PROBATUM EST

И. А. Караваева

ВАЛИДНОСТЬ ПЕДАГОГИЧЕСКИХ ТЕСТОВ

Ижевск 2004

УДК 378. 146. 3 ББК
74. 580. 284 К 21

Серия основана в 2004 году

Редакционный совет серии:
И.Б. Ворожцова (председатель), Н.И. Леонов, Л.Б. Рылова,
С.Ф. Сироткин (отв. ред.), А.Н. Утехина,
Г.А. Ушаков, В.С. Черепанов

Рецензенты:
А.А. Мирошниченко, доктор пед. наук, профессор;
Т.А. Снигирева, канд. пед. наук, доцент

Караваева И.А.
К21 Валидность педагогических тестов: Монография. Ижевск: Издательский дом «Удмуртский университет», 2004. 150 с. («Probatum est.»)

ISBN 5-7029-00944

Монография посвящена актуальной в тестологии проблеме обеспечения валидности тестовых результатов при дидактическом контроле. В работе теоретически обоснованы две системы валидности тестов, построенных на основе структурного и целостного теоретических конструктов. Автор отдает предпочтение построению дидактических тестов на основе целостной модели контролируемого предметного содержания, что позволяет обосновать и разработать достаточно простой и технологичный способ конструирования валидных дидактических тестов, опирающийся на метод экспертных оценок и не требующий трудоемких статистических обработок тестовых результатов.

Работа предназначена для тестологов, преподавателей вузов, студентов и аспирантов педагогических специальностей, учителей и тех, кто интересуется тестовым контролем знаний.

ISBN 5-7029-00944

ББК 74. 580. 284

© ИА Караваева, 2004
© Издательский дом «Удмуртский университет», 2004

ОГЛАВЛЕНИЕ

Введение 5
Глава 1. Система валидности педагогического теста
1.1. Основные классифицирующие признаки видов валидности исследования в аспекте объективности результатов 9
1.2. Характеристика видов валидности теста в аспекте типологизации объектов исследования 24
1.3. Технологизация валидности педагогических тестов 40
Выводы по первой главе 59
Глава 2. Объективность контроля обученности как педагогическая проблема
2.1. Технологичность конструкта обученности как основа объективности педагогического контроля 61
2.2. Критерий валидизации целостного конструкта обученности 81
2.3. Внутренняя валидность как условие объективности результатов тестового контроля обученное™ 92
Выводы по второй главе 99
Глава 3. Экспериментальная апробация процедуры определения валидности теста для измерения обученности
3.1. Конструирование валидного педагогического теста для измерения обученности на основе метода экспертных оценок 101
3.2. Сравнительный анализ результатов эксперимента по измерению обученности 112
Выводы по третьей главе 120
Заключение 121
Литература 128

ПРИЛОЖЕНИЯ

Приложение 1. Анкета. Анализ качества содержания тестовых заданий 143
Приложение 2. Анкета. Анализ качества формы тестовых заданий 144
Приложение 3. Вариант экспериментального теста 145
Приложение 4. Результаты эксперимента 147
Приложение 5. Перевод тестового балла централизованного тестирования (вузовского) в пятибалльную шкалу оценок в 2002 г. 150

ВВЕДЕНИЕ

К важнейшим задачам модернизации образования относятся обеспечение системного реформирования содержания образования, дальнейшая разработка и внедрение государственных стандартов для всех ступеней и форм образования, развитие национальной системы оценки качества образования и на этой основе создание условий для объективного контроля качества знаний обучающихся.

Возникшая на сегодняшний день необходимость обеспечения единого образовательного пространства России потребовала введения единого государственного экзамена в тестовой форме в масштабе всей страны. В этой связи особое внимание уделяется проблемам разработки и использования качественных и объективных тестовых измерителей. В традиционных методиках проверки качества обученности учащихся, как правило, отсутствуют четкие объективные критерии оценки, преобладает субъективный характер оценочных суждений.

Анализ научной литературы позволяет утверждать, что объективная оценка качества обучения возможна при использовании, прежде всего, валидных и надежных педагогических тестовых измерителей. Однако процедуры валидизации педагогических тестов, встречающиеся в научной литературе, строятся по аналогии с более сложными инструментами психолого-педагогических и социологических исследований, объекты которых латентны и слабо формализуемы. Объективность результатов таких исследований подтверждается через привлечение сложного математико-статистического аппарата, использование которого в рамках текущего дидактического контроля нетехнологично. Для рубежного

и итогового контроля тестологи предлагают использовать сертифицированные тесты, построенные на основе методов теории моделирования и параметризации педагогических тестов, которая является основой для построения надежных тестовых измерителей результатов обучения. Валидность таких тестов базируется на статистических характеристиках тестовых заданий, зависящих от свойств выборки стандартизации, что не вполне обеспечивает выполнение требований образовательных стандартов по учебному предмету и не гарантирует сохранение единого образовательного пространства.

Однако многие исследователи не всегда используют теоретически обоснованные процедуры определения валидности педагогических контрольных материалов (в том числе и тестов), не определяют полную систему валидности для них. В связи с этим возникает противоречие между растущей потребностью использования в учебном процессе качественных тестов, объективно контролирующих результаты обучения, и недостаточной разработанностью процедуры определения валидности для обеспечения их качества.

При этом возникает несколько задач исследовательского характера. В частности, следует выявить характеристики видов валидности педагогического теста в аспекте типологизации объектов педагогического контроля и объективности тестовых результатов; обосновать способы определения системы валидности педагогических тестов; определить критерии для валидизации тестовых измерителей обученности; установить показатели качества теста для обеспечения объективности результатов при педагогическом контроле обученности; провести экспериментальное исследование разработанной процедуры определения валидности тестовых измерителей.

Ввиду того, что в педагогической тестологии пока нет однозначных трактовок ряда определений и терминов, в работе используется следующий терминологический аппарат: система валидности педагогического теста (общая валидность теста), понимаемая как комплексная характеристика, состоящая из четырех основных ви-

дов (теоретической, содержательной, эмпирической и валидности теста по форме) и двух типов (внутренней и внешней).

К *внешнему типу валидности* относится вид валидности, направленный на обеспечение экстраполяции тестовых результатов, полученных на выборке нормирования, на всю генеральную совокупность тестирующихся. *Внутренняя валидность* предназначена для обеспечения связи независимых и зависимых переменных, причины и эффекта эксперимента. *Валидизация* трактуется как качественная процедура определения валидности теста, основанная на методе экспертных оценок.

Теоретическая валидность представлена *валидностью конструкта,* или *конструктной валидностью,* которая задает адекватное объекту теоретическое представление предмета измерения. *Содержательная валидность* отражает адекватность соответствия содержания теста теоретическому конструкту. Подвидом содержательной валидности является внешняя *куррикулярная валидность,* отражающая степень соответствия содержания теста структурному теоретическому конструкту. *Эмпирическая валидность* направлена на эмпирическое подтверждение гипотезы исследования. К ее подвидам относятся: *прогностическая, критериальная, сопоставительная, соответствия, текущая валидность,* а также *факторная* и ряд других. К *валидности теста по форме* относится *техническая,* или *композиционная, валидность,* предназначенная для установления соответствия структуры и формы теста принятым в тестологии нормам.

Тип валидности зависит от характера объекта измерения. Под *латентным объектом педагогического контроля* в педагогическом тестировании предлагается понимать уровень обученности. *Структурный теоретический конструкт* определяется как репрезентативное поэлементное представление контролируемого содержания учебного предмета. Под *открытым объектом педагогического контроля* понимается результат обучения (обученность), Обученность представляется *целостным теоретическим конструктом* предметного содержания, который определен структурой предмета и ограничен теоретическими подходами к его изложению.

Для решения поставленных в исследовании задач и проверки гипотезы была использована совокупность общенаучных и педагогических методов: *теоретические* (изучение и анализ философской, психолого-педагогической и методологической литературы по проблеме; системный подход, аналогия, моделирование, гипотетико-дедуктивный метод конструирования объекта измерения); *эмпирические* (тестирование, метод групповых экспертных оценок); *математико-статистическяе* (обработка эмпирических данных).

В структуру работы включены три главы. Глава первая «Система валидности педагогического теста» посвящена анализу современного состояния проблемы обеспечения качества тестирования как экспериментального педагогического исследования; объективности тестовых выводов, их соответствия объекту исследования.

Глава вторая «Объективность контроля обученности как педагогическая проблема» посвящена поиску путей получения объективных тестовых результатов в аспекте их соответствия конструированному объекту измерения через внутреннюю содержательную валидность теста; выбору критерия валидизации теста и принципа истинности, лежащего в основе валидности; построению на этой основе процедур определения валидности педагогического теста.

В третьей главе «Экспериментальная апробация процедуры определения валидности теста для измерения обученности» иллюстрируется конструирование валидного педагогического теста и проводится анализ объективности тестовых результатов проводимого на его основе эксперимента.

Глава 1
СИСТЕМА ВАЛИДНОСТИ ПЕДАГОГИЧЕСКОГО ТЕСТА

1.1. Основные классифицирующие признаки видов валидности исследования в аспекте объективности результатов

В современных гуманитарных науках, таких как социология, психология и психодиагностика, педагогика, большое внимание уделяется качеству экспериментальных данных, полученных в ходе исследований сложного объекта, которым оказывается человек или группа людей. Несмотря на различие в аспектах исследования, науки гуманитарного цикла имеют общий объект исследования, что позволяет иногда применять общие методы в прикладных исследованиях. Однако, как мы полагаем, такое расширение не всегда оправданно. Чем сложнее исследуемые процессы, тем более специальными должны быть методы их познания. Особенно сложными для анализа оказываются внутренние психические процессы. Они недоступны для непосредственного наблюдения, латентны, а значит, требуют для своего изучения специфических методов, которые возникают в той или иной прикладной дисциплине.

Как известно, эксперимент в прикладных науках остается одним из основных методов исследования, а качество результатов, полученных опытным путем, определяет дальнейшее теоретическое развитие и практическое применение этих областей знания. Качественным может быть только такой эксперимент, который обеспечивает получение истинных данных о сущности объекта.

Объективность результатов прикладного исследования становится общей проблемой для целого ряда гуманитарных наук.

Объективность или истинность результатов экспериментального исследования латентного объекта гуманитарного знания нельзя определить как однозначное соответствие, ввиду сложности и недоступности объекта исследования для непосредственного наблюдения и уникальности внутренних психических процессов. В связи с этим в прикладных разделах гуманитарных наук вводится понятие валидности исследования как гаранта адекватности, достоверности, а также надежности полученных результатов. Надежность и валидность метода становятся центральными понятиями прикладных исследований в гуманитарных науках. В современной научной литературе встречаются различные концепции и методы определения этих понятий (А. Анастази [9], Д. Кэмпбелл 165], Я. тер Лаак [66], Э. Менстид, Г. Симин [98) и др.). Однако основной для их понимания оказывается проблема объективности результатов исследования.

Объективность результатов обеспечивает достоверность, или истинность, исследования, которая рассматривается в классических философских монографиях по теории познания (ВА Лекторский [73], Э.М. Чудинов (1761, ВА Штофф [180]). Смысл, вкладываемый в понимание истинности, неизбежно определяет понятие валидности результата исследования и отсылает нас к философским проблемам познания, а именно к неклассической гносеологии (эпистемологии) (Ц. Кэмпбелл [65], К. Левин [70], К Мангейм [85], Л А Микешина [89] и др.). Проследим это влияние.

Сбор и интерпретация информации о предмете являются основными этапами прикладного исследования, в том числе тестирования. В процессе его проведения возникают ошибки измерения: случайные и систематические. Их возможными источниками могут быть: сам человек, его утомляемость, забывчивость; среда, влияющая на испытуемого; методики и процедуры исследования; сами исследователи и т. д. Путем стандартизации измерения можно избежать систематических ошибок или, по крайней мере, их контролировать. Случайные ошибки не контролируются. Как отмечает

Я. тер Лаак, «в той мере, в какой можно учесть или устранить эти факторы, они не являются источниками «ненадежности» теста, но всегда остаются иные, неконтролируемые, ошибки» [66, с. 89]. Здесь под надежностью исследования понимают:

1) релиабильность или точность измерения;
2) однородность полученных результатов.

По его мнению, в эмпирических исследованиях «понятие "надежность" показывает степень свободы результатов тестирования от ошибок измерения, а также степень согласованности и повторяемости полученных результатов» [Там же, с. 88]. В теоретических моделях не существуют неконтролируемые, случайные ошибки. Надежность теста характеризуется точностью как ошибкой измерения и однородностью как коэффициентом надежности эмпирического исследования. Традиционно под коэффициентом надежности понимают корреляционный коэффициент двух рядов результатов тестирования, полученных различными способами.

Коэффициенты надежности теста можно рассматривать в нескольких аспектах. Соотношение двух рядов результатов выявляет внутреннюю согласованность содержания тестовых заданий как *однородность содержания* теста, повторяемость тестовых результатов как их *однородность во времени,* внешнюю согласованность мнений экспертов как *однородность их интерпретации.*

В классической теории тестов ошибка измерения выражает расхождение тестовых баллов, полученных у одного тестируемого. Если величина расхождения невелика, то измерение признается надежным. Постулируется постоянство «истинных» баллов испытуемых во времени. «Истинный» балл отражает то или иное устойчивое латентное свойство. «Следовательно, — как утверждает Я. тер Лаак, — можно говорить об "истинных" баллах (истинных в платоновском смысле, т. е. подлинных, идеальных сущностей), в соответствии с которыми все изменения во времени являются только "видимостью"» {Там же, с. 91]. Расхождение в баллах приписывается свойству процедуры и инструмента тестирования при повторном тестировании одним и тем же тестом одного и того же испытуемого. Так можно определить надеж-

ность теоретически. На практике применяются иные способы установления надежности:

- параллельные тесты — теоретические конструкты-двойники Спирмена;
- повторное тестирование — статистические двойники;
- деление теста пополам — внутренняя согласованность;
- согласованность мнений экспертов.

В дальнейшем исследователи отказались от постулата о неизменности латентного признака, но традиция измерения надежности тестов осталась. Это привело к тому, что в классической теории тестов «отсутствуют средства и процедуры измерения феномена динамики и изменений» [Там же, с. 92]. В научном анализе возник диссонанс между современным представлением о динамичности объекта исследования и классическим методом статичного определения надежности.

Обычно в классической теории тестов надежность и валид-ность рассматривают независимо. Надежность оказывается необходимым, но не достаточным условием валидности результатов исследования. Возникает проблема получения валидного знания или вывода прикладного исследования, которая решается по-разному. Известно, что валидность вывода зависит не только от цели исследования, выбранных методов, но и от принятого в данном случае понятия валидности в смысле истинности, сравнимости и соответствия знания об объекте самому объекту.

В прикладных науках исследователь имеет дело с так называемой экспериментальной моделью «черного ящика», суть которой состоит в закрытости измеряемого объекта, а следовательно, невозможности непосредственного наблюдения и измерения его характеристик. Об измеряемых характеристиках латентного объекта судят по изменениям данных, регистрирующихся на «входе» и «выходе», т. е. по изменениям состояний объекта до и после измерения. Эти данные служат основой для вывода о латентных свойствах объекта. Истинность вывода основывается на теоретических представлениях современной науки о возможных процессах, происходящих «внутри» объекта — «черного ящика». Возникает понятие валидности вывода как адекватности современных теоретических представлений об объекте реальному латентному объекту. Доказательством соответствия служит понятие валидности прикладного исследования. Другими словами, истинность заменяется валидностью.

Итак, поскольку объект в прикладных науках носит латентный характер, постольку возникает проблема обеспечения соответствия результатов эксперимента реальным, т. е. истинным характеристикам исследуемого объекта. Исследователь не имеет возможности проследить реальные свойства и связи, происходящие «внутри» исследуемого объекта, а значит, его результаты и выводы необходимо имеют опосредованный и вероятностный характер.

Я. тер Лаак [66], анализируя зарубежные работы по психодиагностике, вышедшие до середины 90-х гг. 20 в., приводит три типа дифференциации видов валидности, встречающиеся в прикладных исследованиях.

Первая дифференциация видов валидности заключает в себе прогностическую валидность по внешнему критерию, конструктную валидность и содержательную валидность. Конструктная валидность, как указывает Я. тер Лаак, «имеет отношение к теоретическому конструкту самому по себе и включает в себя поиск факторов, объясняющих поведение при выполнении теста» [Там же, с. 104]. Объект исследования должен быть представлен для последующего измерения в виде теоретического конструкта, т. е. как модель объекта исследования, истолкованная в терминах некоторого теоретического подхода. На конструктную валидность возлагается контроль за теоретическим определением объекта исследования, где содержательная валидность «оценивает соответствие содержания теста... измеряемой области поведения» [Там же]. В тест могут входить все без исключения тестовые задания, равновероятно представляющие измеряемую область поведения. Содержание теста должно выражать измеряемый предмет, представленный как теоретический конструкт. Результаты тестирования сопоставляются с каким-либо внешним критерием, т. е. социально значимым поведением: успеваемостью, профессиональной деятельностью, социальной

адаптацией и др. Если они взаимно соответствуют, то заданный теоретический конструкт способен предсказывать поведение и определять прогностическую валидность теста. «Прогностическая валидность теста оценивается по тому, насколько данный прогноз получает подтверждение в будущем или в настоящее время (соответственно прогностическая и конкурентная валидность)» [Там же]. Таким образом, прогностическая валидность теста опосредованно через содержание теста указывает на соответствие теоретически описанного и реального поведения исследуемого объекта. Она связывает предмет исследования и результаты тестирования, являясь эмпирическим, опосредованным подтверждением истинности гипотезы исследования или прогноза.

Вывод о валидности результатов тестирования рассматривается Я. тер Лааком как соответствие *критерию когеренщи* [73; 176; 180]. | «В этом случае какое-либо утверждение должно соответствовать не] реальной действительности, а другим утверждениям. В некоторой единой системе утверждения должны быть связаны между собой» [66, с.98]. Соответствие утверждения системе утверждений подтверждает вывод, несоответствие — либо опровергает вывод, либо опровергает систему. Я. тер Лаак отмечает, что при установлении соответствия некоторого утверждения системе утверждений могут появиться некоторые трудности. Во-первых, сложно логически доказать противоречие между системой и утверждением, если оно не достаточно формализовано. Ввиду неоднозначности и латентности исследуемого объекта это зачастую происходит в психодиагностике и других социальных науках. Во-вторых, отсутствие соответствия между системой и утверждением не противоречит существованию другой системы «связных утверждений, относящихся к той же области реальности» [Там же, с. 99].

При условии, что соответствие достигнуто, можно делать вывод об однородности системы утверждений. Однородность системы указывает на надежность вывода в смысле повторяемости и согласованности результатов. Именно согласованность, или связность, а следовательно, надежность двух или нескольких результатов измерения одного и того же предмета исследования, получен

ных различными методиками, подтверждают валидность исследования, основанного на критерии когеренции, в данной системе видов валидности. На наш взгляд, использование критерия когеренции для определения валидности тестов говорит, скорее, о, надежности получения результатов, т. е. о надежности инструмента исследования — теста, чем о валидности как подтверждении прогноза во временном интервале.

Существование прогностической валидности теста в данной системе видов валидности служит эмпирическим подтверждением гипотезы исследования, т. е. указывает на наличие связи между предметом исследования и тестовыми результатами, представленными на основе принципа когеренции. Как мы полагаем, соответствие результатов тестирования внешнему критерию, сходному предмету исследования, порождает ряд вопросов:

- не является ли данный тест дублирующим уже известные тесты, (надежно) измеряющие тот же предмет;
- измеряет ли данный тест выделенный предмет исследования или же близкий с ним;
- не является ли тривиальной связь независимой и зависимой переменных конструкта;
- возможна ли экстраполяция результатов тестирования на большую, чем экспериментальная выборка, группу тестируемых;
- исходит ли данная система утверждений из единого основания (в том числе из единого конструкта)?

Система видов валидности, основанная на первом типе дифференциации видов валидности — прогностическая, конструкт-ная, содержательная, — не позволяет ответить на поставленные вопросы.

Таким образом, три выделенных вида валидности относятся к валидности теста, скорее, как надежного инструментария и не затрагивают валидность эксперимента в целом, решая проблему валидности опосредованно через соответствие результату другого исследования, валидность которого оказывается неопределенной. По нашему мнению, это связано со сложным и латентным характером объекта исследования, а также с выбором понятия валидности

на основе критерия когеренции. Кроме того, указанные виды валидности не решают проблему экстраполяции результатов эксперимента на внеэкспериментальную группу тестируемых. Здесь можно говорить о валидности полученных при однократном тестировании данных как о возможности подтвердить или опровергнуть гипотезу исследования, заложенную в теоретический конструкт, использованный только для данной выборки испытуемых и проверенный через подтверждение полученных результатов данными другого исследования (надежность). Вопрос о валидности всей теоретической системы в целом, а следовательно, о возможности валидности исследования, полученной на основе сопоставления различных результатов, остается открытым.

Поскольку объект исследования в психодиагностике и других, социальных науках является латентным, постольку обнаруживаются две различных логики сопоставления результатов различных измерений одного и того же объекта исследования при анализе валидности теста. Во-первых, теоретически возможно сопоставление полученных в ходе тестирования результатов с истинным, т. е. абсолютным результатом. Такой подход к валидизации основан на принципе *корреспонденции* [73; 176; 180], т.е. соответствия полученных результатов реальному объекту измерения. Практически такой подход нельзя реализовать из-за латентности объекта исследования.

Во-вторых, можно сопоставить результаты измерения латентной характеристики объекта, полученные при тестировании, с другими независимыми и социально значимыми результатами исследования того же объекта. Как уже отмечалось, в этом случае оценка валидности происходит на основе критерия когеренции, применение которого приводит к надежности результата исследования, но не к валидности вывода. Надежность вывода на основе критерия когеренции действительно может привести к валидному результату, если второе измерение характеристики объекта валидно. Возникает вопрос о том, действительно ли социальная значимость результатов исследования равнозначна валидности этого исследования.

Социальная значимость результатов имеет свое логическое обоснование. Она рассматривается как характеристика валидности результатов исследования, основанная на критерии полезности или *функциональности* [73; 176; 180]. Я. тер Лаак обращает внимание на то, что при использовании подобного внешнего критерия «речь идет не о соответствии аспектам реальности и не о соответствии другим утверждениям. Здесь подразумевается прежде всего то, что утверждения (содержащие в себе некоторую информацию) и идеи дают человеку возможность с большим или меньшим успехом взаимодействовать с реальной действительностью» [66, с. 99]. Валидность исследования становится целезависимой характеристикой, где выбор внешнего критерия задается его полезностью и функциональностью при достижении конкретной цели. «Нормативные модели» должны максимизировать заданный критерий, детерминируя его в соответствии с поставленной целью. Такой определяется вторая дифференциация видов валидности, объединенная под названием «специфическая» [66].

В отличие от первой, она не образует системы валидности теста, а лишь представляет различные типы валидности, выбор одного из которых при валидизации тестов зависит от цели исследования и (или) метода построения теста, а следовательно, выбора способа доказательства валидности эксперимента для достижения поставленной цели.

1. *Синтетическая валидность* является разновидностью прогностической валидности и определяется как возможность теста прогнозировать поведение не по одному, а по нескольким различным критериям.

2. *Факторная валидность* имеет место в случае, когда тест «имеет высокую нагруженность по известным факторам». Ее относят к методам создания теста на основе многомерного анализа и к прогностической валидности.

3. *Конгруэнтная валидность* задается корреляцией результатов тестирования по схожим тестам, т. е. является разновидностью прогностической валидности,

4. *Инкрементная валидность* выявляется через приращение знания в ходе эксперимента, которое подтверждается прогностичес-

кой валидностью, основанной на демографических и социально-экономических данных.

5. *Внутренняя валидность* определяется как нетривиальность корреляционной связи между теоретическими конструктами.

6. *Очевидная валидность,* или валидность «с первого взгляда», оценивается как соответствие содержания теста предмету исследования и, следовательно, является разновидностью содержательной валидности.

Третья дифференциация валидности, по Я. тер Лааку, — это валидность, определяющаяся в экспериментах и квазиэкспериментах: внутренняя и внешняя валидности, валидность статистического вывода и конструктная валидность.

Как известно, особенность психологического и социологического эксперимента или квазиэксперимента заключается в особенностях объекта исследования, т. е. в его латентности. «Объектом психологического исследования, например, является человек или группа людей» [107, с. 621]. Исходя из существующей теории понимания латентной характеристики субъекта как объекта исследования, строится теоретический конструкт, описывающий латентную характеристику в рамках теории, и выдвигаются теоретические следствия или проявления латентной характеристики, вытекающие из теории. Построение конструктов — сложная теоретическая задача, которой свойственны высокая степень обобщения уже имеющихся эмпирических данных описания рассматриваемого феномена, включая частные случаи и предельные положения. Ее отличает строгая формализация в рамках теории описываемого латента, позволяющая предсказывать его поведение. «Психологи соглашаются, — констатирует К. Левин, — что ценность конструктов и теорий в эмпирической науке зависит в конечном счете от их плодотворности при «объяснении» известных фактов и предсказании неизвестных» [70, с. 411.

После определения конструкта начинается процесс планирования эксперимента. Выделяются независимая и зависимая переменные, выбирается валидный план эксперимента. «Переменная — термин используется для обозначения репрезентации конструкта»

[98, с. 125]. Иначе говоря, отбираются свойства латента, теоретически каузально связанные. Манипулирование одним из них (независимой переменной) приводит к изменению состояния другой (зависимой переменной), которое фиксируется в ходе эксперимента. Контроль за установлением каузальности данной связи в ходе эксперимента возложен на внутреннюю валидность эксперимента. «Внутренняя валидность *(Internal validity)* — обозначает валидность заключения о том, что изменения независимой переменной имеет результатом изменения зависимой переменной» [Там же, с. 123]. Так, опосредованно, через переменные (свойства латента), проясняются свойства теоретического конструкта. При определении валидности конструкта выясняется, как выбранная для манипуляции независимая переменная и контролируемая зависимая переменная соотносятся с гипотетическими понятиями. «Поскольку в экспериментах крайне важно выяснить, является ли связь между двумя переменными каузальной, *необходимым* условием для выведения заключения о каузальности является отбрасывание альтернативных интерпретаций связи между переменными (т. е. устранение угроз для внутренней валидности), а не отбрасывание альтернативных интерпретаций операций с причиной и следствием (т. е. угроз для конструктной валидности)» [65, с. 43].

Д. Кэмпбеллом в работе [65] описаны валидные планы эксперимента, использование которых дает возможность проконтролировать факторы, угрожающие внутренней и внешней валидности. Здесь внутренняя валидность представляется в аспекте каузальности связи экспериментальных переменных, а внешняя валидность задается через возможность экстраполяции результатов эксперимента на большую, чем экспериментальная, область. «Внешняя валидность *(External validity)* — относится к возможности распространения результатов исследования на других людей, кроме тех, кто участвовал в эксперименте в качестве испытуемых» [98, с. 123]. Проблема обеспечения внешней валидности решается в зависимости от выбранного плана эксперимента. Д. Кэмпбелл отмечает, что проблема обеспечения внешней валидности не может быть решена абсолютно. Это связано, во-первых, с постоянной необходимостью экстрапо-

ляции результатов эксперимента на более широкую область, чем та, на которой они были получены. Во-вторых, к контролю побочных факторов, влияющих на результаты эксперимента, количество которых может быть неограничено. Практически невозможно контролировать все факторы, влияющие на результаты измерения. Необходимо создать для участников эксперимента одинаковые или близкие условия его протекания. В спецификации теста отражается качество экспериментаиьной выборки и унификация процесса тестирования [65].

При подготовке эксперимента важное значение обретает выбор плана эксперимента. Так, Д. Кэмпбелл приводит 16 валидных экспериментальных планов, использование которых позволяет экспериментатору достигать более валидных результатов. Он отмечает, «поскольку существуют валидные планы, в которых не используется предварительное тестирование, ...такие планы более предпочтительны, так как обеспечивают большую внешнюю валидность или возможность обобщения» [Там же, с. 72]. Необходимым условием обеспечения внешней валидноети эксперимента оказывается учет побочных факторов, влияющих на результаты эксперимента: во-первых, унификация процедуры эксперимента, во-вторых, контроль наиболее значащих факторов и пренебрежение второстепенными.

Для получения валидных результатов при планировании эксперимента «нашим идеалом является выбор таких экспериментальных планов, которые обеспечивают оба типа валидноети» [Там же, с. 46]. Следует учитывать, что увеличение одного типа валидноети может привести к уменьшению другого типа.

Наиболее полно внешняя валидность обеспечивается в так называемом истинном рандомизированном эксперименте *(True randomised experiment)* — «эксперимент, в котором в различных экспериментальных ситуациях участвуют случайные выборки из числа испытуемых» [98, с. 124]. Абсолютное и однозначное распространение результатов эксперимента на большую, чем экспериментальная, область невозможно даже в таком эксперименте. Для квазиэксперимента или эксперимента в полевых условиях рандомизация неосуществима из-за присутствия факторов, не поддающихся контролю. Экстраполяция результатов и внешняя валидность эксперимента всегда носят вероятностный характер (по определению). Для повышения вероятности вывода используют простую случайную выборку, в которую с равной вероятностью может попасть каждый участник эксперимента.

Наличие статистической связи между независимой и зависимой переменными, причиной и эффектом подтверждает каузальность связи. Д. Кэмпбелл заостряет внимание на том, что «воздействие или воздействия должны быть статистически связаны с эффектом, так как, если возможная причина и эффект не связаны друг с другом, одно не может быть причиной другого» [65, с. 40]. Однако статистические показатели не однозначно определяют ковариацию. Даже при жестких критериях (например, $p < 0{,}05$) и строгом применении математического аппарата статистические показатели могут выдавать ложные связи за действительные, а действительные за ложные. Д.Кэмпбелл ввел валидность статистического вывода на основе статистических показателей, которые призваны подтвердить или опровергнуть гипотезу исследования. Валидность статистического вывода зависит от вариативности матрицы распределения результатов, качества репрезентативной выборки, условий протекания эксперимента и контроля над способом предъявления независимых переменных, условий работы контрольной и экспериментальной групп и т. д. [66, с. 107]. Как мы полагаем, и это важно отметить, валидность статистического вывода следует отнести прежде всего к области качества проведения эксперимента, а не к области качества инструмента эксперимента — теста. Именно поэтому в дальнейшем нас будет интересовать валидность инструментария педагогического эксперимента (тестирования), а не валидность процедур управления экспериментом. Здесь педагогическое тестирование мы понимаем как разновидность эксперимента в социальных науках, где педагогический тест как инструмент измерения должен обеспечивать получение валидных результатов. Это значит, что валидный педагогический тест необходимо должен обладать внутренней и внеш-

ней валидностью как инструмент измерения. При отсутствии внутренней валидности измерение теряет смысл и невозможным становится достижение цели тестирования. При отсутствии внешней валидности не обосновано применение теста как инструмента измерения.

Итак, при измерении латентной характеристики объекта исследования валидность понимается как адекватность результатов тестирования истинным характеристикам объекта и обеспечивается следующими видами валидности теста:

1) *конструктной валидностью,* которая задает теоретическое представление предмета исследования;

2) *содержательной валидностью,* отражающей соответствие содержания теста теоретическому конструкту;

3) *прогностической валидностью,* эмпирически подтверждающей или опровергающей гипотезу исследования;

4) *внутренней валидностью,* представляющей каузальность связи между причиной и эффектом;

5) *внешней валидностью,* позволяющей экстраполировать результат на внеэкспериментальную область;

6) *валидностью статистического вывода,* характеризующей адекватность интерпретаций результатов исследования на основе применения статистических характеристик.

Отсюда следуют выводы. Во-первых, специфика объекта исследования социальных наук заключается в невозможности непосредственного наблюдения за латентными характеристиками исследуемого объекта и получения абсолютно истинных результатов прикладного исследования. Недостижима объективность результатов в смысле их соответствия объекту исследования, а понятие истинности результатов заменяется понятием их валидности. Валидность становится вероятностной характеристикой результатов измерения латентного объекта.

Во-вторых, понятие истины в абсолютном смысле достижимо лишь для теоретических систем, а для эмпирических исследований тот или иной аспект истинности постулируется, определяя выбор понятия валидности исследования и его результатов. Другими словами, понятие валидности исследования зависит от выбора критерия истинности, положенного в его основание.

