

377
181
A-181

САНКТ - ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

На правах рукописи

АВАНЕСОВ ВАДИМ СЕРГЕЕВИЧ

МЕТОДОЛОГИЧЕСКИЕ И ТЕОРЕТИЧЕСКИЕ ОСНОВЫ
ТЕСТОВОГО ПЕДАГОГИЧЕСКОГО КОНТРОЛЯ

Специальность 13.00.01 - Теория и история педагогики

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
доктора педагогических наук

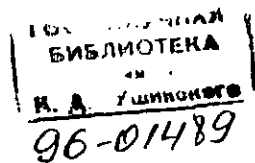
САНКТ - ПЕТЕРБУРГ
1994

OKX
Работа выполнена в Исследовательском центре проблем качества подготовки специалистов при Московском институте стали и сплавов (технологический университет) Государственного Комитета Российской Федерации по высшему образованию.

Официальные оппоненты: доктор педагогических наук,
профессор В.И. ГИНЕДИНСКИЙ

доктор педагогических наук,
профессор В.В. КАРПОВ

доктор психологических наук,
профессор В.Л. МАРИШУК



Ведущая организация - НИИ ВЫСШЕГО ОБРАЗОВАНИЯ
ГОСКОМИТЕТА РОССИЙСКОЙ ФЕДЕРАЦИИ ПО ВЫСШЕМУ ОБРАЗОВАНИЮ.

Защита состоится "14" декабря 1994 года в 12
часов на заседании специализированного совета Д 063.57.48 по
защите диссертаций на соискание ученой степени доктора педа-
гогических наук при Санкт-Петербургском государственном уни-
верситете (199034, Санкт - Петербург, набережная Макарова,
д. 6, факультет психологии).

С диссертацией можно ознакомиться в Научной библиотеке
им. А.М. Горького Санкт - Петербургского государственного
университета (С - Петербург, Университетская наб. 7/9.

Автореферат разослан "6" ноября 1994г.

УЧЕНЫМ СЕКРЕТАРЬ
специализированного совета
доктор педагогических наук

Г.И. МИХАЛЕВСКАЯ

5, 1, 11, 20

377
A 181

ОБЩАЯ ХАРАКТЕРИСТИКА ДИССЕРТАЦИИ

Разработка государственных образовательных стандартов, внедрение новых технологий обучения и контроля знаний вновь, после длительного перерыва, привлекли интерес широкой педагогической общественности к тестам. Являясь частью многих педагогических новаций, тесты позволяют получить объективные оценки уровня знаний, умений, навыков, проверить соответствие требований к подготовке выпускников заданным стандартам, выявить пробелы в подготовке студентов. В сочетании с персональными ЭВМ и программно - педагогическими средствами тесты помогают перейти к созданию современных систем адаптивного обучения и адаптивного контроля - наиболее эффективных, и к сожалению, наименее применяемых у нас форм организации учебного процесса.

Столь значимые для образования возможности педагогических тестов могут проявиться только при условиях изменения общей организации учебного процесса: перехода от привычных групповых форм занятий к индивидуальным, автоматизированным, создания системы стимулов, повышающих персональную ответственность студентов и преподавателей за результаты своего труда. Еще одно условие - это преодоление распространенного мнения о кажущейся простоте педагогического теста и о простоте методов его создания. Заблуждения такого рода являются причиной появления множества псевдотестов, называемых тестами либо по недоразумению, либо из-за сложившейся у нас привычки называть вещи не своими настоящими именами.

Путь к созданию тестов лежит через овладение нетрадиционным для отечественной педагогики направлением, называемым педагогическое измерение (от англ. Educational measurement). Это научное направление возникло в начале нашего века, и с той поры усилиями многих ученых оно достигло заметных успехов. Среди зарубежных авторов, сделавших много полезного для развития Educational Measurement - A. Anastasi, K. Bereiter, A. Birnbaum, W. Brown, L. J. Cronbach, R. L. Ebel, R. Fisher, H. Gulliksen, R. Hamblin, G. F. Kuder, F. M. Lord, W. Masters, J. Nunnally, G. Rasch, M. W. Richardson, G. Sax, C. Spearman, M. Stone, E. L. Thorndike, R. L. Thorndike, D. Weiss и др. В западных странах, особенно в США, к настоящему времени удалось накопить обширный теоретический и фактический материал. В облас-

ти педагогических измерений там ежегодно выпускаются монографии, специальные научные журналы, пособия; с помощью тестов проверяется подготовленность миллионов школьников и студентов, всех военнослужащих, большей части производственного и управленческого персонала.

Попытки развить педагогические измерения и разработать на этой основе тесты предпринимались неоднократно и в нашей стране. Первые попытки относятся к началу 20-х годов; они связаны с такими авторами как П.П. Блонский, А.А. Болтунов, М.С. Бернштейн, Н.А. Бухгольц, С.Г. Геллерштейн, Е. В. Гурьянов, А.Б. Залкинд, М.И. Зарецкий, Г.С. Костюк, С. М. Ривес, А.М. Мандрыка, А. А. Смирнов, М.В. Соколов, А.А. Толчинский, Н.К. Удовиченко, Б.А. Шевырев, А.М. Шуберт, Т.А. Эри и др. Эти исследователи сделали первые попытки определения теста, исследования формы тестовых заданий, критериев качества тестов. Однако их работа в начале тридцатых годов была свернута: вначале с помощью сильно идеологизированной критики, а затем, в 1936 году, официальным запретом Совета Народных Комиссаров. Вторая попытка возрождения культуры педагогических измерений отмечается с 60х годов. Она проходила на волне интереса к программированному обучению и программированному контролю. Среди авторов этого периода выделяются В.И. Огарелков, М.А. Соколова, Г. А. Сатаров, В. С. Черенков и др. По педагогическим тестам опубликовали работы М. С. Бернштейн, В.П. Беспалько, Б.З. Гинзбург, К. М. Гуревич, М.Н. Королев, Н.М. Розенберг, И.А. Цатурова и другие, в том числе - автор этой диссертации.

В последующие годы оживился интерес к общим вопросам научной организации контроля знаний студентов и школьников, к вопросам методологии и теории педагогики, управления учебным процессом. Среди авторов, успешно занимающихся этими вопросами - В.И. Андреев, Ш.А. Амонашвили, С.И. Архангельский, Г.М. Афонина, Г. А. Балл, В.С. Безрукова, Ю.К. Васильев, А.А. Вербицкий, В.И. Гинецинский, Г.Г. Городничий, И.А. Данилов, В.И. Журавлев, В.И. Загвязинский, И.И. Зимняя, В.В. Карпов, М.И. Катханов, В.В. Краевский, И.И. Кобыляцкий, А.В. Кочергин, И.И. Кувшинова, Н.В. Кузьмина, Ч.Куписевич, М.М. Левина, И.Я. Лернер, В.П. Мизинцев, Г.И. Михалевская, А.Т. Молибог, Н.Д. Никандров, В.М. Полонский, В.М. Рогинский, Н.А. Селезнева, М.Н. Скаткин, В.А. Сластенин, Я.С. Турбович, А. И. Уман.

Т.И. Шемова, Н.М. Шамаев, В. А.Якунин и мн. др. Разработкой и применением психологических тестов успешно занимались Ю.З. Гильбух, К.М. Гуревич, В.Л. Марищук Л.Н. Собчик, Л.Т. Ямпольский и мн. др. Сколь бы, однако не были значимы исследования упомянутых авторов, в них проблема методологических и теоретических основ тестового педагогического контроля, можно сказать, не затрагивалась.

Актуальность исследования вытекает из необходимости разработки научно - обоснованных тестов для проверки соответствия знаний образовательным стандартам. Введение последних в практику управления образовательными учреждениями предусмотрено законом Российской Федерации об образовании. Разрыв, образовавшийся между запросами реформирующегося образования в качественных тестах объективного контроля подготовленности студентов и сложившейся традицией оперирования упрощенными схемами субъективного выставления пятибалльных отметок, отсутствие достаточного числа научных разработок по ключевым вопросам методологии и теории тестового контроля является еще одним указанием на актуальность исследования. Научный аспект актуальности исследования вытекает из различий в развитии тестовой науки: статистические аспекты теории тестов хорошо разработаны на Западе и слабо изучены у нас. Сравнительно меньше там изучены педагогические вопросы организации процесса тестового контроля и мало оказались изученными вопросы методологии тестового контроля.

Объект исследования - процесс обучения и контроля подготовленности студентов высшей школы. В этом процессе текущие и итоговые оценки рассматриваются как эмпирически фиксируемые показатели результатов совместной учебной деятельности студентов и преподавателей по формированию необходимых в профессиональной подготовке.

Предмет исследования - обоснование ключевых сторон тестового контроля: возникновение и развитие тестов, форма и содержание тестовых заданий, предмет, цели, функции, принципы, критерии качества и эффективности тестового контроля.

