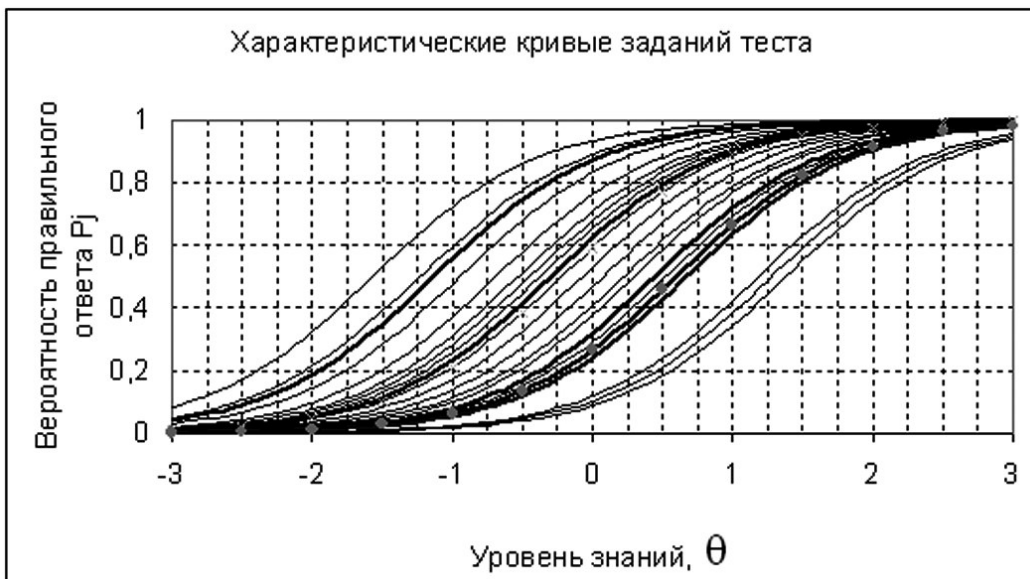


Рис. 2.

Возьмём пример применения однопараметрической модели, где значение параметра a_j принимается равным 1.0 (модель Г. Раша), а значения β_j , предположим, как уже упоминалось, равным +1,0. Для удобства строится вспомогательная таблица 5. Эта таблица содержит результаты вычислительного эксперимента по определению вероятности правильного ответа испытуемых различного уровня подготовленности на задание с параметром трудности $b = 1.0$

Первый шаг: Находим $L = (\theta - \beta_j)$. Для самого низкого уровня подготовленности $L = (-3,0 - 1,0) = -4,0$. (второй столбец табл. 5.)

Второй шаг: Вычисляется значение

$$e^{-L} = 2,71828 - (-4,0) = 54,59801$$

Третий шаг: находится значение знаменателя формулы (2)

$$1 + 54,59801 = 55,59801$$

Четвёртый шаг: Находится вероятность правильного ответа при использованных данных

$$P_j(\theta) = \frac{1}{1 + e^{-L}} = \frac{1}{55,59801} = 0,018315.$$

Интерпретация полученного результата: для испытуемых очень низкого уровня подготовленно-

Таблица 5

Предполагаемый уровень подготовленности испытуемых, θ_j (здесь берётся с шагом 1 логит)	$L = j(\theta - \beta_j)$	e^{-L}	$1 + e^{-L}$	$P_j(\theta)$
-3,0	$(-3-1) = -4$	54,598	55,598	0,018
-2,0	$(-2-1) = -3$	20,086	21,086	0,047
-1,0	$(-1-1) = -2$	7,389	8,389	0,166
0	$(0-1) = -1,0$	2,718	3,716	0,269
1,0	$(1-1) = 0$	1	2	0,500
2,0	$(2-1) = 1,0$	0,368	1,368	0,731
3,0	$(3-1) = 2$	0,135	1,135	0,881

сти, равного $-3,0$ логита, вероятность правильного ответа на задание уровня трудности $1,0$ логит равна $0,018315$, что вполне согласуется с естественной педагогической логикой: правильный ответ малоподготовленного испытуемого на задание среднего уровня трудности маловероятен. Ещё менее вероятным может стать правильный ответ такого испытуемого на трудное задание.

Расчёты для остальных значений приводятся в таблице 5.

Если посмотреть внимательно на строки таблицы 5, то можно заметить, что например, уровень подготовленности испытуемого $\theta_j = 1,0$, который в точности равняется уровню трудности задания $\beta_j = 1,0$, то вероятность правильного ответа равна $0,5$. С увеличением уровня подготовленности испытуемых растёт и вероятность правильного ответа на это задание. Если по полученным значениям $P(\theta)$ построить график, то он будет выглядеть как на рис. 3.

Графические образы заданий

Для предсказания тестовых свойств заданий и для первоначальной ориентации в их потенциальных возможностях профессиональные разработчики теста используют построение геометрических и графических образов заданий.

Известны два метода построения графических образов заданий.