В-третьих, валидность измерения на основе критерия когеренции понимается как надежность, т. е. соответствие теоретической системе утверждений. Валидность самой теоретической системы оказывается необоснованной. Она становится системой с бесконечными характеристиками.

В-четвертых, валидность, определяемая на основе критерия функциональности, трактуется в аспекте практической применимости результатов и соответствия их социальным нормам. Утрачивается валидность, или истинность, как соответствие объекту исследования.

В-пятых, возможность измерения характеристик латентного объекта, в терминах современного знания об его природе, определяется теоретическим конструктом, описывающим измеряемые характеристики. Здесь валидность конструкта понимается как соответствие конструкта объекту исследования, а содержательная валидность — как степень соответствия содержательного наполнения инструмента исследования теоретическому конструкту. Прогностическая, или эмпирическая, или критериальная, валидность трактуется как степень соответствия результатов эксперимента результатам другого, заведомо валидного исследования, которое выбирается в качестве критерия истинности. В результате анализа классифицирующих признаков видов валидности, таким образом, прояснилось, что триада конструктной, содержательной и прогностической валидностей содержит достаточное количество видов валидности для получения вероятностных валидных результатов в определенном аспекте истины единичного измерения статичного признака объекта исследования, а создание инструмента многоразового использования предполагает, что исследование необходимо должно обладать внешней валидностью.

В-шестых, вероятностное соответствие содержания инструмента исследования объекту исследования прослеживается в конструктной и содержательной валидности, а вероятностное соответствие критерию валидизации — в прогностической и внешней

валидности. Однородность и непротиворечивость системы обеспечиваются надежностью результатов исследования и критерием валидизации. Здесь объект исследования оказывается за пределами теоретической системы. Признание соответствия системы объекту исследования необходимо достигается через внутреннюю валидность исследования.

1.2. Характеристика видов валидности теста в аспекте типологизации объектов исследования

Валидность эксперимента в социальных науках понимается как получение истинных в определенном аспекте результатов измерения и выводов исследования латентного объекта. Очевидно, что невозможно получить валидные результаты эксперимента, используя невалидный инструментарий. Использование валидного инструмента в эксперименте является необходимым условием, обеспечивающим валидность экспериментальных выводов. Проблемы качества тестового инструментария освящены в работах известных отечественных и зарубежных ученых (В.С.Аванесова [7], А. Анастази [9; 10], В.А. Дюка [39], К.Ингенкампа [44], А. С. Каза-ринова, А.А. Мирошниченко [46; 47], А.Н. Майорова [80], Д. Ман-гейма, Р. Рича [84], Ю.М. Неймана [106], В.А. Хлебникова [160], М.Б. Челышковой [166] и др.). Тест используется как инструмент эксперимента наряду с наблюдением, анкетами и т. п. Валидность теста — это основная его характеристика в смысле адекватности полученных с его помощью результатов тестирования объекту исследования. По мнению В. А. Дюка, «в отличие от надежности вапидность — мера соответствия тестовых оценок представлениям о сущности свойств или их роли в той или иной деятельности» [39, с. 112]. В определении утверждается, как мы полагаем, возможность двойного представления понятия валидности теста.

Первый аспект рассмотрения валидности теста как адекватности результатов тестирования критерию истинности основан на *принципе функциональности.* Валидность теста здесь предъявляется как его способность манифестировать социально значимые результаты или, другими словами, соответствовать социально значимым нормам. В педагогике, а также в психодиагностике и социологии в качестве социально значимой нормы чаще всего выступает положение о расположении оценок латентной характеристики объекта по нормальному закону распределения статистических величин. Прогностическая валидность понимается как соответствие эмпирических данных тестирования внутреннему критерию, норме — нормальному закону распределения статистических показателей измеряемой характеристики — и выступает в роли эмпирической или критериальной валидности. Тесты, построенные на основе принципа функциональности, получили название *нормативно-ориентированные.* В них эмпирическая валидность несет на себе основную нагрузку при подтверждении гипотезы эксперимента, а содержательная и конструктная валидности носят вспомогательный характер.

Второй аспект рассмотрения валидности теста основан на *критерии когеренции.* Валидность понимается как соответствие тестовых результатов теоретическому конструкту, который создается как описание объекта исследования, заданного определенными принципами той или иной современной научной теории. Определение прогностической валидности теста основывается на сравнении тестовых и других заведомо валидных эмпирических результатов измерений того же самого объекта как теоретического конструкта. Валидные эмпирические результаты выполняют роль критерия истинности, а прогностическая валидность выступает в роли эмпирической или критериальной валидности и также становится наиболее значимым видом валидности, сдвигая на второстепенный план валидность конструкта и содержательную валидность. Эмпирические результаты измерения того же или близкого по смыслу теоретического конструкта, признанные валидными, тоже могут служить критерием валидности. Определение степени валидности теста устанавливается как соответствие результатов тестирования выбранному внешнему критерию. Такие тесты получили название *критериально-ориентированные.*

Из этого следует, что измерение различными способами объекта как общего конструкта, с одной стороны, позволяет определить «эталонные» его результаты и принять их за критерий тестирования, ввиду их заведомой валидности как адекватности скрытым от наблюдения свойствам объекта. С другой стороны, критериальная валидность эмпирически подтверждает внутреннюю связь между результатами валидизируемого и «эталонного» тестирования, но не доказывает ее. Доказательством связи латентных характеристик объекта как конструкта с тестовыми результатами измерения будет установление внутренней валидности последних.

Таким образом, выбор опосредованного внешнего или внутреннего критерия для определения прогностической валидности не является существенным, но он связан со способом доказательства адекватности результатов тестирования, которое, в свою очередь, основано либо на критерии когеренции, либо на критерии функциональности. Нормативно-ориентированные и критериально-ориентированные тесты различаются в аспекте интерпретации понимания истинности результатов измерения, полученных в зависимости от цели тестирования и выбранного критерия. К сожалению, в целом ряде современных работ по педагогическому тестированию нет четкого различения данных видов тестов в зависимости от критерия истинности, выбранного как основание понимания валидности результатов [80; 133 и др.].

Как мы показали, наличие прогностической валидности теста подтверждает адекватность результатов измерения объекту как конструкту, но не доказывает валидность теста из-за отсутствия внутренней валидности. Во-первых, открытым остается вопрос о доказательстве связи между результатами теста и измеряемым теоретическим конструктом, т. е. связи между независимой и зависимой переменной, причиной и эффектом. Другими словами, остается не решенным вопрос о внутренней валидности теста. Во-вторых, сохраняется проблема внешней валидности теста как задача возможной экстраполяции интерпретации результатов измерения на иную, большую, чем экспериментальная, группу испытуемых.

Поскольку содержательная валидность связывает содержание теста и внешние теоретически определенные свойства или проявления конструкта, постольку она подтверждает соответствие между конструктом и результатами измерения. Как утверждают Л.Ф. Бур-лачук и С.М. Морозов, «содержательная валидность характеризует степень репрезентативности содержания заданий теста измеряемой области психических свойств» [Цит. по: 39, с. 112]. Другими словами, с помощью содержательной валидности проясняется вопрос о том, насколько содержание теста, как образ, есть отражение теоретического конструкта как своего прообраза. Репрезентативность в статистике и экспериментальных исследованиях понимается как «показательность каких-либо наблюдений; соответствие характеристик, полученных в результате частичного (выборочного) обследования какого-либо объекта, характеристикам этого объекта в целом, позволяющее распространять выводы частичного обследования на весь изучаемый объект» [145, с. 527]. Иначе говоря, под репрезентативностью содержания теста можно полагать соответствие любого тестового задания содержательной области измеряемого теоретического конструкта.

Содержание теста образуется совокупностью тестовых заданий. Она не взаимно-однозначно описывает содержательную область конструкта, а лишь представляет некоторые его проявления. Тем не менее условие репрезентативности, накладываемое на тестовые задания, позволяет распространять результаты тестирования на всю область предметного содержания теоретического конструкта.

Таким образом, можно сделать вывод о том, что:

• содержательная валидность относится к внешней валидности эксперимента, являясь ее частью, так как позволяет экстраполировать результаты эксперимента, полученные с помощью теста с определенным содержанием, на весь теоретический конструкт;

• через содержательную валидность контролируется отображение или представление теоретического конструкта в содержании теста, другими словами, показывается, в какой степени определенное тестовое содержание предъявляет конструкт, оказываясь од-

ним из множества возможных практических, высказанных на языке действия, вариантов описания теоретического конструкта, где последний выражается на языке теории и понимается в качестве теоретического инварианта;

• через содержательную валидность соотносится содержание теста с результатами тестирования, и через них — с внешним критерием;

• через содержательную валидность устанавливается связь результатов тестирования и конструкта посредством тестового содержания.

Согласно определению В.А. Дюка, «конструктная валидность — это валидность теста по отношению к психологическому концепту — научному понятию (или их совокупности) об измеряемом психическом свойстве (состоянии). Она выражает степень обоснованности индивидуальных различий, обнаруживаемых тестом, с позиций современного теоретического знания» [39, с. 113]. Другими словами, данным видом валидности показывается, в какой мере зависимая переменная, отраженная в результатах тестирования, соотносится с теоретически описанным прогнозом проявления конструкта через независимую переменную конструкта, представленную в тестовом содержании. В валидности конструкта теста проявляется связь зависимой переменной, представленной в результатах тестирования, с конструктом, а связь результатов тестирования с независимой переменной устанавливается через тестовое содержание. Как мы полагаем, валидность конструкта теста оказывается необходимой частью внутренней валидности эксперимента.

Если удается содержательную валидность представить как часть внутренней валидности, то, возможно, отпадет необходимость **в** обеспечении прогностической валидности теста, следовательно, исчезнет надобность в сложном и трудоемком процессе эмпирического подтверждения объективности результатов тестирования через их соответствие внешнему критерию, а исследование валидности станет более технологичным.

Известно, что внутренней валидностью эксперимента обеспечивается каузальность связи между независимой и зависимой переменными теоретического конструкта. Содержательной валидностью теста задается отношение между независимой переменной, представляющей конструкт, и содержанием теста. Это значит, что валидностью конструкта устанавливается соответствие между конструктом и зависимой переменной, проявляющейся в результатах тестирования. Во внутренней валидности, таким образом, должна предъявляться связь между содержанием теста и результатами тестирования. Однако между содержанием теста и его результатами полагается процедура тестирования, осуществляющаяся в определенных условиях, под влиянием множества побочных внешних факторов, и субъект тестирования. Объективность результатов измерения становится сомнительной ввиду того, что латентные характеристики субъекта принципиально не контролируются во всех его особенностях. Невозможно обеспечить внутреннюю валидность тестовых результатов через контроль тестируемого. Кроме того, процедура тестирования оказывается зависящей от принципиально открытой системы внешних факторов, которые никогда невозможно учесть во всем их многообразии.

Итак, по нашему мнению, для усиления объективности результатов тестирования, а значит, увеличения их валидности необходимо:

во-первых, установить связь между экспериментальными результатами измерения и заведомо валидными результатами измерения того же теоретического конструкта;

во-вторых, максимально объективировать процедуру тестирования посредством учета наибольшего числа внешних факторов, влияющих на результаты измерения;

в-третьих, построить технологический процесс тестирования путем элиминирования характеристик субъекта из процедуры тестирования.

Традиционно проблема достижения *внутренней* валидности тестовых результатов решается первыми двумя способами. **Первый способ** подразумевает использование эмпирической, или прогностической, валидности. «Эмпирическая валидность — совокупность характеристик валидности теста, полученных с помощью сравнительного статистического анализа. Показатель эмпирической валидности

выражается количественной мерой статистической связи между результатами тестирования и внешними по отношению к ним критериям оценки диагностируемого свойства» [Там же]. Известно, что процедура получения численного значения эмпирической валидности теста заключается в получении коэффициента корреляции между результатами валидизируемого эксперимента и результатами, полученными в ходе другого экспериментального исследования другого конструкта. Тест обладает эмпирической валидностью при условии, что второй, «эталонный», конструкт тесно связан с исследуемым конструктом и результаты обоих исследований взаимно коррелируют. Такой способ валидизации является *конвергентным* [84]. Если второй конструкт по смыслу противопоставлен исследуемому конструкту и сравниваемые результаты не коррелируют, то валидизация оказывается *дискриминантной* [Там же].

Таким образом, в социологии и психодиагностике объективность эксперимента по измерению латентного объекта обеспечивается:

внутренней валидностью конструкта, отвечающей за соответствие независимой и зависимой переменных;

внешней содержательной валидностью, отвечающей за соответствие содержания теста конструкту (через независимую переменную);

внешней эмпирической валидностью, подтверждающей связь результатов валидизируемого и эталонного тестирования.

Однако выделенные три вида валидности эксперимента в социальных науках образуют незамкнутую систему валидности экспериментальных результатов, так как эмпирическая валидность является внешней. Вследствие незамкнутости системы валидности связь тестовых результатов с измеряемым теоретическим конструктом оказывается вероятностной и приближенной.

Именно прямое перенесение экспериментальных психологических и социологических методов в педагогику, а тем более в образование привело к таким же трудностям в обеспечении внутренней валидности педагогического, эксперимента: применение численных статистических методов лишь подтверждает каузальность связи, но

не доказывает ее. Так, Д. Кэмпбелл указывает, что в социально-психологических экспериментах «нулевая корреляция снижает правдоподобие гипотезы. При высокой корреляции оно возрастает, ибо гипотеза избежала опровержения. Иначе говоря, корреляция не обязательно указывает на наличие причинной связи, но каузальный закон, поскольку он связан с различием между средними значениями, предполагает наличие корреляции» [65, с. 173]. Соответственно, для доказательства внутренней валидности в педагогическом исследовании требуется применение дополнительных качественных методов ее анализа. Практически же исследователи ограничиваются вычислением статистического коэффициента как коэффициента эмпирической валидности. Статистические методы задают только надежность результатов исследования, но их объективность оказывается попрежнему недостижимой.

Тем не менее в педагогической практике решение проблемы внутренней валидности за счет прогностической (эмпирической) валидности до сих пор остается основным способом построения педагогических тестов наравне с другими прикладными науками. Хотя система валидности исследования, традиционно состоящая из трех ее видов («святая троица», по выражению Я. тер Лаака), оказывается принципиально незамкнутой и обладает только вероятностной степенью объективности в аспекте соответствия результатов измерения теоретическому конструкту. Известно, что классификация видов валидности педагогического теста, встречающаяся в исследовательской литературе, сводится к выделению таких ее разновидностей: валидность содержания, конструкта (теоретическая), эмпирическая (соответствия или прогноза) [44]; содержательная и критериальная (то же, что и эмпирическая) [106]; содержа тельная [5]; содержательная, техническая (связанная с улучшением структуры содержания тестовых заданий), критериальная (эмпирическая) в форме валидности соответствия или прогноза [166]; критериальная, конструктная и содержательная [93]. Система пред ложенных видов валидности оказывается принципиально незамкнутой точно так же, как и в других областях гуманитарного знания, на основе которых строилась классификация.

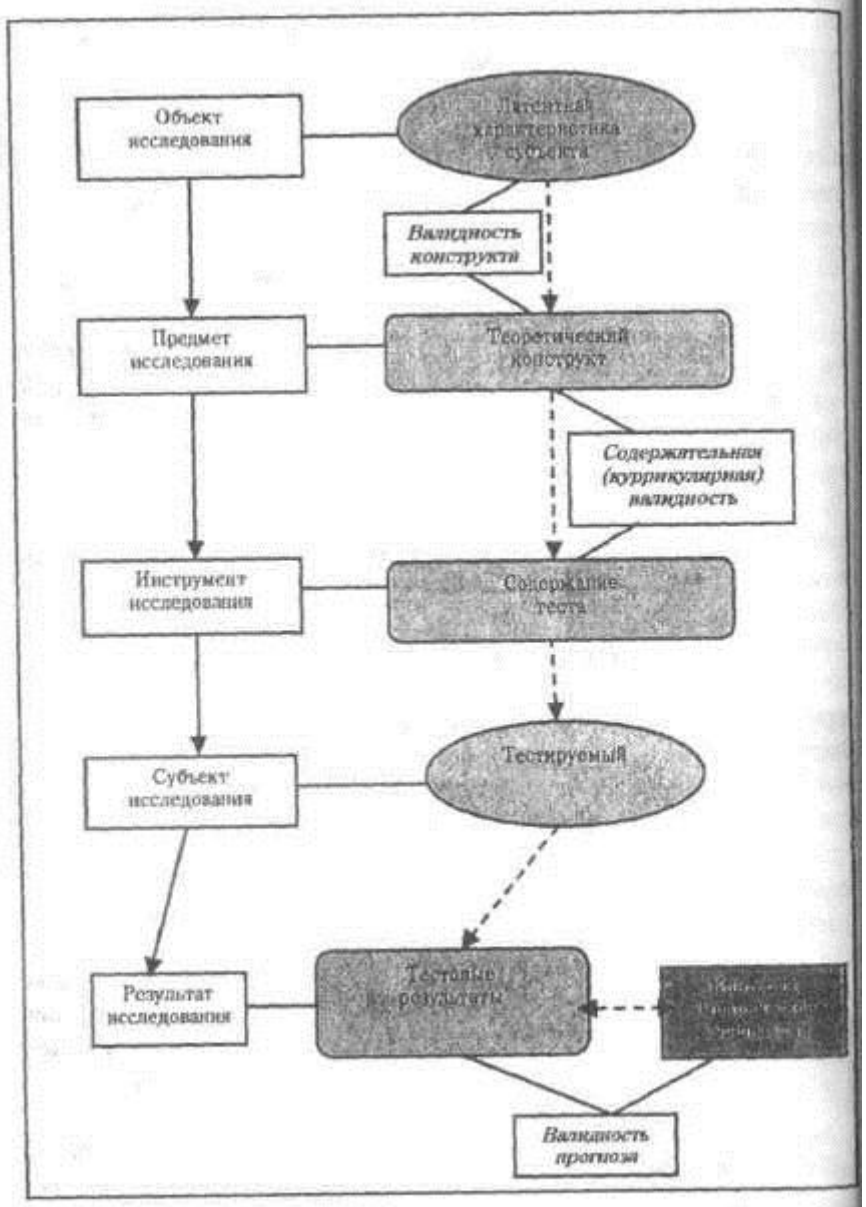


Рис. 1. Классическая схема валидизации латентного объекта

На рис. 1 представлена схема классического определения валидности единичного эксперимента. Пунктиром обозначены вероятностные и неоднозначные связи между элементами валидизации, указывающие на отсутствие внутренней валидности и, следовательно, на незамкнутость представленной системы валидности теста: конструктная, содержательная, прогностическая. Принцип истинности (когеренции или функциональности) лежит в основе выбора критерия эмпирической валидизации, считается валидным и определяет основу интерпретации тестовых результатов, но не влияет на общую схему валидизации измерения латентного объекта.

Вторым способом повышения объективности тестовых результатов оказывается улучшение качества построения процедуры эксперимента через использование валидных экспериментальных планов, позволяющих держать под контролем источники невалидности, угрожающие внутренней валидности эксперимента [65]. В работе Д. Кэмпбелла выделяется восемь источников невалидности как переменных, угрожающих внутренней валидности, и устанавливается возможность контроля над ними через 16 валидных экспериментальных планов. «Если эти переменные не контролируются в экспериментальном плане, то они могут дать эффекты, которые смешиваются с эффектом экспериментального воздействия» [Там же, с. 46]. А.С. Казаринов ввел подход Д.Кэмпбелла к повышению валидности эксперимента (валидацию) в педагогику [46].

Невозможность контроля над всеми факторами, угрожающими валидности, приводит к неограниченному росту числа новых видов валидности педагогического теста. Вследствие этого, А.С. Казаринов отмечает, что «валидность конструкта необходимо рассматривать как неисчерпаемый источник новых видов валидности: латентные свойства дня сегодняшнего, имеющие теоретическую направленность, завтра перестанут быть скрытыми и приобретут направленность сугубо практическую» [Там же, с. 61]. Тогда валидация тестов становится неограниченным процессом, а валидность теста оказывается недостижимой. Таким образом, для второго способа обеспечения внутренней валидности характерно неограничен-

ное увеличение количества видов валидности педагогического теста. Так, например, в работе Е.Н. Лебедевой [68] приводится шесть видов валидности: целеполагания (содержательная), соответствия и прогноза (эмпирические), латентная (конструкта), композиционная, технологическая (организационная), вводится понятие общей валидности теста, состоящей из линейной комбинации коэффициентов шести видов валидности с весовыми коэффициентами. В работах других авторов [46; 48; 92 и др.] список видов валидностей дополняется адаптационной валидностью как обобщенным понятием валидности инструментария [47; 48], психологической — как обобщенным понятием валидности педагогического теста [92] и др. Общая валидность исследования представляется как открытая неограниченная система ее видов.

На рис. 2 показана схема валидизации структурного конструкта. Пунктиром обозначены опосредованные и неоднозначные связи, как внешние, так и внутренние. Основным источником невалидности является тестируемый, влияние субъективных свойств которого провоцирует жесткую регламентацию процедуры тестирования и введение новых видов валидности для контроля над, ними. На схеме показана зависимость выбора типа тестового инструментария (нормативно-ориентированного или критериально-ориентированного) от критерия истинности (когеренции или функциональности соответственно), положенного в основу выбора валидного критерия эмпирической валидизации.

Иначе говоря, объективность тестовых результатов не обеспечивается открытой системой валидности, состоящей из неограниченного количества ее видов. Валидность эмпирического исследования, а следовательно, объективность его результатов недостижима. В практике применения тестов, как правило, ограничиваются вероятностным значением коэффициента валидности каждого вида, что, по нашему мнению, еще более усложняет процедуру определения валидности исследования, а его полная валидизация так и не достигается. При этом заметно снижается технологичность процедуры определения валидности исследования.

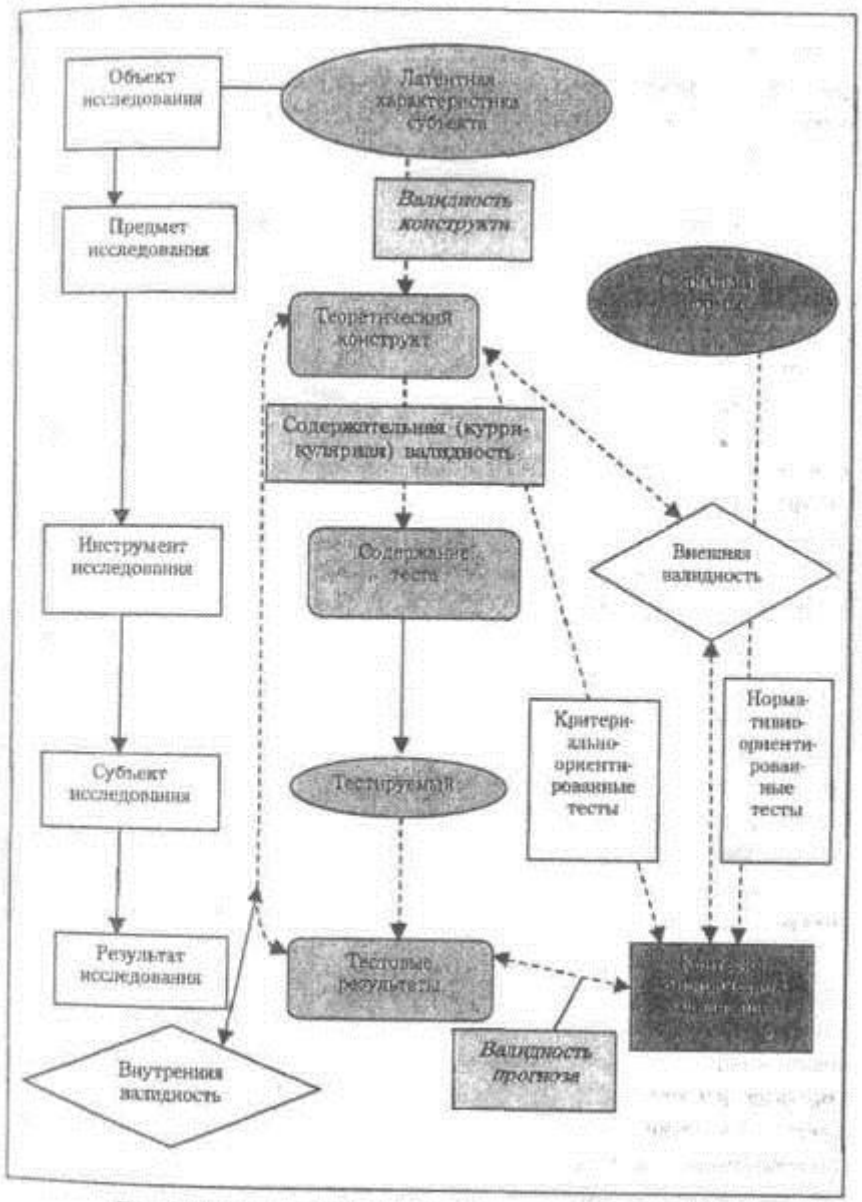


Рис. 2. Схема валидизации структурного конструкта

На наш взгляд, возможен третий, более технологичный способ достижения внутренней валидности тестовых результатов педагогического исследования. Для этого необходимо при определенных условиях соединить содержание педагогического теста и теоретический конструкт обученности в конструкт предметного содержания теста как целого. Такая возможность возникает в связи с тем, что при измерении обученности субъекта в роли объекта измерения выступает содержание обучения — определенное, хотя и многовариативное.

Определяя объект данного исследования как частный случай объекта дидактической тестологии, следует выделить границы и задать пределы применения метода измерения в рамках контроля обученности, а также разделить контроль обученности и диагностику обучаемости. Эта дифференциация обусловлена тем, что обучаемость является психологической характеристикой субъекта обучения и потому содержит в себе неопределенные латентные параметры, а обученность понимается как дидактическое понятие, выражающее определенные и контролируемые результаты процесса обучения. Понятие обучаемости несет в себе характеристики личности определяется как «индивидуальные показатели скорости и качеств усвоения человеком знаний, умений и навыков в процессе обучения» [61, с. 94], а следовательно, является предметом изучена психодиагностики и психолого-педагогических исследований. Обученность же необходимо понимать как результат обучения, заключающий в себе понимание системы предметного содержания как целого, представленного в разных аспектах или хотя бы в одном из них. Результат обучения, как мы полагаем, необходимо рассматривать только как целостное понимание предметного содержания как теоретического конструкта. Наша позиция принципиально отличается от репродуктивных систем обучения. Так, в рамках классического знаниевого подхода к обучению обученность понимается в аспекте воспроизводимости знаний, умений и навыков. В бихевиористском подходе обученность трактует как деятельностная воспроизводимость алгоритмов. Основной критикуемый недостаток репродуктивного подхода в образовании состоит в постоянном и неограниченном наращивании объема предметной информации, предназначенной для воспроизведения. Целостный подход в обучении требует представить предметное содержание как ограниченную замкнутую систему, организованную по внутренним теоретическим законам. Обученность достигается не увеличением числа воспроизводимых структурных единиц предметного содержания, но пониманием внутренних теоретических законов организации предметного содержания. Это интенсифицирует процесс обучения.

Понимание предмета измерения как теоретического конструкта целостного предметного содержания позволяет:

- перейти от латентного объекта исследования к определенному, представленному для контроля; заменить модель эксперимента как «черного ящика» моделью «прозрачного ящика»;
- построить ограниченную, замкнутую и целостную систему валидности теста;
- построить технологичную процедуру определения валидности теста, исключающую субъект измерения в целях получения объективных тестовых результатов;
- перейти в обучении от контроля над репродуктивными знаниями к контролю над системными.

Структура валидизации целостного конструкта представлена на рис. 3. Основной источник невалидности тестовых результатов — субъект тестирования — посредством выбранного критерия валидизации, не зависящего от его субъективных характеристик, выносится за пределы процедуры определения валидности. Тем самым снимается зависимость валидности тестовых результатов от статистических показателей и валидность теста перестает быть внешней.

На наш взгляд, мы доказали, что отличие объекта исследования в процессе контроля результатов обучения в образовании от объектов психологии, социологии и педагогики (в общем смысле) делает нецелесообразным (методологически ошибочным) прямой перенос методов психологического и социологического исследования в дидактический контроль.

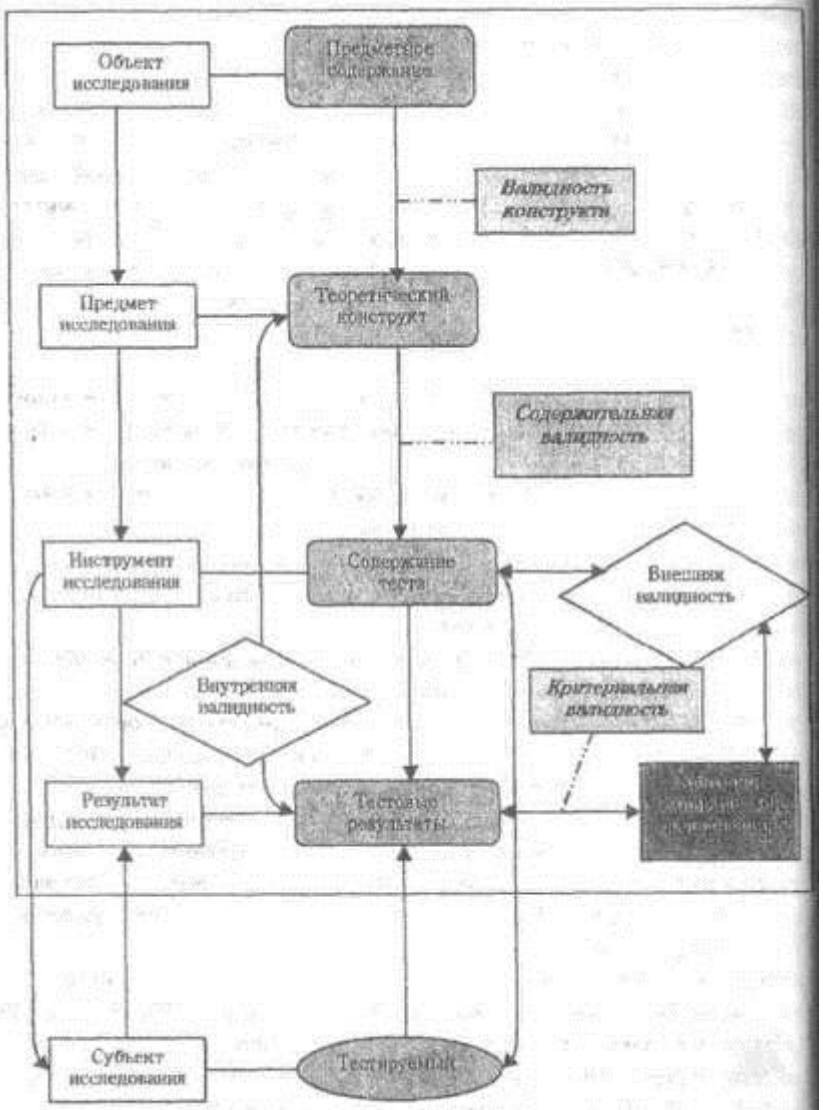


Рис. 3. Схема валидизации целостного конструкта

Определяемый нами подход вапидизации на основе теоретического конструкта как целостного предметного содержания при определенных условиях будет обеспечивать внутреннюю валидность тестового измерения. При наличии внутренней валидности измерения объективность тестовых результатов будет пониматься в аспекте их соответствия теоретическому конструкту, а значит, и определенному объекту измерения. Количество видов валидности в данном случае ограничено, так как нет необходимости контролировать бесконечное количество факторов, угрожающих валидности. Общая, комплексная валидность педагогического теста будет состоять из:

валидности конструкта,

валидности содержания,

эмпирической валидности, подтверждающей теоретическую валидность;

валидности теста по форме (технической), учитывающей специфические требования, предъявляемые к педагогическим тестам.

Таким образом, необходимо сделать следующие выводы. Эмпирическая валидность в прикладных науках решает проблему связи теоретического конструкта и результатов тестирования, причины и эффекта через привлечение внешнего критерия. В основе данной проблемы лежит латентность объекта исследования, что приводит к невозможности создания содержания теста, взаимно-однозначно представляющего конструкт и отвечающего требованиям внутренней содержательной валидности теста.

Другое решение проблемы связи конструкта и результатов приводит к необходимости обеспечения жесткой объективизации процесса тестирования и контроля за качеством проведения эксперимента. Это решение приводит к возникновению неограниченного количества видов валидности и необоснованно усложняет процесс валидизации теста.