Исследовательская установка заключается в стремлении определить систему взаимосвязанных понятий, дать эффективное определение теста, ввести в научный оборот новые принципы, провести философско - педагогический анализ вопросов формы и содержания тестовых заданий, проверить правдоподобность ги-

потез, выдвигаемых при разработке тестов. **Гипотезы**, проверяемые в эмпирических разделах исследования: 1. Оценки экспертов - преподавателей, участвующих в эксперименте по валидации теста, конкордантны между собой и коррелируют с тестовыми баллами. 2. Предлагаемая модель оценки прироста тестовых результатов соответствует задаче объективной оценки прироста результатов обучения.

Цель - провести периодизацию процесса возникновения и применения педагогических тестов, определить методологические основы научного исследования и эффективного преобразования практики тестирования, включающие вопросы статуса, сущности и цели педагогических измерений, сформулировать основные функции и принципы тестирования, положения теории организации процесса тестового педагогического контроля, определить систему критериев качества и эффективности тестов и тестовых заданий, создать метод объективной оценки прироста результатов обучения.

Задачи: 1. Определить и обосновать актуальные методологические и теоретические вопросы тестового контроля.

2. Провести исследование по истории возникновения и применения педагогических тестов, возникновения и развития современных моделей педагогического измерения;

3. Сформулировать эффективное определение педагогического теста, рассмотреть вопросы формы и содержания;

4. Провести методологический анализ основных целей и функций педагогического измерения, сформулировать систему принципов научной организации тестового контроля.

5. Систематизировать научный аппарат основных критериев качества тестов, обосновать критерий эффективности тестов и тестовых заданий.

6. Разработать метод объективного сравнения прироста результатов обучения.

Методологической основой исследования является материалистическая диалектика как учение о связи явлений, о ответственности в развитии общественных явлений. С диалектикой неразрывно связан диалектический метод познания. Диалектический метод рассматривается как основа исторического подхода к объяснению и оценке взаимосвязи формы и содержания, общего и частного в содержании тестов. На уровне общенаучных принципов и форм исследования в качестве методологической основы

исследования выступают логика научного исследования, с ее формальным аппаратом, системный подход и философские принципы измерений. На уровне конкретно - научной методологии в качестве основы выступают методология педагогической науки и принципы проведения педагогических измерений.

Источники исследования - отечественная и иностранная литература по методологии и теории педагогики, по теории тестов и по методике тестирования, актуальные потребности в создании новых тестов, практика обучения профессорско - преподавательского состава вузов методам тестового контроля знаний и разработки тестов по вузовским дисциплинам.

Этапы исследования. Первый (1971-1978): ставились и решались задачи изучения истории и классической теории тестов, разработки тестов и учебных спецкурсов, подготовки публикаций по тестовой тематике.

Второй (1979-1988): проводилось дальнейшее изучение теории тестов, готовились спецкурсы, научные публикации, монография, разрабатывались тесты и тестовые задания.

Третий (1989-1993): проводилось изучение новой теории тестов, разрабатывались программы для ЭВМ, позволяющие моделировать педагогические тесты, вести статистическую обработку данных и объективно оценить прирост учебных результатов, велась работа над методологической литературой и проблемой, над написанием ряда учебных пособий, разрабатывались и читались спецкурсы по тестовой проблематике для руководящего состава высшей школы - проректоров по учебной работе, спецкурсы для преподавателей вузов, техникумов, педагогических училищ, школ, военных, пограничных и др.

Методы исследования:

- Диалектический метод как всеобщий метод познания;
- Общенаучные методы теоретического исследования - классификация, сравнение, аналогия, описание и объяснение, индукция и дедукция, научное доказательство, анализ и синтез, обобщение, абстрагирование, формализация, математизация, исторический метод и другие.
- Общенаучные методы эмпирического исследования - наблюдение, измерение, эксперимент;
- Статистические методы - расчет средних арифметических, показателей вариации, проверка нормальности распределения тестовых результатов, корреляционный и дисперсионный анализ;

- Специально-педагогические методы - педагогическое измерение, математические модели контроля знаний.

Научные положения, выносимые на защиту:

1. История возникновения и функционирования тестового контроля разделяется на периоды: первый - **предыстория** - с древности до конца XIX века, когда были распространены домашние формы контроля знаний и способностей; второй период - **классический** - продолжался с начала 20-х до конца 60-х годов, в течение которого создавалась классическая теория тестов; и третий период - **технологический** - начавшийся с 70-х годов - время разработки методов адаптивного тестирования и обучения на основе развития новой, Item Response Theory, представляющей, по сути, методологию эффективной разработки тестов и тестовых заданий для параметрической оценки испытуемых по измеряемому латентному качеству.

2. **Основные понятия: I. Педагогический тест:** - это оптимальная система фасетных заданий специфической формы, определенного содержания, возрастающей трудности, позволяющая качественно оценить структуру и эффективно измерить уровень знаний, умений, навыков и представлений по учебной дисциплине. **Тестовое задание** - это дидактически и технологически эффективная единица контрольного материала, часть теста, которая отвечает требованиям предметной чистоты содержания (или одномерности), содержательной и логической правильности, правильности формы, приемлемости геометрического образа задания, а также отвечает требованиям аксиомы локальной независимости, технологичности, эффективности, известной трудности и коррелируемости задания с критерием. **Задание в тестовой форме** - это задание, к которому, помимо содержания, предъявляются требования тестовой формы: одинаковость инструкции для всех испытуемых; адекватность инструкции форме и содержанию задания; краткость, формулирование заданий в логической форме высказывания - истинного или ложного, зависящего от ответа; правильность расположения элементов задания, краткость, фасетность (варьирование содержания в пределах укрупненной единицы знания), определенное место для ответов, одинаковость оценки в рамках принятой формы.

3. **Основные формы тестовых заданий:** это задания закрытой и открытой форм, задания на установление соответствия и правильной последовательности. Каждая форма позволяет отра-

зитель специфические для нее элементы содержания контрольного материала. Выбор и разнообразие используемых форм заданий зависит от цели тестирования, содержания, технических возможностей и от уровня обученности профессорско - преподавательского состава в вопросах тестового контроля.

4. Процесс тестового контроля может быть назван научно обоснованным, если он регламентируется группой специфических принципов научной организации: связи контроля с образованием, обучением и воспитанием; принципами объективности, справедливости и гласности; научности и эффективности; систематичности и всесторонности. Основные функции - диагностическая, обучающая, организующая и воспитывающая.

5. Применение **модифицированного метода справедливой оценки учебных достижений** на основе предложенной модели и учета индивидуальных параметров обучаемости позволяет объективно оценить прирост результатов в зависимости от уровня начальной подготовленности и индивидуальных способностей, прогнозировать учебные достижения.

Новизна заключается в попытке разработки педагогических основ научной организации тестового контроля - определении целей тестирования на современном этапе, системы принципов контроля, фундаментальной роли педагогических измерений и методологии тестового контроля. Уровень теоретической новизны определяется тем, что в работе сделана попытка восполнить имеющиеся в мировой литературе пробелы в вопросах возникновения и определения тестов, дополнить известные критерии оценки тестов критерием эффективности, ввести принципы композиции тестовых заданий, сформулировать новое определение педагогического теста, дать периодизацию истории тестового контроля, привести ранее отсутствовавшие в литературе результаты исследования философско - педагогических вопросов соотношения содержания и формы тестовых заданий, обосновать и ввести в научный оборот модифицированный метод объективного сравнения и справедливой оценки прироста учебных достижений. Уровень практической новизны определяется готовностью результатов исследования к внедрению в работу образовательных учреждений и уже полученными результатами довольно широкого внедрения в работу вузов и других учебных заведений.

Методологическая и теоретическая значимость вытекает из

решения ключевых вопросов, связанных с разработкой основ специально - научной, для педагогики, методологии тестового контроля: целей и задач, обоснования роли педагогических измерений, формирование принципов научной организации тестового педагогического контроля в высшей школе, создание метода справедливой оценки прироста учебных достижений в зависимости от уровня начальной подготовки; переосмысление таких положений теории тестов, как определение теста и тестовых заданий, их формы, критериев эффективности и качества. Методологическая и теоретическая значимость исследования состоит также в возможности применения результатов к другим, менее исследованным формам контроля подготовленности студентов.

Практическая значимость диссертации заключается в научном обеспечении разрабатываемой сейчас работы по созданию методов объективного контроля соответствия знаний образовательным стандартам, в разработке практических вопросов научной организации обучения и контроля знаний на основе использования тестов, тестовых заданий и заданий в тестовой форме, к конструированию тестов на основе введенного здесь понятия профессионально подготовленного теста и критерия его эффективности. Кроме того, рассматриваемые в работе модели измерения позволяют практически моделировать тест с заданными критериями качества без сбора дополнительной эмпирической информации. Проведенная классификация форм тестовых заданий и разработка требований к каждой форме помогает преподавателям в практической работе по формулированию тестовых заданий. Выдвинутые в работе принципы организации тестирования позволяют приступить к созданию внутривузовских и региональных систем тестового контроля знаний, создать новые учебные курсы для обучения профессорско - преподавательского состава, аспирантов и студентов педвузов.