Первый метод — построение графиков долей правильных ответов испытуемых на задание j ,

в зависимости от уровня подготовленности испытуемых. При построении графика каждого задания желательно,

чтобы число испытуемых было как можно больше. При этом условии появляется возможность разделить всё множество испытуемых на так называемые балльные группы, с достаточным числом испытуемых в каждой из них. Значения балльных групп откладываются на оси абсцисс. Создаются отдельные группы тех, кто имеет только один правильный ответ, два, три и т.д. Соответственно, на педагогическом языке их можно называть

группами единичников, двоечников, троечников, четверочников, пятерочников, шестерочников и т.д.

В каждой такой балльной группе подсчитывается доля правильных ответов. Значение этой доли в каждой балльной группе и является проекцией на ось ординат. В итоге на плоскости откладываются точки⁵¹.

По этим точкам можно подобрать кривую, наилучшим образом описывающую полученные данные⁵² (Рис. 5).

Второй метод построения графических образов задания — аналитический, аналогичный методу, представленному в таблице 5. При этом могут использоваться различные функции — одно, двух или трёхпараметрические. И хотя все три функции-модели нередко используются в разработке педагогических тестов, наибольшее предпочтение отдаётся модели измерения Г. Раша.

При анализе характеристических кривых обращают внимание на три характеристики: на расположение точки перегиба графика относительно оси X , на место пересечения с осью ординат и на меру крутизны кривой.

Чем правее на плоскости располагается кривая задания, тем оно труднее для испытуемых.

Чем выше точка пересечения начала графика при значениях $\theta < -3$ с осью ординат, тем выше потенциальный или реальный уровень угадывания правильного ответа на задание j .

Чем больше крутизна функции $P(\theta)$, тем задание j точнее измеряет интересующее свойство испытуемых на своём интервале измерения подготовленности испытуемых.

В педагогическом тесте предпочтительно иметь задания с равной крутизной, потому что пересе-

⁵¹ Baker, F.B. The Basics of Item Response Theory. 2 ed. Hieneman, Portsmouth, New Hampshire, 2001.

⁵² Там же.

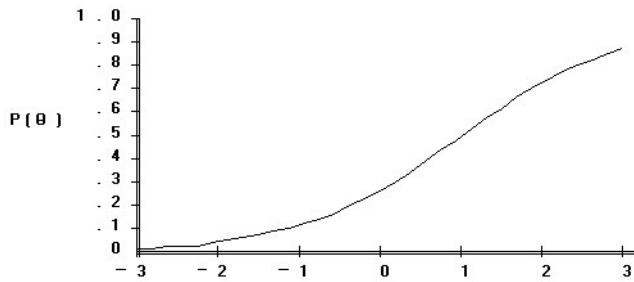


Рис. 3. График функции $P(\theta)$ при уровне трудности задания $\beta_j = 1$

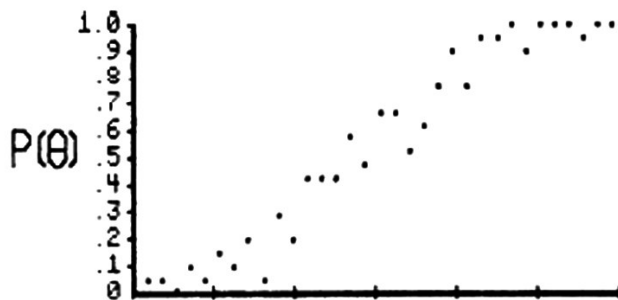


Рис. 4. Эмпирически полученные значения долей правильных ответов на задание теста в каждой балльной группе испытуемых

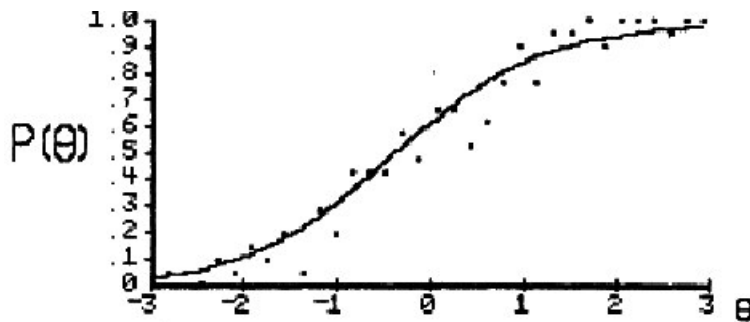


Рис. 5. Функция задания, подобранная по эмпирическим данным ответов испытуемых на задание теста

чение кривых является дополнительным источником погрешностей измерения. Примерное число заданий по одной не очень большой учебной дисциплине – 30, что обеспечивает достаточную точность педагогических измерений. Лучше, чтобы общее время тестирования не превышало более сорока минут. Увеличение времени, как и его уменьшение, снижает дис-

персию тестовых результатов, а следовательно ухудшает дифференцирующую способность всего теста. Как следствие, снижается и общий показатель точности измерений. □