Заимствование методов построения эксперимента, в том числе тестирования из прикладных наук в образование без учета специфики объекта исследования привело к неоправданному усложнению процесса получения адекватных результатов эксперимента и построению валидных тестовых измерителей. А также к затрудне-

ниям при выборе системы валидности теста, обеспечивающей адекватность результатов тестирования.

Изменение объекта исследования на модель «прозрачного ящика» позволяет решить проблему объективации результатов тестирования посредством внутренней содержательной валидности теста.

1.3. Технологизация валидности педагогических тестов

Стратегия получения валидных выводов эмпирического исследования выбирается в зависимости от типа объекта исследования, определяя связь теоретического конструкта с результатами тестирования. Латентность объекта исследования подразумевает либо привлечение внешнего критерия, либо установление контроля над внешними побочными факторами, влияющими на валидность выводов. В обоих случаях валидность вывода вероятностна и опирается на статистические показатели. Открытый и определенный тип объекта исследования позволяет определить содержательную валидность теста как внутреннюю валидность. В современной педагогической литературе разрабатываются проблемы определения системы валидности прикладного педагогического исследования (В.С. Аванесов [6], К. Ингенкамп [44], А.С. Казаринов [46], А.Н. Майоров [80], Е.А. Михайлычев [92] и др.) и способы валидизации (Дж. Глас, Дж. Стенли [32], М.И. Грабарь, К.А. Краснянская [33], О.Ю. Ермолаев [41], В.И. Михеев [95], Ю.М. Нейман, В.А. Хлебников [106], Е.В. Сидоренко [139] и др.).

Учитывая, что педагогическое тестирование является частым случаем прикладного исследования в социальных науках, проанализируем несколько общеизвестных определений педагогического теста с целью определения способа получения валидных результатов через установление достаточного количества видов валидности педагогического теста. Необходимость этого процесса определяется, по нашему мнению, как возможность становления первого технологического этапа в процедуре валидизации педагогического теста.

По мнению К.Ингенкампа, «тесты школьной успеваемости являются методом педагогической диагностики, с помощью которого результаты планируемого и определяемого куррикулумом учебного процесса могут быть максимально объективно, надежно и валидно измерены, обработаны, интерпретированы и подготовлены к использованию в педагогической практике учителями (частично также учащимися) или консультантами» [44, с. 125]. Данное определение представляет педагогический тест как валидный метод исследования. Он реализуется через валидность плана проведения эксперимента и валидность инструментария. Организация эксперимента является частью управления процессом тестирования и не входит в область наших исследовательских интересов. Предметом рассмотрения становится валидность педагогического теста, без которой нельзя получить адекватные результаты измерения.

В.С. Аванесов предлагает определение теста как инструмента. «Педагогическим тестом называется система заданий специфической формы, определенного содержания, возрастающей трудности — система, создаваемая с целью объективно оценить структуру и качественно измерить уровень подготовленности учащихся» [2, с. 5]. Из него следует, прежде всего, то, что валидным может быть признан тест, отвечающий специфическим требованиям к его форме. Во-первых, в тест должны входить только задания в тестовой форме, во-вторых, тест должен быть построен по возрастанию трудности каждого последующего тестового задания. Для создания валидного педагогического теста следует обеспечивать его определенную форму и структуру. Возникает первый вид валидности теста — *валидность теста по форме* (табл. 1). Выполнение параметров структуры валидного теста непосредственно связано с содержанием тестовых заданий.

Содержание тестовых заданий должно «находиться в соответствии с содержанием учебного материала. Другое содержание в тест не включается. Нарушение этого требования приводит к ухудшению качества теста, вплоть до того момента, когда могут возникнуть вопросы о цели теста...» [Там же, с. 5—6]. Следовательно, необходимо осуществлять контроль над содержанием тестовых заданий. *Валидность теста по содержанию* (см. табл. 1) задается выбором тестовых заданий с определенным содержанием и является вторым видом валидности.

Таблица 1

Система валидности педагогического теста

Необходимые компоненты	Виды валидности	Тип валидности
1. Теоретическая валидность	*Конструктная валидность* (задает адекватное объекту исследования представление теоретического конструкта как предмета исследования): — для целостного теоретического конструкта; — для структурного теоретического конструкта	— внутренняя; — внешняя
2. Валидность теста по содержанию	*Содержательная валидность* (отражает адекватность соответствия содержания теста теоретическому конструкту): — содержательная (для целостного теоретического конструкта); — куррикулярная; целеполагания; адаптационная (для структурного теоретического конструкта)	— внутренняя; — внешняя
3. Эмпирическая валидность	*Критериальная* (прогностическая): (направлена на эмпирическое подтверждение гипотезы исследования. Зависит от критерия валидизации)	В зависимости от типа объекта и критерия валидизации
4. Валидность теста по форме	*Техническая* (соответствие формы и структуры теста принятым в тестологии нормам): — композиционная; — технологическая	—

Валидный тестовый измеритель должен давать валидные результаты тестирования. Ю.М. Нейман и В.А. Хлебников дают определение, в котором под педагогическим тестом понимается «некоторое множество специально составленных заданий специфической формы, позволяющее по ответам испытуемых объективно измерить на определенной шкале их уровень подготовленности по конкретным разделам определенной области знания» [106, с. 6]. В этой

дефиниции кроме формы и содержания теста внимание уделяется распределению результатов тестирования на некоторой числовой шкале. Для их правильной интерпретации необходимо убедиться в адекватности переводной шкалы. Соответствием распределения результатов тестирования по данной шкале внешним критериям или нормам будет подтверждаться адекватность интерпретации. Если данный тест измеряет конструкт адекватно другим измерениям того же конструкта, то данное соответствие окажется эмпирическим подтверждением гипотезы исследования или прогноза.

Третий вид валидности педагогического теста определяется интерпретацией результатов тестирования и адекватностью переводной шкалы. Адекватность интерпретации результатов и переводных шкал достигается посредством сравнения эмпирических результатов тестирования с эмпирическими результатами, полученными при помощи других форм контроля. Кроме того, можно использовать валидные тесты, оценивающие ту же характеристику объекта (принцип когеренции). Наконец адекватность интерпретации реализуется при сравнении тестовых результатов измерения принятым нормам (принцип функциональности). При соответствии двух рядов результатов можно сделать вероятностное предположение об адекватности интерпретации результатов тестирования и, в частности, об адекватности переводной шкалы.

Этот вид валидности зависит от эмпирических данных, а значит, относится к внешней валидности. Адекватность интерпретации результатов и переводной шкалы всегда оказывается относительной. Условно обозначим его как *эмпирическую валидность* (см. табл. 1).

Из рассуждений Ю.М. Неймана и В А Хлебникова следует, что целью тестирования является определение уровня подготовленности учащихся по некоторому предметному содержанию. Здесь под «уровнем подготовленности» понимается уровень обученности учащегося по некоторому определенному предметному содержанию, а также его умения и навыки. При этом авторы отмечают, что выделенная характеристика «является, к сожалению, латентным параметром (т. е. недоступным для непосредственного измерения)»

[Там же]. Как мы полагаем, для того, чтобы измерение стало возможным, «уровень подготовленности» должен быть представлен как теоретический конструкт определенного предметного содержания. Содержание теста должно быть адекватно теоретическом конструкту «уровня подготовленности», который измеряется. Так возникает следующий, четвертый вид валидности теста — *теоретическая валидность* (см. табл. 1)-

Итак, можно выделить четыре основных вида валидности, необходимых для обеспечения валидности педагогического теста как инструмента измерения (см. табл. 1):

1) валидность теста по форме;
2) валидность теста по содержанию;
3) эмпирическая валидность;
4) теоретическая валидность.

В зарубежной и отечественной литературе валидность как показатель качества теста определяется с точки зрения двух подхо-дов: качественного [44; 65; 80] и количественного, как коэффици- ент, отражающий степень соответствия результатов выбранному критерию или социальной норме [46; 68; 77; 80; 127; 166].

Проблема обеспечения технологической валидности теста решается при помощи привлечения для ее оценки специалистов-тестологов на основе метода групповых экспертных оценок [68; 171]. Она носит достаточно специфический характер и в той части, которая связана с правильностью формы тестовых заданий [8], не являетеся полноценной валидностью. Ее не относят к валидности как адекватности, взятой в том или ином аспекте. Но та часть технологической валидности, которая связана со структурой теста, т. е. с самим тестовым содержанием, оказывает непосредственное влияние на валидность результатов. Зачастую технологическая валидность обеспечивается еще на этапе создания теста. Например, она задается при построении тестов на основе применения теории IRT (Item Response Theory), т. е. валидность как бы «зашита» в конструировании теста. Здесь применяется количественный анализ.

Известно, что коэффициент валидности теста есть число, показывающее степень связи двух характеристик некоторого признака

или двух признаков, представленных количественно. Для вычисления коэффициента валидности теста (количественной оценки валидности) на основе корреляционного и факторного анализа используют различные коэффициенты корреляции: коэффициент ассоциации Пирсона, коэффициент «фи», коэффициент бисериальной корреляции, коэффициент Спирмена, коэффициент Кендалла и др. (более 50 коэффициентов). Выбор коэффициента корреляции зависит от того, в какой шкале получены количественные характеристики [32; 80; 106].

Для вычисления коэффициента корреляции необходимо иметь два ряда числовых значений (оценок) количественных характеристик. Первый ряд чисел составляют количественные оценки, полученные в результате применения теста. Вторым рядом являются количественные оценки, полученные от внешних или внутренних по отношению к тестированию источников, «эталонных» критериев и норм.

Получение удовлетворительного значения (0,68—1) коэффициента корреляции подтверждает наличие связи между сопоставляемыми характеристиками, но не доказывает ее. Е.В. Сидоренко обращает внимание на то, что наличие корреляционной связи двух характеристик, представленных численно, «подразумевает любые согласованные изменения, которые могут объясняться сотнями причин. Корреляционные связи не могут рассматриваться как свидетельство причинно-следственной связи, они свидетельствуют лишь о том, что изменениям одного признака, как правило, сопутствуют определенные изменения другого, но находится ли причина изменений в одном из признаков или она оказывается за пределами исследуемой пары признаков, нам не известно» [139, с. 201]. Для доказательства связи необходим качественный анализ валидности эксперимента (тестирования) и его результатов. Получение положительного заключения такого анализа определяется удовлетворительным по значению коэффициентом корреляции, т. е. коэффициентом валидности, который будет подтверждать валидность теста.

Для определения валидности теста с точки зрения качественного подхода используется метод экспертных оценок. В своей

работе эксперты анализируют, насколько достижима цель тестирования при использовании того или иного теста. Выделяются характеристики измеряемого признака, имеющие определяющее значение при достижении цели данного тестирования, в аспекте валидности. Другими словами, экспертами выбирается теоретический конструкт измеряемого признака (в рассматриваемом случае «уровень подготовленности») и проводится анализ содержания теста на соответствие теоретической модели измеряемого признака, а также устанавливается соответствие заданий теста форме тестовых заданий, принятой в педагогической тестологии [2]. Валидность как качественная характеристика, взятая с использованием экспертной оценки, выражается определенной степенью: высокая, средняя, низкая [80] или крайними точками градации теста: валиден — невалиден [65].

В настоящее время при оценке валидности педагогических тестов составители, как правило, ограничиваются формальным вычислением коэффициентов различных видов валидности. Тест считается валидным, если полученные коэффициенты валидности принимают высокие значения. Считается, что получением высоких численных результатов бесспорно доказывается высокое качество теста. Д. Кэмпбелл, являясь сторонником количественного экспериментального подхода к исследованию, обращает внимание на то, что «цифры, выдаваемые вычислительной машиной, часто создают псевдонаучный фасад, не имеющий на деле валидного качественного основания» [Там же, с. 267]. Никакой высокий количественный результат не может вызывать доверия, если он не подтвержден качественным анализом. А.Н. Майоров также указывает на то, что «часть показателей валидности может быть определена статистическими методами — корреляционным и факторным анализом. Однако судить только по ним о валидности теста некорректно. Для оценки валидности должен быть рассмотрен весь комплекс возможных показателей» [80, с. 213].

Таким образом, в ходе качественного анализа выясняется:

1) теоретическое обоснование адекватности конструкта через качественный анализ валидности конструкта;

2) теоретическое обоснование адекватности содержания теста конструкту посредством качественного анализа содержательной валидности;

3) возможность установления валидности теста по форме;

4) выбор критерия для количественной проверки адекватности тестовых результатов (эмпирической валидизации).

Выбор критерия для корреляции определяет способ эмпирической валидизации. А. Анастази отмечает, что «валидность по критерию показывает, насколько по результатам теста можно судить об интересующем нас аспекте поведения индивида в настоящем или будущем. Чтобы определить ее, выполнение теста соотносят с критерием, т. е. с непосредственной и независимой мерой того, что должен предсказать тест» [9, с. 131]. Выбранный критерий может быть внешним, не связанным с содержанием валидизируемого теста, или внутренним, согласованным с тестом (табл. 2).

Таблица 2

Классификация тестов на основе выбора критерия эмпирической валидизации

Эмпирическая валидность по внешнему критерию (в зависимости от цели применения теста)	Внутренний критерий эмпирической валидизации (в зависимости от цели интерпретации тестовых результатов)	
	предметная норма (интерпретация результатов по содержанию)	статистическая норма (интерпретация результатов по группе)
соответствия	критериально-ориентированные предметные тесты	нормативно-ориентированные предметные тесты
прогноза	тесты профессиональной пригодности	тесты дифференциации успешности обучения

В зависимости от выбранного внешнего критерия для коэффициента корреляции, валидность разделяют на консенсусную, когда результаты тестирования коррелируют с результатами, полученными от внешних экспертов, и эмпирическую, когда результат тестирования коррелируют с результатами применения других источников.

Валидным признается тест, результаты которого положительно коррелируют с результатами других валидных измерений того же или близкого по смыслу теоретического конструкта (конвергентная валидизация). Он может быть валидным и тогда, когда его результаты не коррелируют с результатами валидных измерений другого по смыслу теоретического конструкта, не связанного с валидизируемым (дискриминантная валидизация) [37; 46; 65; 80; 92].

Нецелесообразно использование дискриминантной валидизации в дидактическом контроле. Неясным является выделение конструкта, не связанного с латентной характеристикой, и получение на его основе адекватного, т. е. валидного критерия. В дидактическом контроле латентной характеристикой оказывается уровень подготовленности субъекта по определенному предметному содержанию. Тогда в качестве эталонного измерения должен выступать уровень неподготовленности по данному предметному содержанию. Но, во-первых, измерение уровня неподготовленности столь же сложно, что и измерение уровня подготовленности. Во-вторых, если даже предположить, что измерения все-таки произведены, возможна обратная зависимость результатов. Кроме того, в качестве эталонных могут быть взяты результаты измерения характеристики, которая не зависит от измеряемой, т. е. характеристики другого предметного содержания, «как, например, навыки лабора-торно-химических анализов при проверке грамотности по родному языку» [92, с. 263]. Отсутствие значимой корреляции не подтверждает, что измеряется именно грамотность родного языка, а наличие корреляции не может ее опровергать, так как возможно существование более общего психологического конструкта, напри- мер, обучаемости или способности субъекта к обучению. Таким образом, при дидактическом контроле над результатами обучения

невозможна дискриминантная валидизация тестов. Нашим выводом подтверждается необоснованность переноса методов построения тестов, из прикладных наук в сферу образования и дидактического контроля его результатов.

При конвергентной валидизации оба теста строятся из общего конструкта, либо из близких по содержанию дидактических конструктов. Наличие корреляции между результатами измерения близких конструктов не гарантируют качества валидизируемого теста. Рассмотрим, например, тесты по темам: «Исследование квадратичных функций» и «Решение неравенств второй степени». Наличие корреляции может свидетельствовать:

1) о наличии общей части конструкта (но это очевидно);
2) одинаковой степени усвоения материала;

или разной чувствительности тестов как инструментов и др.

Отсутствие корреляции свидетельствует, например, о разной степени усвоения материала, о том, что решение неравенств второй степени базируется на выученном алгоритме более общего метода интервалов, а не на фактических свойствах квадратичной функции, т. е. на внешних по отношению к тесту факторах, учесть которые проблематично. Следовательно, при конвергентной валидизации наиболее обоснованным является привлечение эталонных результатов, базирующихся на общем конструкте с валидизируемым тестом. Возникают следующие трудности выбора критерия валидизации.

1. Если критерием служит другой заведомо валидный тест, то в результате может получиться параллельный тест. А. Анастази отмечает, что «если новый тест не является более простым или более коротким вариантом уже существующего теста, то использование последнего в качестве критерия недопустимо» [9, с. 136].

2. Если критерием служит контрольная работа, то возникает проблема валидности и надежности ее результатов. «Традиционный подход, при котором главным внешним критерием является только успеваемость по учительским оценкам, издавна критикуется за субъективность, ненадежность, и невалидность» [92, с. 273].

3. Если критерий получен на основе экспертной оценки, то для повышения валидности необходимо устранить контаминацию критерия, т. е. «совершенно необходимо, чтобы лицам, производящим критериальную оценку, ничего не было известно о тестовых результатах испытуемого» [9, с. 132].

4. Если критерием валидизации абитуриентского теста является успеваемость студента в 1-м семестре, то возникает проблема сдвига критерия из-за изменения выборки испытуемых и изменения конструкта. Иногда ставится под сомнение релевантность критерия. Хотя на первый взгляд данный критерий соответствует цели такого тестирования. «Если этот критерий складывается из субъективных оценок преподавателей различных специальных высших школ, то он лишь в малой степени отвечает научным требованиям» [44, с. 41—42].

Валидность теста — понятие целезависимое, поэтому выбор критерия во многом зависит от цели тестирования. В зависимости от выбора внешнего критерия для валидизации эмпирическая валидность теста подразделяется на два подвида: соответствия и прогноза [44] (см. табл. 2).

«Валидность соответствия помогает установить, в какой мере результаты, полученные посредством использования исследовательского инструментария, соответствуют данным, полученным другими способами» [Там же, с. 41]. Часто в педагогической литературе используются синонимы валидности соответствия, такие как текущая валидность [9; 10; 80], диагностическая валидность [79], конкурентная валидность [82; 106], сопоставительная валидность [19].

«Валидность прогноза позволяет вычислить взаимосвязь между данными, полученными в ходе ранее проведенного исследования, и поведением, наблюдаемым позднее» [44, с. 41]. А. Анастази обращает внимание на то, что хотя текущая валидность (соответствия) теста применяется при оценивании характеристики, сложившейся к данному моменту, а прогно-стическая спустя некоторое время, но основное различие между ними «основано не времени, а на цели тестирования» [9, с. 132].

К. Ингенкамп отмечает, что при доказательстве валидности инструментария проблематично выделить «всеми признанный критерий». Неудовлетворительное качество имеющихся методик оценивания характеристики субъекта побуждает к созданию новых измерителей, и, следовательно, первые не могут и не должны использоваться в качестве критерия при определении валидности второго [44].

Если целью тестирования является выявление уровня подготовленности по определенному предметному содержанию, то валидность соответствия призвана обеспечивать сопоставимость результатов, полученных разными способами. Поскольку получить валидный критерий практически невозможно, постольку объективность тестовых результатов относительна, т. е. зависит от других способов оценивания того же содержания, считающихся валидными. В практике педагогического тестирования, как правило, в качестве внешнего критерия выбирается, например, экспертная оценка знаний по тому же разделу, результаты письменной работы того же самого содержания, результаты других тестов, полученные на той же выборке. Полученный в процессе исследования коэффициент корреляции является количественным показателем валидности соответствия, основанной на принципе когерен-ции. Это значит, что тестовые результаты не противоречат другим формам контроля, а соответствуют им.

Для достижения соответствия важное значение имеет шкала перевода тестовых баллов в оценку. Переводные шкалы строятся на основе мнений экспертов или опираются на нормальный закон распределения тестовых баллов, либо вводятся декларативно. Адекватность переводной шкалы означает, что результаты тестирования соотносятся с результатами, полученными при помощи других форм контроля, выбранными в качестве критерия. Именно поэтому валидность соответствия, основанную на принципе когеренции, можно рассматривать как адекватность интерпретации тестовых результатов другим форм контроля или адекватность переводной шкалы.

Если же целью тестирования оказывается установление дальнейшей успешности обучения или профессиональной деятельнос-

ти, то рассматривается валидность прогноза: «прогностическая валидность имеет значение в дидактическом тестировании при оценке возможной успешности обучения... на следующей ступени образования» [92, с. 284]. Ее численное значение есть коэффициент корреляции тестовых результатов с результатами дальнейшем отсроченной по времени деятельности. В качестве критерия могут выступать полученные различными способами оценки другого предметного содержания, другой деятельности. Успешность иной деятельности оценивается относительно присущих ей социально- значимых норм. Другими словами, валидность прогноза показывает, в, какой степени полученные тестовые показатели измерения латентной характеристики способны прогнозировать успешность иной деятельности. Результаты тестирования сравниваются с результатами измерения отсроченного по времени иного признака иного предметного содержания, иного конструкта. Но условием конвергентной валидизации является общность конструктов обоих измерений. Следовательно, конвергентная прогностическая валидизация не существует, если в качестве критерия выбираются результаты измерения другого конструкта, отсроченные по времени. Или, другими словами, невозможно проведение конвергентной прогностической валидизации на основе принципа когеренции При прогностической валидизации в качестве критерия выступает социально значимая норма, основанная на принципе функциональности (см. табл. 2).

Если результаты измерения уровня обученности сопоставляются с социально значимыми нормами, то при их соответствии можно сделать вывод о прогнозе успешности применения предметного содержания в дальнейшей деятельности и невозможно строит предположение об уровне обученности по определенному предметному содержанию.

Можно заключить, что эмпирическая валидность обеспечивается соответствием результатов тестирования результатам измерения того же признака, полученным другими способами. В зависимости от выбора внешнего критерия она подразделяется на валидность соответствия и валидность прогноза. Кроме того, эмпирическая валид-

ность является доказательством значимости ранее доказанной теоретической валидности. Практическое использование эмпирической валидности обусловлено возможностью получения относительно адекватных переводных шкал (см. табл. 2).

Удобно объединить все эти виды валидности под общим названием критериальная или эмпирическая валидность, так как они определяются выбором критерия для сопоставления результатов [44; 80; 113]. Они являются разновидностями прогностической валидности, используемой в социологии и психодиагностике для эмпирического подтверждения прогноза исследования, и зависят от свойств выборки стандартизации, а следовательно, относятся к внешней валидности теста.

Общим для данных подвидов валидности является не только присутствие внешнего критерия, но и участие в процессе валидизации определенной и постоянной выборки — выборки стандартизации. Как мы отмечали, основной задачей внешней валидности служит распространение результатов эксперимента на всю генеральную совокупность тестируемых. Основным условием, позволяющим проводить экстраполяцию результатов, становится репрезентативность выборки стандартизации. Репрезентативность выборки задается стратификацией, т. е. включением «представителей всех наиболее значимых страт, реально существующих в генеральной совокупности потенциальных участников тестирования, причем в той же пропорции» [106, с. 38]. Здесь под стратами понимаются определенные слои генеральной совокупности: сельские или городские жители, обучающиеся по углубленной или обычной программе по предмету тестирования и т. п. Полученная репрезентативная выборка используется для установления тестовых норм и называется выборкой стандартизации. «Результаты тестирования (в частности, математическое ожидание, дисперсия) на образованной таким образом выборке называют тестовыми нормами (т. е. нормальными, усредненными). Сопоставление результатов относительного тестирования с нормами позволяет ослабить зависимость тестового балла от индивидуальных особенностей той группы людей, в которой проводилось тестирование, и тем самым

стандартизировать результат» [Там же]. Для определения места или рейтинга каждого тестируемого результаты тестирования генеральной совокупности сравниваются с тестовыми нормами. Использование теста с определенной выборкой стандартизации дает качественные и валидные результаты в случае применения его на генеральной совокупности с соответствующими данной выборке характеристиками страт, указанными в спецификации теста. При несоблюдении условия становится невозможной экстраполяция результатов, а полученные данные перестают быть валидными. Важно отметить, что этот аспект контроля за качеством получаемых тестовых результатов переводит проблему репрезентативности выборки из задачи создания валидного инструмента в проблему валидности (адекватности) применения теста. Другими словами, создание качественного теста становится проблемой качественного управления тестированием. Она не входит в область нашего исследования. Таким образом, наше рассуждение доказывает, что участие выборки стандартизации при определении тестовых норм переводят данный вид валидности в разряд внешней валидности. Однако этот вывод сохраняет проблему установления внутренней валидности педагогического тестирования, без которой невозможно получение объективных тестовых результатов в аспекте их соответствия теоретическому конструкту. Для решения этой задачи следует обратиться к анализу выбора возможных внутренних критериев, которые могут задать искомую валидность.

Существуют два подхода к выбору внутреннего критерия. Пер-вый основан на принципе когеренции. В литературе по тестированию на его основе рассматриваются методы внутренней согласованности и факторной валидизации [9; 10; 46; 92]. Оба они в качестве критерия используют внутренний тестовый критерий, связанный с содержанием теста. Принцип когеренции применяется как соответствие содержания каждого тестового задания или содержания группы тестовых заданий содержанию теста в целом. Внутренним критерием является суммарный показатель теста и проверяется его связь с результатами тестирования по отдельным частям теста. Принципом группировки отдельных частей служит

либо принадлежность тестовых заданий к одному разделу программы (внутренняя согласованность), либо к значимым для обоснования конструкта факторам (факторная валидность).

Выбор внутреннего критерия для валидизации, обусловленного содержанием самого теста, позволяет, на первый взгляд, говорить о внутреннем характере данного вида валидности. Но при этом очевидно, что содержание теста должно описывать измеряемую предметную область и как можно более полно и разносторонне приближаться к ней, т. е. содержать задания различной сложности, нацеленные, как принято в педагогическом контроле, на проверку различных элементов структуры измеряемой предметной области. Чем выше степень такого приближения, тем выше содержательная валидность теста. Справедливо и обратное положение. Чем выше содержательная валидность, тем больше в тесте заданий различной сложности, представляющих различные элементы структуры измеряемого содержания. Тестовые задания не должны быть однородными, описывающими какой-то один аспект измеряемого предметного содержания.

Однако наряду с однородностью необходима внутренняя согласованность содержания тестовых заданий. «Тест называется надежным, если он является внутренне согласованным» [80, с. 205], т. е. каждое задание теста должно коррелировать с тестом в целом или (и) с другими заданиями теста [1; 80; 166]. Повышение надежности происходит посредством включения в тест более однородных по содержанию заданий, что снижает содержательную валидность теста. Эта взаимосвязь называется парадоксом Ф. Лорда [5]. В практике создания тестов, как правило, ограничиваются конструированием относительно надежного и относительно валидного теста. МБ. Челышкова отмечает: «В целом при создании гомогенного теста к однородности содержания следует стремиться, но в тех пределах, которые не нарушают правильные пропорции содержательных элементов в итоговом тесте» [166, с. 302].

Следует отметить, что надежность теста как в аспекте релиабильности, так и в смысле внутренней согласованности экспериментально определяется при пилотажном тестировании. Она

зависит от свойств выборки стандартизации, а следовательно, является внешней характеристикой качества теста. Выбор внутреннего критерия для подсчета корреляционного коэффициента не относится к валидности, так как «при отсутствии информации, внешней по отношению к тесту, мало что можно сказать о том, что он измеряет» [9, с. 143]. Таким образом, выбор критерия, связанного с содержанием самого теста, для вычисления коэффициента корреляции приводит к появлению надежности теста как внутренней согласованности или однородности, т. е. еще к одной внешней характеристике теста, но не к внутренней валидности. Выбор внутреннего критерия валидизации на основе критерия когеренции не позволяет конструировать тесты с внутренней валидностью, но обеспечивает их надежность. Надежные тесты позволяют получать объективные тестовые результаты в аспекте их содержательной однородности или точности измерения представленного тестового содержания, но не в аспекте их адекватности теоретическому конструкту измеряемого предметного содержания, т. е. нельзя быть уверенными в том, что надежные результаты гарантируют соответствие объекту, исключая субъективность через однородность теста. Любая интерпретация тестовых результатов оказывается относительной в аспекте измеряемой характеристики объекта.

Чем выше степень однородности теста, тем точнее, надежнее такой тест измеряет определенный элемент структуры содержания, на проверку которого направлена данная совокупность заданий. Целью тестирования становится выявление степени владения данным содержательным элементом, а не определение рейтинга или места учащегося относительно всей группы тестируемых. Несмотря на то, что они относительны по содержанию, представляемому тестом, их называют абсолютными. «Абсолютное тестирование можно назвать критериально-ориентированным, если указывается некоторое пороговое значение уровня подготовленности в четко определенной области знаний, и цель тестирования состоит в том, чтобы проверить, преодолел ли каждый участник тестирования это значение... или не преодолел» [106, с. 39]. Высокая

степень однородности содержания критериально-ориентированного теста достаточно точно описывает измеряемую латентную характеристику, показывая, что тест надежен. Но высокая надежность не обеспечивает валидность теста. Сохраняется необходимость установления связи однородного содержания теста, надежно измеряющего некую характеристику, и предмета измерения — теоретического конструкта.

Отсутствие внутренней валидности надежного теста предполагает его дальнейшую эмпирическую валидизацию для подтверждения теоретической валидности. Для валидизации выбирается тот или иной внешний критерий, зависящий от цели применения теста. Возникает возможность классифицировать способы валидизации по целям и критериям (см. табл. 2). Если высокая однородность содержания сочетается с валидностью соответствия по внешнему критерию, описывающему тот же содержательный элемент, то такой критериальный тест следует отнести к тестам достижений. Если тест помимо высокой однородности обладает прогностической валидностью, то такой тест скорее является аналогом тестов на профессиональную пригодность. Тогда проблема надежности критериально-ориентированного теста как реабильности, или точности, перестает быть актуальной. Решается парадокс Лорда: чем более однородно описывает измеряемый конструкт содержание теста, тем точнее описан конструкт в разных аспектах, т. е. тест валиден по содержанию, и тем более надежные выводы можно сделать о свойствах латента. С этим связано установление достаточно высокого проходного порога успешного выполнения критериально-ориентированного теста — 90-95 %.

Итак, выбор внутреннего критерия на основе принципа когеренции позволил установить степень надежности тестовых измерителей как внешнюю характеристику качества теста, но он не обеспечил внутреннюю валидность результатов тестирования.

Второй подход к выбору внутреннего критерия основан на принципе функциональности, который основан на выборе внутренней тестовой нормы в качестве критерия валидизации. Целью тестирования является соотнесение результатов тестирования с некоторой

признанной нормой, с помощью которой эти результаты сравниваются между собой, и становится возможным определение места или ранга учащегося относительно всех тестирующихся. Такое тестирование называется относительным. «При относительном тестировании важно добиться значимого расхождения результатов как можно большего числа испытуемых. Следовательно, желательна большая дисперсия баллов и приближенно нормальное их распределение» [106, с. 40]. Здесь внутренней нормой становится нормальный закон распределения тестовых результатов.

Если эмпирическое распределение результатов принципиально отличается от нормального закона, то это «говорит, скорее всего, о том, что тест составлен неудачно, и надо менять трудности заданий теста» [Там же, с. 136], а также следует перейти к «принудительной нормализации» [Там же]. Добиться нормального распределения результатов помогает включение в тест заданий, обладающих специальными характеристиками: параметр трудности, параметр дифференцирующей способности, «оценки которых должны обладать свойством инвариантности относительно уровня подготовленности тестируемой группы» [166, с. 55]. При помощи таких заданий добиваются поставленной цели — дифференциации учащихся по степени владения представленным тестовым содержанием. Другие задания в тест не включаются, так как считаются невалидными, не обладающими основными статистическими характеристиками.

Так строятся тесты в современной теории IRT (Item Response Theory), или в русском языке — ТППТ (теория моделирования и параметризации педагогических тестов) [106]. «IRT является инструментом (для нормативно-ориентированного случая — главным). Собственно тест является объектом применения IRT, которая позволяет получить математически корректный тест («надежный»). Вопрос же о его «внешней» валидности остается открытым» [131, с. 15].

Таким образом, при выборе внутренней нормы в качестве критерия получаем нормативно-ориентированный тест, обладающий высокой степенью надежности как релиабильности. Надежность есть характеристика внешняя, зависящая от статистических свойств выборки стандартизации.