Достоверность и обоснованность научных результатов обеспечивается преимущественной опорой на основные и оригинальные источники научной информации ведущих авторов в области теории и практики тестирования - источники, поступившие по заказам автора в российскую государственную библиотеку из главных библиотек развитых, в тестовом отношении, стран мира. Другим источником достоверности и обоснованности результатов является использование современной технологии разработки тестов, включающей, как рефлексивный механизм, сово-

купность методов проверки и перепроверки качества тестовых данных, соответствие результатов измерения критериям надежности, валидности и эффективности. Использование статистических методов и математических моделей давало возможность абстрагироваться от случайных явлений и оперировать в процессе исследования устойчивыми закономерностями. При обосновании соответствия избранной модели эмпирическим данным использовались статистические критерии χ^2 Пирсона и F-критерий Фишера. При определении достоверности различий оценок экспертов использовался дисперсионный анализ.

Апробация результатов исследования проводилась многократно в процессе докладов, лекций и научных разработок: - на научных конференциях и семинарах - вузовских, республиканских, всесоюзных, международных, в том числе российско - французском и российско - американском семинарах (1993); - на всесоюзных совещаниях ректоров (1990г.), проректоров вузов (1990-1991 гг.); - на всесоюзном совещании - семинаре зам. председателей учебно-методических объединений вузов Гособразования СССР (1991); - на занятиях с профессорско - преподавательским составом вузов и с преподавателями школ.

Внедрение результатов в педагогическую практику проводилось в процессе преподавания спецкурсов по тестам и по методам статистической обработки тестовых результатов в ИИК АПН СССР (1971-1990 гг), в ходе занятий с аспирантами АПН СССР (1985- 1987 гг.), в Учебном центре при Исследовательском центре Гособразования СССР (1989-1992 гг.), в институтах повышения квалификации и при проведении выездных занятий с профессорско-преподавательским составом вузов России, Казахстана, Белоруссии, Украины. Под руководством автора вопросам теории и методики тестового контроля обучено свыше четырех тысяч преподавателей, доцентов, профессоров, деканов и проректоров по учебной работе. Проведены многочисленные занятия с преподавателями и руководителями отдельных вузов, с профессорско - преподавательским составом учебно - методических объединений медицинских, педагогических, технических и многих других вузов, с преподавателями школ и техникумов в разных городах России. Недельные циклы интенсивных занятий по теории и методике тестового контроля проведены со студентами физического факультета Поморского педагогического университета (Архангельск) и для учителей сельских школ.

Результаты исследования внедрялись в практику посредством издания учебных и методических пособий по тестовой проблематике, научных статей и тезисов конференций, в том числе международных. Все эти результаты представлены в работах автора. Диссертация состоит из введения, шести глав текста с выводами по каждой главе, заключения, практических рекомендаций и списка литературы. Список содержит 212 источников на русском языке и 152 - на английском.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Первая глава - "ТЕСТОВЫЙ ПЕДАГОГИЧЕСКИЙ КОНТРОЛЬ КАК ПРЕДМЕТ МЕТОДОЛОГИЧЕСКОГО И ТЕОРЕТИЧЕСКОГО ИССЛЕДОВАНИЯ: ПРОБЛЕМНАЯ СИТУАЦИЯ, ПОСТАНОВКА ПРОБЛЕМЫ И ЗАДАЧ" начинается с исследования проблемной ситуации. Истоки сложившейся ситуации прослеживаются с момента, когда тесты запретили как "буржуазные и вредные". С тех пор в вузовской и школьной практике обучения тестам не уделялось достаточного внимания.

Сравнение с западными странами показывает, что они опережают нас по масштабам работы, по числу публикаций, по подготовке научных кадров, уровню и качеству развития теории тестов, по технической и программно - вычислительной оснащенности тестового процесса. Наиболее развитые в тестовом отношении страны - Нидерланды, США, Англия, Япония, Дания, Израиль, Канада, Австралия, Новая Зеландия - оторвались от нас дважды. В первый раз, после 30-х годов, они развивали классическую теорию тестов и основанную на ней практику тестирования миллионов студентов и школьников. Во второй раз - с конца 70-х годов, когда классическая теория тестов стала уступать свои позиции Item Response Theory /IRT/, не имеющей пока адекватного русского перевода. По тестовой проблематике там были защищены тысячи диссертаций, опубликованы книги и монографии, сотни тысяч статей, налажен выпуск специализированных научных журналов, созданы внутривузовские системы тестового контроля знаний. И сейчас с помощью тестов там проверяется подготовленность миллионов школьников и студентов, военнослужащих, управленческого персонала.

Проблемная ситуация. В практике эту ситуацию порождает противоречие между возрастающими запросами общества в качественном и экономически эффективном массовом образовании и невозможностью добиться такого без объективного тестового контроля. В теории проблемную ситуацию определяет наличие

пробелов в разработке понятийного аппарата тестового педагогического контроля, слабая исследованность вопросов формы тестовых заданий, содержания, соотношения формы и содержания, отсутствие принципов научной организации тестирования. В методологии тестового контроля проблемную ситуацию порождает фактическая неизученность процессуальной стороны не только тестового, но педагогического контроля вообще, вопросов научного обоснования его самого. По мнению В.М.Полонского, не раскрыт процесс оценки знаний, его компоненты и связь между ними, почти нет исследований, рассматривающих соотношение между целями и методами контроля системы знаний, нуждаются в дальнейшей разработке методы контроля системы знаний, способы определения диагностической ценности вопросов, индивидуальные и коллективные формы проверки. Н.М.Розенберг считает неисследованными методологические и методико - исследовательские основания теории дидактических измерений. Все перечисленное в еще большей мере относится к методологии тестового контроля.

Проблема методологического и теоретического обоснования тестового контроля охватывает два уровня знаний: 1) теорию организации процесса тестового контроля; и 2) научное обоснование этого процесса. Показано, что в западных странах хорошо разработана математико - статистические теории тестов, но недостаточно развита педагогическая теория тестов. Главные причина - прагматизм, присущий американскому подходу к тестированию. Проблема методологического и теоретического обоснования определена в диссертации как философско - социально - педагогическая. Философско педагогической ее можно считать потому, что решение проблемы зависит от использования достижения наук, относящихся к философским - это логика, гносеология. Социально - педагогической она названа потому, что в ней содержатся признаки, присущие общественным проблемам, такие как влияние результатов тестового контроля на большое число учащихся и их родителей, зависимость образования от качества тестов, от уровня педагогической культуры, от ценностей общества др. Педагогические аспекты этой проблемы - связь тестового контроля с педагогической наукой, определение предмета деятельности профессорско - преподавательского состава, специализирующегося по вопросам тестового контроля; выявление причин отличающегося отношения к тестам

в отечественной и в зарубежной педагогической науке; соотношение понятийного аппарата педагогики и тестологии, спорность названия последней, методы отбора содержания учебной дисциплины для разработки тестовых заданий.

Состояние научной проблемы подытожено в выводах по материалам первой главы: 1. Мало изучен самый древний период - предистории тестов, не хватает материалов по классическому периоду становления тестовой культуры, мало информации о современных моделях педагогического измерения, нет периодизации истории развития тестов. 2. В литературе нет аргументированного определения понятия "педагогический тест" и сопряженной с ним системы понятий, мало изучены вопросы содержания теста, и совсем мало исследована форма, нет принципов композиции тестовых заданий. 3. Педагогическое измерение и научно обоснованная оценка знаний не стали в отечественной педагогике основой для организации системы педагогического тестового контроля. 4. В литературе нет единства в вопросах о целях, функциях и принципах тестового педагогического контроля. 5. Ощущается потребность в расширении числа критериев качества тестов. 6. В отечественной литературе практически не ставится вопрос об объективном сравнении прироста тестовых результатов в процессе обучения, а имеющиеся в иностранной литературе методы такого сравнения оказываются недостаточно удовлетворительными. Каждый вывод положен в основу исследовательской задачи, решаемой в диссертации.