Возникает возможность классифицировать нормативно-ориентированные тесты в зависимости от цели тестирования выбора внешнего критерия для валидизации (см. табл. 2). Если надежный нормативно-ориентированный тест обладает валидностью соответствия, то его следует отнести к тестам, дифференцирующим учащихся по степени владения содержанием теста, или тестам, определяющим ранг учащегося. Если такой тест обладает валидностью прогноза, то он определяется как тест успешности дальнейшей деятельности.

ВЫВОДЫ ПО ПЕРВОЙ ГЛАВЕ

В результате проведенного анализа проблемы обеспечения валидности тестовых результатов можно сделать следующие выводы.

1. Валидность тестовых результатов как адекватное отражение характеристик объекта педагогического исследования невозможна без обеспечения внутренней валидности исследования. При латентном характере объекта педагогического исследования обеспечение внутренней валидности эксперимента невозможно через получение валидного инструмента измерения — теста, обладающего внутренней валидностью. Четыре основополагающих вида валидности педагогического теста, такие как конструктная, содержательная, критериальная и технологическая, являются внешними видами валидности теста. Они не могут обеспечить контроль над тем, что именно измеряет тест.

Другие виды валидности педагогического теста, выделяющиеся исследователями, оказываются разновидностями четырех основных ее видов и отличаются выбором того или иного критерия эмпирической валидизации в определенном аспекте истинности.

2. Отсутствие внутренней валидности теста не позволяет создать полной системы видов валидностей, обеспечивающей валидность тестовых результатов. Внутренняя валидность возможна в ходе эксперимента или педагогического измерения характеристик

латентного объекта при контроле над всевозможными факторами, угрожающими внутренней валидности. Их количество неограничено, а значит приводит к проблеме неограниченного увеличения количества видов валидности исследования, которые задают бесконечную систему валидности тестового измерения. Процесс валидизации тестов становится нетехнологичным.

3. Прямой перенос общих методов эмпирических исследований из социальных наук в педагогику и, в частности, в образование приводит к проблемам получения качественных результатов исследования. Их можно избежать, учитывая особенности исследуемого педагогического объекта. В области педагогического контроля результатов обучения становится возможной замена латентного объекта исследования на открытый, что обеспечит внутреннюю валидность теста через внутренний характер содержательной валидности.

Открытость и определенность объекта исследования задает достаточную и ограниченную систему внутренней валидности педагогического теста, состоящую из конструктной, содержательной и критериальной валидностей, которая дополняется технической.

5. Открытый и определенный характер объекта исследования обусловливает способность представлять объективность тестовых результатов в аспекте их соответствия объекту измерения, так как субъект тестирования исключается из процедуры создания валидного теста. Становится возможным построение технологичной процедуры определения валидности педагогических тестов, ввиду того, что процесс валидизации перестает зависеть от субъектов тестирования и состоит из ограниченного количества этапов определения валидности, равных количеству видов валидности теста.

Глава 2
ОБЪЕКТИВНОСТЬ КОНТРОЛЯ ОБУЧЕННОСТИ КАК ПЕДАГОГИЧЕСКАЯ ПРОБЛЕМА

2.1. Технологичность конструкта обученности как основа объективности педагогического контроля

Традиционный педагогический эксперимент с использованием дидактического теста начинается с процесса построения теоретического конструкта объекта измерения. В настоящее время ученые считают, что уровень обученности учащегося как объект измерения является латентной характеристикой (К. Ингенкамп [44], А.С. Казаринов [46], А.Н. Майоров [80], ЕА Михайлычев [92] и др.). Построение теоретического конструкта такого объекта непосредственно связано со способами структурирования предметного содержания. Организации содержания посвящены труды В.И. Байденко [17], А.Г. Каспржак, М.В. Левита [51], В.С. Леднева [72], А.А. Мирошниченко [91], Б.У. Родионова, А.О. Татура [127], Ю.Н. Семина [136], А.М. Сохора [146], В.С. Цетлина [162] и других авторов.

Для достижения объективности результатов эксперимента в аспекте их соответствия объекту необходимо обеспечить внутреннюю валидность исследования. Однако использование структурного конструкта предметного содержания не позволяет создавать тесты, обладающие внутренней валидностью. Достижение внутренней валидности и, как следствие, получение объективных тестовых результатов становится задачей объективного проведения эксперимента.

Устранение интерсубъективных воздействий в ходе эксперимента возможно при контроле над всеми внешними факторами, угрожающими внутренней валидности, что не представляется возможным ввиду их неограниченного количества. Другими словами, всегда найдется внешний неконтролируемый фактор. В данной ситуации неограниченным становится количество соответствующих видов валидности. Необходимо возникает иерархия видов валидности по степени их значимости. В зависимости от цели тестирования, выбора критерия истинности и условий тестирования отбираются наиболее значимые для данного эксперимента виды валидности. Возникает понятие адаптационной валидности результатов эксперимента как степень адаптации к латентному объекту и понятие валидации как неограниченного процесса повышения валидности [46]. Чем выше степень адаптации, тем выше степень точности результатов тестирования и ниже степень ошибки измерения. Степень точности — это аспект надежности измерения в смысле его релиабильности, и определить ее возможно с заданной долей вероятности (надежности, точности), применяя математико-статистический аппарат, в том числе корреляционные коэффициенты, разработанные и заимствованные из психодиагностики и социологии.

В этой связи латентность объекта и соответствующий ему структурный характер предмета исследования, по нашему мнению, создают ситуацию неопределенности при выборе ограниченного количества видов валидности, участвующих в валидизации как процессе определения валидности. Такой выбор всегда субъективен, а значит, объективность исследования становится сомнительной, внутренняя валидность оказывается недостижимой, что, в свою очередь, приводит к нетехнологичности процесса валидизации. Отсутствие внутренней валидности отражается на тестовых результатах так, что невозможно выявить каузальность связи между независимой и зависимой переменной, причиной и эффектом. В конечном итоге, нельзя установить, что именно измеряет тест. Возникает задача обеспечения внутренней валидности не через внешний элемент эксперимента — его организацию и проведение, а через такой внутренний элемент, который поддается контролю, т. е. через инструмент — тест.

Современное состояние тестологии, в том числе теория IRT, по мнению А.И. Самыловского, позволяет создавать и широко использовать достаточно надежные тесты, притом, что вопрос об их содержательном наполнении остается открытым [131]. Проблема содержательного наполнения дидактическою теста, а следовательно, и проблема обеспечения его содержательной валидности выходит на первый план в современной тестологии. Мы считаем, что содержательная валидность должна иметь внутренний характер для того, чтобы измерение наполнилось смыслом, стало возможным, так как именно содержание теста связывает независимую переменную конструкта и зависимую переменную, отраженную в результатах тестирования. Но латентность объекта исследования и структурный характер теоретического конструкта позволяют обеспечить лишь внешний аспект содержательной валидности и не обеспечивают внутреннюю содержательную.

В образовании, а именно в его части контроля над результатами обучения в настоящее время в качестве объекта исследования рассматривается уровень подготовленности или уровень обученности учащегося по предмету или его разделу. Считают, что он служит личностно-психологической характеристикой учащегося как субъекта обучения и «является, к сожалению, латентным параметром (т. е. недоступным для непосредственного измерения)» [106, с. 6].

Заданный объект исследования подразумевает наличие иерархии соподчиненности уровней обученности, а значит, теоретический конструкт данного объекта будет необходимо иметь структурный характер. Ю.М. Нейман и ВА. Хлебников под обобщенным термином «уровень подготовленности» понимают «уровень обученности испытуемых по указанным разделам, совокупность их умений и соответствующих навыков» [Там же].

Близкое по смыслу определение конструкта и конструктной валидности предлагает А.С. Казаринов: «конструктами, или латентными переменными, называют недоступные для непосредственного наблюдения комплексные производные признаки. К таким переменным... относится «истинная» обученность: валидность конструкта оценивает, насколько данные контрольные материалы соответствуют определению требуемого уровня обученности» [46, с. 60].

Е.А. Михайлычев определяет ее следующим образом: «конструктная валидность дидактического теста — это отражение в нем теоретической модели курса, его структурных пропорций и основных компонентов на всех уровнях — вплоть до учебных элементов» [92, с. 247].

Из приведенных определений следует, что:

1) объект исследования носит латентный характер;

2) в качестве конструкта (предмета измерения) выбирается структурная модель предметного содержания;

3) валидность конструкта связывает структуру предметного содержания с содержанием самого теста.

Иначе говоря, теоретический конструкт, описывающий латентные свойства объекта измерения, является структурной моделью предметного содержания. Другими словами, в качестве теоретического предмета измерения выступает структура содержания по определенному разделу. Это значит, что при измерении тестом уровня подготовленности учащегося проверяется знание определенной структуры предмета.

Знание элементов структуры выражается в умении применять их на практике, демонстрации навыков владения соответствующими элементами содержания. Структурная организация конструкта подразумевает проверку всех навыков, связанных с каждым элементом структуры. В связи с этим возникает ряд проблем. Во-первых, знание всех элементов структуры не есть знание предмета (согласно принципу системности сумма частей не равна целому). Во-вторых, ограниченность теста не позволяет проверить знание всех компонентов на всех уровнях — вплоть до учебных элементов [92]. В-третьих, повышение уровня подготовленности происходит за счет приращения элемента структуры и доведения умений владения данным элементом до навыков, а затем демонстрация навыков при контроле уровня обученности. Все это провоцирует репродуктивный способ обучения, развитие экстенсивной системы образования и, как следствие, возникновение перегрузок в подготовке учащихся к занятиям.

Относительное решение второй проблемы возможно при сохранении структурного характера конструкта. Здесь можно ограничиться ранжированием учащихся по степени владения предметным содержанием, предъявляемым тестом. «При этом не ставится задача оценить уровень подготовленности каждого участника тестирования персонально, надо лишь сравнить их относительно друг друга или относительно подготовленности некоторой специально подобранной группы лиц, принятой в качестве нормативной... выборки» [106, с. 37]. В этом случае можно говорить о сопоставительном уровне подготовленности учащегося относительно всех тестирующихся. Тестирование не определяет «истинного» значения обученности, а валидность не является внутренней.

Другое относительное решение проблемы состоит в том, чтобы разбить конструкт на элементы, поддающиеся контролю, и непрерывно наблюдать за каждым элементом конструкта в ходе мониторинга уровня подготовленности. Здесь становится возможным измерение уровня подготовленности как латентной характеристики. Содержательная валидность всей тестовой батареи может стать внутренней, но модель обучения останется репродуктивной. ЕА Михайлычев отмечает, что «в системе многоэтапного контроля знаний дидактическое тестирование (как компьютерное, так и традиционное) является только одним из многих средств контроля из-за ограниченных возможностей современных дидактических тестов — выявления в основном репродуктивного уровня знаний» [92, с. 7]. Оба представленных варианта решения проблемы измерения уровня обученности на основе структурного конструкта используются в тестологической практике. Однако латентность измеряемого объекта предопределяет сомнительный, либо вероятностный характер валидности вывода.

Определение степени валидности конструкта осуществляется в количественном и качественном подходах. Качественный подход реализуется экспертными методами.

При использовании количественного подхода вычисляют коэффициент корреляции результатов, полученных при использовании валидизируемого теста, и результатов других измерений, выступающих в роли внешнего критерия. Подтверждением валидности теста является получение значимого коэффициента корреляции, который принимается как коэффициент конструктной валидности тес-

та. Он подтверждает связь между зависимой переменной (результатом) и конструктом, опосредуя ее через другое валидное измерение. Значимый коэффициент является эмпирическим подтверждением гипотезы исследования, выступая в роли коэффициента прогностической валидности. «Если эмпирические данные подтверждают гипотезу, то тем самым подтверждается концепция, положенная в основу теста, и способность теста служить инструментом измерения данного конструкта» [80, с. 216]. Заметим, что здесь в качестве одного числового ряда выбираются результаты валидного измерения, а в качестве другого — результаты, полученные при тестировании валидизируемым тестом и зависящие от содержания дидактического теста. Содержательная валидность теста связывает независимые переменные конструкта с содержанием теста. Тест, облагающий конструктной валидностью, должен, прежде всего, иметь содержательную валидность.

Принято говорить о содержательной валидности теста при проверке адекватности содержания теста измеряемому признаку. «Валидность содержания является достаточно проблематичным критерием. Для того или иного конкретного поведения существует содержательное описание. Разрабатываются определенные задания, выявляющие признаки этого поведения. С помощью оценок экспертов устанавливается, в какой мере задания и обусловленное ими поведение содержательно совпадают с предварительным описанием поведения» [44, с. 40]. Содержание теста должно отражать некоторые ранее описанные признаки владения измеряемым содержанием — независимые переменные, а признаки должны соответствовать некоторой теоретической модели измеряемого содержания — конструкту. Здесь представляется структурная модель предметного содержания, в которой последовательно отражены элементы структуры измеряемого содержания как в учебном плане или куррикуле. Содержание теста необходимо должно отражать все элементы его структуры.

Признаком модели является отображение элемента структур в определенном тестовом задании. Основным условием валидности содержания теста оказывается степень репрезентативности структуры измеряемого содержания. «При разработке педагогических тестов, конечно, на первом плане находится содержательная валидность, которая определяется как характеристика репрезентативности содержания теста по отношению к запланированным для проверки знаниям и умениям» [166, с. 319]. Сходные определения содержательной валидности приводятся в ряде работ [19, с. 10; 47, с. 41; 106, с. 61; 127, с. 46]. Репрезентативное представление измеряемого предметного содержания в тесте актуально только для структурной содержательной модели, в которой на первый план выходит поэлементная представленность структуры в тесте, заданная в соответствующих пропорциях [106].

В других содержательных моделях условие репрезентативности не является основным, соответственно, указанные определения описывают частный случай содержательной валидности для структурной модели измеряемого предметного содержания (структурного теоретического конструкта). На наш взгляд, удобно именовать такую валидность куррикулярной. «...Куррикулярная валидность является особым случаем валидности содержания», — указывает К. Ингенкамп [44, с. 40] (см. табл. 1).

Куррикулярной является валидность содержания в адаптивных тестах как тестах с переменным содержанием [163]. Целью адаптивного тестирования является определение с заданной точностью характеристик латентной переменной каждого тестируемого посредством итерации через последовательное предъявление тестовых заданий с заранее известными статистическими характеристиками. Адаптивные тесты обладают высокой, заранее определенной точностью измерения, а следовательно, и высокой надежностью. Однако повышение надежности приводит к снижению его валидности по содержанию, при повышении однородности содержания (парадокс Лорда). Высокая надежность оказывается основным условием, позволяющим экстраполировать результаты тестирования, полученные с помощью ограниченного круга тестовых заданий с заданными статистическими характеристиками, на всю генеральную совокупность измеряемого предметного содержания

данного уровня [181]. Из этого следует, что куррикулярная валидность в адаптивных тестах, разработанных на основе статистического анализа, носит внешний характер (см. табл. 1).

В работе Е.Н. Лебедевой наиболее весомым видом валидности педагогического теста является валидность целеполагания: «Валидность целеполагания — содержательная валидность. Определяет степень соответствия содержания ПКМ (педагогических контрольных материалов) цели тестирования» [68, с. 72]. Различные цели тестирования (текущего или итогового контроля) подразумевают различия в контролируемом содержании, что необходимо учитывается при разработке валидного теста как теста, достигающего поставленной цели тестирования. Различие в контролируемом содержании описано в спецификации теста и принимается во внимание при разработке тестовых заданий.

Анализ теста на установление валидности целеполагания проводится методом групповых экспертных оценок. Экспертам предлагается ответить на вопрос: «Насколько каждое *i*-е задание соответствует сформированной цели диагностики?» [Там же, с. 73]. Однако здесь не уточняется, с чем эксперт должен сравнивать содержание каждого тестового задания. «Содержание тестового задания» и «цель тестирования» — категории несоизмеримые и сравнению не подлежат. В работе Е.Н. Лебедевой приводится «Сертификат дидактического теста для проведения итогового контроля обученности студентов экономических специальностей по курсу информатики». В пункте 3 сертификата «Названия и коды диагностируемых модулей» перечисляются элементы структуры содержания курса информатики, диагностируемые данным тестом. Можно предположить, что экспертами проводится сравнение содержания тестовых заданий с элементами структуры диагностируемого предметного содержания. В этом случае валидность целеполагания оказывается внешней куррикулярной валидностью (см. табл. 1).

Таким образом, понятие содержательной валидности по смыслу шире, чем понятие куррикулярной валидности. Первая отражает соответствие содержания теста признакам предварительно описанного поведения [44], но в практике педагогического тестирования не рассматривается. Термин «куррикулярная валидность» более точен в использовании педагогических тестов, основанных на структурной модели измеряемого предметного содержания, и задан посредством репрезентативного отображения элементов структуры в содержании тестовых заданий. Итак, куррикулярная валидность совпадает с валидностью целеполагания, адаптационной валидностью или содержательной валидностью адаптивных тестов в зависимости от цели тестирования и вида теста.

Для педагогических тестов, построенных на основе структурного конструкта, характерна репрезентативная модель предметного содержания, поэлементно отражающая структуру предмета. Содержание теста «определяется как отображение фрагмента учебной дисциплины в тестовой форме» [6, с. 140]. Структурному конструкту соответствует куррикулярная валидность. Репрезентативная модель содержания теста не может взаимно-однозначно отображать структурный конструкт ввиду ограниченности теста. «Чтобы оценить уровень подготовленности человека в определенной области знания, необходимо проверить правильность выполнения этим человеком, вообще говоря, бесконечно большого количества специально составленных заданий различной трудности» [106, с. 60]. Связь между конструктом и независимыми переменными, представленными в содержании теста, оказывается односторонней. Это значит, что куррикулярная валидность не может быть внутренней.

Однако отсутствие внутренней валидности теста делает измерение относительным. Примерами относительного тестирования являются нормативно-ориентированные, критериально-ориентированные и адаптивные тесты. Нормативно-ориентированные тесты признаются относительным инструментом потому, что здесь происходит сравнение тестирующихся относительно друг друга, а интерпретация тестовых результатов зависит от полноты или репрезентативности выборки стандартизации. Критериально-ориентированные и адаптивные тесты являются относительными из-за неполноты предъявляемого ими тестового содержания. Кроме того, качество интерпретации результатов тестирования зависит от степени репрезентативности измеряемой предметной области в содержании

теста. Как мы полагаем, имеющиеся на сегодняшний день дидактические тесты не обладают внутренней валидностью и являются инструментом относительного тестирования.

Известно, что валидность теста — характеристика целезависимая. Цели тестирования и выбор критерия валидизации подразделяют педагогические тесты на критериально-ориентированные и нормативно-ориентированные.

Целью нормативно-ориентированного тестирования является дифференциация тестируемых по степени успешности владения предъявляемым содержанием теста. Основной характеристикой нормативно-ориентированного теста является его различающая способность в целом и каждого тестового задания в отдельности. К отбору тестового содержания предъявляются три основных условия, Во-первых, тестовые задания не должны выходить за пределы содержания измеряемой предметной области. Во-вторых, тестовые задания должны репрезентативно представлять структуру измеряемого содержания (куррикул). В-третьих, тестовые задания должны обладать хорошей различительной способностью.

Первые два условия являются необходимыми для отбора содержания тестовых заданий. Они обеспечиваются методом экспертных оценок на этапе проектирования теста с привлечением опытных преподавателей и методистов по данному предмету в качестве экспертов. Измеряемое содержание представляется в виде структурной технологической матрицы. Эксперты проводят отбор тестовых заданий, соответствующих данной матрице, и принимают участие в дальнейшей разработке теста. Третье условие является основным и достаточным для качественного отбора содержания нормативно-ориентированных тестов. Оно обеспечивается после проведения пилотажного тестирования и основывается на статистических характеристиках каждого тестового задания и теста в целом: сложности, трудности и различительной способности (дискриминативности). Все статистические характеристики тестовых заданий непосредственно зависят от свойств выборки стандартизации, но не от содержания тестового задания и его отношения к теоретическому конструкту. Так, М.Б. Челышкова утверждает, что «полагаться следует исключительно на эмпирические оценки, а не на свое видение трудности заданий теста» [166, с. 281]. В тест включаются задания, прошедшие экспертизу и обладающие хорошими статистическими характеристиками. Остальные задания из него удаляются. В результате получается тест, обладающий высокой различительной способностью, позволяющий достичь ранжирования тестирующихся как цели нормативно-ориентированного тестирования. Более того, «показатель дискриминативности обычно отождествляют с характеристикой валидности задания теста, т. е. с пригодностью задания служить поставленной цели измерения» [Там же, с. 289], Тест, составленный из «валидных» тестовых заданий, таким образом, не обязательно обладает содержательной валидностью.

Однако исследование валидности готового теста по содержанию не проводится. Как отмечает А.Н, Майоров: «сама процедура создания тестов школьных достижений в том случае, если она не нарушается, дает хороший валидный инструмент, поскольку все этапы определения содержательной валидности «зашиты» в процедуру» [80, с. 215]. Но в данном случае в процедуру построения теста встроена куррикулярная валидность как репрезентативное отображение измеряемого содержания (конструкта) в содержании теста. Для обеспечения внутренней валидности должна рассматриваться содержательная валидность как соответствие содержания теста теоретической модели содержания (репрезентативной модели, основанной на структурном подходе). Тем не менее такой анализ содержания теста отсутствует. «Имеется, однако, очевидная опасность признать валидными по содержанию тесты, которые все как действительно валидны, так и невалидны... Такой фактор при нормативно-ориентированном тестировании не контролируется» [131, с. 14]. Не доказанным остается положение о том, что тест обладает внутренней валидностью., Вследствие этого, валидность готового теста по содержанию удовлетворяет только одному условию — не выходит за пределы куррикула. Валидность теста по содержанию оказывается внешней, так как содержание теста не описывает всю структуру измеряемого предметного содержания, т.е. является не внутренней содержательной, но

внешней куррикулярной валидностью. Однако отсутствие внутренней валидности теста вовсе не означает, что он инвалиден. Даже при этих условиях нормативно-ориентированный тест способен достигать цели тестирования, а значит, быть валидным в общем смысле. Это значит, что основное значение обретает не столько содержание тестового задания, сколько способ конструирования теста, его структуры и формы.

Традиционно полагают, что основной задачей построения нормативно-ориентированных тестов является создание теста с правильным профилем [80; 106; 166] или теста с возрастающей сложностью [5], где на первое место выходит валидность техническая (ее композиционная часть), т. е. соответствие формы теста правильному профилю. Реально техническая валидность в нормативно-ориентированных тестах не исследуется, а соответствие теста заданной форме достигается за счет технологии его создания, основанной на статистических методах. Применение математического аппарата и теории IRT позволяет создавать тест с заранее заданными характеристиками формы. Нормативно-ориентированные тесты являются тестами-формой, а их содержание служит лишь наполнением формы и не играет существенной роли. Валидность теста заменяется его надежностью. Здесь валидность нормативно-ориентированного теста подтверждает достижение цели тестирования как ранжирования тестирующихся по степени владения предъявляемым тестовым содержанием. Но она не показывает способность теста измерять латентный уровень подготовленности учащегося по предмету.

При создании нормативно-ориентированных тестов вычисляется коэффициент критериальной валидности как коэффициент корреляции результатов тестирования с результатами применения других оценочных средств. Данный вид валидности отражает соответствие интерпретации результатов тестирования другим методам оценки и призван обеспечивать адекватный перевод тестовых баллов в принятую в средней школе пятибалльную шкалу. В педагогической практике, в частности, в ходе централизованного тестирования, перевод тестовых баллов в пятибалльную шкалу осуществляется декларативно. Например, в централизованном тестировании принято оценку «5» выставлять 5—6 % лучшим результатам из всех тестирующихся, а оценку «2» — 24—26 % худшим. Это значит, что, во-первых, оценка не показывает действительного уровня владения тестирующихся определенным содержанием, во-вторых, не предъявляет минимальный уровень владения стандартом по предмету, в-третьих, один и тот же тестовый балл ежегодно имеет различное содержательное наполнение.

Итак, в нормативно-ориентированном тестировании содержательная валидность стоит на третьем месте после технической валидности и валидности интерпретации, зависит от статистических характеристик и выступает в роли внешней валидности теста. Тем не менее тест остается валидным, так как достигается основная цель нормативно-ориентированного тестирования — ранжирование тестируемых. Следовательно, с помощью нормативно-ориентированных тестов не возможно измерять обученность. Они могут служить только для сравнения или оценивания положения тестируемых относительно друг друга.

Принято считать, что целью критериально-ориентированных тестов является определение степени владения тестируемым объемом измеряемого содержания. На наш взгляд, это не совсем так. Проследим этапы построения критериально-ориентированных тестов.

Построение критериально-ориентированных тестов проводится по аналогии с построением нормативно-ориентированных тестов. Отличие заключается в интерпретации результатов тестирования и смещении акцентов исследования со структуры теста при нормативно-ориентированном тестировании на содержание теста при критериально-ориентированном тестировании. Основой построения критериально-ориентированных тестов является определение «абсолютного критерия», выделяемого в измеряемом содержании экспертными методами. Он должен обеспечиваться содержанием теста через репрезентативную модель измеряемого содержания предметной области.

В критериально-ориентированных тестах так же, как и в нормативно-ориентированных не проводится исследование содержания теста на предмет адекватного отражения им репрезентативной

модели измеряемого содержания предметной области. Попытки построить валидное отражение структурной модели содержания приводят к неограниченному удлинению теста. Никакой тест, с его ограниченностью, не может репрезентативно отражать измеряемое содержание и носит лишь вероятностный характер. Эксперты, исследуя куррикулярную валидность теста, могут говорить только о мере вероятной представленности в тесте измеряемого содержания. Можно констатировать, что содержательная валидность теста отсутствует как внутренняя часть валидности критериально-ориентированных тестов, а куррикулярная валидность несет в себе, исключительно, характеристики внешней валидности теста.

Следующим по значимости видом валидности для критериально-ориентированных тестов является критериальная валидность, представленная как мера корреляции тестовых результатов пилотажного тестирования критериально-ориентированного теста с результатами, полученными с помощью других форм контроля той же самой содержательной области. Выбор внешнего источника для проведения корреляции определяет критериальную валидность как часть внешней валидности.

Для критериально-ориентированных тестов не имеет большого значения исследование технической валидности теста как соответствия структуры теста правильному профилю, т. е. композиционной валидности. Техническая валидность проявляет себя как технологическая валидность, т. е. как соответствие тестовых заданий принятой в тестологии форме. По определению, валидность теста по форме не является внутренней частью валидности теста, хотя и не зависит от статистических характеристик.

Отсутствие внутренней части в критериально-ориентированных тестах объединяет их с нормативно-ориентированными тестами. Измерение объема предметного содержания становится невозможным. Цель тестирования может быть достигнута только при взаимнооднозначном представлении в содержании теста репрезентативной модели содержания, что не представляется возможным. Критериально-ориентированные тесты оказываются примером тестов, относительных по содержанию.

Адаптивные тесты как тесты с переменным содержанием необходимо представить, вместе с критериально-ориентированными тестами, как тесты, относительные по содержанию. Целью адаптивного тестирования является установление индивидуального порога, отделяющего знание и незнание измеряемого содержания для каждого тестируемого в отдельности. Здесь выстраивается индивидуальная, зависящая от предшествующих звеньев, цепочка тестовых заданий. Она описывает определенное содержание, которое служит подмножеством измеряемого предметного содержания. Его модель выстраивается на основе структурного подхода. Содержательная валидность, по-прежнему, остается внешней по отношению к измеряемому содержанию.

Как мы полагаем, невозможность измерения результата обучения по предметному содержанию, которым владеет тестируемый, объясняется выбором теоретической модели содержания предмета — репрезентативной модели измеряемого содержания, основанной на структурном подходе к описанию содержания учебного предмета. В рамках данной модели содержания необходимый объем знаний достигается за счет введения дополнительных содержательных блоков, т. е. за счет роста структуры содержания. Единственным ограничением увеличения содержания является фактор учебного времени, фиксированный в стандарте. Ограничение времени для усвоения все большего содержания приводит к увеличению нагрузки учащихся, возникновению хронических перегрузок и, как следствие, ухудшению здоровья. По-видимому, следует отказаться от структурного подхода в конструировании предметного содержания, который оказывается не достаточно эффективным, так как опирается на экстенсивный подход к организации обучения.

Экстенсивность обучения и возникающие в связи с ней учебные перегрузки, относительность имеющихся измерителей не замыкают круг проблем, порожденных структурным конструктом контролируемого предметного содержания. Отсутствие внутренней валидности теста не позволяет достоверно контролировать результаты обучения. Не решенной остается наиболее значимая проблема — сохранение единого образовательного пространства.

Единство образовательного пространства в СССР соблюдалось при условии унификации процесса обучения, достигаемого через единство программ, учебников и методических пособий, рекомендованных МО СССР. В настоящее время многообразие учебников, программ и методик преподавания, применяющихся в процессе обучения, порождают множество разнообразных результатов обучения, которые невозможно привести к общим критериям через внешний декларативный контроль над процессом обучения. Возникает актуальная необходимость создания единых учебных стандартов школ России и проведения единого государственного экзамена, контролирующего уровень подготовленности учащихся по тому или иному предмету.

Инструмент ЕГЭ — контрольно-измерительные материалы (КИМ) — конструируются на основе теории IRT и призваны оценивать уровень подготовленности учащихся по предмету [105]. Уровень подготовленности как объект контроля имеет латентный характер, а его конструкт оказывается структурным. Куррикулярная валидность теста, как мы доказали, носит внешний характер, соответственно, КИМ является измерителем, относительным по содержанию. Такой инструмент ЕГЭ не обеспечивает объективное измерение в аспекте адекватности результатов тестирования конструкту. Здесь в качестве конструкта для обеспечения единства образовательного пространства выступает учебный стандарт.

Возникает необходимость в ином способе конструирования содержания предмета и определения эффективных характеристик обучения. Нужно определить в рамках иного подхода к содержанию и другие признаки, отделяющие знание предмета от незнания и описанных в куррикуле (программе, стандарте), а не только умение решать определенный тип задач, как это принято в современной организации контроля результатов обучения.

Следует отметить, что валидность педагогических тестов, как адекватность ранжирования тестируемых по степени владения содержанием, может быть достигнута при построении тестов в теории IRT только тогда, когда соблюдается принцип единого образо-

рательного пространства, т. е. когда логика построения содержания предмета у разных авторов различается незначительно.

Мы имеем в виду границы применения теории IRT в том варианте, в котором она практически сложилась на данный момент. Как показано в работе И.И. Нурминского, Н.К Гладышевой «Статистические закономерности формирования знаний и умений» [109], статистическая сложность учебного задания значительно варьируется в зависимости от учебника, по которому велось преподавание. Это значит, что успех или неуспех результата обучения, т. е. обученность не корректно ставить в зависимость от свойств выборки нормирования теста. Тем более следует учитывать, что в современной практике преподавания существует многообразие учебников с различной логикой построения содержания, рекомендованных Министерством образования.

Необходимо возникает проблема выделения наиболее общих, не зависящих от нормативной выборки характеристик логики владения тем или иным содержанием. Ее решение равносильно реализации внутренней валидности теста, дополненной способом построения такой модели учебного содержания, которая бы не зависела от статистических характеристик и была бы отличной от модели репрезентативного содержания.

Содержательная валидность, будучи внутренней валидностью, должна обеспечивать соответствие содержания теста некоторой теоретической модели измеряемого содержания учебной дисциплины способом, отличным от используемого в структурном подходе.

В проекте учебных стандартов школ России [157] за 1998 г. впервые появился раздел «Содержательные линии курса математики», в котором делается акцент на пять стержневых линий всего школьного курса математики: «Числа и вычисления», «Выражения и их преобразования», «Уравнения», «Функции» и «Геометрические фигуры. Измерение геометрических величин». Они соответствуют традициям преподавания математики в нашей стране и практически полностью совпадают с мировой практикой. Далее, в этом разделе, в зависимости от ступени обучения раскрывается содержание каждой линии. В стандарте обращают внимание на то, что в современном преподава-

нии математики «центральными с точки зрения общего образования являются числовая, функциональная и геометрическая линии, концентрирующие в себе математические знания, которыми должен обладать каждый человек в современном обществе. ... Необходимость овладения всеми школьниками содержанием двух остальных линий определяется самой природой математической науки: оно ориентировано на формирование математического аппарата, без которого не возможно ни рассмотрение внутриматематических проблем, ни решение задач практического и прикладного характера» [Там же, с. 27]. Дальнейшее изложение посвящено раскрытию содержания этих линий и структуре изложения на различных ступенях обучения.