Вторая глава - ИСТОРИЯ ВОЗНИКНОВЕНИЯ И РАЗВИТИЯ ТЕСТОВОГО КОНТРОЛЯ начинается с вопроса о времени возникновения тестов. Выделено три периода. **Первый** в западной тестологии отождествляется с различными испытаниями. Соответственно считается, что тесты появились более четырех тысяч лет назад (Т.Л. Kelly, Р.Н. DuBois). В диссертации приводятся дополнительные факты испытаний в школах писцов Древнего Вавилона середины III тысячелетия до н.э, жрецов в храмах Древнего Египта и др. Возникновению тестов способствовали письменные контрольные работы и экзамены, впервые появившиеся в Древнем Китае. Применение этих форм в странах Запада даже в значительно поздний средневековый период наталкивалось на трудности психологического характера. В Англии сто лет ушло на слом консервативных традиций, препятствовавших применению письменных контрольных работ в учебном процессе, и столько

же, чтобы последние начали использоваться на приемных и выпускных экзаменах. Их ценность быстрее других оценил Орден Иезуитов, который распространил применение письменных работ во многих странах. В 1762 г. состоялось открытое выступление студентов Йельского университета против использования письменной формы экзаменов. В США потребность в создании общественно - государственной системы контроля знаний была осознана в 1885 году; в штате Нью-Йорк начала действовать экзаменационный совет, приложивший немало усилий для разработки методов объективного контроля. Начало общегосударственной оценки знаний положило создание там в 1900 году комитета по проверке знаний абитуриентов колледжей. В России конца XIX - начала XX века вопрос о тестах не обсуждался; в 1872-1875 гг. в Санкт - Петербурге шли дискуссии о полезности и вредности выставляемых ученикам баллов.

Второй период становления тестов проходил под знаком роста научности. Этот период охватывает время с начала XX века до конца семидесятых годов. Импульс к появлению первых тестов дала проблема исследования индивидуальных различий. В наибольшей мере этому способствовали F. Galton (1875), J. McKeen Cattell (1890) и др. Появление тестов учебных достижений связывается с J. M. Rice, опубликовавшего свою работу в 1897 году. На становление теории тестов заметно повлияло создание теории корреляции. Вторая теория легла в основу первой. Корреляция стала использоваться для обоснования надежности и валидности тестов; посредством коррелирования тестовых результатов с внешним критерием стало возможным получение информации о валидности теста. Предложенные K. Pierson статистические методы стали использоваться в большинстве других наук. Это был первый, но не последний вклад зарождавшейся теории тестов в общенаучную, исследовательскую и методическую культуру. Решающий вклад сделал C. Spearman, соединивший теорию физических измерений с корреляционными методами и с идеей психологического измерения, заложив, тем самым, фундамент классической теории тестов. В тридцатых годах нашего столетия возросла конфронтация в обществе и в научной жизни. Педагогика и психология обвинялись в схоластике, узком академизме, в отрыве от практики. Педологи и психотехники осуждались за эмпиризм, подражание западным образцам и увлечение тестами. В условиях авторитарно управ-

ляемого общества любой случай неправильного применения тестов идентифицировался с ошибочностью самого тестового метода. Наступивший в нашей стране застой в разработке педагогических тестов и их применении продолжался до 60-х годов.

Третий период в развитии тестового контроля начинается с семидесятых годов, когда стали появляться публикации по новой, математически более оснащенной и практически более эффективной теории, называемой по - английски Item Response Theory. Разработка этой теории, проходила в США, Дании, Нидерландах и в других странах, имеющих развитую высшую школу. Рассматривая матрицу тестовых данных как результат сопряжения множества испытуемых с множеством заданий, G.Rasch выдвинул идею зависимости результатов этого сопряжения от соотношения уровня подготовленности каждого испытуемого с мерой трудности каждого задания. Реализация этой идеи потребовала создания сопоставимой шкалы знаний и шкалы трудности заданий. Именно ему удалось ввести логарифмическую меру - логит трудности задания β_j , равный $\ln q_j/p_j$, где q_j - доля неправильных ответов на задание j , а p_j - доля правильных ответов; обе доли получаются из матрицы. Симметрично, была введена логарифмическая мера - логит знаний студента θ_i , равный P_i/q_i , где p_i - доля правильных ответов студента i на задания теста; q_i , соответственно, доля неправильных ответов. Вероятность правильного ответа на задания G.Rasch выразил посредством логистической функции

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

где $x_{ij} = 1$, если ответ любого испытуемого i на j -ое задание правильный; θ - уровень знаний, латентная переменная; β_j - уровень трудности j - го задания теста, измеряемой на латентном континууме знаний. Соответственно, вероятность правильного ответа на задание j , обозначаемая (Q_j), и равная, в теории вероятностей, $1 - P_j$, выражается так:

$$Q_j\{x_{ij} = 0 | \beta_j\} = 1 - \exp(\theta - \beta_j) / (1 + \exp(\theta - \beta_j))$$

Симметрично, возникла и модель, описывающая вероятность правильного ответа студентов с уровнем знаний θ_i на задания различного уровня трудности:

$$P_i\{x_{ij} = 1 | \theta_i\} = \exp(\theta_i - \beta) / (1 + \exp(\theta_i - \beta))$$

A. Birnbaum создал двухпараметрическую модель педагогического измерения за счет введения второго, после β_j , пара-

метра, дающего информацию о задании j с точки зрения крутизны характеристической кривой. Обозначив этот параметр буквой a_j , модель измерения приобрела вид

$$P_j(1|\beta_j, a_j) = \exp a_j(\theta - \beta_j) / (1 + \exp a_j(\theta - \beta_j))$$

Добавление параметра c_j , ассоциируемого с вероятностью угадывания правильного ответа в заданиях закрытой формы дало трехпараметрическую модель педагогического измерения:

$$P_j(1|\theta, \beta_j, a_j, c_j) = c_j + (1 - c_j) \exp a_j(\theta - \beta_j) / (1 + \exp a_j(\theta - \beta_j))$$

Все эти модели являются основным теоретическим итогом истории развития тестов; они сейчас используются в адаптивном обучении и адаптивном тестировании, основанные на оценке значений разности $(\theta_i - \beta_j)$ и вытекающей из нее оценки вероятности овладения учебными материалами различного уровня трудности. Они используются в создании банка тестовых заданий, в профессиональном отборе, аттестации студентов, в международных исследованиях уровня знаний и др.

В третьей главе - ОСНОВЫ ТЕОРИИ ТЕСТОВОГО ПЕДАГОГИЧЕСКОГО КОНТРОЛЯ - исследованы три главных вопроса: определение теста, содержание теста и форма тестовых заданий.

Определение теста рассматривается как самый важный вопрос теории. Анализ отечественной и иностранной литературы указывает на пробелы в этом вопросе, что проявляется в нетерминологичности имеющихся определений и в несущественности выделяемых в них признаков теста. Тест представляется, как "увлекательная игра", как "совокупность вопросов", или "серии вопросов и заданий, при выполнении которых можно ограничиться расстановкой цифр, односложными ответами" и т.п. Дано новое определение теста, содержащее термины - понятийные индикаторы, как критерии демаркации теста от псевдотеста. В качестве таковых включены термины: задание теста, содержание, форма, трудность и эффективность тестовых заданий. В исследовании предпринята попытка совершить отход от традиции рассмотрения теста как элементарного средства проверки, пробы, испытания и перейти к рассмотрению теста как системного педагогического средства контроля знаний, метода педагогического измерения. Помимо теста, в диссертации даны определения задания в тестовой форме и тестового задания, которые приведены в основных положениях, выносимых на защиту. Из перечисленных там требований вытекает обязательность эмпирической проверки тестовых заданий. Если судить по содержанию

проанализированных публикаций, то в нашей стране нет сейчас ни тестов, ни тестовых заданий, обоснованных, в достаточной степени, эмпирически и с точки зрения педагогической теории.

Анализ содержания теста. Содержание теста в диссертации определено как оптимальное отображение содержания учебной дисциплины в системе тестовых заданий. Чем глубже и полнее отображение, тем увереннее можно судить о содержательной валидности теста. Однако стремление к валидности теста за счет естественного расширения числа тем, разделов учебной дисциплины и за счет увеличения числа заданий в тесте нельзя признать рациональным, так как невозможно создать тест, содержание которого вбирало бы в себя все содержание учебной дисциплины. При создании теста ставится задача отобразить в нем то главное, что студенты должны знать в результате изучения. Содержание учебной дисциплины шире содержания тестовых заданий и потому не все основные знания, навыки и представления входят в тест; некоторые из них заметно связаны между собой, и могут, следовательно, замещаться. Отображение содержания образования в содержание тестов проводится посредством выделения укрупненных единиц знаний.

Уточнена система принципов отображения содержания учебной дисциплины в содержание теста: это принципы значимости, оптимальности, научной достоверности, возрастающей трудности заданий, комплексности и сбалансированности содержания теста, взаимосвязь формы и содержания, соотносённость с современным состоянием науки.

Исследование формы тестовых заданий. В создании тестов, как и в создании произведений искусства, успех зависит не только от оригинальности идеи, замысла и сюжета, но и от владения формой. Вот почему профессионально сделанный тест — это удачное соединение содержания с формой. Содержание не бесформенно; оно понимаемо только тогда, когда для него подобрана адекватная форма представления. Начиная с Аристотеля, форма представляется как идеальный принцип организации элементов. Идея организации плюс идеи упорядочения и существования содержания в каких-либо упорядоченных формах легли в основу еще одного нашего определения: форма тестовых заданий — это способ организации, упорядочения и существования (сохранения) содержания теста. В настоящей работе тестовая форма рассматривается как инвариант, позволяющий сконцентри-

ровать, в одном предложении, и содержание задания, и содержание ответа студента. Плодотворность такого подхода подтверждена в процессе одновременного обучения композиции тестовых заданий преподавателей различных дисциплин.

Проведена классификация форм и видов тестовых заданий. Выделены четыре основные формы тестовых педагогических заданий: 1) закрытой формы, в которой студенты выбирают правильные ответы; 2) открытой формы, где ответы дают сами испытуемые; 3) на соответствие, в которых элементам одного множества требуется поставить в соответствие элементы другого множества; и 4) на установление правильной последовательности, в которых устанавливается требуемая заданием последовательность действий, операций, вычислений - любой фрагмент алгоритмизируемой учебной и профессиональной деятельности.

Задания закрытой формы разделены на два варианта - с одним правильным ответом из числа предлагаемых и с несколькими правильными ответами; каждому варианту дана своя инструкция. Кроме того, выделено пять основных видов заданий закрытой формы: - с двумя, тремя, четырьмя, пятью и с большим числом ответов. На множестве заданий этой же формы автором определены следующие принципы композиции тестовых заданий: противоположности, альтернативы, градуированности, сочетания, классификации, удвоенной альтернативы, кумуляции. Выделена группа псевдотестовых заданий; к ним отнесены так называемые цепные задания, тематические и текстовые. В качестве критериев проведенной классификации выступали требования к тестовым заданиям.

В четвертой главе - МЕТОДОЛОГИЧЕСКИЕ ОСНОВЫ ТЕСТОВОГО ПЕДАГОГИЧЕСКОГО КОНТРОЛЯ - исследованы главные вопросы научного исследования и эффективного преобразования практики. Педагогическое измерение определяется как научно обоснованное отображение концептуально определяемой системы знаний в эмпирическую шкалу, представляющую собой числовую систему с отношениями. Цель измерения - получение численных эквивалентов интересующего свойства. Результатом процесса измерения является шкала, посредством которой испытуемым даются тестовые баллы. Процесс измерения раскрыт с позиции соединения теоретического и эмпирического подходов. Он начинается с определения основного понятия "Знание учебной дисциплины", выделения тем, разделов, выступающих в роли понятийных

каторов, затем с разработки заданий в тестовой форме, на-
мечаемых на роль эмпирических индикаторов. При этом использу-
ется операционализация понятий, позволяющая уточнить и дета-
лизировать основное понятие, поставить в соответствие поня-
тию понятийные и эмпирические индикаторы. Структурирование
последних методами факторного анализа дает новую эмпириче-
скую структуру понятия, обычно не совпадающую с концептуаль-
ной. Именно таким образом тесты позволяют закрепить теорию с
эмпирией, что делает их наиболее развитым методом педагогического
контроля. В качестве единицы тестового задания, в качестве единицы педа-
гогического измерения" - логит.

Педагогическое оценивание (анг. evaluation) представля-
ет собой процесс формирования и численного выражения мнения
об интересующем свойстве личности. Как отмечалось в нашей
работе 1978 г., для оценивания обычно используются шкала на-
именований и шкала упорядоченной классификации. Автором обо-
снованы две шкалы оценивания: - десятибалльная и одиннадца-
тибалльная. Последняя получена посредством добавления по
одному интервалу в 0,5 стандартного отклонения справа и
слева к известной в литературе (Guilford J.P. и др.) шкале
стандартных девяти единиц, что дает возможность лучшей диф-
ференциации испытуемых, важной при проведении научных иссле-
дований и рейтинга. Рейтинг определен в работе как метод
обоснованного упорядочения студентов с использованием шкалы
порядка. Распространенное у нас определение рейтинга как
кумулятивного текущих оценок и выставления итоговых пятибал-
льных оценок в данной работе не поддерживается.

Основные цели и функции тестирования рассмотрены в свя-
зи с целями образования. Выделена группа социально значимых
целей тестирования. Это комплектование кадрового состава по-
средством использования тестов в профессиональном отборе,
сравнительная оценка качества образования в различных вузах,
городах, республиках и регионах страны, а также сравнение
качества образования в разных странах мира. Решающая роль
отводится внутривузовскому контролю, основная цель которого
- обеспечение объективной оценки подготовленности абитуриен-
тов и выпускников вуза на основе педагогического измерения,
сравнение качества работы преподавателей кафедр, факульте-
тов, и на этой основе - создание стимулов улучшения работы.

Этим целям поставлено в соответствие итоговое и обучающее тестирование. Педагог - практик заинтересован в обучающем контроле знаний студентов; задания в тестовой форме он использует для выявления пробелов в знаниях и коррекции учебного процесса. Тестолог заинтересован в итоговом контроле. Драматическую историю тестового контроля частично объясняется противостоянием сторонников этих двух видов контроля. Наиболее зримо противостояние проявилось в критике заданий закрытой формы, в которой логика итогового автоматизированного контроля подменялась требованиями контроля обучающего. Основная литература по тестам не проясняла различия в целях и видах контроля. Подавляющее большинство работ по тестам написано с позиций итогового контроля, в то время как интересные возможности применения заданий в тестовой форме для обучения педагогикой почти не изучались.

Основные функции контроля - диагностическая, обучающая, организующая и воспитывающая. **Диагностическая функция** вытекает из самой сущности контроля, нацеленного на выявление уровня и структуры подготовленности. Эта функция дает начало педагогической диагностике. Цель диагностики - получение валидной информации. **Обучающая функция** реализуется при использовании заданий в тестовой форме для выявления пробелов, дифференциации знания от незнания: этому способствует разделение знаний на предлагаемые, приобретаемые, проверяемые, применяемые, сохраняемые и забываемые. Содержание теста, тестовых заданий и заданий в тестовой форме образуют проверяемые знания. Перечисленные знания не совпадают по широте, глубине, конкретности, обобщенности, осознанности и прочности. Выявление несовпадающих элементов образует рефлексивную и мотивационную основу обучающей функции тестового контроля. **Организирующая функция** контроля проявляется в проведении учебного процесса. В зависимости от результатов контроля принимается решение о необходимости проведения дополнительных занятий и консультаций, об оказании помощи неуспевающим студентам, о поощрении хорошо работающих студентов и педагогов. Заметную роль в реализации организующей функции контроля играет активизация учебной деятельности студентов посредством разработки современных средств обучения. **Воспитывающая функция** контроля связана с повышением учебной мотивацию личности студента, формирования ответственности за результаты своего

учебного труда, формирования установок на сотрудничество студентов и преподавателей в процессе приобретения знаний, на самоорганизацию процесса учения и др.

Принципы научной организации тестового контроля - это основные требования, которыми руководствуются преподаватели в своей деятельности; они возникают в результате обобщения передового опыта, а также в процессе развития методологии и теории. Соответственно, формулируются принципы практической и теоретической деятельности. Одна из задач методологии тестового контроля - разработка принципов научной организации этой деятельности. В педагогическую теорию и практику принципы вводятся как совокупность исходных правил, способствующих эффективности учебного процесса. Помимо известных в литературе общих принципов обучения и воспитания, в диссертации представлены следующие принципы:

Принцип связи тестового контроля с образованием и обучением. Он признан ассоциацией директоров американских школ, считающих что обучение без последующего тестирования - немыслимо; только по итогам применения тестов можно судить о необходимой обратной связи от контроля к обучению, и знать - что достигнуто и в каком направлении дальше двигаться;

Принцип объективности нацеливает на устранение субъективизма и предвзятости. Традиционный путь - формирование коллегиальной оценки, для чего создаются комиссии. Получаемые при этом оценки нередко называются объективными, хотя они продолжают оставаться субъективными, а точнее, интерсубъективными. Второй путь повышения объективности - это использование стандартных тестовых программ и технических средств;

Принцип справедливости и гласности означает одинаково доброжелательное отношение ко всем, без исключения тестируемым, открытость всех этапов контроля, своевременность ознакомления с результатами тестирования. При отсутствии системы объективного педагогического тестирования менее принципиальные и требовательные преподаватели нередко выглядят лучше тех, кто ответственно относится к оценке знаний студентов. Стремление приукрасить несовершенные показатели связано с формализмом, при котором главным становится не работа, а несовершенные показатели работы. Подмена такого рода приводит к работе на показатель, его овеществлению, реификации.