На этой основе, анализируя структуру учебников, мы выделяем два фундаментальных сложившихся подхода к логике изложения содержания школьного курса математики (алгебры и начал анализа). В основе первого подхода лежит число. Логика изложения содержания курса математики здесь базируется на свойствах чисел, дискретных числовых множеств и операциях над ними. По нашему мнению, имеет смысл обозначить его как числовой подход к изложению содержания. При переходе от дискретных множеств к непрерывным множествам эмпирическим путем возникает понятие функции, затем рассматриваются свойства функции и операции над ними. Далее изложение содержания предъявляется в рамках функционального подхода. При этом полагают, что любое число становится частным случаем функции, как значение некоторой функции при некотором значении аргумента. Возникает возможность переосмысления ранее изученных понятий с точки зрения другого, более общего, а именно функционального подхода к содержанию.

Теоретически возможной может быть и логика разворачивания содержания при движении от функционального к числовому подходу. На практике она не обнаружена нами в общих учебниках как путь раскрытия школьного курса математики.

Обученность как способность увидеть и (или) выделить в незнакомой математической структуре задания знакомое число или знакомую функцию есть признак теоретического мышления, признак владения системой содержания предмета в той или иной традиции

(теории). Таким образом, обученность открыта как объект контроля за результатами обучения, определена через теоретические подходы к предметному содержанию и, следовательно, не латентна.

Минимальное владение содержанием предмета, т. е. необходимая обученность определяется как владение хотя бы одним теоретическим подходом: числовым или функциональным. Поскольку две возможные логики изложения содержания в учебниках оказываются не равнозначными в практике преподавания, постольку минимальное владение содержанием, по нашему мнению, можно определить как владение числовым подходом. Достаточным условием обученности оказывается владение предметным содержанием с точки зрения функционального подхода.

Определенный и открытый объект контроля служит предпосылкой для создания целостного конструкта предметного содержания. Содержание курса математики как целостный теоретический конструкт (или предмет контроля) в аспекте выделенных подходов может быть представлено тремя основными типами заданий:

1) решаемые только в рамках числового подхода;
2) решаемые либо числовым, либо функциональным подходом;
3) решаемые только в рамках функционального подхода.

Задания первого и третьего типов, провоцирующие применение определенного подхода, можно назвать моделирующими заданиями. Если они включены в тест, то смогут выступать в роли определяющих характеристик (признаков) владения подходом или мерой владения логикой содержания предмета — теоретическим знанием. Задания второго типа будут выполнять роль показателей уровня технической сложности усвоенного содержания, или прикладным знанием.

Необходимая обученность, следовательно, конструируется через моделирующие задания числового подхода, а максимальная —через моделирующие задания функционального подхода. Предметом контроля является целостный конструкт предметного содержания, или целостная модель предметного содержания. Она, таким образом, описывает необходимую и достаточную границы обученности как объекта исследования.

Итак, содержание школьного курса математики, на наш взгляд, можно представить, по крайней мере, системой двух основных подходов к изложению содержания. Системный подход к построению содержания предмета позволяет построить модель его содержания, отличную от репрезентативной, что является необходимым основанием для присутствия внутренней валидности теста.

Таким образом, содержательная валидность теста обеспечивает соответствие содержания теста теоретической модели измеряемого содержания учебного предмета, раздела и т. п. Исследование педагогического теста на содержательную валидность определяется выбором модели содержания учебного предмета как теоретического конструкта, существование которого является необходимым условием наличия внутренней валидности теста.

Можно выделить две модели организации содержания предметной области, различающиеся выбором подхода к описанию содержания учебного предмета и подвергающейся измерению. В первой модели за основу конструирования выбирается структурный подход, а во второй — целостный.

В настоящее время в определении предметного содержания преобладает структурный подход, отличительными особенностями которого являются описательный характер знаний, умений и навыков, определяющих объем владения содержанием; определенная последовательность изложения элементов структуры содержания; жесткая регламентация по времени изучения разделов. В структурном подходе моделью содержания измеряемой предметной области является репрезентативная модель содержания, поэлементно отражающая структуру предмета. Основной ее чертой является полнота отображения объема измеряемого содержания предметной области, которая достигается посредством репрезентативного представления элементов структуры. Увеличение объема измеряемого содержания возможно лишь путем прироста элементов структуры. Модель содержания оказывается экстенсивной. Для нее невозможно обеспечить внутренний характер валидности педагогического теста.

Построение теста на основе целостного конструкта определяет внутренний характер содержательной валидности теста. Внутренняя

валидность устанавливает соответствие результатов тестирования теоретическому конструкту. Это условие позволяет получать объективные тестовые результаты в аспекте их адекватности целостному конструкту. Четыре основных вида валидности образуют систему, необходимую и достаточную для установления внутренней валидности педагогического теста, контролирующего обученность. Переход от структурного теоретического конструкта к целостному позволяет организовать технологичный процесс валидизации теста.

При использовании тестов с целостным конструктом предметного содержания для целей контроля результатов обучения возможным становится определение знания предмета как целого, а не только определение знания предмета на уровне владения навыками по отдельным его частям. На его основе можно создавать целостные содержательные модели, обладающие внутренней валидностью.

Если в качестве целостного теоретического конструкта выступает предметный стандарт, то при использовании такого теста для контроля над результатами обучения измеряется обученность стандарту. Решается проблема сохранения единого образовательного пространства.

2.2. Критерий валидизации целостного конструкта обученности

Обоснование выбора критерия для определения эмпирической валидности авторы работ по педагогической тестологии ставят в зависимость от цели исследования и выбранных методов валидизации (Е.Н. Лебедева [68], А.Н. Майоров [80], Е.А. Михайлычев [92], М.Б. Челышкова [166] и др.). Проблема соответствия выбранного критерия постулируемому критерию истинности, положенного в основу валидизации, не рассматривается. Однако от этого зависит качество результатов педагогического контроля, в частности, качество их интерпретации. Проблема обеспечения качества педагогического контроля над результатами обучения освящена в трудах В.П. Беспалько [22], А.А. Вербицкого [25], В.А. Кальней, С.Е. Шишова [49], В.В. Краевского [62], М.М. Поташника [118],

М.Н. Скаткина, В.В. Краевского [142], Н.Ф. Талызиной [150], С.Е. Шишова, В.А. Кальней [178] и др.

Выбор критерия для валидизации теста во многом определяемся также характером исследуемого объекта. В нашем случае объектом исследования оказывается обученность. Обученность в отличие от уровня подготовленности является определенным, открытым объектом с заданными характеристиками предметного содержания, которое всегда ограничено теоретическими подходами. Все измерительные процедуры будут производиться с открытым, определенным объектом.

Если целью исследования является измерение уровня подготовленности учащегося, то объект исследования латентен. Теоретический конструкт представляется в виде структурной модели предметного содержания. Содержание теста не может быть моделью конструкта ввиду ограниченности теста. Из этого следует, что содержательная валидность отсутствует как адекватное соответствие содержания теста теоретическому конструкту, а имеющаяся куррикулярная валидность носит внешний характер. Возникает проблема обеспечения внутренней валидности как установления каузальных связей. Возможное ее решение заключается в выборе значимого внешнего критерия для валидизации и вероятностного определения коэффициента эмпирической валидности.

Если целью исследования становится измерение обученности, то меняется объект исследования. Он перестает быть латентным. Покажем, что объект исследования определен, ограничен и открыт.

С точки зрения целостного подхода к определению конструкта, под обученностью будем понимать результат обучения, заключающийся в понимании системы предметного содержания как целого, представленного в разных аспектах или хотя бы в одном из них. Тогда объектом исследования будет результат обучения, предъявленный во всевозможных аспектах, а предметом оказывается целостный конструкт предметного содержания, или предметное содержание как целое.

Предметное содержание как целое определено куррикулом (программой, стандартом). Оно ограничено существующими подходами

и, следовательно, открыто, хотя и многовариантно, а не латентно. Если объект исследования перестает быть латентным, то исчезает необходимость привлекать внешний критерий для опосредованной валидизации теста. Критерий становится внутренним.

Внутренний критерий может быть как связан с содержанием теста, если валидизация основана на критерии когеренции, так и не связан с содержанием самого теста, если основанием для валидизации будет служить критерий функциональности. Для дидактического теста, как мы показали, существует только конвергентная валидизация, основанная на общем конструкте. Внутренний критерий основан на том же конструкте, что и сам тест. Общим конструктом оказывается предметное содержание как целое. Критерий, следовательно, так же, как и содержание, должен представлять собою конструкт, т. е. являться определенной содержательной моделью.

Теоретический конструкт, таким образом, определен, ограничен, инвариантен относительно структуры предмета и логики его изложения. Одновременно теоретический конструкт вариативен в аспекте наполнения модели конструкта предметным содержанием. Другими словами, существует неограниченное количество представлений теоретического конструкта в виде содержательной модели целостного предметного содержания. Каждая из таких моделей, в том числе и содержание самого теста может выбираться в качестве внутреннего критерия для валидизации.

Если критерием валидизации, основанной на принципе функциональности, выбирается внутренний критерий, не связанный с содержанием самого теста, то процедура валидизации будет заключаться в следующем.

1. Объект исследования представляется в виде теоретического конструкта предметного содержания. Для этого экспертная группа выделяет предметное содержание и основные подходы к описанию содержания предмета, в которых определяются границы применимости данных подходов к выделенному содержанию.

2. Экспертная группа делится на две части и каждая подгруппа на основе представленного конструкта строит целостную модель содержания, тем самым, определяет содержание теста.

3. Оба теста применяют на одной группе учащихся. Результаты сравниваются.

Высокое соответствие результатов обоих тестов будет говорить о надежности полученных результатов или повторяемости результатов. Следовательно, оба теста могут признаваться надежными. При этом оба они не будут параллельными в общепринятом смысле, а значит, не будут тестами-двойниками, написание которых всегда проблематично. Они будут параллельными в аспекте равнозначного отображения теоретического конструкта.

Высокое соответствие результатов обоих тестов будет говорить о валидности теста. Соответствие различных моделей одного и того же конструкта, представленными на основе экспертных оценок различными экспертами, подтверждает адекватность теоретического конструкта. Назовем такую процедуру перекрестной валидизацией. Затруднения, связанные с перекрестной валидизацией, заключаются в установлении соответствия шкал для результатов обоих тестирований. Оба теста не являются двойниками, следовательно, установление общей для них шкалы становится проблематичным.

Если в качестве критерия валидизации, основанной на принципе когеренции, выбирается внутренний критерий, связанный с содержанием теста как целого, то процедура валидизации будет заключаться в следующем.

1. Объект исследования задается как теоретический конструкт предметного содержания. Для этого экспертная группа выделяет предметное содержание и основные подходы к описанию содержания предмета, определяющие границы применимости данных подходов к выделенному содержанию.

2. Экспертная группа делится на две части. Одна подгруппа экспертов на основе составленного конструкта строит целостную модель содержания, тем самым определяет границы содержания теста. Другая подгруппа анализирует тестовое содержание и вырабатывает критерий валидизации.

3. Тест применяют на группе учащихся. Результаты сравниваются с критерием (рис. 4).

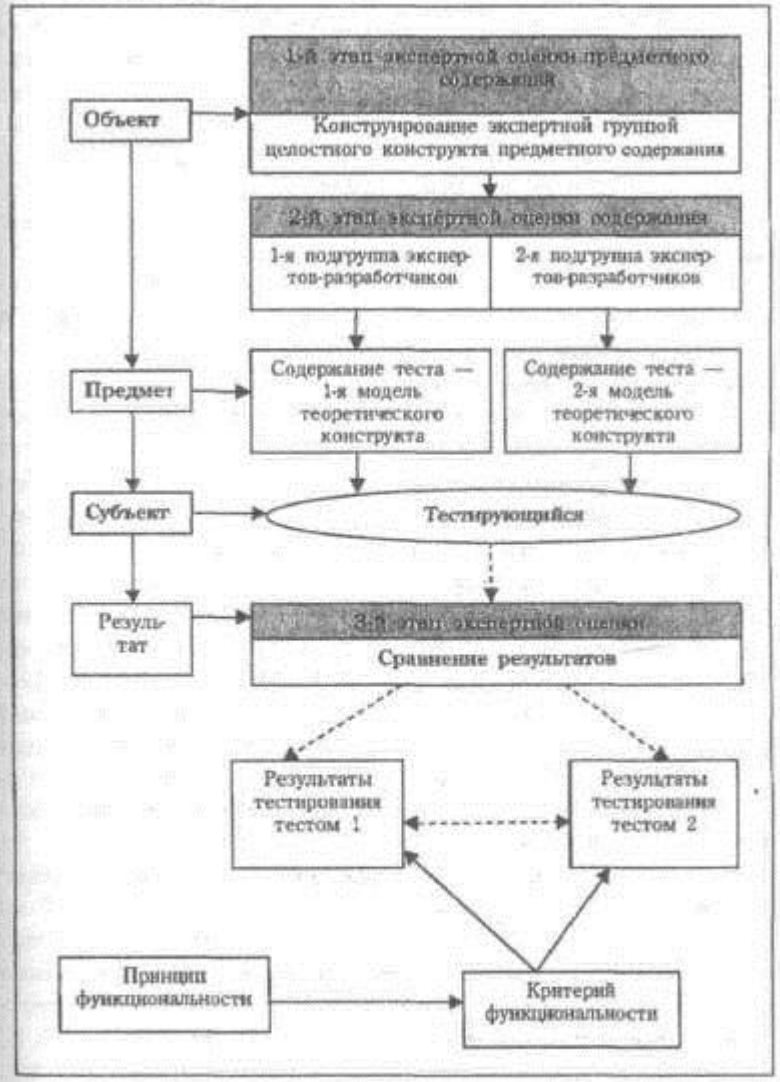


Рис. 4. Блок-схема алгоритма валидизации педагогического теста с целостным конструктом (на основе принципа функциональности)

При такой процедуре валидизации можно построить шкалу для одного теста. Однако невозможным оказывается сравнение результатов, полученных после применения тестов-двойников, без привлечения внешнего критерия.

Проследим оба приведенных процесса валидизации как два возможных решения проблемы установления шкал для тестов с целостным конструктом и внутренним критерием валидизации в аспекте обеспечения внутренней валидности теста.

Во-первых, выбор внутреннего критерия, не связанного с содержанием самого теста, допускает соответствующую валидизацию, основанную на критерии функциональности. Следовательно, к построению шкалы результатов тестирования применим нормальный закон распределения статистических величин. Однако эмпирическая (критериальная) валидность перестает быть внутренней характеристикой. Основная задача построенной таким образом шкалы — это экстраполяция результатов тестирования на большую, чем экспериментальная, выборку тестирующихся. Эмпирическая валидность становится внешней.

Для валидизации целостного конструкта мы выбрали внутренний критерий, опосредованно связанный с содержанием самого теста через общий конструкт (см. рис. 4). Содержание критерия и содержание валидизируемого теста есть параллельные модели одного и того же целостного конструкта. И, как следствие, содержательная валидность валидизируемого теста необходимо становится внешней. В результате, внутренняя валидность теста отсутствует точно так же, как и при валидизации тестов со структурным конструктом. Измерение теряет смысл.

Во-вторых, при выборе содержания теста в качестве внутреннего критерия возможна валидизация теста, основанная на принципе когеренции. Устанавливается непротиворечивость результатов тестирования и внутреннего критерия как целого. Результаты тестирования содержат зависимую переменную. Независимая переменная включена в содержание теста. Определить внутреннюю валидность — это установить соответствие между содержанием теста и результатами тестирования. Но между ними в процессе измерения находится субъект с его внутренней неопределенностью и латентностью. Вновь возникает

модель эксперимента по типу «черного ящика». Это значит, что для валидизации следует привлекать внешний критерий. Можно констатировать, что невозможно обеспечить внутреннюю валидность при валидизации, основанной на принципе когеренции (рис. 5).

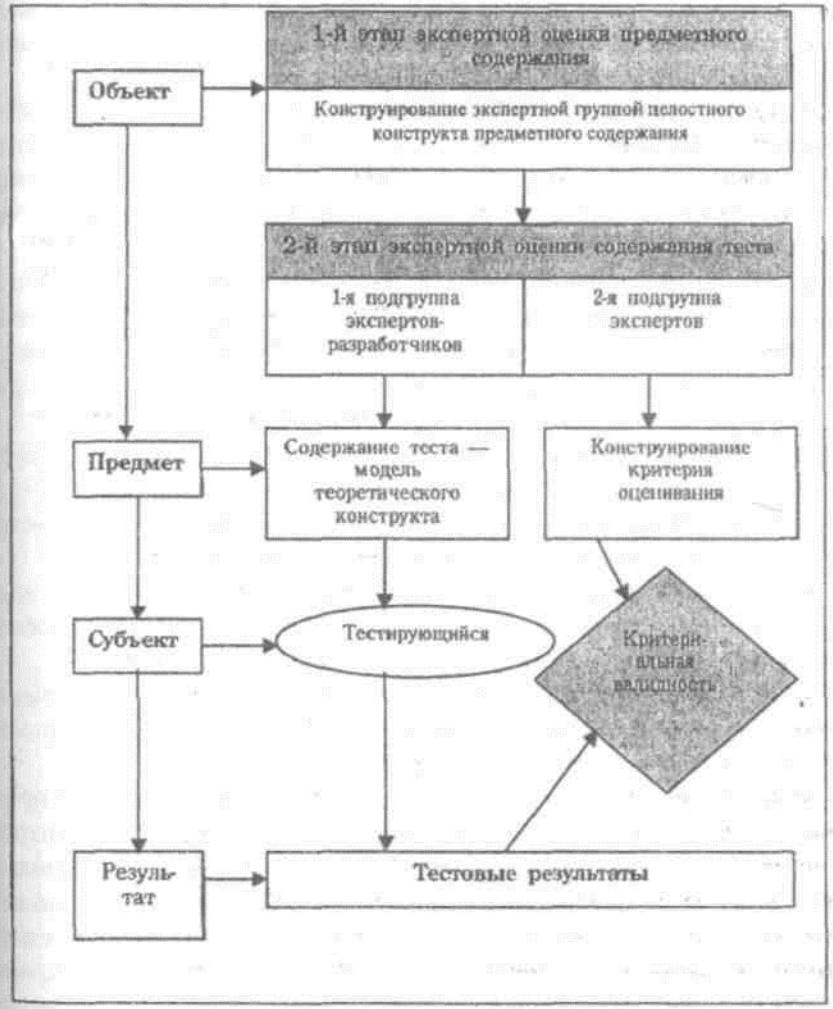


Рис. 5. Блок-схема алгоритма валидизации педагогического теста с целостным конструктом (на основе принципа когеренции)

На наш взгляд, возможен третий способ валидизации тестов, построенных на основе целостного конструкта, который, в отличие от первых двух, обеспечивает внутреннюю валидность. Он базируется на принципе межличностного соглашения, или консенсуса. Здесь внутренний критерий оказывается связанным с самим тестовым содержанием, т. е. критерием является целостная модель содержания. Смысл валидизации будет заключаться в следующем.

1. Объект исследования представляется в виде теоретического конструкта предметного содержания. Для этого экспертная группа выделяет предметное содержание и основные подходы к описанию содержания предмета, определяющие границы применимости данных подходов к выделенному содержанию.

2. Экспертная группа делится на две части. Одна подгруппа экспертов — подгруппа разработчиков — на основе составленного конструкта строит целостную модель содержания, ограничивая содержание теста. Она же определяет критерий обученности применительно к данному тестовому содержанию.

3. Другая подгруппа — подгруппа экспертов — анализирует тестовое содержание и строит на его основе свой критерий.

Группы объединяются и в ходе диалога приходят к единому критерию для данного теста.

Если в ходе диалога по определению критерия консенсус достигнут, то построенный тест считается валидным. Иначе необходимо вернуться к пункту 1.

Блок-схема алгоритма технологии валидизации педагогического теста с целостным конструктом на основе принципа межличностного соглашения представлена на рис. 6.

Здесь консенсус, ориентированный на достижение истинности, представляется как диалог экспертов и понимается как коммуникативный акт, к которому предъявляют следующие требования. «Истинными считаются утверждения, описывающие внешнее положение вещей. Утверждения истинны, если намерения, чувства, потребности не скрываются, а открыто выражаются в процессе коммуникации. Кроме этого, утверждения должны соответствовать нормам и ценностям общества. Имея такую информацию, слушатель принимает или не принимает утверждение» [66, с. 100].

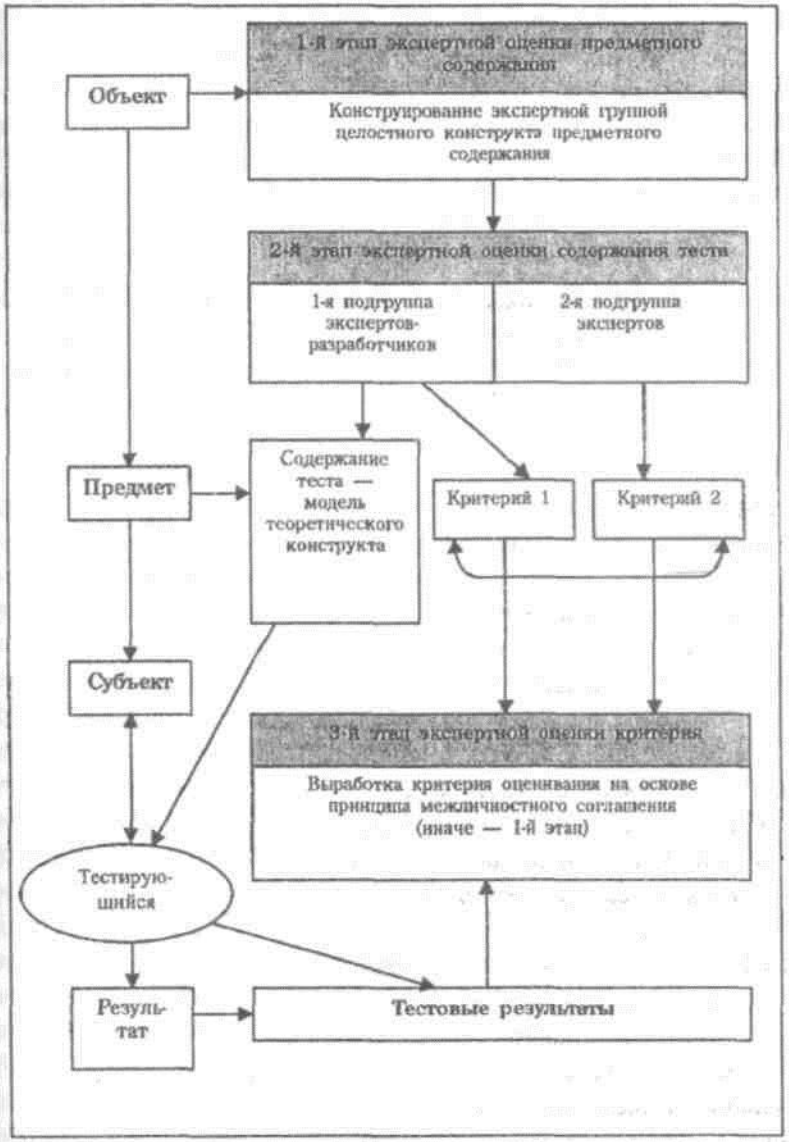


Рис. 6. Блок-схема алгоритма валидизации педагогического теста с целостным конструктом (на основе принципа межличностного соглашения)

Внутренний критерий, выработанный в результате межличностного соглашения экспертов, является основой валидизации теста. Если выработанный критерий определен, а его смысл вскрыт, то результатом работы экспертов окажется внутренний критерий как целостная модель содержания, представленная в виде консенсусной нормы. Поскольку сравнение с ней результатов тестирования носит внешний характер, постольку здесь происходит оценивание. Оценивание проясняет степень соответствия результатов тестирования выработанному критерию, или норме. Внутренний критерий, базирующийся на целостной модели предметного содержания, представляющий целостный конструкт в разных аспектах, позволяет производить валидное измерение и делать вывод об обученности тестируемого.

Валидность педагогического теста, построенного на фундаменте целостного конструкта, отличается от валидности психологических и педагогических тестов, заданных на основе структурного конструкта. Валидность его конструкта инвариантна содержательной валидности. Содержание теста отражает целостность конструкта и является одним из его вариантов. Содержательная валидность инвариантна критериальной валидности, а критерием оказывается само содержание теста, представляющее различные аспекты целостного конструкта. Поскольку независимая и зависимая переменные не разделены субъектом тестирования, постольку валидность теста носит внутренний характер. Независимая переменная, заключающаяся в конструкте, отражается в содержании теста, а зависимая переменная характеризует внутренний критерий, описывая различные аспекты обученности.

Субъект тестирования элиминируется из процедуры создания и валидизации теста. Процесс его конструирования становится технологичным. Содержание теста соответствует объекту измерения, объективность возрастает, вследствие определенности объекта измерения. Надежность теста в аспекте однородности его содержания задается внутренним характером выбранного критерия. Однородность понимается как соответствие каждого тестового задания целостному конструкту. Надежность теста в смысле рели-

абильности отсутствует. Для получения высокой точности измерения обученности необходимо проводить мониторинговые исследования, состоящие из серии тестирований, построенных на основе репрезентативной модели содержания. Измерение обученности с использованием валидных тестов, построенных на основе целостного конструкта и внутреннего критерия валидизации, требует только однократного тестирования, и, следовательно, значение надежности теста смещается с аспекта релиабильности на аспект однородности (см. рис. 6).

Тест, построенный на фундаменте целостного конструкта, обладает и внешней валидностью. Ее характер заключается не в экстраполяции результатов тестирования на большую, чем экспериментальная, выборку, а в распространении результатов тестирования с содержания самого теста на всю измеряемую предметную область для каждого тестирующегося, где содержание теста становится определенным содержательным вариантом целостного конструкта. Другими словами, если учащийся демонстрирует владение целостной моделью предметного содержания, заключенной в тесте, то он также успешно освоил и все измеряемое содержание.

Следует отметить, что изменение принципов конструирования содержания тестов, построенных на основе теории IRT, и, прежде всего, изменение теоретического конструкта со структурного на целостный повысит качество нормативно-ориентированных тестов, позволяя создавать тесты, обладающие высокой надежностью (релиабильностью) и внутренней валидностью.

Таким образом, при выборе содержания теста в качестве внутреннего критерия валидизации, основанного на принципе межличностного соглашения, можно сделать следующие выводы. Определенный характер объекта измерения и соответствующий ему Целостный конструкт предметного содержания позволяют выбрать в качестве внутреннего критерия валидизации содержание самого теста, которое является моделью теоретического конструкта.

Выбор такого конструкта задает содержательную валидность теста как внутреннюю валидность. Устанавливается каузальная связь теоретического конструкта с результатами тестирования.

Возможным становится построение шкалы, свободной от статистических характеристик, но зависящей от представленности подходов, определяющих целостность предъявляемого предметного содержания. Валидизация технологична, вследствие элиминации характеристик субъекта тестирования из процесса создания теста. Тест становится более объективным. Его внешняя валидность рассматривается в аспекте экстраполяции результатов тестирования на содержание большее, чем тестовое, а надежность, направленная на более точное описание конструкта, представляется как однородность содержания теста. Возникают условия конструирования теста, обладающего высокой надежностью и внутренней валидностью одновременно.

2.3. Внутренняя валидность как условие объективности результатов тестового контроля обученности

Основным преимуществом тестового контроля знаний перед иными формами контроля принято считать объективность тестовых результатов. Под объективностью традиционно понимается исключение из тестовых результатов интерсубъективных воздействий или их минимизация (К. Ингенкамп [44], А.С. Казаринов [46], Т. Кук, Д. Кэмпбелл [187], Я. тер Лаак [66], Е.Н. Лебедева [68], А.Н. Майоров [79], Е.А. Михайлычев [92], Ю.М. Нейман, В.А. Хлебников [106], М.Б. Челышкова [163]). К. Ингенкамп выделяет объективность проведения измерений, объективность обработки данных и объективность интерпретации как условия обеспечения объективности методики измерения. С этой точки зрения, объективность «является только предпосылкой надежности и валидности измерения» [44, с. 36].

Мы считаем, что это узкое толкование объективности связано с латентностью объекта исследования. Его неопределенность ограничивает возможность представления объективности как соответствия результатов тестирования объекту исследования. Во-первых, результаты тестирования и латентная характеристика субъекта тестирования как объект исследования несоразмерны друг другу. Во-вторых, латентность объекта исследования не позволяет исключить характеристики субъекта исследования из процесса создания теста. Тест как относительно объективный инструмент заключает в себе субъективные характеристики. В-третьих, латентность объекта исследования позволяет строить предположения о соответствии результатов тестирования объекту исследования на основе вероятностного заключения о валидности теста как инструмента, а также валидности исследования в целом.

Переход с латентного объекта исследования на определенный и открытый позволяет добиться сравнения результатов тестирования и объекта исследования. Они открыты и определены в терминах предметного содержания. В данном аспекте объективность результатов тестирования состоит не только в отсутствии субъективных воздействий или их минимизации, но и в соответствии тестовых результатов объекту исследования. Условием объективности результатов оказывается объективность теста как инструмента исследования, а именно соответствие его содержания объекту исследования. Рассмотрим в данном аспекте факторы, необходимые для обеспечения объективности результатов тестирования.

Объективность педагогического тестирования как исследования, или измерения, необходимо представить с трех сторон. Во-первых, объективное измерение — это процесс, свободный от субъективных факторов. Проблема устранения внешних субъективных воздействий при проведении тестирования в тестологии решается через унификацию процедуры тестирования: строгое следование инструкциям при проведении тестирования, обеспечение равных условий для всех тестирующихся и т. п. Данный аспект объективации измерения не относится непосредственно к инструменту измерения (тесту), но переводит проблему объективности измерения из теории создания тестов в область управления процессом измерения (образования). Объективность процедуры измерения есть необходимое, но не достаточное условие для получения объективных результатов. Следует отметить, что данная грань объективности не ориентирована на соответствие результатов объекту исследования.

Объективность процесса проведения тестирования не зависит от выбранного объекта измерения и является универсальным требованием для всех видов тестов.

Во-вторых, получение объективных результатов тестирования невозможно без использования качественного инструмента. Основными характеристиками качества теста являются его надежность и валидность. Эти понятия непосредственно связаны с характеристиками объекта исследования. Мы условно разделяем тесты с латентным объектом, конструированные на базе репрезентативной модели предметного содержания, и тесты с открытым объектом, построенные на основе целостной модели предметного содержания. Понятия внешней и внутренней валидности, а также надежности в данных тестах расходятся. Проследим, как они связаны с объективностью результатов.

Проанализируем тесты с репрезентативной моделью предметного содержания. В них объект исследования латентен. Для получения адекватных результатов необходимы несмещенные оценки одной и той же характеристики объекта, например, уровня обученности. Необходимым условием объективности является стабильность полученных результатов. Другими словами, тестовый инструмент необходимо должен обладать высокой степенью надежности в смысле релиабильности. Проведение объективного измерения невозможно без использования надежного инструмента. «Под надежностью, или релиабильностью, измерения понимается степень надежности, или точности, с какой может быть измерен тот или иной конкретный признак» [44, с. 37]. Степень надежности выражается коэффициентом надежности. Он служит коэффициентом корреляции и проясняется в пилотажном тестировании на выборке нормирования теста. О надежности теста можно говорить при условий применения его в группе тестируемых, которая идентична по своему составу выборке стандартизации [80; 106; 166]. Обеспечение соответствия между выборкой стандартизации и генеральной совокупностью тестирующихся переводит проблему получения объективных результатов в плоскость управления процессом тестирования, а также выводит надежность

теста в основное условие получения объективных результатов для таких тестов. Иначе говоря, надежность оказывается не столько характеристикой инструмента, сколько свойством качества процедуры тестирования.

Примером технологии построения тестов с высокой степенью надежности, построенных на основе латентного объекта исследования, является теория IRT. Использование математического аппарата при конструировании тестов в теории IRT позволяет создавать надежные тесты. Становится возможным объективное проведение измерения. Однако открытым оказывается главный вопрос, какой именно объект измеряет тест [131]. Нельзя вполне достоверно утверждать, что результаты тестирования относятся к измеряемому объекту. Соответствие или несоответствие результатов измерения объекту определяется интерпретацией.