Принцип научности и эффективности предписывает необходимость разработки тестов на научной основе, проверки результатов на надежность и валидность. Особое значение для повышения качества контроля имеет методология и теория. Методология педагогического контроля определяется как учение об основных положениях, формах, методах, принципах научного исследования и эффективной организации практики. В круг основных методологических проблем входят формулирование главных идей, целей и задач, определение содержания, сущности и принципов тестового педагогического контроля. Наиболее эффективный путь влияния методологии на практику - через разработку теории научной организации тестового контроля знаний. Именно в организационной теории преломляются методологические регулятивы, которые помогают превратить практическую работу в научно - практическую.

Принцип систематичности и всесторонности. Им подчеркивается необходимость согласования целей и результатов текущего, рубежного, тематического и итогового контроля, его регулярности. Всесторонность акцентирует внимание на необходимости репрезентативного охвата основного содержания учебного курса, подлежащего контролю.

В пятой главе - **ИССЛЕДОВАНИЕ ОСНОВНЫХ КРИТЕРИЕВ КАЧЕСТВА И ЭФФЕКТИВНОСТИ ПЕДАГОГИЧЕСКИХ ТЕСТОВ** - рассматриваются критерии надежности, валидности и эффективности. Первый раздел этой главы - **НАДЕЖНОСТЬ РЕЗУЛЬТАТОВ ТЕСТОВОГО КОНТРОЛЯ** - позволил провести систематизацию концептуальных вопросов надежности тестов, отделить прикладные аспекты теории надежности от фундаментальных вопросов точности измерений, доказать ряд основных формул

Теоретическая часть содержит основы надежности тестов и обоснование исходного равенства $s_x^2 = s_t^2 + s_e^2$, где s_x^2 - дисперсия тестовых баллов, s_t^2 - дисперсия истинных компонентов измерения и s_e^2 - дисперсия ошибочных компонентов измерения. В отличие от имеющихся, в диссертации дано педагогико - статистическое определение параллельного теста: параллельными называются тесты, которые имеют сходное содержание, равные средние арифметические тестовых результатов, равные дисперсии и равные индексы корреляции. Далее приведено обоснование научного аппарата: это определение надежности теста $r_{xx} = s_t^2 / s_x^2$, где левая часть обозначает значение коэффициента на-

дежности, а правая - отношение дисперсии истинных компонентов измерения к общей дисперсии теста; определение дисперсии истинных компонентов измерения $s_t^2 = r_{xx} s_x^2$; дан вывод классической формулы определения надежности теста по P. J. Rulon $r_{xx} = 1 - s_e^2 / s_x^2$. Среди рассмотренных методов оценки надежности тестов: повторное тестирование испытуемых; использование параллельных тестов; подсчет числа ошибочных элементов в профилях испытуемых, с последующим использованием формулы L. L. Guttman и др. Проведено доказательство формул стандартной ошибки измерения $s_e = s_x \sqrt{1 - r_{xx}}$ и индекса надежности теста $r_{tx} = s_t / s_x$.

Прикладная часть раздела содержит результаты исследования таких вопросов как: а) предсказание, на основе регрессионной модели, истинных компонентов измерения T_i по эмпирическим данным X_i ; б) определение доверительного интервала для истинных компонентов измерения; в) расчет объема выборки для проведения экспериментальных исследований с коррекцией на недостаточную надежность теста; г) коррекция коэффициента корреляции на ненадежность теста и доказательство формулы коррекции на ненадежность теста и критерия.

В разделе **ОБОСНОВАНИЕ ВАЛИДНОСТИ ТЕСТОВ** подтверждена гипотеза конкордантности экспертных оценок в процессе валидации теста. Показано, что валидность педагогического теста - это мера обоснованности результатов измерения того свойства, которое тест измеряет по замыслу его создателей. Если валидность оценивается мерой соответствия содержания теста цели разработки теста, тот это случай обоснования валидности теста по содержанию. Другой случай - когда в основу теста кладется теоретическое понятие или концепция. При этом возникает проблема обоснования валидности теста как эмпирического референта отображаемого понятия; это пример концептуальной валидности теста, определяемой мерой логического соответствия понятия и отображающего его теста. Для прогнозирования результатов испытуемых при профотборе используется понятие прогностической валидности теста, мерой которой служит корреляция теста с критерием. Прогностическая валидность теста зависит не только от надежности теста, но и от надежности критерия. Необходимые для коррекции формулы приведены в диссертации вместе с результатами дисперсионного анализа; там же даны сравнительные значения надежности и валидности

тестов, позволяющие ориентироваться при оценки их качества

Далее исследованы гносеологические затруднения, вызванные низкими значениями надежности и валидности. В педагогическом контроле оперируют эмпирическими показателями в предположении их качественного отображения сущности явлений и стоящих за ними понятий. Но при переходе от теоретических понятий к их эмпирическим референтам - тестам и заданиям - немало теряется из-за возникающих трудностей отображения понятий в эмпирических показателях. Потеря бывает тем большей, чем ниже надежность и валидность тестов, из чего вытекает необходимость рефлексии относительно качества отображения понятий в показателях.

Эффективность тестов и тестовых заданий исследована в третьем разделе пятой главы. Тест эффективен, если он качественнее, быстрее и дешевле измеряет знания студентов. С понятием "эффективность" сопряжено понятие "оптимальность". Последнее трактуется как наилучшее из возможных вариантов, с точки зрения удовлетворения нескольким критериям, взятым поочередно или вместе. В определении эффективности теста обращается внимание на два ключевых элемента - это число заданий теста и уровень подготовленности студентов. Из теста с большим числом заданий можно сделать оптимальный выбор меньшего числа, что дает систему, не уступающую тесту с большим числом заданий. Тест с меньшим числом заданий в таком случае называется более эффективным. Помимо этого, эффективность теста оценивается с точки зрения соответствия уровня его трудности уровню подготовленности студентов. Эту оценку в литературе нередко относят к валидности, имея в виду идею валидности теста по отношению к уровню знаний тестируемых студентов. Эффективность теста можно оценить и с точки зрения дифференцированной точности измерения знаний студентов различного уровня подготовленности. Тест называется эффективным для измерения знаний студентов с уровнем, соответствующим точке θ_1 оси θ , если он обеспечивает в этой точке максимум информации о значении θ при минимуме числа заданий. Эффективность измерений достигается за счет дифференцированного подбора заданий требуемого уровня трудности для каждого студента, имеющего уровень знаний θ_1 . При эффективном тестировании каждый студент последовательно получает такие задания, чтобы на каждом шаге разность $\theta_1 - \beta_j$ уменьшалась по модулю, посте-

ленно приближаясь к нулю. Поэтому эффективный тест - это адаптивный тест.

Самый эффективный тест тот, который по трудности подобранных заданий точно соответствует уровню подготовленности. Соответствие трудности теста уровню подготовленности студентов оценивается показателем количества информации, получаемой в процессе измерения. Вопрос об эффективности тестирования каждого студента определенным набором заданий, в зависимости от уровня его подготовленности, становится одним из главных вопросов в современной теории и практике тестирования. Вот почему в тестовых системах моделируется адаптивный тест, задания которого подбираются в зависимости от текущих ответов испытуемого. Эффективность теста оценивается и с точки зрения дифференцирующей способности. При одинаковом числе заданий тест с большей дисперсией эффективней теста с меньшей дисперсией; отношение большей дисперсии к меньшей, с последующим умножением на сто, может служить в качестве показателя сравнительной эффективности теста. Эффективность тестовых заданий определяется корреляционными, регрессионными и факторными методами, а также расчетом значений информационной функции задания в точке θ_1 . Первые и вторые позволяют оценить вклад каждого задания в общую вариацию тестовых баллов; факторный анализ является хорошим методом проверки гомогенности теста. Значение информационной функции считается по формуле A. Birnbaum.

Шестая глава - **ОБЪЕКТИВНОЕ СРАВНЕНИЕ ПРИРОСТА ТЕСТОВЫХ РЕЗУЛЬТАТОВ В ПРОЦЕССЕ ОБУЧЕНИЯ** - начинается с краткого анализа известных методов объективации прироста учебных достижений. Это расчет разности между конечными и начальными результатами, оценка разности в процентах от начального, конечного и от максимально возможного результата, метод общей точки E. L. Thorndike, статистические и аналитические методы с использованием различных взвешивающих коэффициентов. Использование разности для справедливой оценки прироста результатов обучения затрудняется объективной тенденцией к ее снижению по мере приближения к пределу. Здесь полезна аналогия с приращением результатов прыгунов в высоту: в процессе тренировки новички прибавляют больше и быстрее, чем мастера; каждый сантиметр рекордной высоты последним дается несравненно труднее, чем новичкам. Поэтому при сравнении прироста в вы-

игрыше всегда \leftarrow и это закон для отмеченных случаев - те, у кого ниже начальные результаты, а в проигрыше те, кто начинает обучение (тренировку) с более высокого результата.