Объективное измерение — это измерение с объективной интерпретацией результатов. Основная задача интерпретации — построение переводной шкалы, адекватной объекту тестирования, т.е. уровню обученности тестирующихся. Для этого используют результаты нормативной выборки. При сопоставлении результатов тестирования, полученных на выборке нормирования, и результатов измерения того же содержания, полученных другим способом (внешний критерий), определяются коэффициенты связи (корреляции) двух числовых рядов. Коэффициент корреляции как коэффициент критериальной валидности в данном случае является коэффициентом валидности соответствия. Валидность соответствия опосредованно через внешний критерий связывает полученные результаты и теоретический конструкт, а через него результаты и измеряемую латентную характеристику. Таким образом, объективность интерпретации результатов тестирования, на наш взгляд, обеспечивается внешней валидностью соответствия и носит вероятностный характер.

Одной из основных задач тестирования, как инструмента объективного оценивания учебных достижений, построенного на основе репрезентативной модели предметного содержания, при измерении латентного уровня обученности является «внедрение стандар-

газированных процедур тестирования как процедур объективного оценивания учебных достижений» [160, с. 71]. В понятие стандартизированной процедуры тестирования для необходимого обеспечения объективности результатов тестирования входят:

1) инструкция для тестирующих;
2) инструкция для тестирующихся;
3) указание характеристик выборки стандартизации;
4) коэффициент надежности теста;
5) коэффициент валидности соответствия.

Следует отметить, что объективность измерения в аспекте соответствия результатов тестирования объекту измерения не достигается из-за отсутствия в тестах с латентным объектом исследования внутренней характеристики валидности и, прежде всего, из-за внешнего характера содержательной (куррикулярной) валидности.

Внутреннюю содержательную валидность имеют тесты с целостным конструктом предметного содержания, измеряющие характеристики открытого и определенного объекта исследования. Внутренняя валидность теста, по определению, указывает на соответствие результатов цели тестирования, т. е. измерению обученности. Таким образом, внутренняя содержательная валидность теста является основным условием объективности измерения в аспекте соответствия результатов тестирования объекту исследования.

Надежность таких тестов рассматривается в смысле однородности тестового содержания, его соответствия целостному теоретическому конструкту. Однородность связана с содержанием теста, а значит, является аспектом содержательной валидности. Однородность как внутренняя характеристика содержания теста позволяет экстраполировать полученные результаты на весь теоретический конструкт, т. е. на предметное содержание как целое. Содержание теста является одним из вариантов проявлений целостного конструкта. Его однородность направлена на представление целостности конструкта в содержании теста. Она манифестирована подходами, а владение ими отражено в тестовых результатах. Как мы утверждаем, обученность и есть владение хотя бы одним

из подходов. Внутренняя содержательная валидность позволяет по результатам даже одного тестирования определить владение тестирующимся представленными в тестовом содержании подходами. Наличие надежности теста в аспекте его однородности предъявляет обученность тестирующегося как владение им всем контролируемым предметным содержанием.

В этой связи можно говорить о третьем аспекте объективного измерения. Объективным может быть признано такое измерение, посредством которого могут быть описаны качественные характеристики объекта измерения, в нашем случае обученности.

По нашему мнению, объективность тестового контроля обученности в аспекте соответствия результатов тестирования объекту исследования может быть обеспечена только при наличии внутренней валидности теста, а именно внутренней содержательной валидности. Другими словами, внутренний характер содержательной валидности в тестах с целостным конструктом предметного содержания является необходимым и достаточным условием объективности тестовых результатов в аспекте их соответствия определенному объекту исследования — обученности.

Относительное решение объективности результатов возможно и в тестах с латентным объектом исследования, построенных на основе структурного конструкта предметного содержания. В таких тестах содержательная (куррикулярная) валидность оказывается внешней валидностью по отношению к измеряемому объекту— уровню обученности. Ее внутренний характер достигается при непрерывном контроле над уровнем обученности в процессе мониторинга. Недостаточно однократного тестирования для получения объективных результатов, а объективность в данном случае будет пониматься не как соответствие объекту исследования из-за его латентности, но соответствие результатов тестирования теоретическому конструкту. В данном аспекте объективность результатов тестирования может быть обеспечена батареей валидных и надежных тестов со структурным конструктом и репрезентативной моделью предметного содержания. Валидность таких тестов будет состоять из трех внешних по характеру видов валидности: конструктной,

куррикулярной и критериальной. Надежность батареи тестов в аспекте релиабильности является необходимым условием обеспечения объективности тестовых результатов.

Таким образом, возможность получения объективных тестовых результатов определяется характером объекта исследования и смыслом понятия «объективность». В настоящее время под объективностью тестовых результатов понимается минимизация интерсубъективных воздействий на них. Вне зависимости от характера объекта исследования объективность результатов тестирования в данном аспекте достигается за счет унификации процедуры и применения качественного инструмента тестирования. В связи с этим вводится понятие сертификации тестов и процедуры тестирования.

Понятие качества теста варьируется в зависимости от исследуемого объекта. Для тестов с латентным объектом и структурным теоретическим конструктом предметного содержания основными характеристиками качества теста, влияющими на объективность полученных результатов, являются надежность и внешняя валидность соответствия. Их наличие позволяет делать вероятностный вывод об объективности полученных результатов в смысле их незначительной зависимости от внешних неконтролируемых факторов. Объективность результатов тестирования понимается как точность измерения и сводится к надежности полученных тестовых результатов, а их связь с измеряемым объектом устанавливается опосредованно через привлечение внешнего критерия для определения валидности соответствия. Абсолютное решение проблемы объективности результатов проблематично, ввиду невозможности обеспечить полную валидность выбранного внешнего критерия.

Отсутствие внутренней валидности таких тестов из-за внешнего характера куррикулярной валидности приводит к относительному решению проблемы обеспечения объективности результатов в аспекте их соответствия теоретическому конструкту (структурному характеру предметного содержания), организации непрерывного мониторинга за результатами обучения.

Основой для получения объективных тестовых результатов в тестах с открытым объектом и целостным характером теоретического конструкта является внутренняя содержательная валидность теста, наличие которой позволяет говорить о соответствии тестовых результатов объекту исследования. Объективность результатов, полученных данным инструментом, понимается в аспекте их соответствия объекту исследования.

ВЫВОДЫ ПО ВТОРОЙ ГЛАВЕ

1. Определенность объекта исследования позволяет представить его как целостное предметное содержание, заданное для измерения в виде теоретического конструкта, определенное структурой предмета и ограниченное представленными в нем теоретическими подходами. Целостное предметное содержание является моделью объекта исследования (обученности), которая понимается в аспекте владения тестируемым одним из общих теоретических подходов к изложению предметного содержания.
2. В аспекте целостного конструкта создается тест, обладающий внутренней содержательной валидностью, содержание которого является моделью конструкта и, одновременно, внутренним критерием оценивания. Результаты тестирования сравниваются с критерием, построенным на основе содержания теста методом экспертных оценок.
3. Наличие внутренней валидности гарантирует адекватность результатов измерения характеристикам открытого объекта. Они являются объективными в аспекте их соответствия объекту измерения, т. е. обученности, пониманию содержания предмета как целого, предъявляют владение тестируемым теоретическим знанием.
4. Субъект исследования элиминируется из процедуры создания теста, и конструирование теста становится технологичным. Тестовая технология объективного измерения обученности, основанная на критерии межличностного соглашения с привлечением экспертов, включает в себя следующие этапы:

1) построение теоретического конструкта;
2) построение содержания теста;
3) выработка внутреннего критерия оценивания.

5. Объективный контроль над обученностью становится возможным по итогам однократного тестирования, а не в ходе мониторинга обученности тестами, построенными на основе структурного конструкта.

Глава 3
ЭКСПЕРИМЕНТАЛЬНАЯ АПРОБАЦИЯ ПРОЦЕДУРЫ ОПРЕДЕЛЕНИЯ ВАЛИДНОСТИ ТЕСТА ДЛЯ ИЗМЕРЕНИЯ ОБУЧЕННОСТИ

3.1. Конструирование валидного педагогического теста для измерения обученности на основе метода экспертных оценок

Выбор принципа истинности, положенного в основу понятия валидности педагогического теста для измерения обученности, определяет выбор метода и этапов конструирования теста и определения его валидности. Принцип межличностного соглашения, или консенсуса, предполагает использование метода экспертной оценки. Консенсус достигается в ходе прямого диалога между экспертами. Результаты достигнутого соглашения являются критерием истинности и признаются валидными на основе определяющего принципа. Если в ходе диалога эксперты не приходят к общему мнению, то процедура повторяется, начиная с предыдущего этапа.

Необходимым условием применения в практике дидактического контроля валидных тестовых измерителей, объективно измеряющих обученность по учебному предмету, является алгоритмизация процедуры построения и оценки их качества, теоретически обоснованная в п. 2.2 (см. рис. 6). Предлагаемая процедура определения валидности педагогического теста, основанная на принципе консенсуса с применением метода экспертной оценки, включена в процесс конструирования теста с открытым объектом измерения

(обученностью), целостным конструктом предметного содержания и состоит из трех этапов.

1-й этап. Экспертная оценка предметного содержания.

Целью данного этапа конструирования валидного педагогического теста для измерения обученности является построение валидного теоретического конструкта контролируемого предметного содержания.

2-й этап. Экспертная оценка содержания теста.

Построение содержания теста, обладающего внутренней содержательной валидностью и технической валидностью формы, является *целью* данного этапа.

3-й этап. Экспертная оценка критерия.

На данном этапе *целью* является установление валидного внутреннего критерия, основанного на содержании теста для измерения обученности.

Проследим практическое применение валидизации на примере конструирования рубежного дидактического теста по математике для слушателей подготовительного отделения Ижевского государственного технического университета за 1-й семестр 2001—2002 уч. г. В связи с этим будем различать этапы конструирования и этапы валидизации теста.

Объектом измерения выбирается обученность слушателей содержанию курса математики первого семестра. Обученность как открытый объект измерения определена структурой предметного содержания и ограничена теоретическими подходами к изложению предметного содержания. Предметом измерения становится целостный теоретический конструкт содержания курса математики первого семестра. В этой связи построение теста начинается с определения теоретического конструкта.

1-й этап конструирования. Этап создания целостного конструкта измеряемого предметного содержания.

1.1. Экспертной группе предъявляется структура предметного содержания, определяющая исследуемый объект.

Для организации контроля над результатами обучения за основу структуры предметного содержания взят учебный план по математике за 1-й семестр. В табл. 3 представлена выписка из «Учебной программы по математике подготовительного отделения ИжГТУ», утвержденная на заседании кафедры «Методика довузовского обучения» Института образовательных коммуникаций ИжГТУ протоколом № 1 от 21 сентября 1999 г.

Таблица 3

Структура теоретического конструкта

№ п/п	Тема	Количество часов		
		лекции	практика	всего
1	*Арифметика*	*2*	*6*	*8*
2	*Степень с рациональным показателем*	*4*	*20*	*24*
2.1	Степень с целым показателем, многочлены, преобразование алгебраических выражений	2	10	12
2.2	Степень с рациональным показателем, корень n-й степени, преобразование выражений, содержащих корни	2	10	12
3	*Функция*	*4*	*22*	*26*
3.1	Определение, основные свойства функции. Графики и их преобразования, обратные функции	2	10	12
3.2	Линейная, квадратичная и дробно-рациональная функция, их свойства и графики	2	12	14
4	*Уравнения*	*4*	*22*	*26*
4.1	Линейные, квадратичные и дробно-рациональные уравнения и их системы, иррациональные уравнения	2	10	12
4.2	Уравнения, содержащие параметр и абсолютную величину, решение задач на составление уравнений и систем уравнений	2	12	14
5	*Неравенства*	*4*	*22*	*26*
5.1	Линейные, квадратичные и дробно-рациональные неравенства и их системы	2	10	12
5.2	Иррациональные неравенства, неравенства с параметром и неравенства, содержащие абсолютную величину	2	12	14

1.2. Экспертная группа выбирает теоретический подход для контроля обученности.

На основе экспертной оценки принимаются решения:

— контролировать целостность конструкта через задания, моделирующие функциональный подход к предметному содержанию;

— опираясь на структуру контролируемого содержания, к таким заданиям предъявлять следующие требования:

1) содержание заданий связано с понятием функции в общем виде (определение функции, область определения и область значения, свойства непрерывности и монотонности функции на промежутке и области определения, четность и нечетность функции, построение и свойства графика функции, преобразования графика функции в общем виде);

2) в качестве предметного содержания моделирующих заданий могут выступать линейные, квадратичные, дробно-рациональные функции, а также сводящиеся к ним функции;

3) для композиции функций могут быть использованы также функции арифметического корня второй и третьей степени, функции модуля аргумента, а также зависимости с двумя переменными, сводящиеся к совокупности вышеперечисленных функций;

4) задания должны обладать невысокой технической сложностью (табл. 4).

1.3. Экспертная группа проектирует содержание теста.

На основе экспертной оценки принимаются решения:

1) конструировать тест, состоящий из двух групп заданий: в группу А включать задания закрытой формы с выбором одного правильного ответа с пятью дистракторами; в группу В включать задания открытой окормы;

2) задания разбить на два типа: Т — задания, контролирующие техники решения; Ф — задания, контролирующие владение функциональным подходом;

3) задания разбить на 3 вида по возрастанию уровня сложности техники, применяемой для его решения;

4) за верное решение задания присваивать баллы в зависимости от вида (технической сложности задания) следующим

образом: верное решение задания 1-го уровня сложности оценивается в 1 балл, 2-го уровня сложности — 2 балла, 3-го уровня сложности — 3 балла.

В табл. 4 представлена спецификация теста, составленная на основе структуры теоретического конструкта и утвержденная решением экспертной группы.

Таблица 4

Спецификация теста

№ п/п	Тема	Уровень технической сложности	Количество заданий	
			группа А	группа В
1	Преобразование алгебраических выражений	1	1	
2	Дробно-рациональные уравнения	1	1	
3	Дробно-рациональные неравенства	1	1	
4	Преобразование выражений, содержащих корни	1	1	
5	Иррациональные уравнения	1	1	
6	Иррациональные неравенства	2 3	1	1
7	Определение, основные свойства функции. Графики и их преобразования, обратные функции	2 1 2 2	1	3
8	Уравнения, содержащие параметр и абсолютную величину	2		1

Реализация первого этапа конструирования теста, состоящего из трех подэтапов, приводит к конструированию валидного теоретического конструкта предметного содержания и является практическим исполнением первого этапа технологии валидизации теста.

На первом этапе конструирования теста экспертная группа приходит к соглашению по поводу определения предмета измерения. Посредством описания теоретического конструкта достигается взаимопонимание и взаимная согласованность представления об объекте измерения — обученности. Этот этап является основой создания

теста, а достигнутый консенсус позволяет без угрозы для валидности поделить группу экспертов-предметников на две непересекающиеся подгруппы: подгруппа разработчиков теста и подгруппа экспертов содержания теста. Для обеспечения технической валидности теста на втором этапе конструирования теста необходимо для оценки формы теста привлечь группу экспертов-тестологов.

2-й этап конструирования. Этап создания на основе целостного конструкта измеряемого предметного содержания педагогического теста, контролирующего обученность.

2.1. Экспертная группа разработчиков конструирует тест на основе его спецификации и требований, выработанных экспертами.

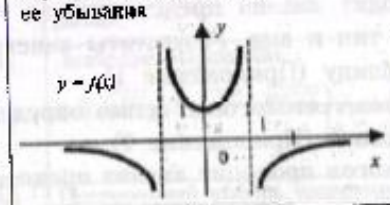
В табл. 5 представлен вариант теста, составленного подгруппой разработчиков. Каждое тестовое задание отнесено решением группы разработчиков к определенному типу (Т — задания, контролирующие техники решения, Ф — задания, контролирующие владение функциональным подходом к изложению предметного содержания). В таблице учтены мнения экспертов-разработчиков относительно оценивания верного выполнения каждого тестового задания.

Таблица 5

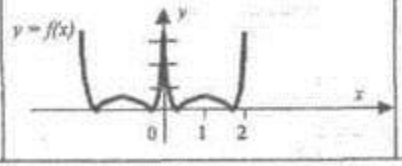
Вариант теста группы разработчиков

№ п/п	Содержание задания	Варианты ответов	Баллы за верное решение	Тип задания
A1	Результат упрощения выражения $\left(\frac{1}{a-3b}-\frac{1}{a+3b}+\frac{6b}{a^2-9b^2}\right)+\frac{b(2a+b)}{a^2-9b^2}$ имеет вид	1) 0; 2) $\frac{1}{2a+b}$; 3) $\frac{12}{2a+b}$; 4) $\frac{12b}{2a+b}$; 5) $12b$	1	Т
A2	Сумма корней уравнения $\frac{x^2-x}{x^2-x+1}-\frac{x^2-x+2}{x^2-x-2}=1$ равна	1) 1; 2) 2; 3) 3; 4) −1; 5) 0	1	Т

Продолжение табл. 5

A3	Число целых решений неравенства $\frac{7x-13}{(x^2-9x+18)(x^2-8x+16)}+\frac{1}{x^2-10x+24}\le 0$ равно	1) 0; 2) 3; 3) 2; 4) 1; 5) 4	1	Т
A4	Значение выражения $(\sqrt{8}\sqrt[3]{3\sqrt{2}})+\left(\sqrt{2\sqrt{3}}\cdot\sqrt[6]{108}\right)$ равно	1) $\sqrt{2}$; 2) 4; 3) $3^{1/6}$; 4) 2; 5) 1	1	Т
A5	Посторонний корень уравнения $\sqrt{x+1}+\sqrt{4x+13}=\sqrt{3x+12}$ равен	1) −4; 2) 4; 3) 2; 4) −2; 5) −1	1	Т
A6	Решением неравенства $\sqrt{x^2-2x+1}+\sqrt{x^2+4x+4}\le 3$ является множество	1) [−1; 2]; 2) (−2; 1); 3) $(-\infty;-2)\cup(1;+\infty)$; 4) [−3; 1]; 5) (−1; 2)	2	Т
A7	Областью значений функции $f(x)=2-\lvert\sqrt{x}-3\rvert$ является промежуток	1) $(2;+\infty)$; 2) $(-\infty;2)$; 3) $(-\infty;2]$; 4) $[0;+\infty)$; 5) $(-\infty;0]$	2	Т
B8	По графику функции $y=f(x)$ определите промежуток ее убывания 		1	Т
B9	Решите неравенство $\frac{x-2}{\sqrt{2x-1}-1}<4$		3	Т
B10	Запишите какую-либо формулу функции $y=f(x)$ если ее областью определения является промежуток $(-2;1]\cup[1;+\infty)$		2	Ф
B11	Решите уравнение с параметром относительно переменной x $\frac{a+1}{ax-2}-1=0$		2	Т

Окончание табл. 5

B12	По данному графику восстановите и запишите формулу зависимости переменной y от x: 		2	Ф

2.2. Экспертная группа разработчиков конструирует критерий обученности по разработанному тестовому содержанию.

В ходе диалога вырабатывается следующая шкала перевода тестовых баллов в отметки: 17—19 баллов — оценка «отлично», 14—16 баллов — «хорошо», 8—13 баллов — «удовлетворительно», 0—8 баллов — «неудовлетворительно».

3-й этап конструирования. Этап анализа тестового содержания на предмет его соответствия теоретическому конструкту.

3.1. Анкетирование экспертов-предметников с целью определения соответствия содержания теста теоретическому конструкту.

Подгруппа экспертов проводит анализ представленных тестовых заданий и определяет их тип и вид. Результаты анкетирования экспертов заносятся в таблицу (Приложение 1).

3.2. Анкетирование экспертов-тестологов с целью определения качества формы тестовых заданий (Приложение 2).

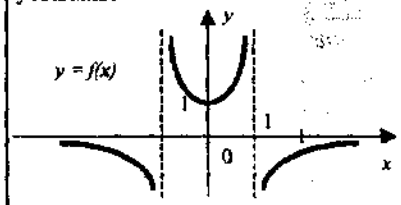
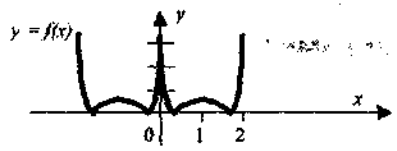
Подгруппа экспертов-тестологов проводит анализ представленных тестовых заданий на предмет соответствия их формы принятым в тестологии нормам. На данном этапе экспертной оценки определяется техническая валидность теста. В результате анализа качества формы тестовых заданий экспертами-тестологами были высказаны замечания по единству формулировок заданий группы А и группы В. Замечания были учтены разработчиками в конечном варианте теста (Приложение 3). Обобщенные результаты экспертной оценки качества содержания теста представлены в табл.6.

Таблица 6

Обобщенные результаты анкетирования экспертов-предметников

№ п/п	Содержание задания	Варианты ответов	Баллы за верное решение	Тип задания
A1	Результат упрощения выражения $\left(\frac{1}{a-3b}-\frac{1}{a+3b}+\frac{6b}{a^2-9b^2}\right)\cdot\frac{b(2a+b)}{a^2-9b^2}$ имеет вид	1) 0; 2) $\frac{1}{2a+b}$; 3) $\frac{12}{2a+b}$; 4) $\frac{12b}{2a+b}$; 5) $12b$	1	Т
A2	Сумма корней уравнения $\frac{x^2-x}{x^2-x+1}-\frac{x^2-x+2}{x^2-x-2}=1$ равна	1) 1; 2) 2; 3) 3; 4) –1; 5) 0	1	Т
A3	Число целых решений неравенства $\frac{7x-13}{(x^2-9x+18)(x^2-8x+16)}+\frac{1}{x^2 10x+24}\le 0$ равно	1) 0; 2) 3; 3) 2; 4) 1; 5) 4	1	Т
A4	Значение выражения $(\sqrt{8}\sqrt{3\sqrt[3]{2}})\cdot\left(\sqrt{2\sqrt{3}}\cdot\sqrt{\sqrt[6]{108}}\right)$ равно	1) $\sqrt{2}$; 2) 4; 3) $3^{1/2}$; 4) 2; 5) 1	1	Т
A5	Посторонний корень уравнения $\sqrt{x+1}+\sqrt{4x+13}=\sqrt{3x+12}$ равен	1) –4; 2) 4; 3) 2; 4) –2; 5) –1	1	Т
A6	Решением неравенства $\sqrt{x^2-2x+1}+\sqrt{x^2+4x+4}\le 3$ является множество	1) [–1; 2]; 2) (–2; 1); 3) $(-\infty;-2)\cup(1;+\infty)$; 4) [–2; 1]; 5) (–1; 2)	2	Т

Окончание табл. 6

A7	Областью значений функции $f(x)=2-\left\|\sqrt{x}-3\right\|$ является промежуток	1) $(2;+\infty)$; 2) $(-\infty;\ 2)$; 3) $(-\infty;\ 2]$; 4) $[0;+\infty)$; 5) $(-\infty;\ 0]$	2	Т
B8	По графику функции $y=f(x)$ определите промежуток ее убывания 		1	Т
B9	Решите неравенство $\frac{x-2}{\sqrt{2x-1}-1}<4$		3	Т
B10	Запишите какую-либо формулу функции $y=f(x)$, если ее областью определения является промежуток $(-2;1]\cup[1;+\infty)$		2	Ф
B11	Решите уравнение с параметром относительно переменной x: $\frac{a+1}{ax-2}-1=0$		2	Т
B12	По данному графику восстановите и запишите формулу зависимости переменной y от x: 		2	Ф

Подгруппой экспертов заданиями, моделирующими функциональный подход, были признаны задания B10 и B12, обладающие вторым уровнем технической сложности. В связи с весомостью смысловой нагрузки, приходящейся на моделирующие задания, обладающие невысокой технической сложностью, экспертами было принято решение, обобщенные результаты которого отражены в таблице (Приложение 2).

Дальнейшее построение теста происходит на следующем этапе конструирования при объединении подгрупп разработчиков и экспертов.

4-й этап конструирования. Этан создания единого критерия на основе результатов анализа тестового содержания.

4.1. В ходе диалога экспертная группа достигает согласия по поводу окончательного варианта содержания теста.

4.2. В ходе диалога экспертная группа конструирует окончательный вариант внутреннего критерия оценивания.

Выработанный критерий оценивания состоит из двух условий:

1) обученность считается достигнутой, если за моделирующие функциональный подход задания слушатель получает 5 баллов;

2) утверждена следующая шкала перевода тестовых баллов в отметки с учетом первого условия: 17—20 баллов — оценка «отлично», 13—16 баллов — «хорошо», 8—12 баллов — «удовлетворительно», 0—7 баллов — «неудовлетворительно».

Второй этап технологии валидизации теста, направленный на определение внутренней содержательной валидности теста, реализуется на этапе 4.1 конструирования теста. Техническая валидность определяется на этапе 3.2 конструирования теста. Третий этап технологии валидизации — конструирование внутреннего критерия оценивания — реализуется на этапе 4.2 конструирования теста, после которого тест признается валидным по критерию.

Таким образом, представленная процедура валидизации педагогического теста для контроля обученности реализуется в процессе конструирования теста с открытым объектом измерения и целостным теоретическим конструктом предметного содержания

на примере содержания курса математики за 1-й семестр при обучении на подготовительном отделении.

В Приложении 3 приводится рубежный тест по математике, контролирующий результаты обучения на подготовительном отделении за 1-й семестр (обученность), базирующийся на целостном конструкте предметного содержания, валидный на основе принципа межличностного соглашения и надежный в аспекте однородного описания тестовым содержанием теоретического конструкта.

3.2. Сравнительный анализ результатов эксперимента по измерению обученности

С целью эмпирического подтверждения большей объективности тестовых результатов при педагогическом контроле обученности с использованием сконструированного в п. 3.1 валидного теста, в 2001—2002 уч. г. на подготовительном отделении ИжПУ проводилось экспериментальное исследование. В основу эксперимента положено сравнение данных дидактического тестового контроля обученности слушателей математики с использованием тестовых измерителей, валидность которых базируется на различных принципах истинности. Планирование, проведение и анализ результатов проводился в следующей последовательности.

1. Анализ проблемы измерения результатов обучения тестами, валидными на основе **различных принципов истинности.**

Целью обучения на подготовительном отделении является успешность поступления в вуз. Критерием успешности в том числе является «отличный» результат по итогам прохождения слушателями централизованного тестирования. Его инструментом является нормативно-ориентированный тест, построенный методами теорий моделирования и параметризации педагогических тестов, обеспечивающими качество через внешнюю куррикулярную валидность и высокую надежность в аспекте точности. Валидность такого теста строится на принципе функциональности, а ее внешний

характер позволяет экстраполировать результаты на большую, чем экспериментальная, выборку. Отсутствие внутренней валидности теста делает измерение на его основе относительным, т. е. происходит ранжирование тестирующихся по степени владения предъявленным тестовым содержанием. То есть успешность тестирования непосредственно зависит от субъективных характеристик генеральной совокупности тестирующихся.

Однако контролировать текущие и рубежные результаты обучения слушателей подготовительного отделения таким инструментом затруднительно, так как для его конструирования необходимо иметь достаточно большую и репрезентативную выборку стандартизации. Это условие невыполнимо для ограниченной и специфической группы слушателей в количестве не более 150 человек. Следовательно, необходимо искать более общие характеристики успешности текущего обучения, не зависящие от субъективных свойств выборки, пригодные для текущего контроля результатов обучения и позволяющие строить объективный прогноз успешности обучения. В качестве такого критерия успешности обучения нами выбирается обученность как целостное владение теоретическими подходами предметного содержания, не зависящая от субъективных характеристик тестирующихся.

Таким образом, для обеспечения сравнительного анализа результатов измерения итогов обучения тестами, валидными на основе различных принципов истинности, необходимо:

— подвергать измерению различными типами тестов по возможности одинаковое предметное содержание;

— в ходе эксперимента стабилизировать субъективные характеристики, влияющие на объективность результатов тестирования.

2. Выдвижение гипотезы эксперимента.

В результате анализа проблемы измерения результатов обучения выдвигается рабочая гипотеза. Измерение обученности, как открытого объекта измерения, педагогическим тестом на основе целостного теоретического конструкта предметного содержания с внутренней валидностью обеспечивает большую объективность тестовых результатов в аспекте их соответствия объекту измерения.

3. Планирование эксперимента.

Объект тестирования — знания слушателей подготовительного отделения (латентный или открытый объект в зависимости от типа инструмента).

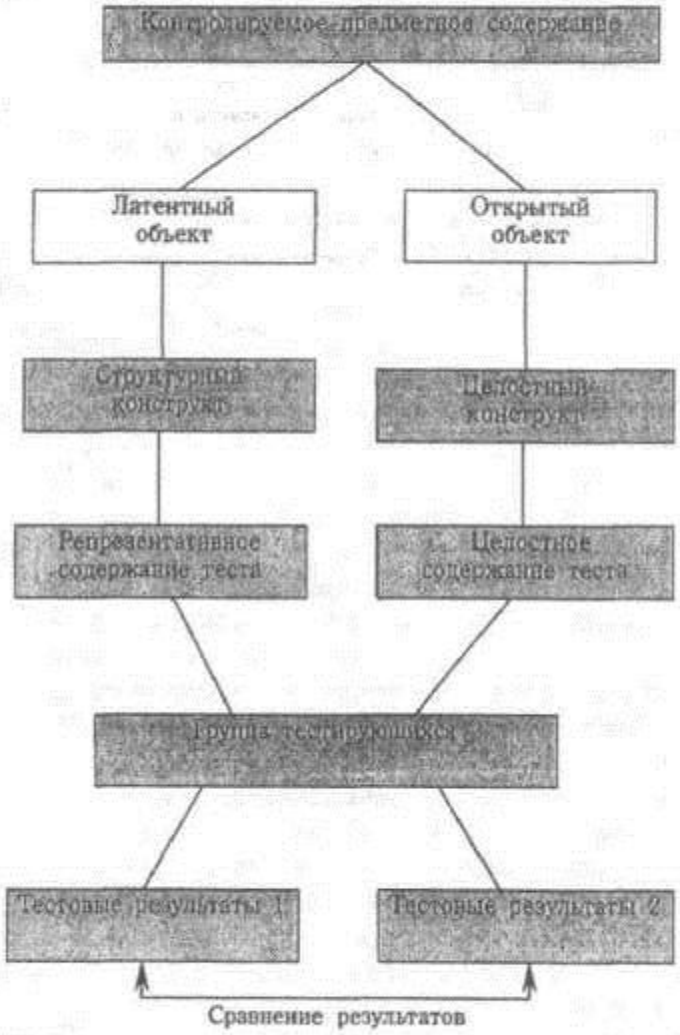


Рис. 7. План эксперимента

Субъект тестирования — группа слушателей подготовительного отделения, состоящая из 100 человек.

Предмет тестирования — предметное содержание, представленное в виде структурного или целостного теоретического конструкта.

Экспериментальный инструмент тестирования: валидный на основе консенсуса тест, обладающий внутренней валидностью и однородностью в аспекте соответствия содержания теста целостному теоретическому конструкту, построенного с целью измерения открытого объекта предметного содержания — обученности.

Контрольный инструмент тестирования: валидный на основе принципа функциональности нормативно-ориентированный инструмент централизованного тестирования, обладающий высокой надежностью и внешней куррикулярной валидностью, построенный на основе структурного конструкта предметного содержания с целью оценивания латентного объекта — уровня обученности. План эксперимента представлен на рис. 7.

4. Проведение эксперимента.

Выборка тестирующихся обоими тестами составила 100 человек (К = 100). В результате эксперимента предметное содержание, контролируемое экспериментальным тестом, включено в содержание, контролируемое в процессе централизованного тестирования. Результаты двух измерений экспериментальным и контрольным тестами представлены в Приложении 4 в порядке уменьшения суммарного балла, полученного в результате тестирования экспериментальным тестом.

5. Анализ результатов эксперимента.

По результатам контрольного тестирования максимально возможные баллы от 81 до 100, соответствующие отметке «отлично» в абитуриентском конкурсе, набрали 19 человек (19 %) (рис. 8). Соответственно 58 %, 20 % и 3 % слушателей по результатам централизованного тестирования получили оценки «хорошо», «удовлетворительно» и «неудовлетворительно». В Приложении 5 приведена выписка из шкалы перевода тестового

балла централизованного вузовского тестирования, рекомендуемая Центром тестирования МО РФ. Таким образом, знания 19 слушателей, получивших оценку «отлично» по результатам абитуриентского тестирования, могли бы быть интерпретированы как обученность школьному курсу математики, или знание предмета как целого.

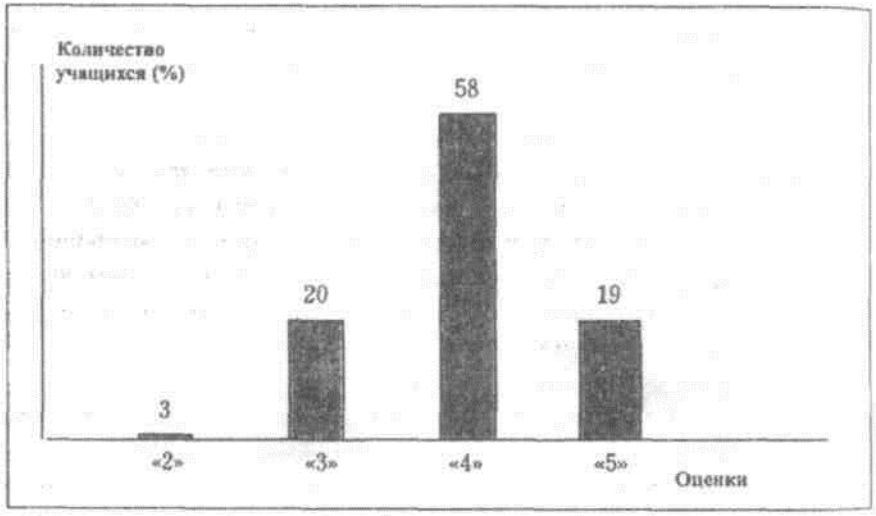


Рис. 8. Диаграмма распределения частот результатов централизованного тестирования

Однако результаты экспериментального тестирования показали, что не все из 19 потенциально возможных слушателей показали знания, которые могут трактоваться как обученность. На основе соглашения экспертов, достигнутого при конструировании теста, владеющими предметным содержанием как целым признаются только те слушатели, которые получили в сумме 5 баллов за моделирующие задания. Обобщенные результаты эксперимента представлены на рис. 9.

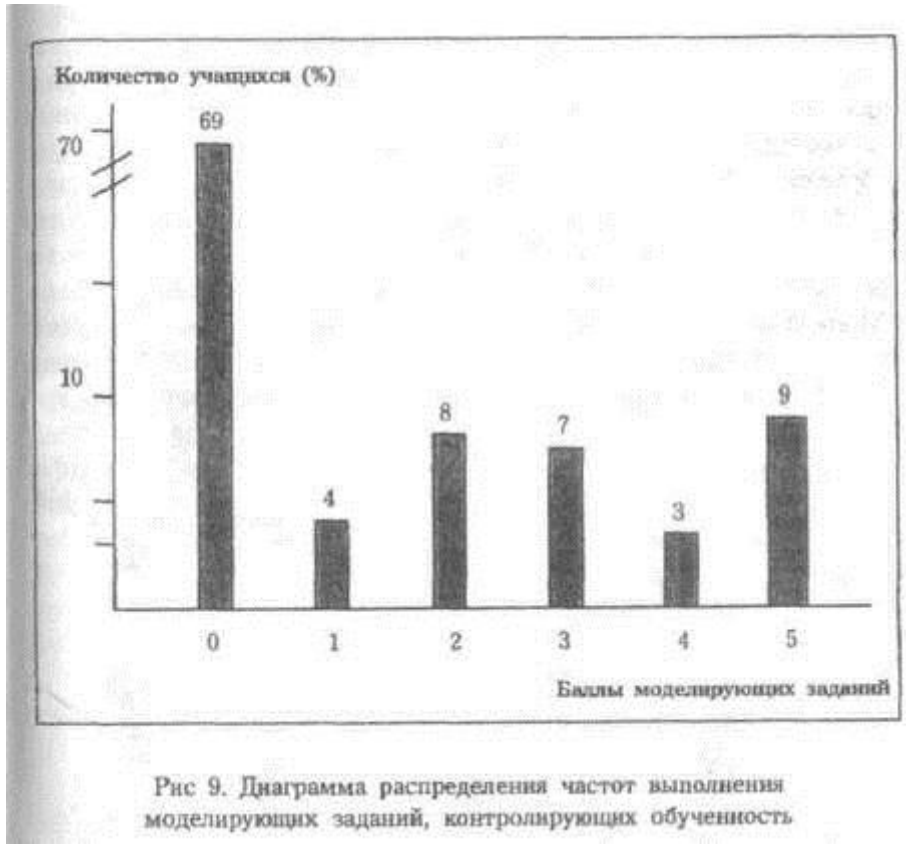


Рис 9. Диаграмма распределения частот выполнения моделирующих заданий, контролирующих обученность

Анализируя их, можно сделать вывод о том, что обученность только 9 человек из 100 (9 %) была признана как целостное понимание предмета. Именно они продемонстрировали максимальное владение функциональным подходом в школьном курсе математики, набрав за представленные в тестовом содержании моделирующие задания 5 баллов из 5 возможных. Количество таких слушателей составляет менее половины из потенциально возможных, получивших оценку «отлично» на централизованном тестировании (19 человек).

С другой стороны, все учащиеся (9%), владеющие предметным содержанием как целым, получили «отлично» на абитуриентском

тестировании. Это позволяет сделать вывод о том, что инструмент контрольного централизованного тестирования не дифференцирует владение предметным содержанием как целым, т. е. является менее объективным инструментом педагогического контроля результатов обучения. В подтверждение данного утверждения на рис. 10 представлена диаграмма распределения частот баллов, полученных по результатам централизованного тестирования, в зависимости от степени владения целостным предметным содержанием. Так, слушатели, набравшие по итогам экспериментального тестирования менее 5 баллов (от 0 до 4 включительно) за контролирующие обученность задания, тем не менее, получили оценку «отлично» на централизованном тес-

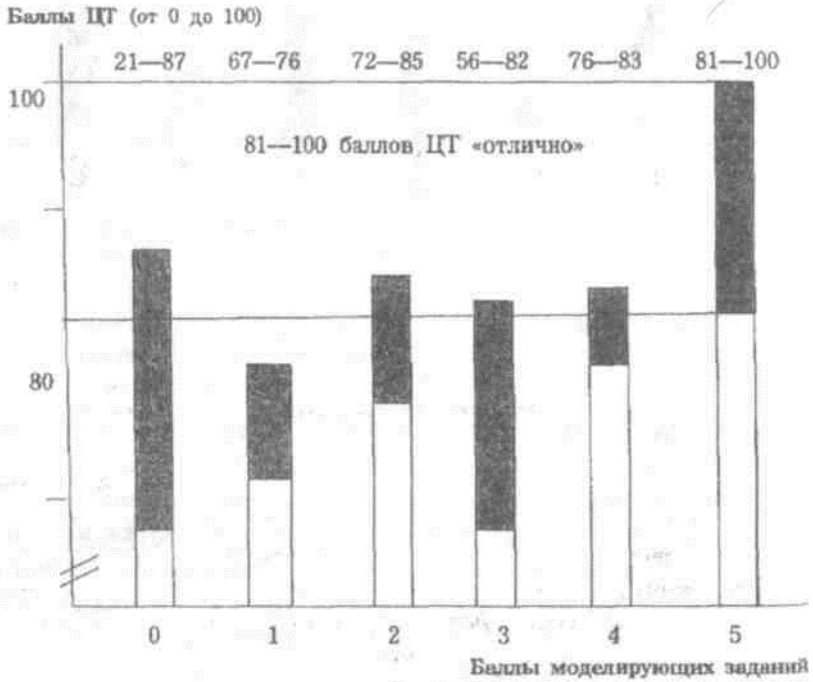


Рис. 10. Диаграмма распределения баллов централизованного тестирования в зависимости от баллов моделирующих заданий

тировании. Таких слушателей 10 человек, что составляет 10% от всех тестирующихся (в Приложении 4 их фамилии выделены курсивом). Среди них 4 балла за моделирующие задания получил 1 человек, 3 балла — 1 человек, 2 балла — 5 человек, 1 балл не получил никто, 0 баллов — 3 человека. Однако все 9 слушателей, получивших по итогам экспериментального тестирования 5 баллов из пяти возможных, подтвердили свои знания на контрольном централизованном тестировании.

6. Подтверждение гипотезы эксперимента.

Интерпретируя результат эксперимента, можно утверждать, что учащиеся, владеющие предметным содержанием как целым, демонстрируют обученность на любом другом конкретном содержательном материале. Анализ результатов эксперимента подтверждает рабочую гипотезу и позволяет сделать следующий вывод. Экспериментальный, валидный на основе принципа межличностного соглашения тест обладает большей различительной способностью определять обученность предметному содержанию как целому, т. е. является более точным инструментом педагогического контроля результатов обучения, а следовательно, и более объективным, чем валидный, на основе принципа функциональности, тест со структурным конструктом предметного содержания.

7. Выводы эксперимента.

Задания, моделирующие функциональный подход, выделяют тестирующихся, владеющих предметным содержанием как целым, т. е. позволяют экстраполировать тестовые результаты на всю предметную область, подвергающуюся измерению.

Знание целостного предметного содержания позволяет прогнозировать успешность абитуриентских испытаний по предмету, т. е. результаты тестирования на основе тестов с целостным теоретическим конструктом являются объективными.

Результаты экспериментального тестирования не зависят от субъективных характеристик выборки, для которой конструировался тест, т. е. удалось элиминировать субъект измерения из процесса построения теста и технологизировать процесс валидизации теста.

ВЫВОДЫ ПО ТРЕТЬЕЙ ГЛАВЕ

1. Технология валидизации педагогического теста для измерения обученности, иллюстрированная на примере построения дидактического теста по математике для рубежного контроля, позволяет добиться внутреннего характера валидности тестового измерителя.
2. Целостный теоретический конструкт предметного содержания является необходимым условием технологичности валидизации теста.
3. Внутренний характер содержательной валидности позволяет экстраполировать тестовые результаты на большее, чем тестовое, содержание.
4. Внутренний критерий валидизации, опирающийся на содержание теста, позволяет элиминировать субъективные характеристики тестирующихся из процесса создания теста и является достаточным условием технологизации процесса валидизации теста.
5. Внутренняя валидность инструмента (конструкта, содержания, критерия) обеспечивает объективность тестовых результатов в аспекте их соответствия объекту измерения — обученности, что подтверждается результатами экспериментального исследования.

ЗАКЛЮЧЕНИЕ

Основными целями введения в педагогическую практику государственных стандартов начального общего, основного общего и среднего (полного) образования являются сохранение единого образовательного пространства и повышение качества образования. Для этого необходимо организовать качественную и, прежде всего, объективную систему контроля над результатами обучения. Создание национальной системы оценки качества образования в России включает в себя теоретическую разработку и практическое использование объективных тестовых средств контроля над результатами обученности на каждом уровне обучения. Применение тестовых технологий для оценивания уровня подготовленности выпускников общеобразовательных учебных заведений с целью аттестации выпускников в рамках единого государственного экзамена и с целью отбора абитуриентов вузов в рамках централизованного тестирования вызывает необходимость получения объективных тестовых результатов и, в этой связи, требует повышения качества используемого тестового инструмента.

Основным критерием объективности экспериментального исследования, в том числе качества тестового измерителя, является его валидность. Категория валидности используется в эмпирическом исследовании для доказательства объективности или истинности его результатов. Она зависит от цели исследования и выбора критерия истинности. В научной литературе по тестологии существует чрезмерно большое количество различных названий видов валидности, их классификаций. В исследовании валидности возникают трудности, связанные с выбором необходимого и достаточного количества их видов в процессе валидизации.

Проведенный в первой главе анализ отечественной и зарубежной научной и методической литературы показал, что для валидизации экспериментального исследования необходимым является обеспечение шести основных видов валидности. Во-первых, валидности конструкта, обеспечивающей соответствие теоретического конструкта объекту исследования. Во-вторых, валидности содержания, отвечающей за репрезентативное представление содержанием теста теоретического конструкта. В-третьих, валидности прогноза (или критериальной валидности), эмпирически подтверждающей соответствие результатов исследования объекту или цели исследования. В-четвертых, внутренней валидности, устанавливающей каузальность связи причины и эффекта. В-пятых, внешней валидности, позволяющей экстраполировать результаты исследования на большую, чем экспериментальная, группу. Наконец, в-шестых, валидности статистического вывода, отвечающей за адекватность интерпретации результатов исследования на основе применения статистических характеристик.

Без обеспечения внутренней валидности невозможно установить объективность тестового измерения в аспекте адекватности его результатов характеристикам объекта. Первые три вида валидности выражают качество планирования эксперимента, а последние задают качество управления экспериментом. Остальные разновидности валидности, встречающиеся в литературе, сводятся к выделенным нами или являются их подвидами. Свое название они получают в зависимости от специфичной цели исследования. Кроме того, их могут маркировать по названию количественного коэффициента корреляции. Имя может задаваться критерием, выбранным при определении степени валидности.

Валидность вывода экспериментального исследования в социальных науках всегда вероятностна, а значит, невозможно получение абсолютно валидных результатов. Вероятность обусловлена латентным характером объекта исследования, который задает невозможность контроля над всеми возможными факторами, угрожающими валидности вывода. Невозможно создать систему видов валидности латентного объекта исследования, которая была бы

достаточной для получения объективных результатов в аспекте их адекватности. Прямой перенос методов валидизации, разработанных и используемых в социальных и психологических науках, без учета специфики объекта исследования в образовании при контроле над результатами обучения, приводит к появлению неограниченного количества видов валидности при установлении валидности результатов педагогического тестирования.

Педагогический тест как инструмент исследования является основой тестового измерения. Содержание теста служит образом объекта измерения. Это значит, что для установления объективности тестовых результатов он должен обладать внутренней содержательной валидностью.

Во второй главе выявлено, что возможность обеспечения внутренней валидности теста непосредственно зависит от характера измеряемого объекта. Современная педагогическая тестология в качестве объекта контроля за результатами обучения рассматривает уровень обученности учащихся, который носит латентный характер. Для реализации возможности измерения объект представляется в качестве структурного теоретического конструкта, отражающего структуру предметного содержания. Содержание теста репрезентативно представляет элементы структуры теоретического конструкта в соответствующих пропорциях. Содержательная валидность такого теста оказывается внешней из-за ограниченности длины теста и называется куррикулярной. Куррикулярная валидность является частным случаем содержательной валидности как соответствия содержания теста структурному теоретическому конструкту. Поскольку она зависит от статистических характеристик тестовых заданий, полученных на выборке стандартизации, постольку является внешней. Статистическая зависимость куррикулярной валидности позволяет экстраполировать тестовые результаты на большую, чем нормативная, выборку и не позволяет экстраполировать результаты на содержание большее, чем тестовое. Внешний характер куррикулярной валидности делает тестирование относительным. Вследствие отсутствия внутренней валидности тестовые результаты оказываются необъективными в аспекте их соответствия объекту измерения.

Для доказательства объективности результатов привлекают критериальную валидность как их соответствие внешнему заведомо валидному критерию. Проблема выбора валидного критерия для валидизации делает процесс нетехнологичным. Абсолютно валидный критерий не существует. В настоящее время в педагогической тестологии используются два относительных решения проблемы, обусловленных выбором внешнего критерия для эмпирической валидизации. Здесь используются количественные методы валидизации — вычисление различного рода коэффициентов корреляции, статистически подтверждающих наличие связи между результатами валидизируемого измерения и внешним критерием.

Если в качестве внешнего критерия выбирается внутренняя норма, а именно нормальный закон распределения статистических величин для выборки стандартизации, то тест называется нормативно-ориентированным. Он признается валидным на основе принципа функциональности и позволяет ранжировать тестирующихся по степени владения тестовым содержанием. Уровень обученности как объект исследования не может быть измерен, ввиду отсутствия внутренней валидности нормативно-ориентированого теста. Главной характеристикой качества становится точность как свобода от ошибки измерения, или надежность НОТ. Зависимость тестовых результатов от субъективных свойств выборки стандартизации делает процесс валидизации нетехнологичным.

Если в качестве внешнего критерия выбираются иные результаты измерения того же объекта, отвечающие целям тестирования, валидность которого всегда остается непроясненной, то тест называется критериально-ориентированным. Критериальная валидность результатов тестирования понимается как их соответствие результатам другой, не более валидной оценки и в педагогических тестах является валидностью соответствия. Содержательная валидность критериально-ориентированного теста, построенного на основе структурного теоретического конструкта, носит характер внешней куррикулярной валидности теста ввиду его ограниченности. Отсутствие внутренней валидности критериально-ориентированного тестирования делает измерение относительным в аспекте относительного

соответствия содержания теста измеряемой предметной области, а значит, нетехнологичным.

Латентный объект и структурный характер теоретического конструкта таких тестов позволяют проводить поэлементный контроль над измеряемым предметным содержанием, в том числе стандартом. Необходимость контроля над целостным предметным содержанием, в том числе необходимость контроля над выполнением стандарта по предмету приводит к наращиванию количества элементов, подвергающихся измерению. Что, в свою очередь, приводит к увеличению структуры изучаемого предмета, потере целостностности образования и, как следствие, экстенсивному характеру обучения. Относительность и необъективность контроля не способствуют сохранению единого образовательного пространства на территории РФ при условии сохранения вариативности образовательных программ.

Отсутствие внутренней валидности теста со структурным конструктом и репрезентативным тестовым содержанием, зависимость результатов тестирования от субъективных свойств выборки стандартизации потребовали выявления наиболее общих признаков, характеризующих результаты обучения. В них должны отражаться целостность знания предмета, стандарта, которые не обусловлены статистическими характеристиками и структурой изложения предмета. Наиболее общим признаком обученности является знание теоретических подходов, представленных в содержании предмета, или знание содержания как целого.

В образовании при контроле над результатами обучения объектом измерения служит предметное содержание, определенное структурой предмета и ограниченное теоретическими подходами к его изложению. Изменение характера объекта контроля с латентного на определенный позволяет представить предметное содержание в виде целостного теоретического конструкта, но не структурной, а целостной модели объекта. В этом случае результат обучения — обученность, как открытый объект измерения, не требует проверки владения всеми элементами структуры предметного содержания. Достаточным для проверки обученности оказывается владение хотя

бы одним теоретическим подходом. Содержательная валидность отражает характеристики объекта, т. е. является внутренней. Результаты тестирования становятся объективными в аспекте их соответствия объекту измерения — обученности.

Целостность конструкта и внутренний характер содержательной валидности позволяют в качестве критерия валидизации выбрать содержание самого теста как внутренний критерий, не зависящий от статистических характеристик тестовых заданий. Определение экспертной группой тестового содержания осуществляется в соответствии с принципом межличностного соглашения. В содержании теста сохраняется зависимость от теоретического конструкта, что наделяет тестовые результаты внутренним характером.

Результаты эксперимента по применению валидных тестов для измерения обученности учащихся, созданных с применением технологии их валидизации, подтверждают обоснованность теоретических положений об объективности тестовых результатов. В 100 % случаев результаты тестирования, полученные с помощью валидного теста, построенного в соответствии с целью измерения, подтвердили прогноз успеха на абитуриентских испытаниях при традиционном контроле тестами централизованного тестирования. Этот результат устанавливает высокую степень различительной способности тестов, построенных на основе технологичного целостного конструкта.

Возможность объективного измерения тестами, созданными на основе целостного конструкта, позволит, на наш взгляд, повысить качество обучения, не увеличивая объема структуры предмета, интенсифицировать процесс обучения, т. е. перейти от знаний, умений, навыков, предъявляемых в рамках деятельностного подхода, к системным теоретическим знаниям. Возможность объективного контроля; не зависящего от статистических характеристик, создает реальные перспективы обеспечения образовательного стандарта и сохранения единого образовательного пространства страны, используя многообразие программ и методик преподавания, сложившееся к настоящему моменту.

В ходе исследования выявлен ряд проблем. На наш взгляд, нет границ применения таких тестов для целей рубежного и итогового тестирования, однако ограничены возможности их конструирования. Во-первых, не все содержательные единицы и не каждого учебного предмета поддаются построению целостного конструкта. Поскольку целью изучения отдельных дисциплин является не формирование теоретического знания, а закрепление определенных навыков, постольку в ряде случаев необходимо корректировать цели обучения. Во-вторых, соответственно новым целям необходима корректировка содержания ряда предметов, зафиксированного в стандарте. Другими словами, стандарт по предмету необходимо представить в виде целостного теоретического конструкта для интенсификации обучения. В-третьих, обеспечение интенсивности обучения и возможности контроля над теоретическим знанием необходимо требует привлечения ученых-дидактов с целью создания новых методов и методик обучения, базирующихся на целостном конструкте предметного содержания.

ЛИТЕРАТУРА

1. Аванесов В.С. Основы научной организации педагогического контроля в высшей школе. М.: Изд-во МИСиС, 1989. 166 с.

2. Аванесов В.С. Форма тестовых заданий. М., 1991. 33 с.

3. Аванесов В.С. Математические модели педагогического измерения. М.: Исследовательский центр проблем качества подготовки специалистов, 1994. 25 с.

4. Аванесов В.С. Научные проблемы тестового контроля знаний. М.: Исследовательский центр проблем качества подготовки специалистов, 1994. 135 с.

5. Аванесов В.С. Теоретические основы разработки заданий в тестовой форме. М: Исследовательский центр проблем качества подготовки специалистов, 1995. 93 с.

6. Аванесов В.С. Композиция тестовых заданий. М.: Адепт, 1998. 217 с.

7. Аванесов В.С. Определение, предмет и основные функции педагогической диагностики // Педагогическая диагностика. 2002. № 1. С. 41—43.

8. Азальгадов Г.Г. Теория и практика оценки качества товаров (основы квалиметрии). М.: Экономика, 1982. 256 с.

9. Анастази А. Психологическое тестирование: В 2 кн. Кн. 1, М.: Педагогика, 1982. 320 с.

10. Анастази А. Психологическое тестирование: В 2 кн. Кн. 2. М.: Педагогика, 1982. 336 с.

11. Андреев С.С., Серебрякова Л.А. Роль централизованного тестирования при региональном мониторинге качества образования // Развитие системы тестирования в России: Тез. докл. Всерос. конф. Ч. 1. М.: Прометей, 2000. С. 12—14.

12. Архангельский СИ. Учебный процесс в высшей школе, его закономерные основы и методы. М.: Высшая школа, 1980. 368 с.

13. Атаханов Р. Математическое мышление и методики определения уровня его развития / Под ред. В.В. Давыдова. Москва; Рига, 2000. 208 с.

14. Бабанский Ю.К. Интенсификация процесса обучения. М.: Знание, 1987. 80 с.

15. Бабанский Ю.К. Проблемы повышения эффективности педагогических исследований. М.: Педагогика, 1982. 192 с.

16. Бабанский Ю.К., Поташник М.М. Оптимизация педагогического процесса (в вопросах и ответах). Киев: Рад. школа, 1983. 287 с.

17. Байденко В.И. Образовательный стандарт. Опыт системного исследования. Новгород: Изд-во НовГУ им. Ярослава Мудрого, 1999. 440 с.

18. Байденко В.И. Стандарты в непрерывном образовании: современное состояние. М.: Исследовательский центр проблем качества подготовки специалистов, 1998. 249 с.

19. Балыхина Т.М. Словарь терминов и понятий тестологии. М.: Изд-во РУДН, 2000. 164 с.

20. Белых С.Л., Копотева ГЛ. Рекомендации по составлению учебных программ. Ижевск: Изд-во ИУУ, 2002. 28 с.

21. Бершадский М.Е., Гузеев В.В. Дидактические и психологические основания образовательной технологии. М.: Центр «Педагогический поиск», 2003. 256 с.

22. Беспалько В.П. Образование и обучение с участием компьютеров. М.: Изд-во Московского психолого-социального института; Воронеж: Изд-во НПО «МОДЭК», 2002. 352 с.

23. Беспалько В.П. Педагогика и прогрессивные технологии обучения. М.: Изд-во института профессионального образования Министерства образования России, 1995. 336 с.

24. Беспалько В.П. Слагаемые педагогической технологии. М.: Педагогика, 1989. 192 с.

25. Вербицкий А.А. Новая образовательная парадигма и контекстное обучение. М.: Исследовательский центр проблем качества подготовки специалистов, 1999. 75 с.

26. Веселкова Т.С. Междисциплинарные тесты как средство диагностики системности знаний учащихся: Дис. ... канд. пед, наук. Ижевск: ИУУ, 2000. 146 с.

27. Вопросы совершенствования проверки и оценки знаний, умений и навыков учащихся / Под ред. И.И. Кулибабы. М., 1975.

28. Гальперин П.Я. Психология как объективная наука // Избранные психологические труды / Под ред. А.И. Подольского. М.; Воронеж, 1998.

29. Горбушин Ш.А. Инновационная технология обучения физике в профессиональном лицее: Автореф. дис. ... канд. пед. наук. Казань: Институт среднего специального образования РАО, 1997. 43 с.

30. Гершунский Б.С. Перспективы развития теории оптимизации в педагогике // Проблемы развитая и методики оптимизации педагогического процесса: Сб. науч. тр. / Редкол.: М.М. Поташник (отв. ред.) и др. М.: Изд-во АПН СССР, 1989. С. 4—14.

31. Гессен СИ. Основы педагогики. Введение в прикладную философию / Отв. ред. и сост. П.В. Алексеев. М.: Изд-во «Школа-Пресс», 1995. 448 с.

32. Гласс Дж., Стенли Дж. Статистические методы в педагогике и психологии. М.: Прогресс, 1976. 495 с.

33. Грабарь М.И., Краснянская К.А Применение математической статистики в педагогических исследованиях. Непараметрические методы. М.: Педагогика, 1997. 135 с.

34. Гузеев В,В; Планирование результатов образования и образовательная технология. М.: Народное образование, 2000. 240 с.

35. Гуревич КМ. Профессиональная пригодность и основные свойства нервной системы. М.: Наука, 1970.

36. Давыдов В.В. Виды обобщения в обучении: Логико-психологические проблемы построения учебных предметов. М.: Педагогическое общество России, 2000. 480 с.

37. Джарол Б. Мангейм, Ричард К. Рич. Политология. Методы исследования. М: Изд-во «Весь мир», 1999. 544 е.

38. Дьяченко В.К. Новая дидактика. М.: Народное образование, 2001, 496 с,

39. Дюк В.А Компьютерная психодиагностика. СПб.: Изд-во «Братство», 1994. 364 с.

40. Единый государственный экзамен: Материалы Центра тестирования Минобразования России на конкурс контрольных измерительных материалов и технологий по проведению эксперимента в 2001 году. М,: Век книги 3, 2001, 510 с.

41. Ермелаев О.Ю. Математическая статистика для психологов. М.: Московский психолого-социальный институт; Флинта, 2002. 336 с,

42. Жуковская ЗД. Методические основания и технология разработки и функционирования комплексной системы контроля качества подготовки специалистов в вузе: Автореф. дис. ... д-ра пед. наук.СПб.: СПбГУ, 1994.

43. Измерения в педагогике: Сб. науч. труд. / Под ред. В.С. Черепанова. Ижевск: Изд-во ИУУ, 1997. 116 с.

44. Ингенкамп К. Педагогическая диагностика. М.: Педагогика, 1991. 240 с.

45. Исаев В.А., Игнатьева Е.Ю. О возможности применения тестирования при итоговой аттестации выпускников в системе многоуровневого образования // Развитие системы тестирования в России: Тез. докл. Всерос. конф. Ч. 3. М.: Прометей, 2000. С. 32—33.

46. Казаринов А.С. Технология педагогического эксперимента. Глазов: Глазов, 1999. 192 с.

47. Казаринов А.С., Култышева А.Ю., Мирошниченко АА Адаптационная валидизация педагогических тестов. Глазов: Изд-во ГГПИ, 2001. 124 с.

48. Казаринов А.С., Култышева А.Ю., Мирошниченко А.А. Технология адаптационной валидности тестовых заданий. Глазов: Изд-во ГГПИ, 1999. 62 с.

49. Кальней В.А., Шишов СЕ. Технология мониторинга качества обучения в системе «учитель—ученик». М.: Педагогическое общество России, 1999. 86 с.

50. Карпенко Д.С., Карпенко О.М., Шлихунова Е.Н. Автоматизированная система мониторинга эффективности усвоения знаний и качества тестовых заданий // Инновации в образовании. 2001. № 2. С. 69—85.

51. Каспржак А.Г., Левит М.В. Базисный учебный план и российское образование в эпоху перемен. М.: МИРОС, 1994. 144 с.

52. Качество знаний учащихся и пути его совершенствования / Под ред. М.Н. Скаткина, В.В. Краевского. М.: Педагогика, 1978. 208 с.

53. Квалиметрия человека и образования: методология и практика. Теория и практика квалиметрического маниторинга образования в России: Матер. 8-го Всерос. симпозиума. М.: Исследовательский центр проблем качества подготовки специалистов, 1999. В 3 ч. Ч. 1. 99 с; Ч. 2. 120 с; Ч. 3. 70 с.

54. Квалиметрия в образовании: методология и практика. Отечественная квалиметрическая научно-практическая школа в системе

высшего образования: итоги и перспективы развития. Актуальные проблемы анализа и оценки качества образования: Матер. 10-го Всерос. симпозиума. М.: Исследовательский центр проблем качества подготовки специалистов, 2002. Кн. 1. 206 с.

55. Квалиметрия в образовании: методология и практика. Развитие квалиметрии человека и образования: теоретические постановки и практика: Матер. 10-го Всерос. симпозиума. М.: Исследовательский центр проблем качества подготовки специалистов, 2002. Кн. 2. 413 с.

56. Квалиметрия в образовании: методология и практика. Мониторинг и контроль качества образования: развитие методологии и опыт: Матер. 10-го Всерос. симпозиума. М.: Исследовательский центр проблем качества подготовки специалистов, 2002. Кн. 3. 299 с.

57. Кирьяков Б.С. Педагогическая модель интеллектуального воспитания школьников. Рязань: Изд-во «Русское слово», 2002. 208 с.

58. Китаев Н.Н. Групповые экспертные оценки. М.: Знание, 1975.64с.

59. Кларин М.В. Инновационные модели Обучения в зарубежных педагогических поисках. М.: Арена, 1994. 222 с.

60. Кларин М.В. Инновации в обучении: метафоры и модели. Анализ зарубежного опыта. М.: Наука, 1997. 223 с.

61. Коджаспирова Г.М., Коджаспиров А.Ю. Педагогический словарь. М.: Изд-во «Академия», 2001. 176 с.

62. Краевский В.В. Методология педагогического исследования. Самара: Изд-во СамГПИ, 1994. 165 с.

63. Краснянская КА, Кузнецова Л.В. Оценка математической подготовки школьников по результатам международного тестирования. М.: Просвещение, 1995. 96 с.

64. Кулемин НА. Введение в квалиметрию общеобразовательной школы. Ижевск: Изд-во «Алфавит», 1998. 166 с.

65. Кэмпбелл Д. Модели экспериментов в социальной психологии и прикладных исследованиях / Пер. с англ. СПб.: Социально-психологический центр, 1996. 391 с.

66. Лаак Я. тер Психодиагностика: проблемы содержания и методов. М.: Издательство «Институт практической психологии»; Воронеж: НПО «МОДЭК», 1996. 384 с.

67. Лебедева Е.Н. Методические рекомендации по сертификации дидактических тестов для диагностики уровня обученности студентов. Ижевск: Изд-во ИжГСХА, 1998. 37 с.

68. Лебедева Е.Н. Сертификация тестовых измерителей как способ повышения объективизации контроля обученности: Дис. ... канд. пед. наук. Ижевск: Изд-во ИжГТУ, 1998. 151 с.

69. Левин К. Разрешение социальных конфликтов. СПб.: Издательство «Речь», 2000. 408 с.

70. Левин К. Теория поля в социальных науках / Пер. с англ. СПб.: «Сенсор», 2000. 368 с.

71. Левитес Д.Г. Практика обучения: современные образовательные технологии. М.: Изд-во «Институт практической психологии»; Воронеж: НПО «МОДЭК», 1998. 288 с.

72. Леднев В.С. Содержание общего среднего образования: Проблемы структуры. М.: Педагогика, 1980. 264 с.