В разделе - ДИФФЕРЕНЦИАЛЬНАЯ МОДЕЛЬ ОЦЕНКИ СРАВНЕНИЯ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ - поставлена задача построить математическую модель, которая позволяет объективно сравнить результаты студентов, начинающих с разного уровня подготовленности. Введено формальное определение темпа прироста результатов $\mu = du(x)/dx/u(x)$. Теоретически бесспорно, что для определенного типа учебной, трудовой, воинской и прочих видов деятельности, имеющих вполне измеряемый предел достижений, темп прироста тем меньше, чем выше уровень обученности. Уменьшение темпа может носить линейный или, скорее, нелинейный характер. Для упрощения задачи принимается гипотеза линейности уменьшения темпа прироста результатов достижений по мере роста обученности. Математически это выражается в виде дифференциального уравнения $\frac{u'(x)}{u(x)} = A - B \cdot u(x)$, где A, B - два положительных параметра, подлежащие определению в процессе эксперимента. (Математическая часть разработана и опубликована в совместной работе с Ю. Л. Костижовским).

Если заданы начальные условия $u(0) = u_0$, где положительную величину u можно трактовать как начальную подготовку студента, то интегрируя уравнение, будем иметь

$$u(x) = \frac{A \cdot u_0 \cdot e^{Ax}}{A + B \cdot u_0 \cdot (e^{Ax} - 1)}, \quad (A, B > 0)$$

откуда вытекает первое следствие: при достаточно большом количестве проведенных занятий (условимся в этом случае считать $x \rightarrow \infty$), независимо от начальной подготовки i -го студента $u_0 > 0$, $u_i(\infty) = A_i/B_i$. Отношение A_i/B_i можно трактовать как прогнозируемый предел достижений данного студента. Естественно принять дополнительное условие, что начальная подготовленность каждого обучаемого не превышает возможного предела достижений: $u_0 \leq A/B$. Смысл положительного параметра A : положив $u \rightarrow 0$, $A = \lim_{u \rightarrow 0} \frac{u'(x)}{u(x)}$, откуда видно, что параметр A - это темп научения при нулевом уровне подготовленности.

Возникает вопрос - могут ли параметры A и B быть идентифицированы так, что с достаточной точностью теоретическая кривая обучения аппроксимировала бы реальный процесс учения?

Для обоснования ответа на данный вопрос в качестве экспериментального материала были использованы данные E.L. Thorndike и его сотрудников, полученные в опытах по обучению студентов письму левой (непривычной) рукой. Для определения индивидуальных параметров кривой учения - u_0 , A и B введено общепринятое условие минимума среднеквадратического отклонения теоретических данных от экспериментальных. В результате

получается выражение $Q(u_0, A, B) = \sqrt{\sum_{k=1}^N (u_{tk} - E_{tk})^2}$, где $E_{t1}, E_{t2}, \dots, E_{tN}$ - совокупность экспериментально найденных объемов научения i -го учащегося, соответствующих периодам 1, 2, ..., N.

Для поиска минимума функции $Q(u_0, A, B)$ используется метод градиента. Начальные приближения $u_0^{(0)}$, $A^{(0)}$, $B^{(0)}$ выбирались так, чтобы расстояние от начальной точки $(u_0^{(0)}, A^{(0)}, B^{(0)})$ до точки минимума функции было возможно меньшим. Результаты численного определения параметров кривой научения, хорошо аппроксимирующих экспериментальные данные, приведены в таблицах 6.1, 6.2 и 6.3 диссертации.

В разделе - ЭКСПЕРИМЕНТАЛЬНОЕ ОБОСНОВАНИЕ МЕТОДА ОБЪЕКТИВНОГО СРАВНЕНИЯ ТЕСТОВЫХ РЕЗУЛЬТАТОВ - предложенная дифференциальная модель использовалась для проверки гипотезы пригодности модели, для прогнозирования конечных результатов и для создания нового метода сравнения результатов обучения. Если на основании экспериментальных кривых обучения для совокупности учащихся идентифицированы соответствующие каждому из них пары (A_1, B_1) , (A_2, B_2) , ..., (A_R, B_R) , где R - количество учащихся в группе и начальная подготовленность каждого соответственно была u_{10} , u_{20} , u_{R0} , то трудность в оценке индивидуальных достижений каждого учащегося вызвана их различной начальной подготовленностью. При одинаковой начальной обученности $u_1(0) = u_2(0) = \dots = u_R(0) = u_0$ знания параметров дает возможность предсказать объем научения у каждого обучающегося в заданный период N. Для обучаемого i этот объем считается по формуле

$$u_i(N) = \frac{u_0 * A_i * e^{A_i * N}}{(e^{A_i * N} - 1) * B_i * u_0 + A_i}$$

где i - номер испытуемого; u_0 - условная общая точка начальной подготовленности; N - число занятий.

Метод общей точки вводится с той же целью, с какой делал

это E.L. Thorndike, но в усиленном варианте - с использованием дифференциальной модели и с учетом индивидуальных параметров. Для каждой пары студентов с индексами $1 \leq l, m \leq N$ со свойством $u_1(N) < u_m(N)$ естественно считать, что m -й студент достиг лучших результатов, чем l -й. В качестве u_0 лучше брать среднюю арифметическую. Свойство решений дифференциального уравнения таково, что независимо от начальной подготовленности студента u_0 , $u_l(x) \rightarrow A_l/B_l$. Если группа студентов пересчетом приведена к одному и тому же начальному уровню обучения u_0 , то наибольшего успеха достигает тот, кто обладает большим значением $u_{l0}(\infty) = A_{l0}/B_{l0} \geq u_l = A_l/B_l$. Результаты эмпирической проверки применения метода модифицированной общей точки представлены в таблицах шестой главы.

В заключении подведены основные итоги исследования и даны практические рекомендации.

Выводы: I. История развития педагогических тестов - это результат многовекового процесса развития форм и методов педагогического контроля. Выделено три основных периода. Первый период - контроль, основанный на интуиции и здравом смысле. Представленные данные опровергают распространенную на Западе точку зрения о четырехтысячелетней истории тестов. Это период предистории. Второй период назван классическим. Третий период - результат воздействия технологической революции, охватившей ключевые сферы общества, и в том числе - образование, в котором автоматизированный самоконтроль и тестовый контроль знаний начали играть возрастающую роль.

2. Самым существенным фактором современного периода научного развития тестового контроля стала Item Response Theory, которая на Западе применяется для разработки современных тестов, создания адаптивных обучающих и контролирующих программ, профессионального отбора. Математические модели этой теории замыкают длительный путь исторического развития, который прошли тесты с древности до наших дней.

3. Тест отличается от множества других методов проверки знания составом, структурой, а также качеством и эффективностью получаемой информации. Состав теста существенно зависит от цели и времени тестирования, содержания и формы заданий. Тест можно считать научно обоснованным тогда, когда все задания сформулированы в соответствии с установленными требованиями. Включение в определение теста определяющих призна-

ков позволяет отнести это определение к числу эффективных.

4. Основные требования к содержанию теста - это требования предметной и формальной чистоты, логической правильности формулирования тестовых заданий, уровень трудности, степень значимости учебного и контрольного материала в структуре знаний. Сформулированы принципы отбора содержания: научная достоверность, соответствие современному уровню знаний, репрезентативность, возрастающая трудность, вариативность и системность содержания, комплексность и сбалансированность, взаимосвязь содержания и формы.

5. Выделены четыре основные формы тестовых заданий, названные заданиями закрытой и открытой формы, на соответствие и установление правильной последовательности. К псевдотестовым заданиям отнесены цепные задания, тематические и текстовые. Основанием проведенной классификации форм заданий являются педагогические, логические, технологические критерии. В числе сформулированных принципов разработки тестовых заданий закрытой формы - принципы альтернативы, противоположности, сочетания, кумуляции, градуированности, классификации, удвоенной альтернативы.

6. Получена классификация тестовых и псевдотестовых заданий. К тестовым заданиям относятся те, которые отвечают, одновременно, трем требованиям - корректности содержания, формы и наличия системообразующих свойств. Первому требованию отвечают задачи, вопросы, и потому этот признак необходим, но недостаточен. Выполнение только второго требования порождает псевдотестовое задание, являющееся тестовым только по форме. Одновременное соответствие первому и второму требованию позволяет выделить задание в тестовой форме.

7. Фундаментальной основой тестового педагогического контроля является теория педагогических измерений. Анализ форм контроля знаний с точки зрения объективности, измеримости, научности, репрезентативности и технологичности приводит к выводу о преимуществах тестов. Педагогические измерения открывают возможности лучшей дифференциации студентов.

8. Двум основным целям тестирования поставлены в соответствие две формы организации тестового контроля - итоговое и обучающее. Основная часть педагогов в оценке тестов руководствуется обычно логикой обучающего контроля, в то время как педагогические тесты являются средством для прове-

дения массового объективного итогового контроля. Определены основные функции тестового педагогического контроля и принципы регламентации процесса тестового контроля. Дано определение методологии тестового контроля. Все эти результаты приведены в тексте диссертации и реферата.

9. Рассмотрены два традиционных критерия качества тестов и введен в научный оборот третий - эффективность теста и тестовых заданий. Результат измерения можно назвать эффективным, если он дает максимум возможной информации при измерении знаний студентов определенного уровня. Эффективность достигается за счет точного соответствия уровня трудности задания с уровнем знаний испытуемого, что реализуется посредством алгоритмов адаптивного тестирования.

10. Распространенная сейчас практика оперирования педагогическими показателями с неизвестными значениями надежности и валидности создают в отечественной педагогике ситуацию гносеологического и управленческого тупика, выходом из которого является овладение культурой педагогических измерений.

11. Сравнение известных в зарубежной литературе методов объективации учебных достижений выявило необходимость разработки нового метода справедливой оценки прироста результатов. Предложенная здесь дифференциальная модель обучения подтвердила адекватность гипотезе и, кроме того, дала возможность прогнозирования предела индивидуальных достижений и времени, необходимого для достижения этого предела - по нескольким начальным точкам. Точность прогноза возрастает с увеличением числа начальных точек, характеризующих индивидуальный результат процесса учения.

12. Эмпирическая проверка этой модели указывает на ее существенные преимущества в смысле справедливости оценки прироста и точности прогноза конечных результатов. Предложенный метод справедливой оценки учебных достижений может стать одним из реальных средств в решении важной проблемы гуманитаризации образования.

СОДЕРЖАНИЕ ДИССЕРТАЦИИ ОТРАЖЕНО В ПУБЛИКАЦИЯХ:

1. Вопросы объективизации оценки результатов обучения. - М.: НИИВШ, Отдел научной информации. 1976. - 66с.
2. Опыт построения теста для оценки знаний студентов// Научная организация учебного процесса. Вып. 3, ч.1 - М.:МИФИ, 1976. - С.108-117. (в соавт.).

3. Вопросы применения тестов для контроля знаний студентов//Научная организация учебного процесса вып. 3, ч.1 - М.: МИФИ, 1976.- С.102-107.(в соавт.).

4. О количественной оценке эффективности обучающей программы//Проблемы совершенствования методики программированного обучения.- М.,Пед. об-во РСФСР, 1977.-С.181-185(в соавт.)

5. К вопросу о социальных показателях //Современные концепции уровня, качества и образа жизни. Сов.Соц.Ассоц. 1977.-С.128-133.

6. Методические вопросы оценки знаний студентов высшей школы. Вып.19 - М.: НИИВШ, Отдел научной информации, 1978.- 48с.(в соавт.).

7. Проблема психологических тестов//Вопросы психологии. 1978.- № 5.- С.97-107.

8. Опрос экспертов //Прогнозирование социальных потребностей молодежи: опыт социологического исследования. Отв.ред. И.В.Бестужев-Лада.- М.: Наука, 1978.- С.141-153.(в соавт.).

9 Планирование эксперимента и статистическая обработка данных // Пути совершенствования практических занятий по психолого-педагогическим дисциплинам в высшей школе.- М.: НИИОП АПИ СССР, 1979.- С.77-91.

10. Тесты в социологическом исследовании: Монография. М.: Наука, 1982.- 199с.

11. Методологические проблемы профессионального отбора // Проблемы профессионального психологического отбора авиационных специалистов. Материалы военно - научной конф. 22-23 мая 1984 г./Под общей ред. Г.У.Дольникова.- М.: Мин-во обороны СССР, ВВС, 1985.- С.175-182.

12. Критерии эффективности профориентационной деятельности // Актуальные вопросы повышения эффективности профориентации молодежи в условиях осуществления реформы общеобразовательной и профессиональной школы. Тез. док. коорд. семинара-совещ. 17-22 марта.- Пермь, 1985.- С.24-27.

13. Вопросы разработки общей теории профессиональной ориентации школьников // Актуальные проблемы профессиональной ориентации учащейся молодежи в условиях профессионального всеобуча. Тез.науч.конф. -С.6-7. -Глухов, 1985.-110с.

14. Методологические вопросы профессиональной ориентации и профессионального отбора//Проблемы формирования профпригод-

ности специалистов / Отв.ред. Ю.М.Забродин.- М.: Экономика, 1985.- С.9-11.

15. О профессиональном самоопределении школьников // Советская педагогика. 1985.- № 7.- С.25-28.

16. Методологический анализ понятия "Профессиональная ориентация" //Проблемы профессиональной ориентации учащейся молодежи в условиях осуществления реформы. Красноярск, КГПИ, 1985.- С.3-13.

17. Профессиональный отбор: вопросы теории //Школа и производство. 1986.- № 12.- С.9-10.

18. Основы педагогического контроля в высшей школе.-С.273-285. //Основы педагогики и психологии высшей школы: Учебное пособие / Под. ред. А.В.Петровского.- М.:МГУ, 1986.- 303с.

19. Теоретические и методологические проблемы формирования профессиональной пригодности военных специалистов//Проблемы оптимизации подготовки к деятельности военных специалистов. Ч.1.- Минск, 1986.-С.11-23. (в соавт.).

20. Методологические и теоретические вопросы профессиональной ориентации.-С.5-37. //Профессиональная ориентация школьников/ Под ред. А.Д.Сазонова.- М.: 1986.- 303с.(в соавт.).

21. Из истории психологических тестов. С.23-34.//Общая психодиагностика: Учеб.пособие / Под ред. А.А.Бодалева, В.В.Столина.- М.: МГУ, 1987.- 304с.

22. Основы научной организации педагогического контроля в высшей школе: Уч. пособие для слушателей Учебного Центра.- М.: МИСиС, 1987.- 107с.

23. Применение статистических методов и ЭВМ в педагогических исследованиях//Введение в научное исследование по педагогике: Учеб. пособие для студентов пединститутов /Под ред. В.И.Журавлева.- М.:Просвещение, 1988.-С.139 -154.

24. Система тестового контроля знаний, умений и навыков студентов высшей школы//Тез. докл. респ. науч. конф. "Перспективные системы управления качеством подготовки специалистов в высших учебных заведениях: исследования, проектирование, внедрение". Ч.1.- Каунас:КПИ, 1989 - С.72-73.

25. Научная организация тестового контроля знаний студентов высшей школы: постановка проблемы //Методические и организационные аспекты активного овладения профессией. Тез. докл. науч.- метод. конф.- Одесса, 1990.- С.7-8.

26. Применение тестов для формирования кадров высшей

школы//Тез. докл. междунар. семинара. Подготовка, повышение квалификации и аттестация научных и педагогических кадров.- М.: МЭИ, 1990.- С.168-170.

27. Состав, форма и структура педагогического теста // Тез. докл. рег. конф. "Научная организация педагогического контроля.- Семипалатинск: СГМИ, 1990.- С.3-6.

28. Научная организация обучения и контроля знаний студентов//Организация автоматизированного тестового контроля знаний студентов. Тез. докл.- С.3 - 5//Организация автоматизированного тестового контроля знаний студентов/Отв. ред. Д.Ш. Матрос. Петропавловск (Каз.), 1991.- 59 с.

29. Форма тестовых заданий. Учеб. пособие для преподавателей.- М.:МИСиС, 1991.- 36с.

30. Роль тестов в современной организации управления учебным процессом в высшей школе. Тез. докл.//Материалы Международной конференции "Экономика и управление высшей школой".- Красноярск, 12-14 мая 1992.- С.152-153.

31. Композиция тестовых заданий // Химия в школе. 1993.- № 1.- С.24-28.

32. Современная технология разработки педагогических тестов //Тез. докл. Всерос. научно - практ. конф."Новая аттестационная технология абитуриентов".-С.21-23.-М.: МПТУ, 1993.-58с.

33. Математические модели педагогического измерения. Научное издание. М.: Исследовательский центр проблем качества подготовки специалистов, 1994. - 26с.

34. Научные проблемы тестового контроля знаний. Моногр. М.: Исследовательский центр проблем качества подготовки специалистов, 1994. - 135 с.

35. Содержание теста: теоретический анализ // Химия в школе. 1994. - N 2. - С. 30-34.

36. Применение тестов в управлении образованием. Тезисы докл. Рос.- Франц. конф. М.: РГТУ, 1994, С. 79-82.

37. Основные понятия педагогической тестологии//Тезисы докладов участников школы - семинара "Научные проблемы тестового контроля знаний (14-18 марта 1994г.)", М.: 1994. - 108с.

Общий объем публикаций по теме исследования составляет свыше 36 п.л.