73. Лекторский В.А. Субъект, объект, познание. М.: Наука, 1980. 356с.

74. Лернер И.Я. Дидактические основы методов обучения. М.: Педагогика, 1981. 185 с.

75. Лернер И.Я. Качества знаний учащихся. Какими они должны быть? М.: Педагогика, 1978.

76. Любимова О.В. О валидизации «пороговых» оценочных средств // Проблемы вузовской и школьной педагогики: Тез. докл. регион, науч.-практ. конф. «Четвертые Есиповские чтения». Глазов: ГГПИ, 2001. С. 64—65.

77. Люсин Д.В. Критериально-ориентированные педагогические тесты: сущность и методы конструирования: Автореф. дис. ... канд. пед. наук. М.: Исследовательский центр проблем качества подготовки специалистов, 1995. 28 с.

78. Люсин Д.В. Основы разработки и применения критериально-ориентированных педагогических тестов. М.: Исследовательский центр проблем качества подготовки специалистов, 1993. 51 с.

79. Майоров А.Н. Тесты школьных достижений: конструирование, проведение, использование. 2-е изд. СПб.: Образование и культура, 1997. 304 с.

80. Майоров А.Н. Теория и практика создания тестов для системы образования. М.: Народное образование, 2000. 352 с.

81. Майоров А.Н. Мониторинг в образовании СПб.: Изд-во Образование — Культура, 1998. 320 с.

82. Майоров А.Н. Образовательные стандарты Санкт-Петербургской школы. Требования к тестам школьных достижений. СПб.: Образование и культура, 1996. 30 с.

83. Макаров АА Методология и методы системной организации комплексного мониторинга качества образования: Автореф. дис. ... д-ра пед. наук. М.: Исследовательский центр проблем качества подготовки специалистов, 1999. 36 с.

84. Мангейм Д.Б., Рич Р.К Политология. Методы исследования / Пер. с англ. М.: Изд-во «Весь Мир», 1999. 544 с,

85. Мангейм К. Структурный анализ эпистемологии. М.: ИНИОН, 1992. 38 с.

86. Материалы для организации и проведения мониторинга эксперимента / Сост. Г.С. Ковалева, СВ. Иванова. М.: Дрофа, 2001. 224 с.

87. Методические указания по применению экспертных оценок при проверке качества преподавания в вузе / Сост. В.С, Черепанов, А.С. Казаринов. Глазов: ГГПИ, 1989. 20 с.

88. Методы математической статистики в психологии; Методическое пособие для психологов / Сост. АА. Абдулхаков. Ижевск: Изд-во ИУУ, 1999. 68 с.

89. Микешина Л.А. Философия познания. Полемические главы. М.: Прогресс — Традиция, 2002. 624 с.

90. Мирошниченко А.А. Информационно-семантическое структурирование учебного материала (на примере школьного курса физики): Автореф. дис. ... канд. пед. наук. Ижевск: Изд-во УдГУ, 1995. 22 с,

91. Мирошниченко А.А. Профессионально-ориентированные структуры учебных элементов. Глазов: Изд-во ГГПИ, 1999. 68 с.

92. Михайлычев ЕА. Дидактическая тестология. М.: Народное образование, 2001. 432 с.

93. Михайлычев Е.А. Система педагогической диагностики: ключевые понятия и принципы // Педагогическая диагностика. 2002. № 1. С. 44—66.

94. Михеев В.И. Методика получения и обработки экспериментальных данных в психолого-педагогических исследованиях. М.: Изд-во УДН, 1986. 84 с.

95. Михеев В.И. Моделирование и методы измерений в педагогике. М.: Высшая школа, 1987. 198 с.

96. Михеев В.И., Шабунин М.И. О проблеме взаимодействия школьного и вузовского метематического образования в России // Проблемы теории и методики обучения. 1999. № 4. С. 85—88.

97. Мониторинг в образовании: Сб. науч.-метод. материалов. Ижевск: Изд-во ИУУ, 2000. 188 с.

98. Мэнстид Энтони СР., Симин Гюн Р. Методология в социальной психологии: проверка гипотез // Перспективы социальной психологии. М.: Изд-во ЭКСМО-Пресс, 2001. С. 94—126.

99. Навалихина И.А. О технологии измерения адаптационной валидности теста // Развитие системы тестирования в России: Тез. докл. Всерос. конф. Ч. 2. М.: Прометей, 1999. С. 100—101.

100. Навалихина И.А Содержательная валидность теста как основ-| ное условие объективности тестового контроля // Развитие тестовых технологий в России: Тез. докл. Всерос. конф. М.: Центр тестирования МО РФ, 2002. С. 241—242.

101. Навалихина И.А. Проблема соотношения видов валидности в структуре общей валидности теста // Квалиметрия человека и образования: методология и практика: Матер. ГХ симпозиума / Под. ред. Н.А. Селезневой, А.И. Субетто. М.: Исследовательский центр проблем качества подготовки специалистов, 2000. С. 19—21.

102. Навалихина И.А Соотношение внешней и внутренней валидности теста в теории IRT // Квалиметрия в образовании: методология и практика: Матер. X симпозиума. Кн. 2. М.: Исследовательский центр проблем качества подготовки специалистов, 2002. С. 85—87.

103. Навалихина И.А. Об эффективности использования нормативно-ориентированных и критериально-ориентированных тестов в аспекте валидизации теста // Развитие системы тестирования в Удмуртской Республике: Матер, регион, науч.-практ. конф. / Под ред. В,С, Черепанова, Ижевск: Изд-во ИжГГУ, 2001, С, 68—69,

104. Навалихина ЙА. Проблема соотношения трудности и сложности тестовых заданий при измерении качества обучения в аспекте валидности // Cogito. Вып. 5. Психолого-педагогические исследования профессиональной деятельности: Сб. ст. / Под ред. С.Ф. Сироткина. Ижевск: Изд-во УдГУ, 2001. С. 123—130.

105. Нейман Ю.М, Как оценивается уровень подготовленности учащихся по результатам единого государственного экзамена. М.: Центр тестирования МО РФ, 2002. 32 с.

106. Нейман Ю.М., Хлебников В.А. Введение в теорию моделирования и параметризации педагогических тестов. М., 2000. 168 с.

107. Немов Р.С. Психология: Учебник: В 3 кн. М.: Гуманит. изд. центр ВЛАДОС, 2001. Кн. 3. 640 с.

108. Новые возможности в управлении качеством образования: Сб. докл. Ч. 1 // Информационные технологии в управлении качеством образования и развитие образовательного пространства: Серия материалов Всероссийской школы-семинара / Под общ. ред. Н.А. Селезневой, И.И. Дзегеленка. М.: Исследовательский центр проблем качества подготовки специалистов, 2000. 210 с.

109. Нурминский И.И., Гладышева Н.К. Статистические закономерности формирования знаний и умений учащихся. М.: Педагогика, 1991. 224 с.

110. Образовательный стандарт высшей школы: сегодня и завтра / Под общ. ред. В.И. Байденко, НА Селезневой. М.: Исследовательский центр проблем качества подготовки специалистов, 2001. 206 с.

111. Овчинников В.В. Оценивание учебных достижений учащихся при проведении централизованного тестирования. М.: Центр тестирования МО РФ, 2001. 27 с.

112. Осмоловская И.М. Организация дифференцированного обучения в современной общеобразовательной школе. М.: Изд-во «Институт практической психологии»; Воронеж: Изд-во НПО «МОДЭК», 1998. 160 с.

113. Отраслевой стандарт. Педагогические тесты. Термины и определения. М., 2001. 23 с,

114. Оценка качества подготовки выпускников основной школы по математике / Г.В. Дорофеев, Л.В. Кузнецова, Г.М. Кузнецова и др. М.: Дрофа, 2000. 80 с.

115. Панасюк В.П. Научные основы проектирования педагогических систем внутришкольного управления качеством образовательного процесса. СПб.; М., 1997. 297 с.

116. Педагогический мониторинг качества образования / Под ред. В.И. Андреева. Казань: Изд-во КГУ, 1999. 266 с.

117. Подготовка учителя математики: Инновационные подходы / Под ред. В.Д. Шадрикова. М.: Гардарики, 2002. 383 с.

118. Поташник М.М. Качество образования: проблемы и технология управления. М.: Педагогическое общество России, 2002. 352 с.

119. Поташник М.М. Проблемы оптимизации в педагогике // Советская педагогика. 1985. № 2.

120. Проблемы качества образования, его нормирования и управления: Сб. науч. ст. / Под общей ред. Н.А. Селезневой и В.Г. Казановича. М.: Исследовательский центр проблем качества подготовки I специалистов, 1999. 192 с.

121. Проблемы развития теории и методики оптимизации педаго-| гического процесса: Сб. науч. тр. / Редкол.: М.М. Поташник и др. М,: Изд. АПН СССР, 1989. 123 с.

122. Равен Дж. Педагогическое тестирование: проблемы, заблуждения, перспективы. М.: Когито-Центр, 2001. 142 с.

123. Развитие системы тестирования в России: Тез. докл. I Всерос. |конф. (Москва, ноябрь 1999 г.): В 3 ч. М.: Прометей, 1999. Ч. 1. 99 с; 14. 2. 120 с; Ч. 3. 70 с.

124. Развитие системы тестирования в России: Тез. докл. II Всерос. конф. (Москва, ноябрь 2000 г.): В 5 ч. М.: Прометей, 2000. Ч. 1; 78 с; Ч. 2. 79 с; Ч. 3. 100 с; Ч. 4. 108 с; Ч. 5. 57 с.

125. Развитие системы тестирования в России: Тез. докл. III Всерос. конф. (Москва, ноябрь 2001 г.). М.: Прометей, 2001. 254 с.

126. Развитие системы тестирования в России: Тез. докл. IV Всерос. конф. (Москва, ноябрь 2002 г.). М.: Прометей, 2002. 314 с.

127. Родионов Б.У., Татур А.О. Стандарты и тесты в образовании. |М.: Изд-во МИФИ, 1995. 48 с.

128. Родыгина ТА. Диагностика начального уровня обученности I студентов на основе гетерогенных тестов (на примере курса электротехники): Автореф. дис. ... канд. пед. наук. Ижевск: Изд-во УдГУ, 1996. 23 с.

129. Родыгина Т.А., Файзуллина Г.З. Проектирование и применение тестов в профессиональном образовании. Ижевск: Изд-во ИУУ, | 2002. 64 с.

130. Сакаева С.Р. Тестирование как метод повышения эффективности и объективности контроля знаний в общеобразовательной школе (на примере школьного курса физики): Дис. ... канд. пед. наук. Ижевск: Изд-во УдГУ, 1997. 139 с.

131. Самыловский А.И. Тест как объективный измерительный инструмент в образовании // Вопросы тестирования в образовании. 2001. № 1. С. 10—39.

132. Саранцев Г.И. Методика обучения метематике в средней школе. М.: Просвещение, 2002. 224 с.
133. Сафонцев СА Критериально-ориентированное тестирование // Педагогическая диагностика. 2002. № 1. С. 77—80.
134. Семенко Е.А., Дорошенко В.Г., Агрова О.В. О некоторых принципах структурирования педагогических тестов (на примере математики) // Развитие системы тестирования в России: Тез. докл. II Всерос. конф. (Москва, ноябрь 2000 г.): В 5 ч. М.: Прометей,
2000. Ч. 4. С. 13—14.
135. Селезнева НА Качество высшего образования как объект системного исследования: Лекция-доклад. М.: Исследовательский центр проблем качества подготовки специалистов, 2001. 76 с.
136. Семин Ю.Н. Интеграция содержания инженерного образования: дидактический аспект. Ижевск: Изд-во ИжГТУ, 2000. 140 с.
137. Семин Ю.Н. Учебные тезаурусы теоретической и прикладной механики. Ижевск: Изд-во ИжПУ, 2001. 76 с.
138. Серебрякова Л.А. Мониторинг качества общего среднего образования в региональной системе: Автореф. дне. ... канд. пед. наук. Великий Новгород: НГУ, 2000. 21 с.
139. Сидоренко Е.В. Методы математической обработки в психологии. СПб.: ООО «Речь», 2000. 350 с.
140. Сидорина ТВ. Теоретические основы экспертирования образовательных систем: Автореф. дис. ... д-ра пед. наук. Киров: ВГПУ, 2001. 43 с.
141. Скаткин М.Н. Проблемы современной дидактики. М.: Педагогика, 1980. 96 с.
142. Скаткин М.Н., Краевский В.В. Содержание общего образования. М.: Педагогика, 1981. 96 с.
143. Снигирева ТА Диагностика структуры знаний обучающихся на основе тезаурусного и квалиметрического подходов: Автореф. дис. ... канд. пед. наук, Ижевск: Изд-во УдГУ, 2001. 23 с.
144. Снигирева ТА. Методические рекомендации по организации диагностики структуры знаний обучающихся на основе тезаурусного и квалиметрического подходов / Под ред. В.С. Черепанова. Ижевск: Экспертиза, 2000. 34 с.
145. Современный словарь иностранных слов: Около 20 000 слов. М.: Русский язык, 1992. 740 с.

146. Сохор А.М. Логическая структура учебного материала. М.: Педагогика, 1974. 192 с.
147. Субетто А.И. Квалитология образования. СПб.; М.: Исследовательский центр проблем качества подготовки специалистов, 2000. 220 с.
148. Субетто А.И. «Метаклассификация» как наука о механизмах и закономерностях классифицирования (опыт обобщения): В 2 ч. Ч. 2. СПб.; М.: Исследовательский центр проблем качества подготовки специалистов, 1994. 80 с.
149. Субетто А.И., Чекмарев В.В. Мониторинг источников формирования содержания высшего образования. Кострома; М.: Исследовательский центр Госкомвуза, КГПУ им. НА Некрасова, 1996. 239 с,
150. Талызина Н.Ф. Теоретические основы контроля в учебном процессе. М.: Знание, 1983. 96 с.
151. Талызина Н.Ф. Управление процессом усвоения знаний (психологические основы). М.: Изд-во МГУ, 1984. 345 с.
152. Талызина Н.Ф. Педагогическая психология. М.: Издательский центр «Академия», 2001. 288 с.
153. Теоретические основы содержания общего среднего образования / Под ред. В.В. Краевского и И.Я. Лернера. М.: Педагогика, 1983. 352 с.
154. Теоретические проблемы современного школьного учебника: Сб. науч. тр. / Отв. ред. И.Я. Лернер и Н.М. Шахмаев. М.: Изд-во АПН СССР, 1989. 172 с.
155. Требования к знаниям и умениям школьников: Дидактико-методический анализ / Под ред. АА Кузнецова. М.: Педагогика, 1987. 176 с.
156. Управление качеством образования / Под ред. М.М. Поташника. М.: Педагогическое общество России, 2000. 448 с.
157. Учебные стандарты школ России: Государственные стандарты начального общего, основного общего и среднего (полного) общего образования. Кн. 2. Математика. Естественно-научные дисциплины / Под ред. В.С. Леднева, НД. Никандрова, М.Н. Лазутовой. М.: «ТЦ Сфера», «Прометей», 1998. 336 с.
158. Формирование приемов математического мышления / Под ред. Н.Ф. Талызиной. М.: Изд-во ТОО «Вентана-граф», 1995. 231 с.
159. Фридман Л.М. Теоретические основы методики обучения математике. М.: Московский психолого-социальный институт; Флинта, 1998. 224 с.

160. Хлебников ВА, Нейман Ю.М., Парфенов В.С., Самыловский А.И., Шарыгин И.Ф. Объективная оценка учебных достижений // Педагогическая диагностика. 2002. № 1. С. 67—76.

161. Холодная МА. Психология интеллекта: парадоксы исследования. Томск: Изд-во Том. ун-та; М.: Изд-во «Барс», 1997. 392 с.

162. Цетлин В.С. О структуре учебника // Теоретические проблемы современного школьного учебника: Сб. науч. тр. / Огв. ред. ИЯ. Лернер и Н.М. Шахмаев. М.: Изд-во АПН СССР, 1989. С. 22—34.

163. Челышкова М,Б. Адаптивное тестирование в образовании (теория, методология, технология). М.: Исследовательский центр проблем качества подготовки специалистов, 2001. 165 с.

164. Челышкова М.Б. Методические рекомендации по разработке педагогических тестов для комплексной оценки подготовленности студентов в вузе. М., 1995. 77 с.

165. Челышкова М.Б. Разработка педагогических тестов на основе современных математических моделей. М.: Исследовательский центр проблем качества подготовки специалистов, 1995. 32 с.

166. Челышкова М.Б. Теория и практика конструирования педагогических тестов. М.: Исследовательский центр проблем качества подготовки специалистов, 2001. 410 с.

167. Челышкова М.Б., Ковалева Г.С. Основные подходы к оценке качества подготовки обучаемых в нашей стране и за рубежом: Обзорный доклад // Квалиметрия человека и образования: методология и практика: Материалы восьмого симпозиума. М.: Исследовательский центр проблем качества подготовки специалистоЕ, 1999. 62 с.

168. Челышкова М.Б., Ковалева Г.С, Татур АО., Хлебников ВА Концепция аттестационного тестирования выпускников общеобразовательных учреждений на этапе перехода от школы к вузу // Проблемы качества, его нормирования и стандартов в образовании: Сб. науч. ст. М.: Исследовательский центр проблем качества подготовки специалистов, 1998. С. 110—118.

169. Черепанов В.С. Теоретические основы педагогической экспертизы: Дне. ... д-ра пед. наук. М.: АПН СССР, 19&1. 351 с.

170. Черепанов В.С. Экспертные методы в педагогике. Пермь: Изд-во ПГПИ, 1988. 84 с.

171. Черепанов В.С. Экспертные оценки в педагогических исследованиях. М.: Педагогика, 1989. 152 с.

172. Чернова Ю.К. Квалитативные технологии обучения. Тольятти: Изд-во Фонда «Развитие через образование», 1998. 149 с.

173. Чернова Ю.К. Теория и практика проектирования квалитативных технологий обучения: Дис. ... д-ра пед. наук. Казань, 1998. 364 с.

174. Чошанов М.А. Гибкая технология проблемно-модульного обучения. М.: Народное образование, 1996. 160 с.

175. Чошанов М.А. Обзор таксономии учебных целей в педагогике США // Педагогика. 2000. № 4. С. 86—91.

176. Чудинов Э.М. Природа научной истины. М.: Политиздат, 1977. 312 с.

177. Шихова О.Ф. Квалиметрический подход к разрешению проблемы диагностичности образовательного стандарта: Дис. ... канд. пед. наук. Ижевск: Изд-во ИжГТУ, 1997. 171 с.

178. Шишов С.Е., Кальней В.А. Мониторинг качества образования в школе. М.: Педагогическое общество России, 1999. 320 с.

179. Штейнберг В.Э. Дидактические многомерные инструменты: теория, методика, практика. М.: Народное образование, 2002. 304 с.

180. Штофф В.А. Проблемы методологии научного познания. М.: Высшая школа, 1978. 269 с.

181. Шухардина В.А. Адаптивные тесты как средство индивидуализации педагогического контроля качества знаний учащихся: Дис. ... канд. пед. наук. Ижевск / УдГУ, 2002. 174 с.

182. Birnbaum A. Some Latent Trait Models and Their Use in Inferring an Examinee's Ability / EM. Lord & M.R Novick // Statistical Theories of Mental Test Scores. Reading, Mass.: Addison — Wesley, 1968.

183. Bloom B.S. Taxonomy of Educational Objectives // The Classification of Educational Goals, Handbook I: Cognitive Domain. N. Y., 1956.

184. Riecken H.W., Boruch R.E, Campbell D.T., Caplan N,, Glennan T.K, Pratt J., Rees A, Williams W. Social Experimentation: A Method for Planning and Evaluating of Social Intervention. N. Y: Academic Press, 1974.

185. Campbell D.T., Fiske D.W. Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix // Psychological Bulletin. 1959. 56. P. 81—105.

186. Cook T.D., Campbell D.T The Design and Conduct of Quasi-Experiments and True Experiments in Field Settings // Handbook of industrial and organizational Psychology. Chicago: Rand McNally College Publishing Co, 1976.

187. Cook T.D., Campbell D.T. Quasi-Experiments. Chicago: Rand McNally College Publishing Co, 1979.
188. Cronbach L.J., Meehl P.E. Construct Validity in Psychological Tests // Psychological Bulletin. 1955. 52. P. 281—302.
189. Emberley P.C. The Restoration and Renewal of Core Curricula // Providence. Providence, 1996. Vol. 3. № 3. P. 250—270.
190. Linn R.L. Educational Measurement. (Third Edition) // American Council on Education. London: MacMillan, 1989.
191. Lord F.M. Applications of Item Response Theory to Practical Testing Problems. Hillsdale, N-J.: Erlbaum, 1980.
192. Lord F.M., Novick M.R. Statistical Theory of Mental Test Scores. Reading, Mass.: Addison — Wesley, 1968.
193. Rasch G. Probabilistic Models for some Intelligence and Attainment Test. Copenhagen: Danish Institute for Educational Research, 1960.

ПРИЛОЖЕНИЯ

Приложение 1

АНКЕТА
АНАЛИЗ КАЧЕСТВА СОДЕРЖАНИЯ ТЕСТОВЫХ ЗАДАНИЙ

Уважаемый эксперт _____!

Подготовительное отделение ИжГТУ проводит исследование по определению качества тестовых измерительных материалов по математике, предназначенных для рубежного контроля над обученностью слушателей подготовительного отделения.

Просим Вас оценить тестовые задания с точки зрения соответствия их содержания целостному теоретическому конструкту контролируемого раздела математики. С этой целью заполните, пожалуйста, предложенную таблицу:

Анализ качества содержания теста

Группа и № задания	Тип задания: Т — задание, контролирующее технику решения; Ф — задание, моделирующее функциональный подход	Уровень технической сложности: 1-й уровень сложности — 1 балл, 2-й уровень сложности — 2 балла, 3-й уровень сложности — 3 балла	Соответствие задания теоретическому конструкту в целом (однородность)	Уровень сложности заданий, моделирующих функциональный подход (если отличается от уровня технической сложности)
A1...				
B12				

СПАСИБО!

Приложение 2

АНКЕТА
АНАЛИЗ КАЧЕСТВА ФОРМЫ ТЕСТОВЫХ ЗАДАНИЙ

Уважаемый эксперт _____!

Подготовительное отделение ИжГТУ проводит исследование по определению качества формы тестовых измерительных материалов по математике, предназначенных для рубежного контроля над обученностью слушателей подготовительного отделения.

Просим Вас оценить задания теста с точки зрения соответствия их формы принятым в тестологии нормам. Если с Вашей точки зрения форма задания удовлетворительная, то поставьте «+», если форма задания неудовлетворительная, то поставьте «–» и укажите замечания по каждому заданию в отдельности.

Результаты согласования экспертов при анализе качества содержания теста

Предмет экспертной оценки	A1	A2	A3	A4	A5	A6	A7	B8	B9	B10	B11	B12
Тип задания	Т	Т	Т	Т	Т	Т	Т	Т	Т	Ф	Т	Ф
Уровень технической сложности задания	1	1	1	1	1	2	2	1	2	1	2	2
Количество баллов за верное решение	1	1	1	1	1	2	2	1	3	2	2	3

СПАСИБО!

ВАРИАНТ ЭКСПЕРИМЕНТАЛЬНОГО ТЕСТА

Подготовительное отделение ИжГТУ
Экзаменационный билет по математике
1-й семестр
2001—2002 уч. г.

ВАРИАНТ № 1

ЧАСТЬ А

УКАЖИТЕ ИЗ ПРИВЕДЕННЫХ ВАРИАНТОВ ОТВЕТОВ ПРАВИЛЬНЫЙ

1. Результат упрощения выражения $\left(\frac{1}{a-3b}-\frac{1}{a+3b}+\frac{6b}{a^2-9b^2}\right)+\frac{b(2a+b)}{a^2-9b^2}$ имеет вид:

1) 0; 2) $\frac{1}{2a+b}$; 3) $\frac{12}{2a+b}$; 4) $\frac{12b}{2a+b}$; 5) $12b$.

2. Сумма корней уравнения $\frac{x^2-x}{x^2-x+1}-\frac{x^2-x+2}{x^2-x-2}=1$ равна:

1) 1; 2) 2; 3) 3; 4) –1; 5) 0.

3. Число целых решений неравенства

$\frac{7x-13}{(x^2-9x+18)(x^2-8x+16)}+\frac{1}{x^2 10x+24}\le 0$ равно:

1) 0; 2) 3; 3) 2; 4) 1; 5) 4.

4. Значение выражения $(\sqrt{8}\sqrt{3\sqrt[3]{2}})+\left(\sqrt{2\sqrt{3}}\cdot\sqrt{\sqrt[6]{108}}\right)$ равно:

1) $\sqrt{2}$; 2) 4; 3) $3^{1/2}$; 4) 2; 5) 1.

5. Посторонний корень уравнения $\sqrt{x+1}+\sqrt{4x+13}=\sqrt{3x+12}$ равен:

1) –4; 2) 4; 3) 2; 4) –2; 5) –1.

6. Решением неравенства $\sqrt{x^2-2x+1}+\sqrt{x^2+4x+4}\le 3$ является множество:

1) $[-1;\ 2]$; 2) $(-2;\ 1)$; 3) $(-\infty;\ -2)\cup(1;\ +\infty)$; 4) $[-2;\ 1]$; 5) $(-1;\ 2)$.

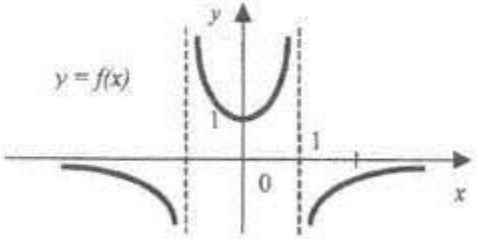
7. Областью значений функции $f(x)=2-\left|\sqrt{x}-3\right|$ является промежуток:

1) $(2;+\infty)$; 2) $(-\infty;\ 2)$; 3) $(-\infty;\ 2]$; 4) $[0;+\infty)$; 5) $(-\infty;\ 0]$.

ЧАСТЬ В

ДЛЯ КАЖДОГО ЗАДАНИЯ ЗАПИШИТЕ ПРАВИЛЬНЫЙ ОТВЕТ

8. По графику функции определите промежуток ее убывания:



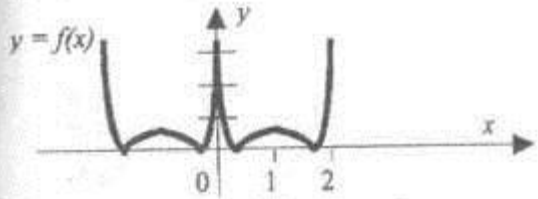
9. Решите неравенство $\dfrac{x-2}{\sqrt{2x-1}-1}<4$.

10. Запишите какую-либо формулу функции $y=f(x)$, если ее областью определения является промежуток $(-2;\ 1]\cup[1;+\infty)$.

11. Решите уравнение с параметром относительно переменной x:

$\dfrac{a+1}{ax-2}-1=0$.

12. По данному графику восстановите и запишите формулу зависимости переменной y от x.



Приложение 4

РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТА

№ п/п	ФИО	Баллы контрольного тестирования (от 0 до 100)	Баллы экспериментального тестирования		
			Всего за тест (от 0 до 20)	Ф-задания (от 0 до 5)	Т-задания (от 0 до 15)
1	Игнатьев	100	20	5	15
2	Закирова	93	20	5	15
3	Залилова	81	19	5	14
4	Перевозчиков	89	18	5	13
5	Емельянова	88	17	5	12
6	Вольнов	84	17	5	12
7	Ярема	82	17	5	12
8	Бердинских	79	17	3	14
9	Вахрамеева	78	17	4	12
10	*Ломаева*	82	17	2	15
11	Дадукова	93	16	5	11
12	Сафина	76	16	4	12
13	Казаков	77	15	3	12
14	Гребенщиков	100	14	5	9
15	*Пыжьянова*	85	14	0	14
16	Ушаков	77	14	2	12
17	*Шайгарданова*	82	14	3	11
18	Кузнецов	72	13	2	11
19	*Ганнага*	84	13	2	11

20	Бахтияров	67	12	1	11
21	*Дулисов*	83	12	4	8
22	*Кожевников*	82	12	0	12
23	Носков	61	12	2	10
24	*Чагина*	85	12	2	10
25	Артемьева	74	11	3	8
26	*Башак*	81	11	2	9
27	Бердникова	64	10	0	10
28	Горохов	56	10	3	7
29	Ерохина	70	10	0	10
30	Филиппова	75	10	0	10
31	Аслинова	77	9	0	9
32	Войцевич	74	9	0	9
33	Зуева	68	9	0	9
34	Князева	69	9	0	9
35	Мельников	78	9	0	9
36	Нурканова	78	9	0	9
37	Паздерина	53	9	0	9
38	Телегин	59	9	0	9
39	*Чуканин*	87	9	0	9
40	Чичканова	67	9	1	8
41	Бутин	71	8	0	8
42	Губайдуллин	76	8	1	7
43	Жулдыбина	68	8	0	8
44	Мокрушин	64	8	0	8
45	Снигирева	64	8	0	8
46	Ходырева	68	8	0	8
47	Демин	69	7	0	7
48	Меньшиков	58	7	0	7
49	Носкова	70	7	0	7
50	Шиляева	54	7	0	7
51	Ахметканова	79	6	0	6
52	Богатырева	70	6	1	5
53	Булдакова	49	6	0	6
54	Вахрушева	62	6	0	6
55	Макарова И.	58	6	0	6
56	Трухина	68	6	0	6
57	Ярославцев	77	6	3	3
58	Бикнева	67	5	0	5
59	Белоглазова	61	5	0	5
60	Васильев	34	5	0	5
61	Минахметов	51	5	0	5
62	Назарова	44	5	0	5

63	Полякова	72	5	0	5
64	Пермякова	56	5	0	5
65	Холмогорова	62	5	0	5
66	Чиркова	63	5	0	5
67	Емельянова	72	4	0	4
68	Карпова	61	4	0	4
69	Коньшев	51	4	0	4
70	Леонтьева	53	4	0	4
71	Пашкеева	72	4	0	4
72	Фархаева	58	4	0	4
73	Якимова	49	4	0	4
74	Быстрова	62	3	0	3
75	*Богомалов*	84	3	2	1
76	Данилов	63	3	0	3
77	Кириллова	53	3	0	3
78	Никитина	46	3	0	3
79	Торхова	55	3	0	3
80	Хохряков	72	3	0	3
81	Иванов	67	2	0	2
82	Калимуллина	46	2	0	2
83	Лапина	64	2	0	2
84	Макарова Н.	60	2	0	2
85	Мурзыханова	55	2	0	2
86	Андреева	41	1	0	1
87	Кожевников А.	44	1	0	1
88	Мерзляков	48	1	0	1
89	Подобедов	47	1	0	1
90	Пашин	40	1	0	1
91	Пушкова	61	1	0	1
92	Сысоев	40	1	0	1*
93	Масленников	66	1	0	1
94	Зворыгин Д.	47	0	0	0
95	Зворыгин Е.	40	0	0	0
96	Малых	28	0	0	0
97	Некрасов	21	0	0	0
98	Наговицин	60	0	0	0
99	Ханнанова	40	0	0	0
100	Чирков	38	0	0	0

Приложение 5

Перевод тестового балла централизованного тестирования (вузовского) в пятибалльную шкалу оценок в 2002 г.

(по рекомендациям Центра тестирования МО РФ)

Предмет	Пятибалльная шкала			
	2	3	4	5
	Интервал тестовых баллов			
Математика	0—37	38—53	54—80	81—100

СЕРИЯ «PROBATUM EST»

Ирина Александровна Караваева

ВАЛИДНОСТЬ ПЕДАГОГИЧЕСКИХ ТЕСТОВ

Монография

Корректор А.В. Соколова
Технический редактор С.И. Зянкина
Компьютерная верстка Л.О. Фольмер

Лицензия ЛР № 020411 от 16.02.97.
Сдано в производство 20.12.03. Формат 60×84 $^1/_{16}$.
Печать офсетная. Усл. печ. л. 8,83. Уч.-изд. л. 9.
Тираж 100 экз. Заказ № 22.

Издательский дом «Удмуртский университет»,
426034, г. Ижевск, ул. Университетская, 1, корп. 2.
Факс 8(3412) 75-21-55; *e-mail: knigi@udm.ru;*
internet: http://www.uni.udm.ru/pubhouse